A Comprehensive Approach to Support Research – Processes in the CRC 1270 ELAINE

Max Schröder[1,2], Frank Krüger[1], Robert Zepf[2], Ursula van Rienen[3], Sascha Spors[1]

[1]Institute of Communications Engineering, University of Rostock
[2]University Library, University of Rostock
[3]Institute of General Electrical Engineering, University of Rostock

Virtual research environments play a central role in today's interdisciplinary research. They provide a secure and convenient way for collaboration and traceable research. While all-in-one solutions exist that enable researchers to collaborate during the entire research data lifecycle, we argue that a flexible virtual research environment can also be composed of off-the-shelf services. In that vein, we introduce the virtual research environment of the CRC 1270 ELAINE and discuss different implementation aspects.

1 Introduction

Managing digital research artefacts (such as data, models and source code) in a convenient, comprehensive, flexible and secure way is an important issue in today's research (Hanson et al., 2011). Many approaches, however, focus on storing, archiving and publishing only the research data. Other artefacts of the research process such as the model, the source code of the analysis, and the computing environment are often not considered. In the Collaborative Research Centre (CRC) (German: Sonderforschungsbereich) 1270 ELAINE supported by the German Research Foundation (German: Deutsche Forschungsgemeinschaft, DFG), instead, we aim at supporting the overall research process that begins at the definition of the hypothesis and ends at the dissemination and publication of the results and findings. While all-in-one software exists that aims at supporting the entire lifecycle, different reasons prevent researchers from using such solutions. Firstly, such all-in-one solutions are not very flexible when it comes to the individual needs of a multidisciplinary and heterogeneous research group. Secondly, if the software will no longer be developed, great effort is required to re-use and migrate the research artefacts to another solution. Furthermore, hosting all-in-one software at one own's institution, often is not easy when it comes to installation, maintenance, and extension. Finally, researchers are often not allowed to use "Software as a Service" offers that are hosted abroad. Reasons for this are for instance data privacy issues and governmental guidelines.

In the CRC ELAINE, the needs of the researchers range from documenting simulation studies and wetlab experiments, to managing and sharing data and models in different lab environments, to supporting versions, provenance and archival of research artefacts.

In contrast to all-in-one software solutions, we propose a Virtual Research Environment (VRE) that builds upon simple individual services, particularly tailored for the specific needs of each research field. This enables the easy replacement of single services if e.g. requirements change or service development is discontinued. In order to determine the requirements of the individual researchers, we performed a systematic assessment by use of a questionnaire (Krüger and Spors, 2018) in addition to visits of the research groups. For the appropriate selection of each service, we additionally interviewed domain experts that are target users of the service.

With respect to the research artefacts, the research process can be structured into three major tiers: 1. Study planning and data collection, 2. Collaborative modelling and data analysis, and 3. Reproducibility, provenance and archival. For each of these tiers, a large number of simple but specialised Open Source web-services are available. By the use of standardised interfaces, the services can be integrated into a comprehensive, flexible and secure VRE. The VRE for the CRC ELAINE which we currently implement is illustrated in Figure 2. The particular services, e.g. electronic lab notebooks, Jupyter notebooks and Docker are result of the requirements analysis with the researchers.

The next section introduces the general idea of VREs. Section 3 presents the details of the VRE within the CRC 1270 ELAINE followed by a discussion of the VRE.

2 Virtual Research Environments

Candela et al. (2013) define VREs as “innovative, web-based, community-oriented, comprehensive, flexible, and secure working environments conceived to serve the needs of modern science.” Thus, they enable more efficient, open, and reproducible research (Barker et al., 2019). Agreeing with this definition, we would like to add another aspect. The overall research process is driven by research artefacts beginning at the formulation of a hypothesis; ending at the dissemination and publication of the data and source code. Research artefacts in this case are data such as spreadsheets, sensor data and survey results. Furthermore, simulation models, source code, and images are also research artefacts, e. g., CT scans. However, not all artefacts are digital, e. g., materials, prototypes, and lab notes. A VRE, then, has to incorporate and track the creation, modification, and share of these research artefacts that are increasingly (re-)used in the research lifecycle (Briney, 2015).

