A Framework Design for Information Management in Heritage Science Laboratories

RUI BORDALO, Universidade Católica Portuguesa, School of Arts, Research Center for the Science and Technology of the Arts, Portugal, and HERCULES Laboratory, University of Évora, Portugal

CARLO BOTTAINI and ANTÓNIO CANDEIAS, HERCULES Laboratory, University of Évora, Portugal

The dataflow in any scientific research laboratory is continuous and considerable even in analytical niches such as heritage science laboratories. This article discusses advantages of using a Laboratory Information Management System (LIMS) for organising and systematising research in an interdisciplinary analytical laboratory. LIMS solutions are proven to be effective in managing laboratories, organising their daily management tasks, improving work conditions and increasing productivity, managing equipment and resources, and managing and safeguarding scientific data. LIMS software is commercially available since the 1990s with over a hundred software packages intended either for general or specific purposes. However, the ones currently available do not fulfil all the requirements for heritage science or are over-the-top, complex, one-fits-all standard solutions. The system here proposed was developed for a heritage science laboratory, considering its unique requirements, having as a case study the HERCULES Lab, from the University of Évora (Portugal). This article documents our approach based on the analysis of the unique requirements for the lab resulting in a proposal of a custom user-centred web-based system. For this, we discuss in detail the workflow, and the required system architecture. We conclude that although it is a small niche market for major manufacturers to dwell on, a custom LIMS is of the utmost importance for the current management of heritage science laboratories across the globe.

CCS Concepts: • **Information systems** → *Information systems applications*; *Enterprise information systems*; *Enterprise resource planning*;

Additional Key Words and Phrases: LIMS, laboratory management, web technology, knowledge management, scientific data management

1 INTRODUCTION

The study of cultural heritage, despite its implicit societal importance, has never been a driving force for the development of new technology but rather its recipient. This has been true for most of the technological advances

that we have witnessed in the past few decades, such as chemical and physical analysis, imaging or 3D scanning among many other hundreds technological applications. With the need to study art and cultural heritage objects in general, there has been a proliferation of heritage science laboratories, usually attached to academic or governmental institutions. Heritage science is a broad interdisciplinary field that combines humanities and sciences, often using analytical techniques to study not only the material composition of objects but to extract technical information to enable important wide-ranging studies such as their degradation and conservation, establishing material sources, dating, or the artist's technique, contributing thus to an improved interpretation of the object. Even though it is a niche area, as an interdisciplinary field these laboratories combine equipment from a large number of areas, most notably from analytical chemistry, physics, engineering, material sciences, or biotechnology. Nowadays, some of these laboratories easily comprise a large number of instruments with a scientific output of data and image files of dozens different formats [Ailamaki et al. 2010], which has created a challenge to manage [Shoshani and Rotem 2009; Marcum and George 2010].

Naisbitt [1982] is widely paraphrased as having said that "we are drowning in data but starved for information". Laboratories are research facilities whose ultimate objective is to produce knowledge and are very dependent from the data-information-knowledge stream [Ackoff 1989]. This has been a problem felt early on big laboratories and so software solutions were developed [Skobelev et al. 2011]. Working independently or in networks, such as European infrastructures [Pallot-Frossard 2016], there are numerous heritage science laboratories across the globe, either focused in specific or wide-ranging materials, each one with different needs but with similar requirements.

Until some time ago, small heritage science laboratories did not have the problem of data management since each researcher used to manage its own research, samples, notes, data outputs, and archives, independently of the analytical techniques used. However, beyond its qualitative attributes, artworks are an immense source of quantitative data. As with any other field, laboratories are faced with the exponential growth of data not only from the analysis of samples but also from the research output [Shoshani and Rotem 2009; Marcum and George 2010], adding complexity levels not always easy to manage.

In order to prevent the laboratory from reaching a breaking point through the loss of resources, money, and time, often hindering research, thesis, and knowledge in general, a promising solution is the use of dedicated management software. Laboratory information management systems [Moore 2001] are software applications that were developed within the life sciences, mostly for the pharmaceutical industry assisting in the development of drugs and other products and in quality control issues. This article proposes a custom-based information management system based on the needs of a typical heritage science laboratory such as is the case of the HERCULES Lab, a research centre at the University of Évora (Portugal) with a staff of 30 PhDs. The article focuses necessarily on the laboratory workflow but acknowledges that it is part of a wider workflow that starts from the heritage object and works towards its conservation, protection and improved interpretation. The implementation of such system will increase the efficiency and quality of the work by establishing norms, standard workflows, and procedures; promoting quality control and teamwork; and assisting project management tasks.

2 LABORATORY INFORMATION SYSTEMS

Under the umbrella term "laboratory information management systems" [Pantanowitz et al. 2007; Moore 2001; Wood 2007], interdisciplinary software solutions have been merging applications that help managing scientific data in the laboratory such as Laboratory Information System (LIS), Laboratory Information Management System (LIMS), Electronic Laboratory Notebook (ELN) [Machina and Wild 2013], and Scientific Data Management System (SDMS) [Barisits et al. 2019]. Historically, the two major types are LIS and LIMS, frequently used interchangeably and often mixed up. In the absence of a hard definition, their distinction relies in the fact that LISs are used to manage clinical and forensic testing in research laboratories, such as hospital settings, where the results are related to people rather than samples (patient-centric), while LIMSs are used in industrial laboratories where results are oriented for batches and individual samples (sample-centric), such as in the pharmaceutical

or chemical manufacturing industries. Nevertheless, this distinction was once clearer since, in order to serve many markets, these systems are currently so full of crossover features that their differences are increasingly unclear and thus merged under the term LIMS. As for ELNs and SDMSs, as the names suggest, the former is a digital version of the traditional researchers' notebook [Voegele et al. 2013], gathering several data related to experiments, analyses, samples, and collaborators, while the latter is a general document management system to capture, catalogue, and archive data by a particular instrument.

