Towards in-situ knowledge acquisition for research data provenance from electronic lab notebooks*

Max Schröder[1,2][0000−0003−1522−494X], Susanne Staehlke[3][0000−0002−0251−8263], Barbara Nebe[3,4][0000−0003−2662−0905], and Frank Krüger[1][0000−0002−7925−3363]

[1] Institute of Communications Engineering, University of Rostock, Rostock, Germany
[2] University Library, University of Rostock, Rostock, Germany
[3] Department of Cell Biology, University Medical Center Rostock, Rostock, Germany
[4] Department Life, Light & Matter, University of Rostock, Rostock, Germany

Abstract. The documentation of wet-lab experiments is essential for the reproducibility of experimental investigations and their results. The semantic representation of such documentation in a machine understandable format allows the implementation of automated understanding and comparison. Objective of this study is to investigate whether a semantic model can be created from the information available in the protocols from the wet-lab. By analysing a protocol from a biomedical wet-lab experiment, we demonstrate that Electronic Laboratory Notebooks could serve as a mechanism for the *in-situ* knowledge acquisition about experimental investigations. The protocol and the model is available at: https://github.com/m6121/Semantic-Modelling-CA-Imaging

Keywords: Provenance · Ontology · Electronic Laboratory Notebook.

1 Introduction

The documentation of research investigations is essential for the reproducibility of their scientific results [7]. Moreover, it plays an important role for the provenance documentation of the research data that was created and modified during these investigations. When it comes to the wet-labs, such as in the biomedicine, chemistry, or physics, documentation is typically done in laboratory notebooks. While analogue lab notebooks are well established for the individual documentation of such investigations, with the digital turn and the rise of (virtual) collaboration, Electronic Laboratory Notebooks (ELNs) become increasingly important. *ELN protocols*, thus, include notes of the researcher, but also measurements and the resulting data. This is in contrast to *experimental protocols* such as published instructions to conduct experimental investigations. ELN protocols can be seen as instances of experimental protocols with measurements, comments and information of a particular execution of the instructions including timestamps, the employed software and the performing researcher.

* This work was financially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1270/1 - 299150580.

In their article, Stocker et al. [29] propose the usage of Virtual Research Environments (VREs) for the curation of scientific information to overcome the document centred scientific communication. They propose the usage of semantic technologies for the representation of such information and illustrate the effectiveness of their approach by modelling the process of statistical hypothesis testing in a machine understandable format. Their approach enables the automatic processing of such for further exploration [4].

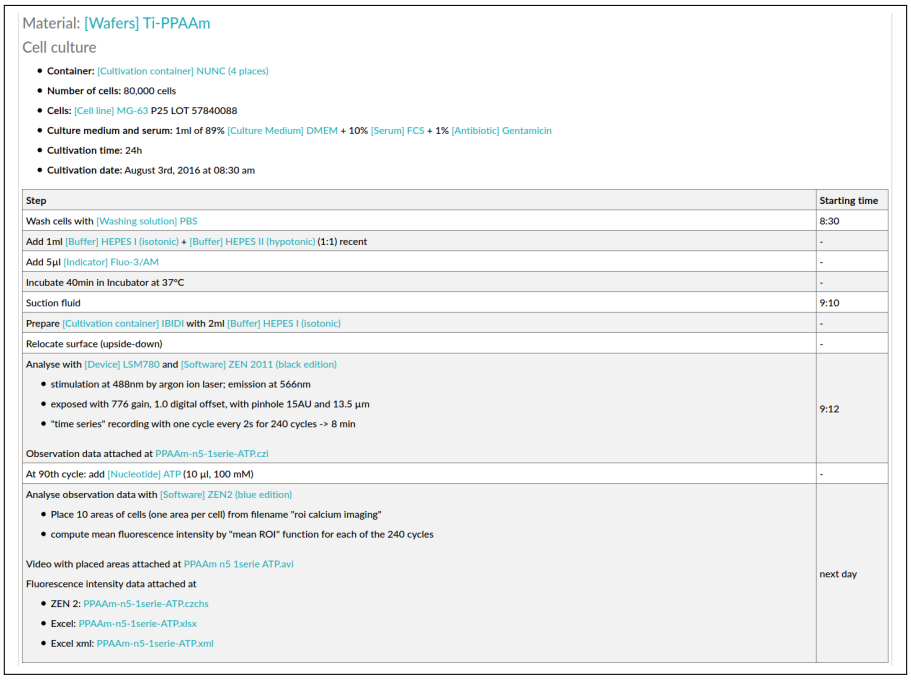
We use the term *in-situ knowledge acquisition* to refer to the systematic generation of knowledge within the VRE. In this work, we investigate, whether it is possible to implement *in-situ* knowledge acquisition about experimental investigations and the raised research data from the information available in ELNs. To this end, we implement a semantic provenance model based on general purpose and domain specific ontologies to incorporate the information from a wet-lab protocol. The ELN protocol describes a particular execution of an experimental protocol for Calcium Imaging [28]. Furthermore, the semantic provenance model is connected to other public Knowledge Graphs (KGs) providing additional sources of semantic knowledge. Thus, we demonstrate that ELNs can serve as a basis for *in-situ* knowledge acquisition about the provenance of research data.

