



# INTEGRATED LABORATORY SYSTEMS

**Keith Johnston**

**TESSELLA SUPPORT SERVICES PLC**

Issue V1.R1.M0

July 2003



## **Characteristics of a modern lab**

The process of conducting scientific research within a laboratory environment is a relatively complex one. The typical workflow for a research scientist will involve numerous steps, in which they generate a hypothesis, design a test, execute the test, interpret their results and then revise their understanding in light of the results. With the advent of modern computing it has been possible to create software tools that help with this process. However, such applications characteristically focus on only one step within the workflow. For example, an application may be designed to record data from experimental hardware, validate experimental design, analyse data or simplify the reporting of results.

Although such tools offer significant benefits, because they are introduced to tackle specific problems they cannot easily be used together. For example, the time that a lab assistant saves by using a software tool to collect data may be lost manually reformatting the results to make them compatible with a given processing tool. Since several tools may be used during an experimental workflow this can have a significant impact on productivity.

This situation, where a great deal of technology is available that cannot easily be integrated, is typical of many labs.

Further secondary issues may also complicate the process. For example, each tool will offer a unique, non standard interface that the user must learn. Furthermore, existing tools may not take proper account of the way that scientists work. The research and development process often relies heavily on collaboration and resource sharing. For example, two research groups may agree to share the cost of an expensive hardware tool that they require. Similarly, several individuals with different areas of expertise may work together to complete a project that would be impossible on an individual basis. In these cases workflows may span multiple collaborating research groups, or a single tool may need to be shared between several research groups. These issues interact to make a research lab a highly dynamic environment that often has a very complex staff infrastructure.

Providing IT resources to a lab in this fashion also makes it difficult to properly address the administration associated with scientific experimentation. With recent legislation this has become a much more significant problem. For example, laboratories are increasingly required to provide robust audit trails, to demonstrate valid quality control procedures, to secure the data that they store and to provide

their results to governing bodies in standard formats. Many current scientific software tools are incapable of fulfilling these requirements.

In summary, because software tools in the scientific workplace are not created with these complex demands in mind they often fail to address the experimental workflow as a whole. Therefore, the net benefit of these solutions is reduced.

### **Key characteristics of an integrated lab system**

A well-designed integrated laboratory system will aim to tackle the shortcomings discussed above. Its goal is to enable scientists to use multiple tools in a coordinated fashion. Furthermore, it should make this process as simple as possible for users to accomplish, so that underlying complexity is transparent. This is primarily achieved by providing a more holistic solution, focusing on the complete scientific process rather than the individual steps or tools. Fundamental to these systems' design is the idea that a typical laboratory is a non-static environment. Workflows will change, resources will be shared, researchers will move between research groups and groups themselves may collaborate. An integrated lab system is designed with flexibility and agility in mind so that it can respond easily to these changes. The solution should be able to provide a set of computing tools, normally termed resources, to a user in such a fashion that they may be transparently connected to form a complex workflow. A resource could be a software application or a hardware device. Despite providing such a high degree of flexibility, integrated lab systems must also be demonstrably robust in order to insure the integrity of the conclusions that they will support. Furthermore, integrated lab systems must often provide backward compatibility to legacy tools. This is important since it allows an organisation to utilise existing software applications within a newly deployed integrated lab system.

One of the most important features of an integrated lab system is that it will allow a user to connect resources into a workflow, irrespective of type. This means that users can easily replicate real-life processes within the system. For example, an experiment may require the retrieval of data from one hardware device, its processing via a software tool and finally use of the processed results as the input to a further hardware tool. Hardware tools in the science industry typically expose proprietary protocols for computer control. An integrated lab system can abstract such protocols to a generic communications interface. Thus software and hardware resources can be used seamlessly together.