Commercially available since the 1980s, nowadays LIMS can be found across fields and have been constantly in high demand [Skobelev et al. 2011]. The current demand for LIMS is mainly driven by the competitive life sciences sector, largely by its biotechnological, pharmaceutical, and healthcare industries, followed by Energy, Petrochemical, Biotechnology, Geology, and Cosmetics [Thurow et al. 2004; Pantanowitz et al. 2007]. All these areas often demand high volume and high-tech investment such as automation of analyses. Nevertheless, there are other fields that also require a data management system but do not require, or even allow, such a high degree of complexity and investment. LIMS, as with any other digital solution, has been following the current software trends and can be found as full web, cloud or SaaS (Software as a Service). Without question, the future points in the direction of LIMS based on Internet of Things [Ma et al. 2017] even though, at the moment, the current scientific laboratory equipment is largely not yet prepared for that end.

LIMS's main objectives may be summarised as follows: to centralise all the data/information/workflow of a laboratory; to provide document management; to ensure the continuity of the samples throughout the workflow, from receiving them to their archival; and to facilitate the quality control of all the samples and projects.

Usually, three types of LIMS solutions can be found in the market, each one having specific advantages and limits that can be summarised as follows:

(1) *Fully Customisable Solution* – Since features are adaptable to the needs of each laboratory, the implementation and configuration of this kind of system to a specific client takes much longer than with other systems;

(2) *General All-Purpose Type of Solution* – Due to the very large number of features offered as standard, with limited or no tailor-made configuration to the client, the client must adapt or compromise their work to the solution offered;

(3) Custom-made solution - often also built in-house, to a specific project or institution, it may be actually a cheaper solution in the short end although in most cases the cost of its future maintenance and support is not immediately considered.

The natural choice for a LIMS in an academic or cultural institution is open source in order to be economically efficient and to be adaptive for the laboratory's particular needs. Unfortunately, there are no open source projects in this field that the authors could find. A search for LIMS and LIS in sourceforge.net and github.com, both open source software repositories, returned over 30 projects, most of them now discontinued. Table 1 summarises the projects that are still active or that were updated within the last five years. The majority of them were made for bio-informatics and healthcare sectors and only a few had stable releases. Since these options are hardly useful for a heritage science laboratory, we propose that the best option for a LIMS for this particular application is a custom-made solution, described in the following sections.

3 THE HERCULES LAB

The HERCULES Lab is a research centre from the University of Évora. Established in 2009, it focuses on the study of cultural heritage through the integration of physical and material sciences methodologies and tools in interdisciplinary approaches. The lab has four main lines of research: (a) Archaeometric approaches to Past Cultures, (b) Science for the Arts, (c) Science for Heritage Conservation, and (d) New materials and tools for Cultural Heritage. To achieve this, it is formed by several labs with state-of-art equipment with the capability to develop

Table 1. Open Source LIMS Found in Sourceforge.net and Github.com, Active or Last Updated Within the Last Five Years, Organised Chronologically by Last Update and Life Span

Product	Area	Licence	Production life	Application	Language	Database	Application Field
MISO	Biomedicine	GPL v3	2012–2020	web	Java	MySQL	Life Science
GNU LIMS (occhiolino)	Healthcare		2013–2020	server	Python 3	n.a.	Health
OpenSpecimen	Bio-Informatics	BSD 3-Clause	2014–2020	cloud	Java	MySQL	Life Science
Leaf LIMS	Generic	MIT	2017–2020	web	Python	PostgreSQL	Generic
SENAITE	Generic	GPL v2	2017–2020		Python (Zope/Plone CMS)	n.a.	
iSkyLIMS	Bio-Informatics	GPL v3	2018–2020		Python	MySQL	Life Science
LabKey Server		ASL	2007–2019		Java	PostgreSQL	
C4G BLIS	Healthcare	Unknown	2010–2019		PHP	n.a.	Health
OpenELIS		MPL v1.1	2012–2019		Java	PostgreSQL	
Baobab LIMS	Biobank	GPL v3	2016–2019	server	Python (Zope/Plone CMS)	ZODB	Life Science
Kalenis	Generic		2017–2019	web	Python	PostgreSQL	Generic
openBIS	Bio-Informatics	ASL v2.0	2007–2018	server	Java		Life Science
MetaLIMS	Metagenomic research	GPL v3	2015–2018		PHP	MySQL	
Screensaver HTS LIMS	Bio-Informatics	GPL v2	2009–2017	web	Java	PostgreSQL	
Clover	Plant biology	GPL	2013–2017		PHP, JavaScript	MySQL	
ClinViro	Bio-Informatics	GPL v3	2017–2017		JavaScript	PostgreSQL	
GNomEx	Genomic Repository		2010–2016		Java, Flex	MySQL	
OpenLab Framework	Bio-Informatics		2010–2016		JavaScript, XML		
ISMeralDa	Biomedicine	GPL v2	2009–2016		C++		
MASTR-MS	Biomedicine	GPL v3	2009–2016		Python	PostgreSQL	
CLIMS		GPL v2	2006–2015		Java	MySQL	
FreeLIMS	Bio-Informatics		2007–2015	cloud	JavaScript, JSP, Java		
LIMS.NET		Artistic License	2009–2015	web	Visual Basic .NET, ASP.NET	Visual Basic .NET, ASP.NET	
Open-LIMS	Biology	GPL v3	2010–2015		PHP	PostgreSQL	
SMITH	Bio-Informatics	MIT	2011–2015	server	Java	MySQL	
Bika LIMS	Healthcare	GPL v3	–2014, active?	web	Python (Zope/Plone CMS)	SQL	Health
adLIMS	Bio-Informatics	GPL v2	2014–2014		Java	PostgreSQL	Life Science

Note: ASL: Apache Software License; BDS: Berkeley Software Distribution; GPL: GNU General Public License; MPL: Mozilla Public License.

ground-breaking research, comprising in-situ non-destructive analysis and imaging, laboratory microanalysis, and high-resolution chemical analysis.