2 Electronic Laboratory Notebooks

ELNs are software tools that provide a convenient method for the textual documentation of wet-lab investigations which in contrast to analogue lab notebooks provide additional features such as *sharing* and *searching*. According to Schnell [21], ELNs can provide a single place for all information about the experiment. This includes research data, such as microscopy images, tabular measurement data, or results of statistical analyses. The use case in this paper has been documented employing the general purpose ELN software *elabFTW*.

Beside the editor for the textual description of investigations, ELNs often comprise an inventory database, which contains detailed information about research objects (see Figure 2) including used materials and devices. Linking to such research object entries allows the re-use of existing information and, thus, increase the quality of the documentation. Devices and materials can be centrally maintained reducing wrong documentation information. Furthermore, the revisions of a protocol provide enhanced provenance information about the experiment. Thus, the textual description with links to the employed research material, and the raised research data makes ELNs the ideal source for provenance modelling of the entire research process.

Experimental protocols are often supported in ELNs by templates. These can be centrally created by experienced researchers and then used for particular ELN protocols. In other words, an ELN protocol describes a concrete investigation including a set of parameters filled into the template. This is in contrast to published experimental protocols e. g., at protocols.io, where the general procedure but not a specific execution of this is documented. Figure 1 illustrates an excerpt



Material: [Wafers] Ti-PPAAm

Cell culture

- **Container:** [Cultivation container] NUNC (4 places)
- **Number of cells:** 80,000 cells
- **Cells:** [Cell line] MG-63 P25 LOT 57840088
- **Culture medium and serum:** 1ml of 89% [Culture Medium] DMEM + 10% [Serum] FCS + 1% [Antibiotic] Gentamicin
- **Cultivation time:** 24h
- **Cultivation date:** August 3rd, 2016 at 08:30 am

Step	Starting time
Wash cells with [Washing solution] PBS	8:30
Add 1ml [Buffer] HEPES I (isotonic) + [Buffer] HEPES II (hypotonic) (1:1) recent	-
Add 5µl [Indicator] Fluo-3/AM	-
Incubate 40min in Incubator at 37°C	-
Suction fluid	9:10
Prepare [Cultivation container] IBIDI with 2ml [Buffer] HEPES I (isotonic)	-
Relocate surface (upside-down)	-
Analyse with [Device] LSM780 and [Software] ZEN 2011 (black edition) • stimulation at 488nm by argon ion laser; emission at 566nm • exposed with 776 gain, 1.0 digital offset, with pinhole 15AU and 13.5 µm • "time series" recording with one cycle every 2s for 240 cycles -> 8 min Observation data attached at PPAAm-n5-1serie-ATP.czi	9:12
At 90th cycle: add [Nucleotide] ATP (10 µl, 100 mM)	-
Analyse observation data with [Software] ZEN2 (blue edition) • Place 10 areas of cells (one area per cell) from filename "roi calcium imaging" • compute mean fluorescence intensity by "mean ROI" function for each of the 240 cycles Video with placed areas attached at PPAAm n5 1serie ATP.avi Fluorescence intensity data attached at • ZEN 2: PPAAm-n5-1serie-ATP.czchs • Excel: PPAAm-n5-1serie-ATP.xlsx • Excel xml: PPAAm-n5-1serie-ATP.xml	next day

Fig. 1: Excerpt from an ELN protocol of a calcium imaging procedure. A log of the experimental steps is given at the left side with time stamps on the right. Links to inventory items or attached research data are represented in blue.

of the ELN protocol of the use case in this paper: a Calcium Imaging procedure. Note that the ELN protocol contains information specific to the actual execution such as timestamps and measurement data.