Integrated lab systems allow users to take advantage of laboratory resources irrespective of physical location. In this regard an integrated lab system facilitates ubiquitous access to valuable laboratory resources. Consider a situation in which a research organisation is spread across multiple locations. It is often impractical or financially unfeasible to provide every type of lab resource in each location. The necessity to utilise a particular resource in another location has typically been a bottleneck in research productivity. Because integrated lab systems provide resources in a ubiquitous fashion they allow physically distinct resources to be accessed more readily. Consequently, they can alleviate the effects of such limitations. The flexibility of this feature is still more valuable if an integrated lab system is implemented atop a wireless network. Together these technologies enable scientists to move through their working environment without losing access to any of the resources that they need to do their job.

Beyond recognising that workflows within a scientific environment can be highly variable, integrated lab systems address the fact that some application requirements apply system-wide. For example, every tool might be required to record an audit trail or it may be necessary to define security permissions for groups of resources together. Such requirements are difficult to achieve in an uncoordinated environment. However, the system-wide viewpoint of an integrated lab system simplifies this goal because such features can be implemented in an aspect-oriented fashion.

It is relatively unusual that an organisation considering the introduction of an integrated lab system has no existing IT provision. Since considerable investment will have been made in these existing systems a further goal of an integrated lab system will be to offer the most cost effective upgrade path. Typically, this will involve the provision for backward compatibility with these existing resources. Because an integrated lab system is likely to be based around a modular design this process is simplified. A proxy module could be created for each existing tool (based, for example, around an adapter pattern) to represent it within an integrated lab system.

At first inspection it may be difficult to discern an integrated lab system from a lab information management system. Both act to improve the flow of information through an organisation; both aim to maximise the value that can be derived from research undertaken; and both can be made to interface with hardware devices. The key difference between these systems is in their focus within the science lab.

LIMS typically take a data-centric view of the lab ensuring the integrity of its data. Integrated lab systems instead take a functional view of the lab focusing on simplifying the research process itself. These systems may also differ in terms of structure. While LIMS typically follow a traditional client-server model, a decentralised node-based architecture is more appropriate for an integrated lab system. Finally, the coordinating role of an integrated lab system means that it will typically operate at a higher level than a LIMS. For example, an integrated lab system may mediate the connection between a LIMS and another lab system (such as a data gathering or data reporting tool) as part of a research workflow.