The analytical infrastructure has over 300m^2, organised in five laboratories: Microscopy and Microanalysis Laboratory, Chromatography Laboratory, Mass Spectrometry Laboratory, Materials Characterisation Laboratory, and Biodegradation and Biotechnology Laboratory. Plus, since HERCULES has research projects all over the country, it has a mobile unit, unique in Portugal, with equipment for in-situ analysis. At any time, HERCULES has a community of around 30 to 50 people at different stages of their career, including staff researchers, students, interns, PhD candidates and Postdocs, and technicians.

While its initial facilities were small, the laboratory has grown massively with the acquisition of new instruments each year. However, the complexity of managing the infrastructure has also grown and it is representative of the main issue any laboratory faces. Concerning its management, facing the steady increment of equipment and researchers, the laboratory moved from equipment' article-based schedules to a free web-based general laboratory management service that, unfortunately, on an update of their website, discontinued the equipment scheduling feature. Since it was the only feature that was of interest to the laboratory, the service is no longer used. A problem that is worth mentioning is that the web-based system had some user resistance due to lack of an intuitive user interface. From there, the natural option was to move on to Dropbox, using a spreadsheet for each equipment, where a reservation can be made directly by the researcher at any time. However, this system relies on good practices since the reservations are not always made in time, especially with portable or less used equipment. The system works as a first-come, first-served but in the case of swaps or urgent tasks, arrangements are done directly between the researchers. The only exceptions are SEM-EDX and XRD since they have the highest demand. In order to accommodate all the requests, the lab's supervisor considers the urgency of the analysis, objectives, time slot requested, week or days preferences, and other limitations to prepare each month's schedule in a spreadsheet.

From an individual standpoint, as in any other laboratory, all researchers are responsible for their own research, managing their own samples, from sample preparation to sample archival, access to the equipment, use of each instrument, interpretation of results, and preparation of manuscripts and publication. This, however, presents some practical disadvantages, both from the equipment and organisational approach. Concerning the equipment, almost every instrument's software, whether bench or portable, works only with Microsoft Windows *operating system (OS)*. This implies that some software only works with a certain OS version, which is a major disadvantage, meaning that each computer has its own ecosystem and that they have to be kept offline to minimise security risks.

Concerning the work's organisational approach, since all researchers manage their own samples, there is no unified system for naming and cataloguing samples; there are no samples' archival standards and no organised or mandatory archive; there is no priority level assigned to each scheduling task, making it arduous to reconfigure the schedule in case of urgent situations and scheduling incompatibilities that may appear, for example, in case of unavoidable last-minute appointments; with few exceptions, every result is saved and kept in each equipment's computer, although there is no systematic mandatory periodic backup; and all researchers are responsible to save, backup, and manage their own instrument results. Although so far there have been no issues, such as missing files or corrupted hard drives, some of these aspects may present serious potential problems in the long run that a unifying system such as a LIMSs could easily solve.

4 WORKFLOW AND REQUIREMENTS

4.1 Laboratory Workflow

The usual workflow at the HERCULES Lab is generally the same as any other laboratory dedicated to heritage science research. After a previous assessment of the study objectives and the corresponding selection of which analytical techniques are required for the study (such as imaging, molecular, or elemental characterisation), the
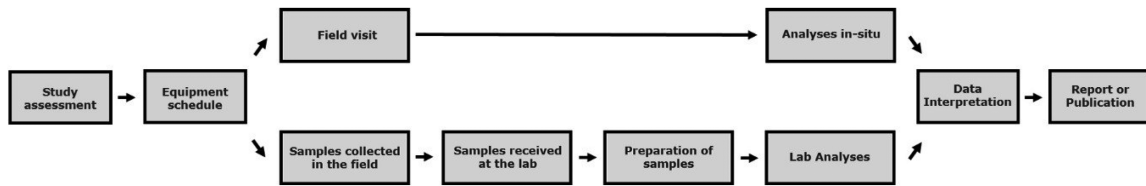
Fig. 1. Heritage Science laboratory workflow diagram.

researcher schedules the use of each equipment at the lab and a visit to the location of the work of art. If needed, the mobile unit and portable equipment, as required, will be used as well.

The practical side of the work usually starts, as shown in Figure 1, with a first examination of the object to be studied, either at the laboratory or more frequently in its location such as a museum, church, conservation workshop, archaeological field, or any other cultural site. The researcher then takes into consideration the objective of the study and considers whether it is possible to perform *in-situ* non-destructive analyses to the object or work of art or if further investigation with laboratorial equipment is required. If so, the researcher, whenever possible and necessary, will collect micro-samples after performing an *in-situ* non-invasive analysis and interpreting the data.

Back at the laboratory, also if necessary, before their preparation, the samples are further divided into smaller samples in order to be analysed by different techniques. Having previously scheduled the necessary equipment, the researcher then proceeds to use it during its allocated time. The researcher is responsible to organise and save all the necessary results from each equipment. Usually, each instrument's computer contains a folder for each researcher and project. Later on, the results of each technique will be interpreted and cross-referenced with results from other techniques and/or colleagues if it is the case, and finally published in the form of a report or a scientific article. This broad workflow is usually followed by all the researchers (including master students, Ph.D candidates, and Postdocs) and technicians working at the laboratories.

4.2 Laboratory Management System Workflow

Based on the practical workflow currently used by the researchers, the workflow using an idealised laboratory information management system is presented here in order to better identify the system requirements from the end-user.

Either in a small study of a single object or in a multi-year publicly funded research project, any researcher may initiate a new study or research project, either allocating other researchers and equipment to do it or not. Preapproval of each project or of access to each piece of equipment by the lab's supervisor or system manager may be considered a logical management step. However, from the end user and a practical point of view, it is deemed unnecessary and may pose more practical drawbacks than managerial benefits. The system should be as direct and easy to use as possible, cutting unnecessary steps whenever possible.