3 Use Case: Detection of Intracellular Ca^{2+} Dynamics

Objective of this study is to represent a complex experimental investigation based on the description available from the protocol in the ELN in a machine understandable form. To this end, we focus on the detection of intracellular calcium ions (Ca^{2+}) dynamics by Ca-imaging, which was previously published by Staehlke et al. [28] in their article "*Enhanced calcium ion mobilization in osteoblasts on amino group containing plasma polymer nanolayer*". The detailed documentation of this *in-vitro* method is decisive for the reproducibility of the research results. A description of the experimental procedure follows.

Cells are able to sense their environment, e. g., material surface modifications, and transmit external stimuli via cellular structures (like integrin adhesion receptors and actin cytoskeleton) and downstream signalling cascades (intracellular biochemical signalling molecules) into the cell nucleus to modulate cell function

Table 1: Public Knowledge Graphs (KGs) providing structured information about topics relevant to research.

Name	Website	Details
Wikidata	https://www.wikidata.org	Knowledge transferred from Wikipedia and other Wikimedia projects
DBpedia	https://wiki.dbpedia.org	Similar approach to Wikidata
ORKG	https://www.orkg.org	Knowledge about papers
RRIDs	https://scicrunch.org/resources	Knowledge about research resources

such as proliferation or differentiation [27]. In their work, Staehlke et al. [28] investigate whether intracellular calcium dynamics are important for the transmission of external stimuli into the cell and reflect the cellular behaviour.

Intracellular Ca^{2+} are responsible as a second messenger system for signal transmission [30], which ultimately control cell functions [27,28]. For example mechanical stimuli or adenosine triphosphate (ATP) [27] induce a short-term Ca^{2+} rise from the smooth endoplasmic reticulum and thus the intracellular Ca^{2+} dynamic, which trigger further signal transmission [12]. To investigate the role of intracellular Ca^{2+} dynamic on different chemical surface conditions 1.) bare titanium (Ti), 2.) Ti with plasma polymerized allylamine (Ti+PPAAm) which was provided by Leibniz Institute for Plasma Science and Technology (INP) Greifswald 3.) Ti with a collagen type-I-layer (Ti+COL) 4.) tissue culture plastic (IBIDI), 24 h-cultured human osteoblasts MG-63 were stained with a common established non-ratiometric Ca^{2+} indicator [30,27,28] and examined by a confocal laser scanning microscope. The global calcium signal was recorded in vital fluo-3-labelled osteoblasts over a period of 240 cycles every 2 s using mode "time series" (Zen2011, black edition, Carl Zeiss). To stimulate intracellular Ca^{2+} dynamic, ATP was added after the 90th cycle. The fluorescence intensity of the stimulated cells was quantified using the ZEN2 (blue edition, version 2.0.0.0) software by "Mean ROI" mode for 10 defined areas of individual single cells per experimental set-up. The important data such as time and mean fluorescence intensity of ROIs were recorded as tabular data. In the final step, statistical analysis was performed using GraphPad Prism7 software (GraphPad Software Inc., LaJolla, CA USA).

To illustrate the effectiveness of the ELN usage, the analogue documentation was digitized together with the authors of the original study. Only information, available from the original protocols were transferred. Figure 1 illustrates an excerpt from the resulting protocol.

4 Semantic Modelling of ELN Protocols

In the following, the semantic modelling process is described. Afterwards, a summary of the model and a discussion of the representation details are given.