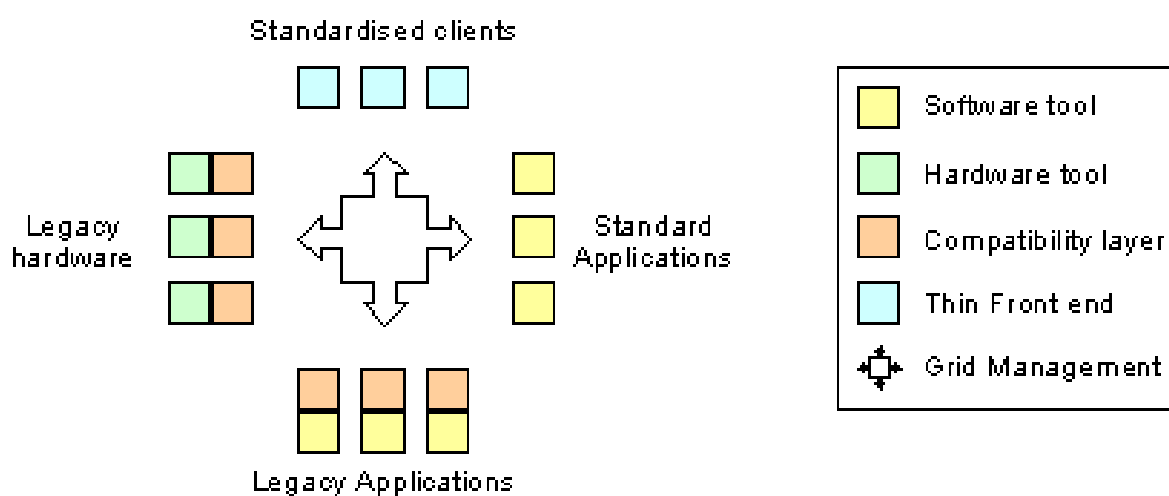
### **Integrated Lab System Design**

As discussed above, integrated lab systems must meet a rigorous set of requirements. They must be flexible, robust, ubiquitous, backward compatible and cost effective. Meeting these requirements is very challenging and a strong design forms the basis for success. In particular, because integrated lab systems must operate across several problem domains, an understanding of numerous technical areas is required. The ideal solution can vary quite significantly depending upon the industry within which it will be used. Often working environments will impose additional requirements on such a system. For example, the presence of sensitive equipment in hospitals precludes the use of standard wireless networks, therefore an alternative solution is required.

Although integrated lab systems solutions are typically problem-specific we can make some general comments about their high level design. Most importantly, the requirements placed on such systems makes them strong candidates for development using a grid architecture. Grid architectures define a high-level infrastructure within which multiple resources may be coordinated, without the need for centralised control, to deliver non-trivial levels of service. Since the goal of integrated lab systems is to provide a mechanism by which a set of tools may be used together in a coordinated fashion, such an architecture is very appropriate. In particular taking such an approach simplifies the processes of defining a workflow and describing how workflows will be constructed. Grid architectures also provide an infrastructure that understands the service-based provision of functionality and the need for service discovery mechanisms. Lastly, a grid architecture is of value since it describes an infrastructure within which system-wide functionality may be implemented. Thus, it provides an elegant solution by which features such as audit trails or resource access can be easily managed across an entire system.

Within a grid architecture we must define mechanisms for both the coordination and connection of resources. Since grids are a relatively new technology standard protocols for these tasks have not yet been determined. In the long term it may become obvious that different solutions are required depending on the problem domain. This may be further polarised within the integrated lab systems field. At present a large variety of nascent technologies occupy a share of the grid solutions market. These range from open-source solutions such as the Globus toolkit to commercial products.

Figure 1 illustrates how an integrated lab system could be designed using a grid architecture. Importantly, software resources, hardware resources and users all appear as nodes within the grid. Nodes interact in order to complete a given workflow. Common interactions would be the discovery, coordination and connection of individual nodes. The grid infrastructure provides a device independent mechanism by which nodes may communicate in order to achieve these interactions. As long as a resource can communicate in this way its underlying implementation is irrelevant – this provides the basis for the use of hardware, software and legacy devices within a single workflow. Constructing a solution in this way also maximises the ability of the system to adjust to new or unusual technologies. For example, wireless networking could be introduced to accommodate resources that must be physically movable. Likewise, a clustering platform could be introduced to support resources with demanding processing requirements.



**Figure 1 – An Integrated Lab System based around a grid architecture**

### Benefits of an Integrated Laboratory System

The benefits of introducing an integrated lab system are significant. Taking a coordinated, holistic approach to providing computing tools within a science lab will simplify their use, allowing scientists to be more effective and therefore more productive. An integrated lab system ensures that the user can access each tool in a standard way, making them easier to learn. Secondly, it will be significantly easier to use multiple tools together to achieve a goal. Such a solution allows several tools to be connected seamlessly together to form a complex workflow, enabling scientists to use existing tools in new and more efficient ways. Significantly, integrated lab systems provide value-added services that allow the description of complex processes involving logical decisions or complex behaviour. Furthermore, an integrated lab system can enable the coordinated use of resources that are present in geographically or electronically distinct locations.