In order to add researchers to the project, it is a precondition that every researcher has a system profile and access to it. The addition of another person to a project will give him or her access to the project profile. For security reasons, the system may contain limitations of interaction with the project management and with limited access to project sections or actions, either by establishing it in each case or with a pre-set access by categories such as PI, co-PI, researcher, intern, and technician. Furthermore, a project may contain other projects or a project may be linked afterwards to a parent project, as in the case of research projects that are lengthy and involve several other researchers for particular tasks. Even though direct research focuses on samples, the system is project-centric, whereas a project can be a work of art, ensemble, or site. This would allow the proper contextualization of the samples, including linking the many analysis that can be made with the same sample as well as analytical work previously performed to the work of art.

Once the project profile is open and filled with information such as description, historic site or work of art, source of the samples, past technical and analytical data if available, funding, supervisor or responsible researcher, among many other types of relevant information, the system will allow access to the schedule module to calendarise the use of equipment. This will include the date and hour of use for equipment at the lab and location, in the case of the mobile unit, to account for the distance between places and the estimation of time of use and cost. This section may include a priority function for urgent jobs and rescheduling with other researchers. All the equipment, including portable kind, needs to be registered. This process should be designed to take a couple of minutes at most. Then, collected samples are registered, numbered, and documented. The samples' code number may be automated or personalised; in this case, the system will account for other similar codes and prevent its use. Each collected sample may be divided and resampled. In the end, the remaining samples will be properly stored at the laboratory archive for future reference. This methodology around the samples and their analysis emulates the functions of a basic ELN system, helping the researcher to manage his or her work.

Given that each instrument software has its own ecosystem, namely working on a different OS and originating several file formats, during each analysis, the researcher may access the system and upload automatically all the files corresponding to each analysis, including the analysis parameters, in case the equipment does not provide this information automatically. In this case, the system will recognise them from the uploaded files. Each equipment or technique may be easily tailored in order to read and present information without the need to use special external software. This will be the case with generic spectral files which will be read automatically and may allow basic control of the spectral data. The researcher may add notes, information, and interpretation of the results. All data will be archived automatically on a central server with remote access to every researcher.

4.3 System Requirements

The system aims to benefit directly the researcher by facilitating the entire process from recording all the experiments and simplifying the organisation and curation of digital resources to the publication of results. To this end, and based on the previous workflow, a number of essential system requirements is summarised in Table 2. Also, it should be highlighted that, even if custom-made, LIMS should implement quality management standards such as the commonly used ISO 9001 [ISO 2008; Silbermann et al. 2011; Hyks and Kolis 2014], a standard for general quality management, ISO 17025 (for testing and calibration laboratories), ISO 15189 (for medical and diagnostic laboratories but applicable to others), and ISO 27001 (information security standard) [Watson and Jones 2013]. On the same note, the system should fully embrace the Open Science paradigm [Burgelman et al. 2019].

5 SYSTEM ARCHITECTURE

A traditional LIMS consists of an application or software package that accesses a database for information. LIMSs uses a client/server model where data is entered and processed by client software on remote machines and stored in a database. In general terms, the system is designed so that data is entered either manually or automatically, organised, analysed, and finally reported. Based on the previously described lab's workflow and the user experience and requirements, a framework was designed to aid the laboratory's management. The system is intended as open source, given its advantages, and as a modular system. Figure 2 shows the main processes for the proposed model. The system presented here contains several modules that assist management, including sample tracking, workload assignment, data organisation and visualisation, calendarisation, standard automated report generation, and archival functions.

The analytical equipment, from any spectroscopic technique to a microscope, is connected to the lab's desktop computers or laptops, as shown in Figure 2, installed with the vendor's proprietary software. These same computers can also access the LIMS's web-based interface and the user can upload the analysis files, with proprietary and/or open-file formats, to the server. Since each computer has a different software ecosystem, with diverse operating systems and versions, drivers, and types of analytical proprietary software, the direct upload

Table 2. An Overview of the System Requirements for a LIMSs

Requirements	Description
General	
Intuitive and user-friendly interface	The system should not be designed, based on managerial or bureaucratic perspectives.
Search capability	Within the system and across readable file formats.
Drag-and-Drop file uploader	A feature to improve interaction and avoid time consuming tasks.
Link to external sources	It should be able to link to external sources such as references and other services.
Reports	Ready-made templates for generating reports in pdf and text processing file formats.
File and image manipulation	It should allow creating and editing text, spreadsheets, and images.
File management	To organise and keep track of files.
Data Analytics	Feature with basic capabilities to provide overview over the workflow and projects.
Audit Logs	For security and compliance.
Standard compliance	ISO and Open Science/Open Data compliance
Core	
Operating System compatibility	Windows, Macintosh, Linux, Android, iOS.
Support multiple DBMS	Either SQL or noSQL.
Single sign-on (SSO)	Access to the system should be made through a secure individual credentials.
Multiple user access	It should allow simultaneous access to the system.
Multiple windows	It should support multiple windows open simultaneously.
Stable URLs or persistent identifiers	To prevent failure to access information and links.
Metadata	This is a requirement for versioning and other features.
Data migration	For incoming data from other systems or towards upload on other systems.
Hosting and Storage	
Cloud / Local / Hybrid	To be considered according to the university infrastructure, cost, and security concerns.
Storage capacity per user	Ideally unlimited, within reason.
Versioning	All entries are recorded and time-stamped for audit purposes.
Data protection	Real-time protection (file redundancy) and backups.
Recovery options	Trashed items can be recovered.
Other	
Documentation	May include FAQs, Wiki, instructions.
Training	May include short demonstration videos for new users.

of the files to the LIMSs avoids any potential compatibility issues and customisation costs. Having previously set the project profile and samples to analyse, the direct upload of selected files enables each user to manage the research per sample. The server can be located on a closed local network or cloud-based, subject to cost choices and infrastructure of the laboratory. Finally, the LIMS has several modules and features, such as equipment scheduling and sample tracker, to help research and information management, ultimately used to produce research outputs. Given that the laboratory may provide external services, an invoicing and accounting external system may also be linked to the LIMS.