Several solutions for specific research fields exists, such as for the structural biology (Morris et al., 2019) and genomics (Codó et al., 2019), astronomy (Herwig et al., 2018), archaeology (Meghini et al., 2017), and earth sciences (Marelli et al., 2016). Especially in highly multidisciplinary research projects, such specifically tailored solutions, however, cannot be employed for all researchers. Thus, a more generally applicable solution is needed. The Open Science Framework[1] is an example for such all-in-one software. Instead of using a single software for the overall life cycle, simple tools tailored for specific use cases integrated into each other are another option. This increases the effort for integration, but at the same time increases flexibility of the VRE. Services can be easily replaced if requirements change or development is discontinued. Khoonsari et al. (2019) follow a similar direction by providing a scalable data analysis platform for applications in metabolomics that can be launched on demand using Docker[2] and Jupyter[3] notebooks.

Central aspects of a VRE are the support for FAIR principles Wilkinson et al. (2016), powerful programming interfaces which are needed in order to integrate the services into the research workflow and existing software systems as well as an effective user interface. Ardito et al. (2016) proposed the creation of a VRE addressing the diversity of users based on the meta-design approach. In order to support the FAIR principles—i. e., findability, accessibility, interoperability, and reusability—Dimitrov and Stoyanov (2018) proposed a discovery service for the CKAN4[4] data repository engine.

[1] see https://osf.io/
[2] see https://www.docker.com/
[3] see https://jupyter.org/
[4] see https://ckan.org/

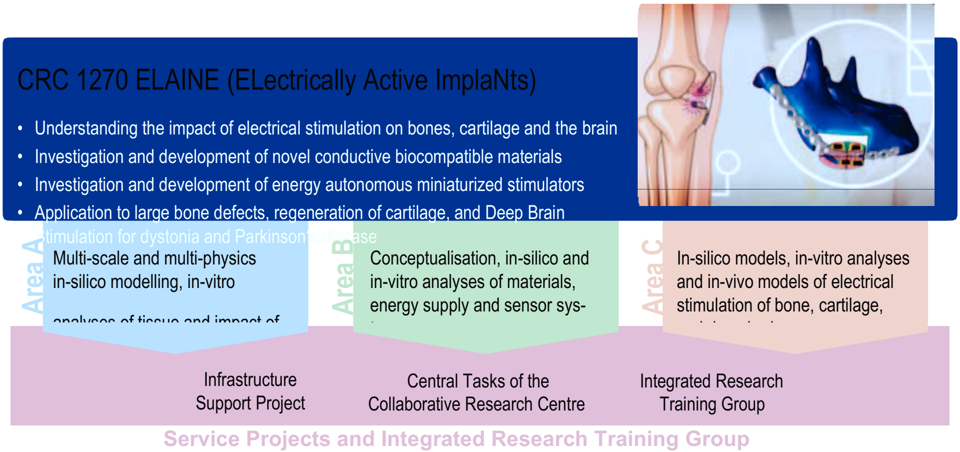

Figure 1: Schematic overview of the research aspects within the CRC 1270 ELAINE

3 The CRC 1270 ELAINE Virtual Research Environment

The next section introduces the research objectives of the CRC 1270 ELAINE in order to better understand the needs and requirements for its VRE. Afterwards, the details of the CRC 1270 VRE are discussed.

3.1 The CRC 1270 ELAINE

A CRC is a large interdisciplinary collaboration project funded by the German Research Foundation (DFG) with a strong focus on fundamental research on a particular topic. Objective of the CRC 1270 ELAINE is to investigate different aspects of electrically active implants and put novel solutions forward. The ultimate goal is to provide implants for electrical stimulation in bone, cartilage, and brain that are energetically autonomous. For this purpose researchers from medicine and biology, electrical and mechanical engineering, material and computer sciences, and physics work in close cooperative on the following aspects:

1. Establish innovative energy autonomous implants that allow a feedback-controlled electrical stimulation,
2. Efficient multi-scale simulation models to enable rapid progress in targeted im- plant improvements and patient-specific therapies, and
3. Analyse the basic mechanisms of electrical stimulation in bone, cartilage and brain, and translate this knowledge into clinical practice.

A schematic overview of these aspects is given in Figure 1.

Typical investigations within the CRC 1270 ELAINE involve experiments and simulations. The results of the experiments are used to calibrate and validate simulation models, whereas the results of the simulation are used to determine optimal parameters for the electrical stimulation of the tissue. As this research is conducted by different research teams at different locations (partly even different cities), the usage of a VRE that enables researchers to share data and results in a traceable way is of central interest. As previously discussed, further aspects of the VRE are usability, security and privacy.