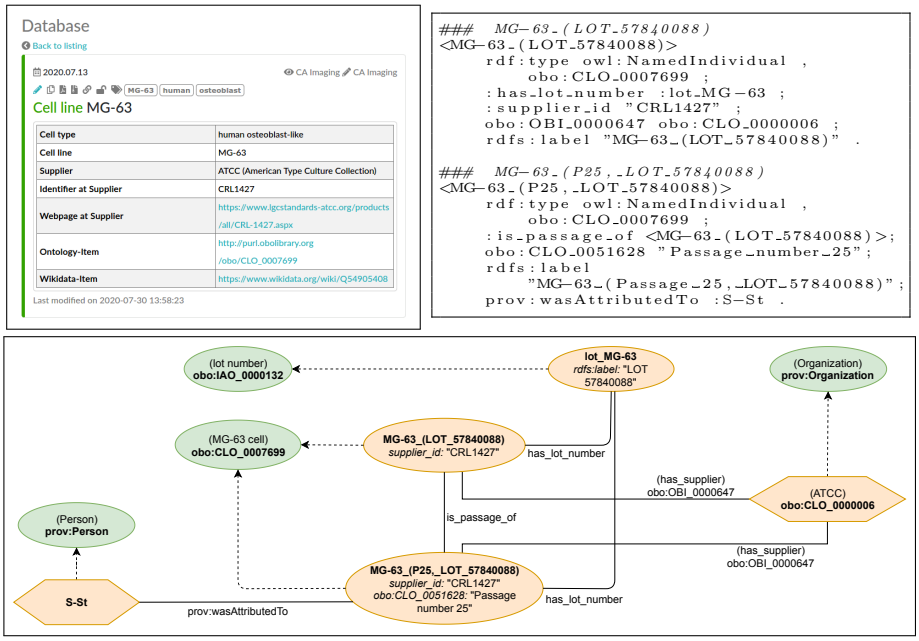


Fig. 2: MG-63 cell line entry from the inventory database **Left**: screen shot from the inventory database; **Right**: excerpt from the model describing the MG-63 cells used in the experiment; **Bottom**: graphical representation: green: ontological classes, orange: instances, solid lines: relations, dotted lines: inheritance.

4.1 Modelling Process

We employed the Protegé (https://protege.stanford.edu/, v5.5.0) software for the formal ontology modelling. The modelling itself was structured as follows:

1. Using the ELN protocol and the elements from the inventory database, we identified relevant ontologies by employing BioPortal (https://bioportal.bioontology.org/) and Ontobee (http://www.ontobee.org/).
2. From these search results, a set of ontologies was selected based on maximising the coverage of terms used in the ELN protocol. Ontologies compatible with the *Basic Formal Ontology* (BFO) [25] from the OBO Foundry Ontologies (http://www.obofoundry.org/) [23] were preferred.
3. Identifier of the ontology classes are included into the inventory database entries in the ELN (see Figure 2) as a reference for the semantic modelling.
4. In the semantic model, instances of these classes were created to represent the particular entities and activities of the protocol and the inventory. Each instance got a persistent id in the local namespace reflecting the individual instance, e. g., http://provenance.elaine.uni-rostock.de/ontologies/ca-imaging/MG-63_(P25,_LOT_57840088) is used to encode passage 25 of the MG-63 cells that were delivered with the LOT number 57840088.

Table 2: Ontologies selected for the semantic model of the use case. Upper rows list general ontologies; the lower rows domain specific ontologies for resources and activities.

Name	Source	Details
BFO	[25]	Basic Formal Ontology
PROV-O	[15]	PROV Ontology
BTO	[9]	BRENDA Tissue Ontology
CHEBI	[6]	Chemical Entities of Biological Interest Ontology
CLO	[20]	Cell Line Ontology
FaBiO	[17]	FRBR-aligned Bibliographic Ontology
OBI	[1]	Ontology for Biomedical Investigations
RO	[24]	Relation Ontology

5. References to other KGs (see Table 1) were partly included by employing the *owl:sameAs* relation. Employing this mechanism, we also include references to Research Resource Identifiers (RRIDs, https://www.rrids.org/). However, when it comes to concrete instances such as the MG-63 cells with a particular LOT number, RRIDs do not reflect these information.

In order to demonstrate the potential of ELN protocols, we concentrated on employing classes that are at the same granularity as the documentation of the experiment in the ELN. Furthermore, despite the references to other KGs, we omitted information in the semantic model that was not present in the ELN.

4.2 Semantic Model of the Calcium Imaging Procedure

The original experiment investigated four different surface conditions: 1.) Ti, 2.) Ti+PPAAm, 3.) Ti+Col, and 4.) IBIDI. Here, we concentrate on the condition Ti+PPAAm, as apart from the material the setting were equal. Table 2 lists the ontologies selected for the semantic representation of the use case. BFO represents the upper level ontology. In addition, the *PROV-O* ontology [15] was employed to provide a formal semantic representation of the provenance of experimental results. The other ontologies provide classes for the entities documented in the protocol respectively inventory database. In summary, the protocol involves eleven steps and 19 inventory items that are represented in about 80 instances of classes by use of six ontologies.

4.3 Re-usable Semantic Structures

During the modelling process, several different semantic patterns have been identified. A brief description of them is provided here. When using existing concepts, the source ontology is given in parentheses.

Inventory Items Documenting the use of biological and chemical resources within an experiment is essential for the credibility of the findings and of particular interest when re-using the data. The ELN inventory items contain information about biological and chemical resources (e. g., the type or a category) as well as information about the supplier (e. g., name, a supplier specific identifier and a web page) which are essential provenance information.