5.1 Web-Based System

Commercial LIMS with internet interfaces have been available since the early-2000s [Thurow et al. 2004]. A web-based system [Zhang 2014], which is a client-server application that runs on browsers, offers the most advantages
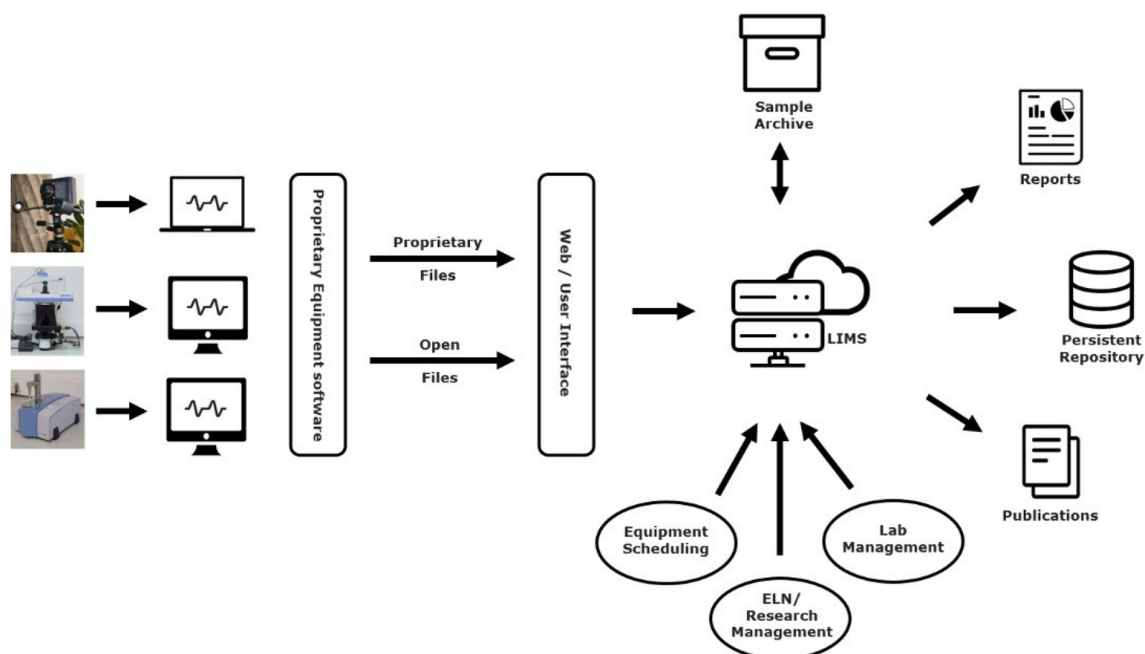
Fig. 2. Overview of the main processes of the proposed LIMS model.

for this system because: (a) is platform independent and works well with heterogenous environments, avoiding the problem of several individual software ecosystems; (b) uses existing infrastructure; and (c) can be accessed remotely, which is crucial given that heritage science research often involves work developed *in situ* with portable equipment, potentially at any location in the world.

5.2 Security, Data Storage, and Backup

Even though cloud-based systems are web-based, some questions concerning privacy and confidentiality issues may arise. The system requires a multi-user and multi-experience level to allow several users to access independently the system at the same time over a secure network. Access to the system is granted through different pre-established access levels (administrator, manager, researcher, technician, assistant). Given that it is a small system, user accounts are set up by the system administrator, depending on each specific case. Projects, datasets, or samples may be set to public or private to other researchers, also due to privacy and confidentiality issues, although they are open to the system administrators. On the subject of security and data permanence, data migration should also be factored in, both to import or to export between systems.

The system should be safeguarded against unauthorised access or accidental deletion of data. A great data management system is of no value if it does not prevent data loss through redundant data storage and automated backups. Redundant data storage may be accomplished through redundant array of disks (RAID), where the data is mirrored and distributed in several independent disks, avoiding thus loss of data, preventing data corruption or disk failure. Periodic automatic backups are important for consistency and reliability of the backups and can be easily setup automatically. Off-site periodic backups may also be established. Redundancy will also be applied to the power supply.

Storage cost is always a decision factor for these systems. However, the price per Gb has been dropping in the last decade and is now in the range of 10 cents, whether for single SSD [Harris 2020] or cloud-based services such as Amazon [Amazon Web Services 2020]. A note should be made since, in these commercial services, costs

may vary according to the overall volume storage space rented per month and number of hours per day of use, with costs billed in per-second increments.

5.3 Files

The files generated by the instrument's proprietary software are often in a proprietary format. Some software, however, do offer the possibility to save the data in open formats such as the standard comma-separated value (CSV). In case this is not possible, parsing scripts may be set up to each equipment software so that, when saving the proprietary format, an open format will also be saved. In the future, this will allow the same data to be opened by multiple compatible external software for data processing. Special attention will be taken for character encoding. The output files, such as the final versions of reports, should be exported and saved as immutable format, such as the portable document format (PDF).

From spectral files, usually of a very small file size, to imaging files, typically up to some dozen Mb, the system should be capable of handling a large number of files. The exception would be individual files of more than 100 Mb, such as some lossless raster images or 3D files.

5.4 System Components and Modules

5.4.1 Research Management. This key component comprises different managing tasks for each access level. For the system administrator, it allows management of the laboratory on a daily basis and project management. This may include data analytics for the use of equipment, calculation of costs, need to order consumables or other materials, and a summary of activities performed by each researcher, for example. For the typical researcher, it allows essential project management functions, such as management of time, resources, and deliverables.