The VRE of the CRC 1270 is also a testbed for novel services that are later de- ployed for all members of the university. To support this target, we are working closely together with the university library and the IT and Media center. To be applicable for all researchers of the university, the VRE service cannot be domain specific.

The VRE of the CRC 1270 ELAINE is built upon a set of microservices, each of them addressing particular objectives of the research data lifecycle. To this end, these services are ordered accordingly, resulting in the three tiers for data collection, data analysis and data archival. A schematic representation of the VRE is given in Figure The selection of the services based on the results of the questionnaire (Krüger and Spors, 2018) that was used as an initial analysis of the requirements within the CRC 1270 and subsequent visits of the research groups. A strong focus during the selection of potential services, however, lay on web based open source software in order to decrease proprietary dependencies and increase convenient usage of the researchers avoiding the install of software on their computers. In the following, an overview of the three tiers is provided, the different aspects are discussed and the specific microservices introduced

3.2 Tier 1: Study Planning and Data Collection

The first tier of the VRE is concerned with the beginning of a new research investigation: the documentation of the research question or hypothesis. Furthermore, also the subsequent documentation of the experiment respectively study as well as the initial data storage has to be afforded at this stage.

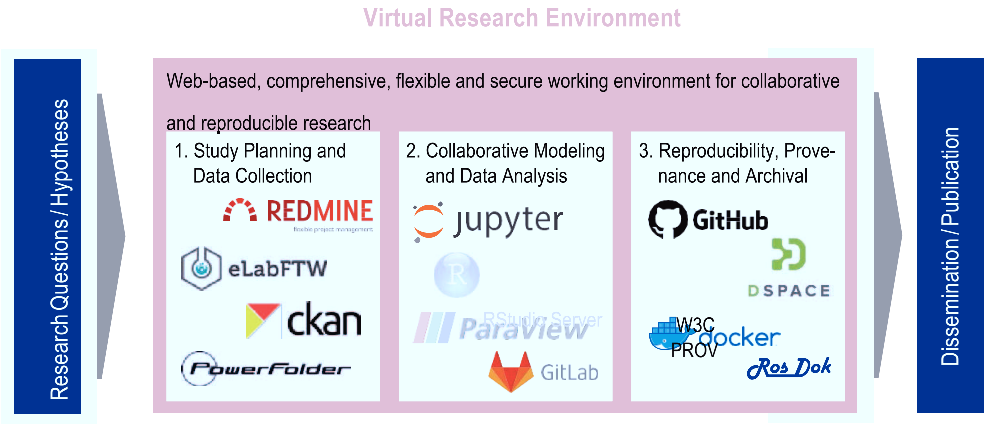

Figure 2: The current VRE in the CRC ELAINE including the main microservices. RStudio Server and ParaView are ideas for further extensions.

General Project Management

A general-purpose project management service is needed for the project management tasks such as central storage of protocols and newsletters, and the documentation of the research questions and hypotheses arising within the CRC. Furthermore, the structure of the CRC should be implementable into this tool so that individual projects of the CRC have particular workspaces that can be shared with other researchers and groups. For the general communication and documentation of the CRC, further workspaces are needed.

In the CRC 1270, we use Redmine[5] for this purpose. Although Redmine was originally aimed at tracking software development, the research project documentation has some similar requirements. Basically, two mechanisms are required: (1) store research artefacts in a versioning system, and (2) store textual documentation. Redmine provides both mechanisms in a simple but functional user interface and provides great community support including a multitude of extensions. Other software that provides similar functionality is e. g., OpenProject[6], basecamp[7], and trac[8].

Experiment Documentation Tool

In laboratory environments such as in the biological and medical research, lab notebooks are common to use in order to document particular experiments

[5] see https://redmine.org/
[6] see https://www.openproject.org/
[7] see https://basecamp.com/
[8] see https://trac.edgewall.org/

respectively studies. Digital solutions for such white paper doc- umentations exists. Their main advantage is searchability and readability of multiple researchers at the same time if configured appropriately. Although general project management tools can be employed for the documentation of laboratory experiments, electronic lab notebook (ELN) solutions might better suited. ELN tools often provide additional features such as materials databases, laboratory management features, and the possibility to define and share Standard Operation Procedures (SOP). Such functionality can also be used to define the set of minimum information that are required to make experiments reproducible (Budde et al., 2019).

