MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190244

Cardiac Tissue Engineering as Use Case to Connect Biomedical Research Laboratories to an Emerging Global Data Infrastructure

Markus Suhr[a], Nadine Umbach[a,c], Tim Meyer[b,c], Wolfram-Hubertus Zimmermann[b,c], Ulrich Sax[a]

[a] Department of Medical Informatics, University Medical Center Göttingen, Göttingen, Germany,
[b] Institute of Pharmacology and Toxicology, University Medical Center Göttingen, Göttingen, Germany
[c] DZHK (German Center for Cardiovascular Research), partner site Göttingen

Abstract

Methods for cardiac tissue engineering and application in experiments are core technologies developed at the Institute of Pharmacology and Toxicology in Göttingen. As is the case in many academic research laboratories data capture and documentation may be improved to latest methods of digital research. A comprehensive information system infrastructure is the foundation of further advances toward automation of lab processes. A data management system concept is proposed and prototypically deployed that enables traceability of assets within the lab and reproducibility of published assays and results. The prototype integrates existing electronic lab notebook, experiment result database, and a newly introduced research data management system by means of a custom developed portal and integration component. The architecture concept and developed integration tools explore connection of routine experimental work in a biomedical research lab to a universal infrastructure of data.

Keywords:

Automation, Laboratory; Data Curation; User-Computer Interface;

Introduction

Biomedical research laboratories in industry and academic environments are challenged with an ever-increasing complexity and quantity of data to be handled. Engineered human tissue models are, for example, pushing into high-throughput drug screening [1], precision medicine [2], and clinical tissue replacement therapy applications [3]. Laboratory automation technology and data science methods are introduced into the biomedical lab domain. Complementary to this bottom-up need for digital transformation, regulatory and political initiatives to improve data management in research demand adherence to quality standards and promote infrastructure development to facilitate an interconnected ecosystem of data [4]. Both the potential benefits and the methodological pitfalls of "big data" applications in biomedical research are well documented [5] with an overall expectation that given expert handling and realistic interpretation of computational models and predictive methods, significant scientific progress is possible. Driven by everyday needs in the development of biotechnological methods for generation and application of engineered myocardium, we investigate the prerequisites and best practices for capturing laboratory process and experimental data, and enabling integration with cross-organizational data-driven research.

Background

The Institute of Pharmacology and toxicology of the University Medical Center Göttingen (UMG) has a long-standing history in the generation of engineered human myocardium (EHM) induced pluripotent stem cells following a well-defined protocol [6]. The primary functional parameter of EHM samples are force and frequency of contraction (FoC, R-R) which directly report on quality and maturity of tissues and are evaluated during *in vitro* compound screening experiments, see Figure 1. The technological challenge currently faced is to develop scalable methods for generation, cultivation and analysis of EHMs that are compliant with Good Manufacturing Practice (GMP). Our key approach is the automation of the entire process by removing human interventions in handling and documentation stepwise one after another.

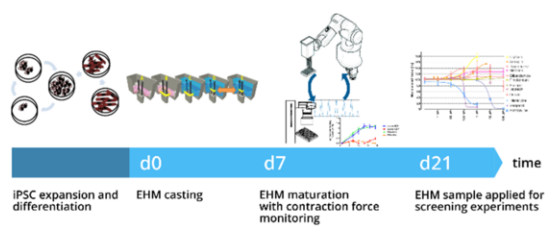

Figure 1 – Simplified process of engineered heart muscle generation and substance screening. (CC-BY-SA 4.0 T. Meyer, M. Suhr)