5.4.2 Equipment Management. This consists of a simple module where each equipment is listed, together with its respective accessories and consumables. Linked to the lab equipment scheduler, it registers the number of hours in use per day, week, and month, and, when applicable, for time-dependent consumables, sends a warning when a new module should be ordered or when a module should be replaced.

5.4.3 Lab Equipment Scheduler. Equipment scheduling is very important for meeting deadlines, keeping equipment downtime to a minimum, knowing the whereabouts of portable equipment, and optimisation of the researchers' time and productivity. Thus, it is a vital module of the entire management system. This module consists of a calendar, divided in 24h slots, with information on each piece of equipment available and accessories required or needed. It is linked to the equipment management module as well, in order to, for example, report issues and control maintenance downtime. In an academic environment, it may also be used for booking rooms, for locating equipment, and for classes.

5.4.4 Chemical Inventory. Each individual laboratory has its own requirements and needs in terms of chemical products, from solvents to resins. These are important for sample preparation and for the use of the equipment, and so must be safely stored and its use controlled. Thus, if the stock is somehow interrupted, it may interfere with the laboratory normal operations or delay analytical work. The module should display at least the following features: type of products, manufacturers, safety data, quantity or number of bottles, validity date, warning system to keep track of the inventory and a warning when to order more, for example, when a minimum quantity is reached.

5.4.5 Graphical Sampling Documentation. Proper graphic documentation is a great tool to assist researchers to know where the samples were taken from in the original object or work of art. This tool is of the utmost importance since it allows researchers to document and assess the importance of the sample and possible implications for its analysis. The module has at least some basic features such as import image, display, and basic image editing and note taking. Once identified, the samples are then linked automatically to other system modules.

5.4.6 Sample Tracking. Another important purpose of the system is managing samples [List et al. 2015]. After proper documentation of sampling and association to the Graphical Sampling Documentation module, the samples should be registered and numbered with an automated or personalised code. This includes areas for note taking, descriptions, and adding images or text files, if necessary. Further on, any given sample may be subdivided for analysis purposes, which will be accounted for. Any analysis that will be made later on any of the samples will be associated to it. QR (Quick Response) codes may be used on each sample's box which, linked to the database, may be checked with any mobile phone. Registration and numbering will allow organising and keeping a sample archive or library, linking permanently in an easy to access way their associated documentation, saving crucial information to the context, collection and analysis of the samples. This allows future referencing and use, independently of the researcher permanence to the institution.

5.4.7 Reports. The possibility to generate automated reports from each project or dataset should not be underestimated. As major advantages, it helps save time, increases the productivity, reduces the bureaucratic work, and helps while preparing for publication. Ready-made templates should be available and may be customised by each researcher if desired.

6 DATA MANAGEMENT

Proper data management is crucial for the success of the system and should cover related areas such as data modeling, data integration, data governance and data quality management. As such, a research data management plan should be prepared describing in detail the data that will be acquired; how it will be described, organised, processed, analysed, maintained, shared (licensing), preserved and archived; as well as consider documentation standards and best practices. In research, one must also consider funding and publisher directives as well as compliance with national and international laws and policies. In cultural heritage, there are a series of national and international documentation standards adopted by cultural institutions that can be followed to describe any heritage object. Thus, the LIMS should follow those same standards adopted by the object's museum to document it. Only then further analytical and other new data can be properly linked and documented. By choosing a widely used or proper standard, and given that documentation content is usually the same, the system may take that into account and import the content into the adopted standard.

One of the most widely used standards and the one used by major cultural heritage institutions is the CIDOC Conceptual Reference Model (CRM), which is also an international standard (ISO 21127:2014). This is a reference model for the exchange of information for cultural heritage documentation through the use of a formal ontology for concepts and information [Le Boeuf et al. 2018; Niccolucci and Felicetti 2018; Castelli et al. 2019]. The importance of the creation of rich metadata [Niccolucci and Felicetti 2018] should also not be understated and is also covered by CIDOC CRM. The use of common vocabularies and ontologies [Daraio et al. 2016] is thus of the utmost importance in order to consider the system part of a wider framework. Although some multilingual general vocabularies are well accepted as standard, little standardization has been achieved for heritage science laboratory work. Therefore, until there will be a wide international agreement on technical terms, the laboratory should create its own vocabulary, based on the researchers experience and feedback, to avoid typos and misinterpretations.

Regarding data openess, concerns towards open science should also be drawn. Open science is inextricably related to the FAIR guiding principles [Wilkinson et al. 2016], an acronym that stands for Findable, Accessible, Interoperable, and Re-usable. They are a set of principles that contribute significantly to good data management helping data creators and publishers to maximize the added-value of curated data. Thus, even though it is not an official standard, compliance with these principles are highly endorsed by international institutions such as the European Commission [2016], among many others. Even when data must remain private or of restricted use, FAIR principles should be considered in the planning phase in order to avoid the system's early obsolescence or a future complete system redesign.

One of the underlying key issues so far is data interoperability [Castelli et al. 2019], which allows data to flow across systems, enabling them to access, exchange, and integrate data without restrictions. This is fundamental to ensure data quality and a precondition to enable open data. Data interoperability is referring to semantic interoperability here, allowing the automatic exchange and interpretation of data in order to maximise results towards the creation of knowledge. In order to achieve this, a semi-structured data model, based on the addition of semantic tags that define a structure for the data within the document, is required. The communication of data between the analytical instruments and the system is here seen as mere data compatibility (as is the case of CSV files, which are purely unstructured data), and not as interoperability. Thus, the use of standards to ensure complete interoperability should start with character encoding in text files, such as UTF-8, one of the most used and that encodes all the European languages, or ISO 8859-1. As for data interchange formats, probably the biggest contenders are JSON (JavaScript Object Notation) [Bourhis et al. 2017], an open standard language-independent text file format, standardized as ISO/IEC 21778:2017, and XML (Extensible Markup Language) [Chituc 2017], an easy to use markup language used across the internet, not only in data exchange but also for communication protocols, being now an industry standard.