Providing such a system using a modular, decentralised architecture ensures that the solution can provide the degree of flexibility required by end users while maintaining the stability and robustness necessary in a scientific environment. For example, redundancy can be introduced to protect against the failure of a resource, or proxy resources could be created to allow backward compatibility with legacy devices. Even relatively unusual requirements such as clustered

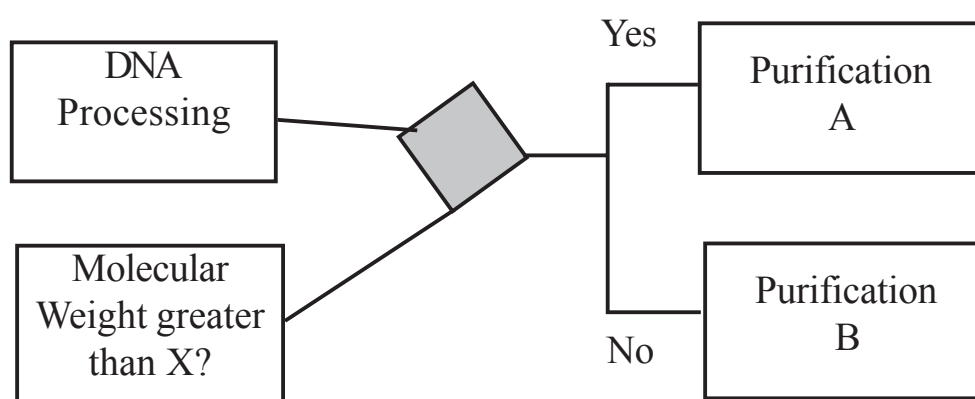
systems could be easily accommodated. Finally, aspect-oriented features such as security permissions, transaction management, productivity monitoring or audit trails can be implemented in a system wide fashion.

Perhaps most importantly, integrated lab systems deliver cost effective solutions. Because they have a modular structure initial cost can be kept low with additional functionality added as and when required. The flexibility of the system is of further value since it maximises its ability to respond to changes in the science industry. Legacy features can easily be removed and compatibility with new standards added. These characteristics increase such a system's life span and therefore its value. Consequently, the risk involved with deploying such a solution is reduced.

### **Example: Genomics Lab**

One area in which integrated lab systems will be of particular value is the biotech industry. For example, while there is a great deal of interest in genomics at present, research is often laborious due to the inadequacy of the tools available. Consequently the productivity of scientists is impeded. Necessary tools typically range from web-based search engines through standalone analysis packages to specialised hardware equipment. The inadequacy of such tools is largely due to a lack of industry standards within this nascent field. Although there are many companies creating tools for this industry they cannot be easily integrated due to incompatibilities between vendor-specific protocols and formats. This is compounded by the fact that a scientist must often use several tools in the course of their research. Consequently, research can be frustrating and time-consuming to perform.

An integrated lab system could address this problem. Take, for example, a typical scenario in which a region of DNA is sequenced and must then be characterised. The general workflow for achieving this includes multiple steps in which a scientist will retrieve sequence output, format the data and then manually process the results. Using an integrated lab system, the scientist could create an electronic workflow that defined how relevant tools should be used together to achieve this goal. The integrated lab system would provide a compatibility layer, transparently connecting each resource required during the process. Therefore the hard work of coordinating these various tools is eliminated. Furthermore, because the system could manage the process, no human input would be necessary and the goal could be achieved more quickly and with fewer errors.



**Figure 2 – Decision making workflows allow Integrated lab systems to model complex processes**

Integrated lab systems also make it possible to use multiple incompatible hardware devices together within a workflow. To continue our example, a workflow could be constructed that allowed the output from a sequencing machine to be analysed for regions of interest, and PCR primers for these regions created automatically using a DNA synthesis machine. Indeed, as the software control of hardware devices improves, the value of an integrated lab system will further increase. This could be extended to robotic control of existing hardware tools by an integrated lab system so that complete control can be provided over any type of resource.