Automation of complete drug screening experiments depends on the same atomic laboratory processes as EHM generation, routine change of culture media and FoC measurement. A video-optical method for simultaneous measuring of 48 EHM samples on a microtiter culture-plate is developed on site, enabling major gains in process efficiency. Currently, EHM generation protocol steps are executed manually by lab personnel and are documented in paper notebooks and the electronic lab notebook (ELN) RSpace [7]. The video-optical FoC measurement is performed by a high-resolution camera attached to a hardware frame-grabber for near real-time transformation of the optical signal into time series data. Raw output from the camera system is processed using a custom MATLAB program to calculate functional parameters. The tool stores results in a MySQL database along with experiment parameters and metadata like EHM sample source materials (base cell line, generation protocol, batch ID, creation date, type and geometry of culture-plate, etc.). The database schema has been designed and frequently expanded to enable fine-grained documentation and planning of EHM generation and

experiments with process automation in mind. Data is stored describing assets like culture-plates, cell lines, chemical substances, and personnel involved as well as the execution of experimental steps.

These laboratory processes initiate a typical scientific data life cycle routine. Beginning with the planning and creation of primary data, several stages of processing, publication, and archiving are traversed before allowing re-use of result data for new investigations [8]. Such research data management (RDM) activities are supported by sophisticated infrastructures and systems for ETL processes, data publication, and data sharing [9]. Goals for technical implementation have been coined the FAIR principles, calling for research data that is findable, accessible, interoperable, and reusable [10]. Advanced RDM tools achieve compliance with the FAIR criteria but usually on a per-project basis and not as a basic tool for everyday documentation and data capture in the laboratory. Considering the overall trajectory towards interconnected systems of FAIR data across organizational and discipline boundaries [4], on-site data management infrastructure is needed that allows for user-friendly, FAIR-compatible, data provenance preserving, and possibly regulation compliant everyday experimental work in a constantly changing ecosystem of experimental settings and lab automation.

Objectives

Based on the EHM use case, we explore the preconditions and available tools for a laboratory RDM infrastructure enabling acquisition, storage, and sharing of raw data and metadata created during the process. Essential quality requirements for a resulting system architecture are usability in a laboratory environment and software sustainability in a highly specific area of biomedical research, i.e. drawing on lean human and financial resources for information technology. Ideally, the concept should be flexible enough to be reused for different experimental scenarios in biomedical research laboratories.

Methods

Literature search is conducted to (a) determine mandatory quality standards for data processing systems in the given experimental context, and to (b) collect reported best practices for RDM in similar biomedical contexts and tools available for prototypical implementation. Through iterative manual screening of articles' title, abstract, and full text, the result set is refined. Based on frequent interviews and discussion sessions with the involved biomedical researchers, laboratory processes are modeled as basis for formation of a system requirements catalog following recommendations from the requirements engineering field [11]. Combining the insights from the literature search, we propose a concept for an architecture of RDM tools to meet the requirements of a biomedical research lab and implement a prototype supporting the EHM generation use case.

Results

Goals and Requirements

Taking into account content and scope of application of multiple "GxP" guidelines [12], we include implications for information systems from GLP in the requirements engineering process, namely: user access control mechanism, audit trailing, metadata storage, data integrity and quality checks, as well as validation of interfaces, archival, and backup procedures [12]. From the analysis of data management best practices and available tools we derive six goals for a comprehensive system: (CG1) to enable research process, (CG2) to maximize usability, (CG3) to ensure information security, (CG4) to favor dynamic extension, (CG5) to maximize sustainability, and (CG6) to increase reporting quality of results. Based on the conducted literature and requirements analyses as well as the defined goals, we defined primary system requirements that imply general usage scenarios for information systems in the given context. This set represents the functionality of a socio-technical system to operate core processes at a biomedical research laboratory.

Tools

The literature search results revealed 10 distinct software tools and frameworks for data management in a research laboratory. Multiple of these are no longer actively developed since original publication. Among open source projects, two at least partially closed-source commercial products are reported (LabMatrix, LabKey Server) which have not been further considered for application in this exploratory case study. We exclude tools that are tailored only for minor or too specific aspects of the system requirements. Final candidate tools that enable both experimental data capture and FAIR-compatible data sharing are openBIS [13] and SEEK [14] from the "FAIRDOM" ecosystem.