For these reasons, we decided to deploy elabFTW[9] as ELN solution in the CRC 1270. This tool is not tailored for a specific domain but rather for the documentation of many kinds of experiments providing a white-paper based documentation style. While the main demand of this tool is for in-vitro and in-vivo documentation, some researchers try this tool for their general research documentation. Other prominent examples of electronic laboratory notebooks include labfolder[10], openBIS[11], and RSpace[12].

Data Storage and Internal Data Sharing

The second major aspect of this tier is the storage and sharing within the research consortium of the raw data created during an experiment respectively a simulation study. Despite the actual storage of arbitrary data formats, several additional requirements have to be met. The provenance of data has to be tracked, which not only includes meta information such as the author of the data and a description; but also revisions of the data and the meta data i. e., versions. The data artefacts have to be findable, accessible, interoperable, and reusable (Wilkinson et al., 2016) for the original creator but also other researchers for whom this dataset is shared. The internal storage platform ideally is integrated into the publication service in order to automatically create a public version of the data set on request. In the CRC 1270 ELAINE, two services for the data storage were selected that com- plement each other: CKAN[13] and PowerFolder[14]. CKAN is a tool that enables the stor- age of meta information along with a set of resources which can be files or urls. The tool then provides a web interface for the search and simple inspection of the data sets.

[9] see https://www.elabftw.net/
[10] see https://www.labfolder.com/
[11] see https://labnotebook.ch/
[12] see https://www.researchspace.com/
[13] see https://ckan.org/
[14] see https://www.powerfolder.com/

It supports a basic versioning of the datasets by providing meta data fields for version information and the possibility of creating relations between datasets. Although this satisfies the requirements for the storage of data from research investigations, the software PowerFolder—a file synchronisation and sharing tool—(along with OnlyOffice[15]) is employed for the collaborative editing of text, spreadsheets and presentations. The PowerFolder service is already deployed by the central IT and Media Center for all members of the university.

3.3 Tier 2: Collaborative Modelling and Data Analysis

After the experimental study has been planned and the raw data collected, the modelling and data analysis phase begins. During this phase, researchers often iteratively develop models or perform data analyses. Different versions of data, models and scripts evolve, which have to be maintained appropriately. Furthermore, during the development and discussion of the data analysis, it is often necessary to share intermediate results and insights. This indicates the need for version control and collaborative environments for modelling as well as data analysis and exploration. Instead of using click-and-point software, the analysis should be developed as source code in order to enable reproducibility. Furthermore, in order to enable other researchers understand the analysis and models, the method of literate programming recently gained attention. This method combines explanatory text and source code in a single document. In the CRC 1270 ELAINE, a Jupyter notebook[16] service is deployed in order to support literate programming. This service provides an interactive web user interface for the development of source code and documentation as well as the execution of the analysis using the computing resources of the server. Similar software include RStudio Server[17] with the R packages knitr[18] or Sweave[19]. The Jupyter service, however, does not track different versions of the model respectively the analysis. To incorporate this functionality in the CRC 1270, a local Gitlab[20] instance is deployed, that provides GIT repository management as core functionality. The Gitlab, furthermore, provides continuous integration functionality that can be used in order to control data quality as well as reproducibility (Pietsch et al., 2019). In combination both services enable to collaboratively work on the data analysis and keep track of modifications and results.

[15] see https://www.onlyoffice.com/
[16] see https://jupyter.org/
[17] see https://www.rstudio.com/products/rstudio/
[18] see https://yihui.name/knitr/
[19] see https://leisch.userweb.mwn.de/Sweave/
[20] see https://about.gitlab.com/

3.4 Tier 3: Reproducibility, Provenance and Archival

This last tier is concerned with the tasks of completing the research investigation artefacts, providing a comprehensive and integrated overview of the investigations, and providing convenient access to them. As such this task is divided into three parts: Reproducibility, Provenance, and Archival.

Reproducibility

To support reproducibility on a technical level a comprehensive doc- umentation, tracking of the research artefacts, and programming of the research analysis has already been discussed. Additionally, it is crucial to document the software stack that is used for the computation of the results i. e., the computing environment. For the comprehensibility of the research investigations it, however, is not enough to provide an executable image of this computing environment. Furthermore, a description of the environment with its software dependencies including the software versions is needed. This enables both easy execution as long as the image is working and a specification enabling to rebuild the computing environment if the virtualisation solution does no longer exist.