Integrated lab systems can allow workflow definitions to be saved for subsequent re-use. Thus a process that is carried out frequently, such as DNA sequence characterisation, could be easily accomplished. Decision-making may also be introduced into a workflow, allowing the steps in a process to be determined as it executes. For example, consider a situation in which two different DNA purification techniques are available, whose suitabilities vary depending on some input condition, such as molecular weight. A workflow could be created that chooses the most appropriate purification technique at runtime, relieving the scientists from making this decision manually (figure 2). Perhaps most importantly, the ability for saved workflows to masquerade as resources within an integrated lab system allows a hierarchy of dependant workflows to be constructed. This is a powerful feature since it enables a complex workflow to be defined in terms of several simpler, low-level workflows. Features such as workflow re-use, workflow decision-making and workflow hierarchies help to provide the high level of flexibility necessary from an effective integrated laboratory system.

**Example: Pharmaceuticals**

Pharmaceutical companies can also achieve significant benefits from implementing an integrated lab system. Modern pharmaceutical research is centred to a very large extent on data. A typical lab will generate vast amounts of research data as it screens compounds that could have the potential to become commercial drugs. A fundamental part of the research process, therefore, is the storage and subsequent analysis of this data. Indeed, the key to success will often be the ability to derive the maximum possible value from the data that the lab generates. However, the activity of managing the flow of information from experimental hardware through analysis and reporting tools (and often back to experimentation tools) is relatively complex. Such a process could be significantly improved by deploying an integrated lab system.

The flow of data through a pharmaceutical lab can be broadly organised into the following steps

- Data capture
- Data storage
- Data processing
- Data analysis
- Analysis reporting

Traditionally data management in pharmaceutical labs has been achieved using a LIMS. As information management requirements have increased over time they have typically been accommodated by adding more and more functionality to existing LIMS. For example, CSCW (computer supported collaborative working) functionality may be required to enable collaborative use of information resources. Alternatively, data mining techniques have become popular, since they allow research organisations to elucidate new relationships within their results. Most recently, HTML front ends have been added to provide web-based information management capabilities. However, evidence suggests that this convergent method of responding to changing user needs is less than ideal. Typically, as more and more functionality is added to an application it will become less stable, less flexible and more difficult to support.

The implementation of an integrated lab system provides an alternative solution by which these needs may be addressed. Importantly the modular architecture of an integrated lab system provides superior flexibility with respect to a standalone LIMS. Rather than adding additional functionality piecemeal into a single application research labs can choose the ideal tool for each step in the data

management process. The integrated lab system provides a mechanism by which these tools may be co-ordinated, in a seamless fashion, to overcome problems of incompatibility. There is also the potential to make integrated lab systems more stable than LIMS since advanced features such as functional redundancy and resource-specific access permissions may be implemented. Perhaps most importantly, systems support can be more rigorous under an integrated lab system. Resources are functionally independent so the impact of one tool's failure is kept to a minimum. Furthermore, rather than having to support one huge application, an integrated lab system can be supported by several 'domain experts' each of whom has specialist knowledge of one or more resources.

Integrated lab systems provide a framework, within which a broad variety of information management tools may be co-ordinated to maximise the value derived from research data. Importantly, no assumptions are made about the type of data or the way in which it will be used. Instead the goal of an integrated lab system is to enable the seamless integration of data management resources to best afford the research process. A typical data management workflow constructed within an integrated lab system is shown in figure 3. The flow of information from data capture to analysis reporting takes place from right to left. However, integrated lab systems can enable an advanced set of behaviours. In this example, the workflow has been constructed so that inferences from a decision support system feed back into a high-throughput screening tool. This allows the results obtained at the initial stages of research to be used to potentially improve the quality of subsequent results.