Architecture Concept

Laboratory information systems are commonly divided into three categories: user-centric ELN, operative data- and process-centric laboratory information management system (LIMS), research result-centric scientific data management system (SDMS) [15]. The requirements specified exceed the functionality offered by any single system especially with regard to the need for extensibility: constantly changing experimental settings and analytical processes require a system that can be flexibly integrated and extended with new modules and functionality while at the same time meeting quality requirements like interoperability and sustainability with given resources.

Machina and Wild discuss three models for system integration in drug discovery research [15] which is very similar to the given environment: "ELN-centric" integration is driven by the idea to unify all user interaction with different subsystems through the interface provided by an electronic lab notebook. Modular integration describes the interconnection of sub-modules from neighboring systems like ELN and LIMS enabled by a distinct integration system. The service-oriented integration approach relates closely to the Service-Oriented-Architecture (SOA) paradigm of software design [16] with multiple independent modules (services) providing encapsulated functionality and communicating through messages. We combine aspects from these integration models into an architecture concept as follows: The available tools for the research lab data management system are self-enclosed software systems with internal modular structure as presumed in the modular-integration approach. To ensure operative flexibility and longevity of the systems, extensibility is maximized by applying the SOA paradigm. We propose a service-oriented integrated system architecture constructed around five logical components:

- **Electronic laboratory notebook** (ELN). A system for manual laboratory work documentation and collaboration in GxP-compliant manner.
- **Laboratory information management system** (LIMS). A leading system for laboratory asset registration and management.

- Multiple **experiment-specific subsystems** providing services for custom settings that cannot be implemented with ELN and LIMS.
- **Data integration and analysis system** (DIAS). A system for integration of data from disparate sources, visualization and analysis.
- **Research data management system** (RDMS). A system for research project planning, high-level documentation, result and data sharing.

To achieve integration of these components, three auxiliary building blocks complete the concept:

- **Portal**. A (web-based) central access point for all functionality that provides a unified graphical user interface.
- **Service integration layer**. The control center for service-based integration, connecting services and subsystems and exposing functionality via API or relaying GUI to the portal service. Central shared services like user management, authentication, and authorization should be realized within the integration layer to harmonize subsystem behavior.
- **External services**. An arbitrary number of external services and systems relayed to the overall system through connector services in the integration layer.

Communication between services and subsystems should be realized through application programming interfaces as the SOA paradigm proclaims. Subsystem user interfaces can be integrated into the portal-based unified GUI. If subsystems do not provide API or GUI, it should be possible to customize services within the integration layer to provide surrogate functionality. Figure 2 depicts a graphical representation of the architecture concept.

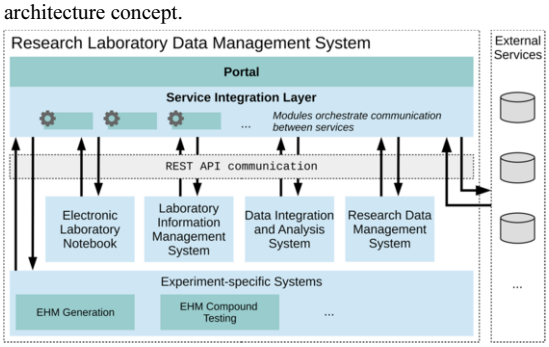

Figure 2 – Architecture concept for a comprehensive research laboratory data management system. (CC-BY-SA 4.0 M. Suhr)

Prototype Implementation

With regard to the goal of enabling especially the EHM experiment-specific workflows, ELN and RDMS are the components to be implemented with priority: ELN because it enables coherent documentation of manually executed workflow tasks; RDMS because the planning and publishing capabilities enable improvements for both organization of work and FAIR-compliant publication of results. The existing RSpace instance is to be integrated and a local instance of SEEK will be set up as part of the prototype system.