The following information was added to the items of the inventory database to prepare the semantic modelling: 1.) *Ontology-Item* refers to the corresponding ontological class we identified for this item, and 2.) *Wikidata-Item* refers to the corresponding wikidata element, which holds information about this item. Figure 2 illustrates the inventory entry for the MG-63 cell line entry and illustrates the resulting semantic model of the MG-63 cells of the use case. The particular lot number and the passage number 25 have been used from the protocol documentation. An instance of the *MG-63 cell* (CLO) has been created that encodes the specific MG-63 cells with lot number *57840088*. This has been connected to the instance of the *lot number* (IAO) with relation *has_lot_number* which was newly created. The manufacturer respectively supplier is modelled as instance of class *Organization* (PROV) and added to the instance of the resource with the relation *has_supplier*. For the modelling of a passage, two instances of the *MG-63 cell* (CLO) are connected with the relation *is_passage_of*. Furthermore, the person creating the passage is connected with the relation *wasAttributedTo* (PROV) to the resource.

Similar to biological and chemical resources, devices and software used in the experiment have been modelled including information about the device respectively software but also the manufacturer respectively developer.

Protocol The documentation of the running example encoded in the ELN protocol is mainly composed of a list of individual instructions tagged with a time stamp. Examples include the washing of the cells, the combination of biological resources into a mixture, or the incubation. These steps can semantically be modelled by corresponding activities, e. g., *washing* (OBI) and *elution* (OBI); *creating a mixture of molecules in solution* (OBI); or *cell line cell culturing* (CLO), *cell fixation* (OBI) and *acclimatization* (OBI). Additionally, these processes are inherited from *Activity* (PROV) in order to encode general provenance information.

This modelling approach can be employed for most steps until the experimental procedure results in the creation of data files that often are created by devices and software (see Section 3). The last three steps in Figure 1 contain the relevant steps in the ELN protocol. The upper two steps describe the creation of the observation data while the last step is the actual analysis which was originally done using Microsoft Excel.

These instructions in general are encoded as the sub-class of *planned process* (OBI). For the data creation, this results in the *image creation* (OBI) activity. The output of this is a *data set* (IAO) of *images* (IAO) containing a file with the proprietary format CZI encoded as *data format specification* (IAO). The data

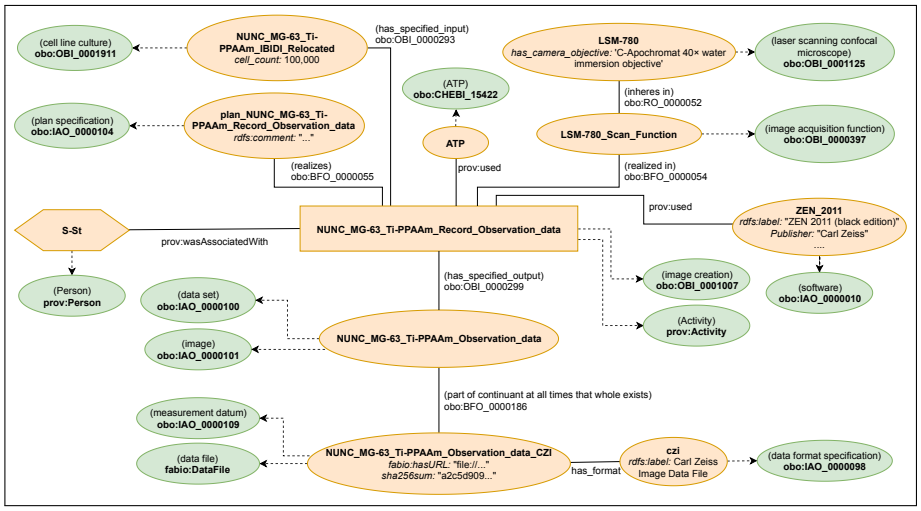


Fig. 3: Graphical representation of the semantic model describing the data recording (see also Figure 2).

file itself is an instance of the class *measurement datum* (OBI). The *laser scanning confocal microscope* (OBI) including its *image acquisition function* (OBI) are related to this process employing *realized in* (BFO). Furthermore, the person executing the experiment *wasAssociatedWith* (PROV) the activity. Figure 3 illustrates the semantic model of this data recording activity.

Afterwards, the data in the file itself might be further modelled. In the case of tables, this information can be directly formalised employing the ontologies. For images as the primary data format in our running example, regions of interest might be identified and the contained information be semantically described. However, for the sake of brevity, we omit this modelling here.