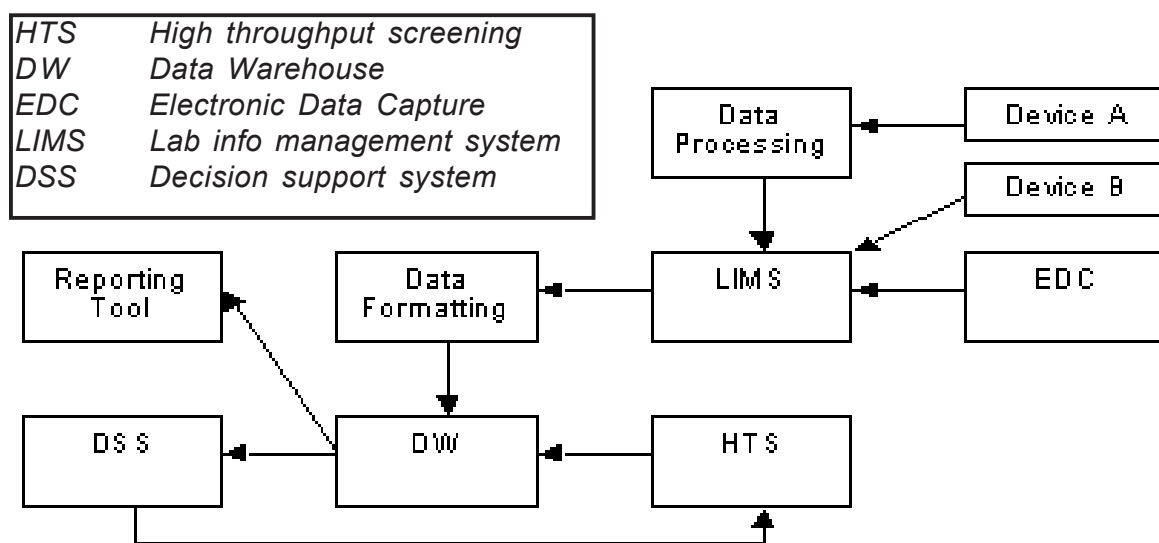


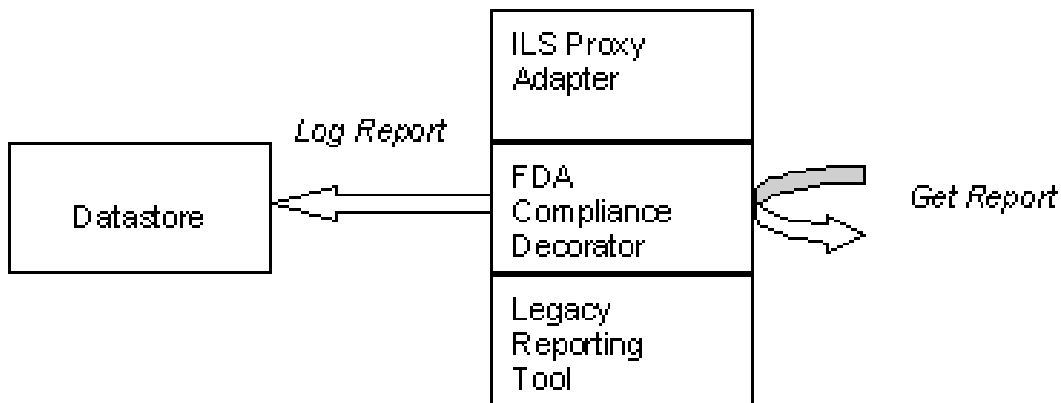
Figure 3 – A typical data management workflow

**Example: FDA Rule 21 CFR Part 11 compliance**

Electronic documents are increasingly replacing paper documents in all industry sectors, where organisations see an opportunity to store and share information far more productively than was previously possible. This is particularly the case within the science industry, where information is such a valuable commodity. As the move from paper based to electronic records has become more widespread the U.S. Food and Drug Administration (FDA) has issued a set of regulations that describes the minimum requirements for electronic records systems. These requirements are described in FDA Rule 21 CFR part 11. In broad terms the requirements cover three key areas:

- ❑ Demonstrating record security – ensure that data can only accessed and/or changed by authorised persons
- ❑ Demonstrating record integrity – ensure that the data itself has not been inappropriately changed and that it is presented in a consistent manner (i.e. the interpretation of the data cannot be inappropriately changed)
- ❑ Demonstrating record authenticity – provide a mechanism by which the source of electronic records may be validated