The MySQL database system developed for the EHM compound testing experimental data capture and storage and the MATLAB program for analysis of video-optical contraction force measurements are heavily interwoven in the current form. Instead of replacing these operational systems in an early phase of the project, they are to be integrated as a first experiment-specific subsystem.

Finally, portal and service integration layer components have been implemented as custom solution for demonstration purposes: The "LabHub" in its first iteration is intended as a central platform for user interaction, resource management, data entry, and access to data from past experiments in the EHM context. Aiming for usability, extensibility, and sustainability we utilize the widespread content management system Drupal as a software development framework to create a set of logically independent modules for user interaction and data processing. Service integration functionality is demonstrated by offering an administrative user interface for service and resource registration. Drupal modules in the LabHub context define service types which again are used by other modules as a resource. Instances of services are registered through GUI using internet protocol. This allows for example, to register multiple instances of the SEEK platform, which in turn can be connected to by an arbitrary number of modules that communicate with the SEEK REST API. "Connector" modules that define service types and "consumer" modules that interact with the respective web-service API are implemented for the SEEK and RSpace systems. API and application layer modules for the existing EHM MySQL database are drafted, enabling role-based management of laboratory asset data and access to experimental result data. Using the integration layer functionality, this experiment-specific system communicates with both RSpace and SEEK instances allowing for user-centric documentation of workflow steps in the ELN and FAIR-compliant publication of FoC measurement data in the RDMS component.

At the current stage of the project, access to both the Drupal and SEEK web servers is restricted to the local network. Developmental stage source code of the LabHub prototype is available through a public code repository (http://hdl.handle.net/21.11101/0000-0007-CADE-C).

Discussion

A challenge implicit in the research question is how to find candidate software for implementing an infrastructure satisfying all the diverse requirements of operative and research data processing and managing in a biomedical lab. This is mostly due to the circumstance that no canonical descriptive name for the category of software exists. "ELN" and "LIMS" are labels well-defined and attached to a multitude of open source, service-based, and commercial products. The term "SDMS" is mentioned in articles and per definition comes very close to the here proposed requirements for research lab data management [17] but is neither topical nor actively promoted. Of the discovered software tools that in description satisfied a considerable subset of our requirements, some are no longer available or maintained. The most challenging task for this project is to not follow these projects into a dead end but to aim for sustainability. Remarkably, some software tools initiated in the years 2007-2011 have succeeded due to innovative concepts and future-proof development processes.

Architecture Concept

Inefficiency of user interfaces is a central barrier of adoption of IT systems in the lab [19]. Overcoming usability barriers in practice is the actual challenge and must be made a focal point of infrastructure design and implementation. Four of our declared six main goals emphasize software quality aspects, addressing this need to improve quality of research data handling while not negatively interfering with operative processes [18].

Appropriate models of laboratory information system architecture and the integration of subsystems exist for decades already with similar challenges pronounced as [19]. Despite this understanding of the problems, formulation of solution models, and the obvious progress of technology, seamless integration of data producing and data processing entities is an ongoing effort. The presented analysis updates the modeling with explicit focus on sustainable operation of the technical infrastructure. Strength of the proposed concept is the extensible integration of existing software systems as well as service modules into an architecture accessible through a central portal component. Combination of integration methods [15] serves to integrate the existing systems as components, as demonstrated here with the RSpace instance. This approach allows for re-use of the framework at other laboratories as it is agnostic of whether or not components have been deployed before. "ELN-centric integration" [15] follows a similar reasoning but for adoption of the approach an extensible and modifiable ELN component is required. In comparison, the framework presented here does not require any existing system to function as the central access and integration component but encourages a modular solution.