For this purpose, containerisation solutions recently are gaining attention. Within the CRC 1270, we use Docker[21] as a containerisation tool. This enables to build the computing environment based on the specification in so-called Dockerfiles. The Docker tool exists for many operating systems providing an easily executable environment on the personal computer of the researchers. Furthermore, continuous integration chains such as in Gitlab support the usage of these container images. Thus, they can be incorporated to automatically check data quality and consistency. Other solutions include Singularity[22].

Provenance

For the reproducibility but also the comprehensibility of the research investigations even after years, it is crucial to document the provenance of research artefacts and their relations (McClatchey, 2018). Provenance information should be integrated into all parts of the VRE so that it is both understandable for other researchers and machine

[21] see https://www.docker.com/
[22] see https://www.sylabs.io/singularity/

interpretable. The latter is crucial in order to provide consistency checks and automatic notifications about new revisions of artefacts.

In the CRC 1270, we incorporate the standardised W3C PROV formalism (Schröder et al., 2019). While a machine interpretable version in the form of turtle and other languages can be maintained, the information can easily be converted into a provenance graph that is understandable by researchers. A prior version of this formalism is known as OPM Provenance Model.

Archival and Publication

A reproducible and comprehensible documentation of the research investigations shall be achieved for a long period of months and years. To support this archival of the research, multiple services need to be employed. All research artefacts need to be archived under consideration of the FAIR principals (Wilkinson et al., 2016). For the archival the same requirements as already for the storage of the artefacts during the research process need to be satisfied. For the archival task we use multiple services within the CRC 1270 depending on the type of artefact. For source code artefacts, we use GitHub[23] in companion with Zenodo[24] as a public archive. Whereas GitHub is a platform for the management of GIT repositories, Zenodo is a public data storage repository for the archival. Zenodo provides DOIs for each version of a data set and can automatically import the resources from a GitHub release. When it comes to datasets the service DSpace[25] comes in and, finally, for the archival of publications, the RosDok[26] platform is used. DSpace is a repository software that enables storage of arbitrary datasets similar to the CKAN data storage, but with a richer customisable meta data set. Similar software includes Dataverse[27] and Invenio[28]. The RosDok service is a local variant of the mycore framework[29], which also is a repository software.

4 Discussion

Beside the service requirements of the VRE, other requirements exists that play an essential role. To those belong convenience, general technical and legal aspects as well as requirements from the particular research artefact types.

[23] see https://github.com/
[24] see https://zenodo.org/
[25] see https://duraspace.org/dspace/
[26] see http://rosdok.uni-rostock.de/
[27] see https://dataverse.org/
[28] see https://invenio-software.org/
[29] see http://www.mycore.de/

Research Artefact Types Not all research artefacts are digital, i. e., a multitude of different media are involved such as prototypes of 3d structures, materials, and sensor data. These different media have to be handled with their corresponding meta information including provenance and version information. Tracking this is essential in order to enable comprehensive understanding of the research experiment. We are still working on an efficient solution tracking real objects and their corresponding digital information.

Convenience Dimensions Usability and training effort are other dimensions that need to be considered. This is highly dependent on the prior knowledge of the re- searchers which are not only different between research fields but also within a re- search group. As such, often a compromise between ease of use, functionality, and training effort has to be found when selecting a service. Often, researchers have to learn many new methods and concepts, especially the concept of versioning is not always clear. Furthermore, the user interfaces are not standardised so that every ad- ditional service requires at least a minimal level learning. A thorough training concept is currently developed and already partly implemented for the researchers in the CRC 1270.

Technical Dimensions Despite the previously discussed aspects of the services, some additional general technical facets were considered: The most important issue is that tools support single-sign-on functionality so that researchers don't need to register for every service individually. Software maintenance costs but also installation and infrastructure maintenance effort are also influencing aspects. Furthermore, the possibility of extending the functionality of a tool is needed in order to adapt the solution to local requirements. A very important issue is security; the services have to enable fine-grained rights management and the possibility to create backups.