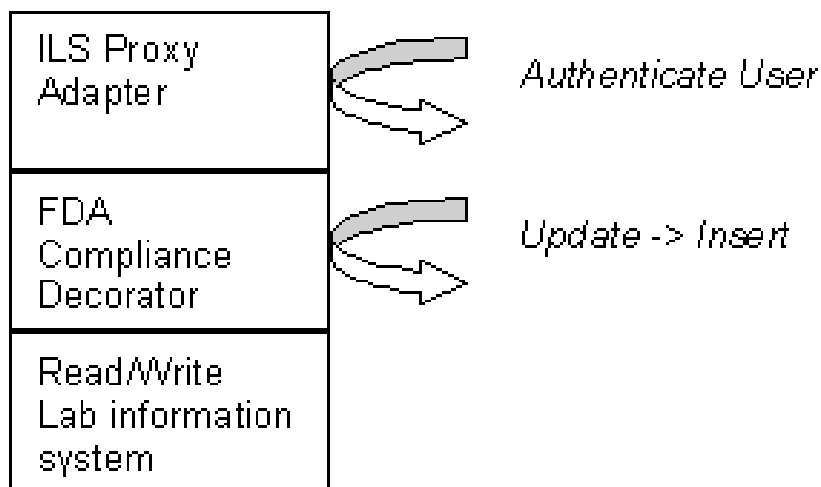
Many currently available research tools do not yet provide native support for the FDA guidelines. This may be for a number of reasons, for example the tool vendor has not yet released a compatible version or the investment required to alter a bespoke tool is too great. In fact compliance often requires changes to an organisation's information process as a whole. Because integrated lab systems take a process-centric view of research they are well placed to support this. Integrated lab systems can provide a well-structured framework within which FDA compliance may be provided to these 'legacy' applications. The provision of such functionality can often be achieved using a decorator pattern. Using such an approach, a component that provides the missing functionality is created to 'wrap' the legacy resource. When messages are sent to the resource from a workflow the decorator intercepts them and the relevant tasks are carried out to maintain FDA compliance. Two examples are shown in the following figures (4 and 5).



**Figure 4 – Providing document logging**

In figure 4 (above), FDA compliance is provided to a reporting tool. Requirements specify that changes to documents over time should be represented as separate documents that show the changes at each stage. If a reporting tool generates documents dynamically the contents of the document will change with the data. Here whenever a copy of the document is dynamically created the FDA compliance decorator logs a static copy in a datastore. This history can later be accessed and thus changes at each stage are recorded.

In figure 5 (below), FDA compliance is provided for a legacy LIMS. The present tool allows existing information to be easily updated by users. However, FDA requirements state that a proper audit trail should be maintained describing every time that data is changed. This can be achieved using a compliance decorator. The decorator intercepts update messages and maps them to insert messages that create a new version of the data. This provides a mechanism by which the existing record is preserved – insuring the necessary audit trail is recorded.



*Figure 5 – Providing audit trailing to a legacy LIMS*

### **Real World or Blue Skies?**

Although the development of an integrated lab system is an ambitious project these solutions can be created now. Perhaps more importantly such applications can be implemented both reliably and cost effectively. The key reason for this is that such systems can be constructed using robust, industry proven, well supported components. Consequently, the major cost of developing an integrated lab system is limited to design and implementation time and not the development of new, untested technologies.

This conclusion is supported by the great many projects already underway. For example, IBM recently demonstrated a ‘web services for the life sciences’ application on their alphaworks website. The demonstration wrapped three genetics tools into distinct web services and showed how they could be used together to provide additional value. Secondly, Manchester University has an active research project creating a virtual lab workbench for life scientists, based on a grid architecture. This prototype highlights the value of integrated lab systems as a whole but in particular illustrates the benefits in productivity gained from a solution that is created with the research and development process in mind.