Considering the broad spectrum of specialized systems discovered in the literature search and the indicated system requirements, only an integrated architecture of multiple systems and services is probable to succeed in the long run. Classic LIMS solutions are reported to expand in functionality and aim towards becoming comprehensive information systems for all laboratory needs [15] but the dynamic nature of method and experiment development at a research lab is not suitably addressed by traditional monolithic systems. Notably, the product STARLIMS is named as an example for an expanding LIMS [15] and is already deployed at UMG service units Biobank and Stem Cell Unit. Investigation of synergistic effects and integration opportunities is planned.

Among RDM projects in medical informatics and biomedicine, the framework discussed here is remarkable for not focusing infrastructure within a collaborative project. Typically, systems are developed and proposed for project-specific use cases on grant-based funding. Here, persistent organizational infrastructure is considered with research projects and the associated data management systems located within the outer system context. Sustainability and extensibility have been exclaimed goals so that the resulting architecture should be more long-lived than project-centric infrastructures. The modularity embedded in the concept is meant to ensure that subsystems can be exchanged and replaced to address changing technical and structural requirements. The more systems are integrated into a complex data management architecture however, the higher becomes the need for capturing data provenance records to prevent data quality from deteriorating [9,20].

LabHub Prototype

The goal dimensions initially declared are addressed by the prototype. Still, the achievements mark only a beginning when assessed against the business goals of enabling research process and improving reporting quality. Most problematic is the question of sustainability. The topic has been at the core of considerations leading to the architecture concept and implementation plan. The prototype created is not inherently sustainable though. Some of the approaches reported in the scientific literature have vanished quickly in spite of support by a professional team of researchers or developers. Re-evaluation of openBIS as a data management system may thus be of interest. Major drawback ascertained in assessment of openBIS for the EHM use case was the user interface design not suited for cross-platform operation. Given the resources to integrate openBIS into a modern, usable, platform-independent front end, reuse of the system's wide spectrum of data management features could be an elegant way to create an integrated platform, as already demonstrated by other projects [21].

Promoting LabHub development to become a data-centric counterpart for efforts to standardize control interfaces for lab automation devices may be considered. We plan cooperation with the Standardization in Lab Automation (SiLA) to advance the automation of EHM generation processes. To our knowledge, no platform for management, control, and scheduling of lab processes using SiLA compatible devices does exist yet, indicating an opportunity to investigate.

Systemic Challenges

Sustainability of a system is less a question of technology than of available resources. Project grants for specialized innovative research data processing systems may be easier to allocate than permanent funding for infrastructures needed for basic operation. Ideally, the operative infrastructure would support FAIR data sharing to promote secondary use of data internally for quality control and meta-analyses, and externally in research projects. Above all, implementing and maintaining a complex integrated system in such a way requires capable personnel. A small dedicated team of software engineers may become an integral asset.

Laboratories and institutes will need to adopt gateway services for inter-organizational collaborative data-driven research projects on (inter-) national scale. Here, the SEEK instance operates as such by allowing FAIR-compatible sharing of data. Projects like the Medical Informatics Initiative in Germany [22] or the European Open Science Cloud (EOSC) [23] promote infrastructural integration and cross-site data exchange. Independent of how the challenge of privacy preserving data sharing will be solved technically, providing internal primary data in FAIR-compliant ways enables integration with the emerging global data infrastructure [4]. Developing and operating a FAIR-compatible infrastructure at an institute or laboratory may prove to be of both scientific and strategic advantage in technological advancement of biomedical research in the coming years.

Conclusions

We describe challenges and requirements that transcend the specific setting of cardiac tissue engineering. The proposed architecture framework promotes extensibility and usability for highly dynamic, constantly evolving operative realities. The topic of innovative research lab data infrastructure is open for collaborative advancements and investigation. FAIR data publication and reporting are explored for the presented use case but are not yet operational and require further investigation and refinement especially with regard to application of semantics. Achieving sustainability of the LabHub software or any integrated infrastructure is a major challenge and requires commitment regarding resources and effort. The development of an integration and control framework for laboratory process automation is promising in terms of scientific and operative potential.