7 DISCUSSION

There are several commercial and open-source LIMSs available in the market. However, none of those solutions seems adequate. Commercial licencing is very expensive and the systems are usually 'one-fits-all' solutions that are either very costly to adapt to new functionalities or constrain the end-users to adopt the software structure, with a new non-flexible work methodology, which is far from acceptable. Open source systems are often less expensive solutions and more adequate for adapting to new specific needs. Unfortunately, no existing open source LIMSs are suitable for heritage science since most of them are specifically designed for other fields, typically to biotechnology and medicine.

The importance of the design and usability of the system cannot be disregarded since a poorly designed system or hard to navigate on will dictate user's resistance to its use. Thus, well designed interfaces, with natural adaptation to the workflow, and easy interaction with the data is crucial.

The implementation of a LIMS in a laboratory is not without issues. One of the main advantages is the procedures' standardisation, improving effectiveness and efficiency, such as the quality and time management of both the research and researchers as well their output. However, this is done at the expense of the staff, who may have to adapt or create new routines, with the undoubtful natural resistance to change, which is the biggest cause of failure of implementing a new system. Indeed, one of the bigger risks of implementing mandatory management systems is that they may be seen as an additional burden by the laboratory collaborators and lower their overall motivation to use the system. This is especially true if it is seen as a passive system, where the user must feed the information without any feedback, reducing thus their productivity. This drawback may be surpassed if the people are directly involved from the beginning in setting up the most efficient workflow to helping transform it in a tool they use on a daily basis to manage efficiently their own data and research. For example, if every step is designed to be as quick as possible by eliminating unnecessary steps, automating processes to avoid information repetition, and implementing time-saving features such as a drag-and-drop file uploader, will minimise the researchers' level of commitment that is required. Thus, if the workflow is set organically and intuitively, the learning curve is minimised which will boost participation and adoption of the new system. Also, a period of adaptation before moving towards the use of the LIMSs while continuing to use article-based or other documentation is important to allow a gradual natural adoption.

Legal issues and repercussions should be also considered for private and public data and permissions. For example, in cases where samples are considered owned by the researcher, the laboratory, or any other institution. The European Union General Data Protection Regulation [2016] should also be considered. Even though patent disputes and fraud claims are unlikely in the heritage science field, these concerns may be anticipated by implementing protocols for time stamping, locking pages, and electronic signatures applied to pages or projects.

The system, even though open-source, has its own cost since it must be custom-made and tailored to the specificity of the laboratory where it is installed. However, it is believed that, in the case of research infrastructure networks, the joint effort would make it possible and affordable.

8   CONCLUSIONS

The main objective of this article is to discuss the requirements of a laboratory information management system, presented here as a productivity software, and to propose a model for a heritage science laboratory, having as case study the HERCULES Lab (Portugal). This type of laboratory has very specific requirements than larger laboratories from other fields, thus making most commercial and open-source LIMS not very efficient and appropriate solutions. Nevertheless, LIMS are a crucial tool for managing a laboratory and its research, presenting several advantages such as increasing productivity, managing equipment and resources, and safeguarding scientific data. Thus, a custom web-based solution that was designed to fit a specific workflow in heritage science is presented here. The framework presented includes the system architecture, security, data storage and backup, file management, and system modules. Its implementation, however, may present risks that should be accounted for during its design, such as failure due to a lack of collaboration of the staff or a loss of productivity. Underlying concerns, mainly with compliance with international documentation standards in the cultural heritage field as well as sharing and re-use of data towards an open science mindset, during the planning phase should be taken in account, or otherwise dooming the project as a standalone solution of little use. The successful implementation of a LIMS software as described here will undoubtedly have a positive effect on both the productivity of the researchers as individuals and on the laboratory's running operations. Also, in the near future, once critical data is collected, the database may also be used as data mining source for further research.

REFERENCES

Russell Ackoff. 1989. From data to wisdom. *J. Appl. Syst. Anal.* 16, 3–9.

Anastasia Ailamaki, Verena Kantere, and Debabrata Dash. 2010. Managing scientific data. *Commun. ACM* 53, 68. DOI: https://doi.org/10.1145/1743546.1743568

AWS - Amazon Web Services. 2020. *Amazon EBS Pricing*. https://aws.amazon.com/ebs/pricing/ (accessed 23-03-20).

Martin Barisits, Thomas Beermann, Frank Berghaus, Brian Bockelman, Joaquin Bogado, David Cameron, Dimitrios Christidis, Diego Ciangottini, Gancho Dimitrov, Markus Elsing, Vincent Garonne, Alessandro di Girolamo, Luc Goossens, Wen Guan, Jaroslav Guenther, Tomas Javurek, Dietmar Kuhn, Mario Lassnig, Fernando Lopez, Nicolo Magini, Angelos Molfetas, Armin Nairz, Farid Ould-Saada, Stefan Prenner, Cedric Serfon, Graeme Stewart, Eric Vaandering, Petya Vasileva, Ralph Vigne, and Tobias Wegner. 2019. Rucio: Scientific data management. *Comput Softw Big Sci* 3, 11 (2019). DOI: https://doi.org/10.1007/s41781-019-0026-3

Patrick Le Boeuf, Martin Doerr, Christian Emil Ore, Stephen Stead (eds), 2018. Definition of the CIDOC conceptual reference model. ICOM/CIDOC CRM Special Interest Group. http://www.cidoc-crm.org/sites/default/files/2018-10-26%23CIDOC%20CRM_v6.2.4_esIP.pdf (accessed 06-05-20)

Pierre Bourhis, Juan L. Reutter, Fernando Suárez, and Domagoj Vrgoč. 2017. JSON: Data model, query languages and schema specification. In *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS'17)*, Chicago, IL, 123–135. DOI: http://dx.doi.org/10.1145/3034786.3056120