Legal Dimensions The legal dimensions for supporting research processes by a VRE are manifold. From a top-down-view, this starts with governmental and institutional laws and guidelines that regulate data handling. Furthermore, especially in biological and medical fields, there are often approvals of experimental procedures that also contain regulations on the handling of research artefacts. Then there are laws about supervising working habits of employees that is technically easy when digital systems are involved. But also the visibility and editability of investigations and research artefacts need to be controlled, especially when it comes to industrial cooperations and data privacy. Data privacy is also crucial with respect to combining multiple sources of research artefacts which may lead to de-anonymisation of them. There are many more legal aspects involved which we, however, cannot analyse here completely.

5 Conclusion

In this paper we argued that a VREs can support researchers in multiple disciplinary teams at different locations during the entire data lifecycle. We presented the VRE of the CRC 1270 ELAINE. We discussed different services and showed how they can contribute to the individual phases of the data lifecyle: 1. study planning and data collection, 2. collaborative modelling and data analysis, and 3. reproducibility, provenance and archival. In favour of more flexibility, the VRE was built from many specialised services. The requirements were collected by conducting interviews with the participating researchers, other research data management experts and early adoption test phases. While the VRE is still in development, a lot of open issues have to be addressed in future work. This starts by systematically assessing information about acceptance and usability from the researchers but also includes better integration of the particular services into each other as well as into the workflow of the researchers. Also the support for automatic workflows from data recording to analysis, including an assessment of the data quality could be of interest. Finally, as the VRE is envisaged as central data and information storage, methods of automatic data or text mining and information retrieval could be employed in order to provide automatic documentation.

Acknowledgement

This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the Collaborative Research Centre 1270 ELAINE.

Literatur

Carmelo Ardito, Maria Francesca Costabile, Giuseppe Desolda, Maristella Matera, and Paolo Buono. A meta-design approach to support information access and manipulation in virtual research environments. In *Lecture Notes in Computer Science*, pages 115–126. Springer International Publishing, 2016. doi: 10.1007/978-3-319-50070-6_9.

Michelle Barker, Silvia Delgado Olabarriaga, Nancy Wilkins-Diehr, Sandra Gesing, Daniel S. Katz, Shayan Shahand, Scott Henwood, Tristan Glatard, Keith Jeffery, Brian Corrie, Andrew Treloar, Helen Glaves, Lesley Wyborn, Neil P. Chue Hong, and Alessandro Costa. The global impact of science gateways, virtual research environ ments and virtual laboratories. *Future Generation Computer Systems*, 95:240–248, jun 2019. doi: 10.1016/j.future.2018.12.026.

Kristin Briney. *Data Management for Researchers: Organize, maintain and share your data for research success*. Pelagic Publishing Ltd, 2015.

Kai Budde, Julius Zimmermann, Elisa Neuhaus, Max Schröder, Adelinde Uhrmacher, and Ursula van Rienen. Requirements for documenting electrical cell stimulation experiments for replicability and numerical modeling. In *2019 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, jul 2019. To appear.

Leonardo Candela, Donatella Castelli, and Pasquale Pagano. Virtual research en vironments: An overview and a research agenda. *Data Science Journal*, 12(0): GRDI75–GRDI81, 2013. doi: 10.2481/dsj.grdi-013.

Laia Codó, Genís Bayarri, Javier Alvarez Cid-Fuentes, Javier Conejero, Romina Royo, Dmitry Repchevsky, Marco Pasi, Athina Meletiou, Mark D. McDowall, Fatima Re- ham, José A. Alcantara, Brian Jimenez-Garcia, Jürgen Walther, Ricard Illa, François Serra, Michael Goodstadt, David Castillo, Satish Sati, Diana Buitrago, Isabelle Brun- Heath, Juan Fernandez-Recio, Giacomo Cavalli, Marc Marti-Renom, Andrew Yates, Charles A. Laughton, Rosa M. Badia, Modesto Orozco, and Josep Ll. Gelp´ı and. MuGVRE. a virtual research environment for 3d/4d genomics. apr 2019. doi: 10.1101/602474.

Vladimir Dimitrov and Stiliyan Stoyanov. Solutions for data discovery service in a virtual research environment. *Scalable Computing: Practice and Experience*, 19(2):181–187, may 2018. doi: 10.12694/scpe.v19i2.1350.

B. Hanson, A. Sugden, and B. Alberts. Making data maximally available. *Science*, 331 (6018):649–649, feb 2011. doi: 10.1126/science.1203354.