In conclusion, a framework for comprehensive research lab data management is proposed for discussion and possibly adaptation. Our findings are a foundation for diverse subsequent investigation and possibly software development. The core challenges require close cooperation and collaboration of biomedical researchers and computer scientists to pave the way for dynamic and persistent socio-technical information systems connected to a global infrastructure of data.

Acknowledgments

This work was funded by the BMBF in the project Data and Information Management (DZHK (German Center for Cardiovascular Research)) (grant 81Z7300173) as well as by the DFG for the Collaborative Research Centers (CRC) 1002 on Modulatory Units in Heart Failure, subproject INF, and CRC 937 on Collective Behavior of Soft and Biological Matter.

References

[1] B. Fermini, S.T. Coyne, and K.P. Coyne, Clinical Trials in a Dish: A Perspective on the Coming Revolution in Drug Development, *SLAS Discov. Adv. Life Sci. R D*. (2018) 12. doi:10.1177/2472555218775028.

[2] E.M. Antman, and J. Loscalzo, Precision medicine in cardiology, *Nat. Rev. Cardiol.* **13** (2016) 591–602. doi:10.1038/nrcardio.2016.101.

[3] M. Tiburcy, J.E. Hudson, P. Balfanz, S. Schlick, T. Meyer, M.-L. Chang Liao, E. Levent, F. Raad, S. Zeidler, E. Wingender, J. Riegler, M. Wang, J.D. Gold, I. Kehat, E. Wettwer, U. Ravens, P. Dierickx, L.W. van Laake, M.J. Goumans, S. Khadjeh, K. Toischer, G. Hasenfuss, L.A. Couture, A. Unger, W.A. Linke, T. Araki, B. Neel, G. Keller, L. Gepstein, J.C. Wu, and W.-H. Zimmermann, Defined Engineered Human Myocardium With Advanced Maturation for Applications in Heart Failure Modeling and Repair, *Circulation*. **135** (2017) 1832–1847. doi:10.1161/CIRCULATIONAHA.116.024145.

[4] P. Wittenburg, and G. Strawn, Common Patterns in Revolutionary Infrastructures and Data, (2018). doi:10.23728/b2share.4e8ac36c0dd343da81fd9e83e72805a0.

[5] H. Fröhlich, R. Balling, N. Beerenwinkel, O. Kohlbacher, S. Kumar, T. Lengauer, M.H. Maathuis, Y. Moreau, S.A. Murphy, T.M. Przytycka, M. Rebhan, H. Röst, A. Schuppert, M. Schwab, R. Spang, D. Stekhoven, J. Sun, A. Weber, D. Ziemek, and B. Zupan, From hype to reality: data science enabling personalized medicine, *BMC Med.* **16** (2018) 150. doi:10.1186/s12916-018-1122-7.

[6] M. Tiburcy, T. Meyer, P. Soong, and W.-H. Zimmermann, Collagen-Based Engineered Heart Muscle, in: M. Radisic, L.D. Black III, and L.D.B. III (Eds.), Card. Tissue Eng., Springer New York, 2014: pp. 167–176. doi:10.1007/978-1-4939-1047-2_15.

[7] S. Macdonald, and R. Macneil, Service integration to enhance research data management: RSpace electronic laboratory notebook case study, *Int. J. Digit. Curation*. **10** (2015) 163–172.

[8] L. Corti, V. Van den Eynden, L. Bishop, and M. Woollard, Managing and sharing research data: a guide to good practice, Sage, Los Angeles, 2014.

[9] C.R. Bauer, N. Umbach, B. Baum, K. Buckow, T. Franke, R. Grütz, L. Gusky, S.Y. Nussbeck, M. Quade, S. Rey, T. Rottmann, O. Rienhoff, and U. Sax, Architecture of a Biomedical Informatics Research Data Management Pipeline, *Stud. Health Technol. Inform.* **228** (2016) 262–266. doi:10.3233/978-1-61499-678-1-262.