J.-C. Burgelman, C. Pascu, K. Szkuta, R. Von Schomberg, A. Karalopoulos, K. Repanas, and M. Schouppe. 2019. Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Front. Big Data* 2. DOI: https://doi.org/10.3389/fdata.2019.00043

Lisa Castelli, Achille Felicetti, and Fabio Proietti. 2019. Heritage science and cultural heritage: standards and tools for establishing cross-domain data interoperability. *International Journal on Digital Libraries*. https://doi.org/10.1007/s00799-019-00275-2 (accessed 06-05-20)

Claudia-Melania Chituc. 2017. XML interoperability standards for seamless communication: An analysis of industry-neutral and domain-specific initiatives. *Computers in Industry* 92–93, 118–136. DOI: https://doi.org/10.1016/j.compind.2017.06.010

Cinzia Daraio, Maurizio Lenzerini, Claudio Leporelli, Paolo Naggar, Andrea Bonaccorsi, and Alessandro Bartolucci. 2016. The advantages of an ontology-based data management approach: Openness, interoperability and data quality. *Scientometrics* 108, 441–455. DOI: http://dx.doi.org/10.1007/s11192-016-1913-6

European Commission, Directorate-General for Research & Innovation. 2016. *Guidelines on FAIR Data Management in Horizon 2020*. Version 3.0. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 06-05-20)

EU General Data Protection Regulation (GDPR): Regulation (EU.) 2016. */679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.* https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed 23-03-20)

Robin Harris. 2020. The $15 per terabyte hard drive is coming [WWW Document]. ZDNet. URL https://www.zdnet.com/article/the-15-terabyte-is-coming/ (accessed 23-03-20).

Ondrej Hyks and Karel Kolis. 2014. Development of the digital forensics laboratory management system using ISO 9001 and ISO/IEC 17025. *IDIMT*, 87–94.

ISO, 2008. *ISO 9001: 2008 Quality Management Systems — Requirements*. ISO, Geneva.

Markus List, Steffen Schmidt, Jakub Trojnar, Jochen Thomas, Mads Thomassen, Torben A. Kruse, Qihua Tan, Jan Baumbach, and Jan Mollenhauer. 2015. Efficient sample tracking with open lab framework. *Sci. Rep.* 4. DOI: https://doi.org/10.1038/srep04278

Yichen Ma, Fuyao Wang, and Zhuozheng Wang. 2017. Intelligent laboratory management system based on internet of things. In *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*. IEEE, 464–467.

Hari K. Machina and David J. Wild. 2013. Electronic laboratory notebooks progress and challenges in implementation. *J. Lab. Autom.* 18, 264–268. DOI: https://doi.org/10.1177/2211068213484471

D. B. Marcum and G. George (Eds.). 2010. *The Data Deluge: Can Libraries Cope with E-science? Libraries Unlimited*, Santa Barbara, Calif.

Reagan W. Moore. 2001. Data management systems for scientific applications. In *The Architecture of Scientific Software, IFIP — The International Federation for Information Processing*. Springer, Boston, MA, 273–284. https://doi.org/10.1007/978-0-387-35407-1_16

John Naisbitt. 1982. *Megatrends: Ten New Directions Transforming Our Lives.* Warner Books, Massachusetts.

Franco Niccolucci and Achille Felicetti. 2018. A CIDOC CRM-based model for the documentation of heritage sciences. In *2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, San Francisco, CA, 1–6, DOI: 10.1109/DigitalHeritage.2018.8810109

Isabelle Pallot-Frossard. 2016. De CHARISMA à IPERION CH et E-RIHS : vers une infrastructure européenne de recherche pour les sciences du patrimoine. *Technè* 46–49.

Liron Pantanowitz, W. H. Henricks, and Bruce A. Beckwith. 2007. Medical laboratory informatics. *Clin. Lab. Med.* 27, 823–843. DOI: https://doi.org/10.1016/j.cll.2007.07.011

Arie Shoshani and Doron Rotem (eds.). 2009. *Scientific Data Management: Challenges, Technology, and Deployment.* CRC Press

J. Silbermann, S. Weinert, C. Wernicke, and M. Frohme. 2011. Quality and information management in the laboratory. In *3rd IEEE International Symposium on Logistics and Industrial Informatics*. 93–98. DOI: https://doi.org/10.1109/LINDI.2011.6031127

D. O. Skobelev, T. M. Zaytseva, A. D. Kozlov, V. L. Perepelitsa, and A. S. Makarova. 2011. Laboratory information management systems in the work of the analytic laboratory. *Meas. Tech.* 53, 1182–1189. DOI: https://doi.org/10.1007/s11018-011-9638-7

Kerstin Thurow, Bernd Göde, Uwe Dingerdissen, and Norbert Stoll. 2004. Laboratory information management systems for life science applications. *Org. Process Res. Dev.* 8, 970–982. DOI: https://doi.org/10.1021/op040017s

Catherine Voegele, Baptiste Bouchereau, Nivonirina Robinot, James McKay, Philippe Damiecki, Lucile Alteyrac. 2013. A universal open-source electronic laboratory notebook. *Bioinformatics* 29 (2013), 1710–1712. DOI: https://doi.org/10.1093/bioinformatics/btt253

David Watson and Andrew Jones. 2013. *Digital Forensics Processing and Procedures: Meeting the Requirements of ISO 17020, ISO 17025, ISO 27001 and Best Practice Requirements.* Elsevier. DOI: https://doi.org/10.1016/C2010-0-69255-8

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3 (2016), 160018. DOI: http://dx.doi.org/10.1038/sdata.2016.18

Simon Wood. 2007. Comprehensive laboratory informatics: A multilayer approach. *Am. Lab.* 39, 20.

Chun Yan Zhang. 2014. Design of laboratory management system based on web. *Appl. Mech. Mater.* 608–609, 367–370. DOI: https://doi.org/10.4028/www.scientific.net/AMM.608-609.367