Tessella have extensive experience of lab systems development. We have implemented bespoke LIMS, electronic data capture, decision support and hardware control systems. We are also able to provide professional consultation on systems integration, FDA compliance and grid architectures / distributed computing. For further information on the benefits of implementing an integrated lab system within your organisation contact Tessella by e-mail at [info@tessella.com](mailto:info@tessella.com) or by telephone on +44 (0) 1235 555511.

**Tessella Support Services plc**  
**Creating Software for Science and Engineering**

Tessella's services range from feasibility studies, through system design, development, implementation and ongoing support. Our expertise includes:

Data Analysis Software  
Data Capture  
Simulation Software  
Advanced Graphics  
Systems Support  
Database Applications

**Other Technical Supplements available include:**

- |   |  |
|---|--|
| <input type="checkbox"/> Archiving of Electronic Info   | <input type="checkbox"/> Object Oriented Programming   |
| <input type="checkbox"/> Active Server Pages            | <input type="checkbox"/> Pocket PC                     |
| <input type="checkbox"/> Automated GUI Testing          | <input type="checkbox"/> Portable GUI Development      |
| <input type="checkbox"/> Bayesian Statistics            | <input type="checkbox"/> Printer Technology Guide      |
| <input type="checkbox"/> Beowulf Clusters               | <input type="checkbox"/> Real Time Systems             |
| <input type="checkbox"/> C++                            | <input type="checkbox"/> Regression Testing            |
| <input type="checkbox"/> Client-Server Technology       | <input type="checkbox"/> Security and the Internet     |
| <input type="checkbox"/> COM                            | <input type="checkbox"/> Simulation                    |
| <input type="checkbox"/> Computational Fluid Dynamics   | <input type="checkbox"/> Soft Computing                |
| <input type="checkbox"/> Computer Image Processing      | <input type="checkbox"/> Software Design Methodologies |
| <input type="checkbox"/> Decision Support Systems       | <input type="checkbox"/> Software Development Cycle    |
| <input type="checkbox"/> Electronic Data Capture        | <input type="checkbox"/> Software Documentation        |
| <input type="checkbox"/> Electronic Lab Notebooks       | <input type="checkbox"/> Software Portability          |
| <input type="checkbox"/> Excel                          | <input type="checkbox"/> Software Re-engineering       |
| <input type="checkbox"/> Extending the Life of Software | <input type="checkbox"/> Software Specification        |
| <input type="checkbox"/> Federal Drug Administration    | <input type="checkbox"/> SQL                           |
| <input type="checkbox"/> FORTRAN 90                     | <input type="checkbox"/> UNIX Inter-Process Comms      |
| <input type="checkbox"/> Grid Computing                 | <input type="checkbox"/> UNIX Systems Performance      |
| <input type="checkbox"/> High Throughput Screening      | <input type="checkbox"/> UNIX Workstations             |
| <input type="checkbox"/> Instrumentation                | <input type="checkbox"/> Visual Basic 6                |
| <input type="checkbox"/> Integrated Lab Systems         | <input type="checkbox"/> WAP                           |
| <input type="checkbox"/> J2EE                           | <input type="checkbox"/> Web Services                  |
| <input type="checkbox"/> Java                           | <input type="checkbox"/> Windows 2000 Services         |
| <input type="checkbox"/> Lims                           | <input type="checkbox"/> XML                           |
| <input type="checkbox"/> Linux                          | <input type="checkbox"/> X Windows                     |
| <input type="checkbox"/> Microsoft Net                  |  |



INVESTOR IN PEOPLE

**Tessella Support Services plc**

3 Vineyard Chambers, Abingdon, Oxon, OX14 3PX, England

Tel: (+44) (0) 1235 555511 Fax: (+44) (0) 1235 553301

E-mail: [info@tessella.com](mailto:info@tessella.com) Web Address: <http://www.tessella.com>