[10] M.D. Wilkinson, M. Dumontier, Ij.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*. **3** (2016) 160018. doi:10.1038/sdata.2016.18.

[11] K. Pohl, Requirements Engineering Fundamentals, 2nd Edition: A Study Guide for the Certified Professional for Requirements Engineering Exam - Foundation Level - IREB compliant, Rocky Nook, Inc., 2016.

[12] S.H. Segalstad, International IT Regulations and Compliance: Quality Standards in the Pharmaceutical and Regulated Industries, John Wiley & Sons, 2008.

[13] A. Bauch, I. Adamczyk, P. Buczek, F.-J. Elmer, K. Enimanev, P. Glyzewski, M. Kohler, T. Pylak, A. Quandt, C. Ramakrishnan, C. Beisel, L. Malmström, R. Aebersold, and B. Rinn, openBIS: a flexible framework for managing and analyzing complex data in biology research, *BMC Bioinformatics*. **12** (2011) 468. doi:10.1186/1471-2105-12-468.

[14] K. Wolstencroft, S. Owen, F. du Preez, O. Krebs, W. Mueller, C. Goble, and J.L. Snoep, The SEEK: a platform for sharing data and models in systems biology, *Methods Enzymol.* **500** (2011) 629–655. doi:10.1016/B978-0-12-385118-5.00029-3.

[15] H.K. Machina, and D.J. Wild, Laboratory informatics tools integration strategies for drug discovery: integration of LIMS, ELN, CDS, and SDMS, *J. Lab. Autom.* **18** (2013) 126–136. doi:https://doi.org/10.1177/2211068212454852.

[16] T. Erl, Service-oriented architecture: concepts, technology, and design, Prentice Hall Professional Technical Reference, Upper Saddle River, NJ, 2005.

[17] H.K. Machina, and D.J. Wild, Electronic laboratory notebooks progress and challenges in implementation, *J. Lab. Autom.* **18** (2013) 264–268.

[18] K.T. Taylor, Evolution of Electronic Laboratory Notebooks, in: Collab. Comput. Technol. Biomed. Res., Wiley-Blackwell, 2011: pp. 301–320. doi:10.1002/9781118026038.ch19.

[19] R.D. McDowall, and D.C. Mattes, Architecture for a comprehensive laboratory information management system, *Anal. Chem.* **62** (1990) 1069A-1076A.

[20] B. Baum, C.R. Bauer, T. Franke, H. Kusch, M. Parciak, T. Rottmann, N. Umbach, and U. Sax, Opinion paper: Data provenance challenges in biomedical research, *It - Inf. Technol.* **59** (2017). doi:10.1515/itit-2016-0031.

[21] C. Mohr, A. Friedrich, D. Wojnar, E. Kenar, A.C. Polatkan, M.C. Codrea, S. Czemmel, O. Kohlbacher, and S. Nahnsen, qPortal: A platform for data-driven biomedical research, *PLOS ONE*. **13** (2018) e0191603. doi:10.1371/journal.pone.0191603.

[22] S.C. Semler, F. Wissing, and R. Heyder, German Medical Informatics Initiative, *Methods Inf. Med.* **57** (2018) e50–e56. doi:10.3414/ME18-03-0003.

[23] B. Mons, C. Neylon, J. Velterop, M. Dumontier, da S. Santos, L.O. Bonino, and M.D. Wilkinson, Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud, *Inf. Serv. Use*. **37** (2017) 49–56. doi:10.3233/ISU-170824.

Address for correspondence

Markus Suhr, University Medical Center Göttingen, Department of Medical Informatics, von-Siebold-Straße 3, 37075 Göttingen, +49 551 39 61519, markus.suhr@med.uni-goettingen.de, https://orcid.org/0000-0002-6307-3253