Falk Herwig, Robert Andrassy, Nic Annau, Ondrea Clarkson, Benoit Côté, Aaron D'Sa, Sam Jones, Belaid Moa, Jericho O'Connell, David Porter, Christian Ritter, and Paul Woodward. Cyberhubs: Virtual research environments for astronomy. *The Astrophysical Journal Supplement Series*, 236(1):2, may 2018. doi: 10.3847/1538-4365/aab777.

Payam Emami Khoonsari, Pablo Moreno, Sven Bergmann, Joachim Burman, Marco Capuccini, Matteo Carone, Marta Cascante, Pedro de Atauri, Carles Foguet, Alejan- dra N Gonzalez-Beltran, Thomas Hankemeier, Kenneth Haug, Sijin He, Stephanie Herman, David Johnson, Namrata Kale, Anders Larsson, Steffen Neumann, Kristian Peters, Luca Pireddu, Philippe Rocca-Serra, Pierrick Roger, Rico Rueedi, Christoph Ruttkies, Noureddin Sadawi, Reza M Salek, Susanna-Assunta Sansone, Daniel
Schober, Vitaly Selivanov, Etienne A Thévenot, Michael van Vliet, Gianluigi Zanetti, Christoph Steinbeck, Kim Kultima, and Ola Spjuth. Interoperable and scalable data analysis with microservices: applications in metabolomics. *Bioinformatics*, mar 2019. doi: 10.1093/bioinformatics/btz160.

Frank Krüger and Sascha Spors. A questionnaire to estimate the needs for research data management. 2018. doi: 10.18453/rosdok_id00002290.

Fulvio Marelli, Mirko Albani, and Helen Glaves. Everest: a virtual research environment for earth sciences. In *EGU General Assembly Conference Abstracts*, volume 18, 2016.

Richard McClatchey. Data provenance tracking as the basis for a biomedical virtual research environment. *CoRR*, abs/1803.07433, 2018. URL http://arxiv.org/abs/ 1803.07433.

Carlo Meghini, Franco Niccolucci, Achille Felicetti, Paola Ronzino, Federico Nurra, Christos Papatheodorou, Dimitris Gavrilis, Maria Theodoridou, Martin Doerr, Dou- glas Tudhope, Ceri Binding, Roberto Scopigno, Andreas Vlachidis, Julian Richards, Holly Wright, Guntram Geser, Sebastian Cuy, Johan Fihn, Bruno Fanini, and Hella Hollander. ARIADNE. *Journal on Computing and Cultural Heritage*, 10(3):1–27, aug 2017. doi: 10.1145/3064527.

Chris Morris, Paolo Andreetto, Lucia Banci, Alexandre M.J.J. Bonvin, Grzegorz Cho- jnowski, Laura del Cano, José Marıa Carazo, Pablo Conesa, Susan Daenke, George Damaskos, Andrea Giachetti, Natalie E.C. Haley, Maarten L. Hekkelman, Philipp Heuser, Robbie P. Joosten, Daniel Kouřil, Aleš Křenek, Tomáš Kulhánek, Victor S.Lamzin, Nurul Nadzirin, Anastassis Perrakis, Antonio Rosato, Fiona Sanderson, Joan Segura, Joerg Schaarschmidt, Egor Sobolev, Sergio Traldi, Mikael E. Trellet, Sameer Velankar, Marco Verlato, and Martyn Winn. West-life: A virtual research environment for structural biology. *Journal of Structural Biology: X*, page 100006, feb 2019. doi: 10.1016/j.yjsbx.2019.100006.

Christian Pietsch, Jochen Schirrwagen, and Vitali Peil. Conquaire: Coupling a local gitlab instance with an institutional repository for instant research data publications. 2019. doi: 10.5281/zenodo.2602714.

Max Schröder, Hendrikje Raben, Frank Krüger, Andreas Ruscheinski, Ursula van Rienen, Adelinde Uhrmacher, and Sascha Spors. Provenance patterns in numerical modelling and finite element simulation processes of bioelectric systems. In *2019 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, jul 2019. To appear.

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Apple- ton, 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 Vel- terop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR guiding principles for scientific data management and stew ardship. *Scientific Data*, 3:160018, mar 2016. doi: 10.1038/sdata.2016.18