5 Related Work

The semantic representation of ELN protocols employs a set of methods and technologies from different domains: 1. semantic modelling and knowledge representation, 2. mechanisms to analyse and foster reproducibility, and 3. process tracking technologies. Related work from each of these domains will be discussed in the following.

Semantic models can technically be represented as a KG. According to Hogan et al., KGs are "*intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities*" [11]. Examples of such KGs are listed in Table 1. To avoid ambiguities and enable automated reasoning of KGs, ontologies can be used. They provide the definition of the terms and concepts of a KG. The OBO Foundry creates and maintains ontologies for the biomedical domain that are

designed to be interoperable and logically well formed [23]. A general purpose ontology for the semantic definition of provenance information is the PROV ontology [15]. Ontology databases such as the BioPortal help in finding suitable ontologies for particular applications.

Soldatova et al. proposed the *EXACT2* ontology and sketched a natural language processing framework aiming at the automatic semantic extraction of knowledge from biomedical protocols [26]. Similarly, the *SMART Protocols* ontology for representing experimental investigations has been proposed by Giraldo et al. [8]. In difference to both, we focus on the *in-situ* knowledge acquisition of ELN protocols, i. e., instances of experimental protocols which include measurements, research data and are often written employing abbreviations and incomplete sentences. Thus, we are not relying on the high quality of the experimental protocols that were published at e. g., protocols.io.

An ontology for the semantic documentation of provenance in microscopy experiments was proposed by Samuel and König-Ries: Reproduce Microscopy Experiments (REPRODUCE-ME) [18]. This ontology uses PROV as the upper level similar to the approach used in this paper. However, in contrast, here the properties and relations for the objects are already existing and defined in the domain-specific ontologies identified from the terms used in the ELN protocol.

Apart from these efforts to represent wet-lab investigations, a multitude of methods and tools have been proposed aiming at the representation of provenance of data analyses. Samuel and König-Ries proposed *ProvBook*, a software extension that tracks provenance in Jupyter notebooks [19], which are used for interactive literate data analysis. *Dataprov* [2] is a general purpose tool enabling the generation of provenance information from the execution of tools. Similarly, the *noWorkflow* tool has been proposed to non-intrusively keep track of such information [16]. Other methods of provenance tracking are also known as lineage tracking or lineage retrieval [3] and workflow tracking systems [5].

6 Discussion and Conclusion

To summarize, in this paper we encoded the documentation of a wet-lab investigation available as ELN protocol including inventory database items as semantic knowledge by employing existing ontologies. ELN protocols often contain domain specific notes including abbreviations, incomplete sentences and comments. On the other hand, these notes encode a log of what the researcher actually conducted in the wet-lab based on experimental protocols. Thus, the ELN protocol can provide further details about the experiment and the involved research data, which enables researchers from the same domain to repeat the experiment under equal conditions. The resulting model illustrates that the information available from the ELN protocol is indeed sufficient to describe the provenance of the resulting research data.

Scientists documenting their experiments in ELNs have to be careful to include all necessary details, e. g., time stamps, quantities and biological and chemical resource details such as lot numbers. In our use case, the necessary infor-

mation was available and transferred from the analogue documentation together with the original authors of the study [28] in order to document future investigations in the ELN. However, no information was added to the protocol. Typically, an experiment is conducted several times, with only minimal changes to the setting. Considering the use case, the investigations of the four different surface conditions represents this. Also, lab specific procedures are often used within the same team, which results in semantically similar, but syntactically different descriptions. By employing templates for experimental protocols and individual procedures, this could be exploited to create easy-to-use documentation framework. By use of RDFa [10], templates for protocols and inventory items could be annotated during curation and the information could later be used for the semantic documentation of provenance.

ELN protocols can be semantically encoded in a highly diverse manner by employing existing ontologies. The large number of ontologies existing as well as their compatibility to each other, but also the selection of appropriate classes are questions arising during the semantic modelling. Domain experts have to find ontologies that cover a large amount of their terms at the same granularity. But also the use of ontologies structuring ELN protocols respectively experimental protocols itself or their reproducibility such as SMART Protocols [8] and REPRODUCE-ME [18] can be employed. At the same time these ontologies ideally are accepted in the community, maintained, and well documented itself. This set of requirements renders this task especially for researchers inexperienced with ontological modelling a complex task. Thus, (semi-)automatic services supporting the selection and mapping are needed. For the identification of the concepts of interest in the ELN protocols, techniques of natural language processing could potentially be employed to extract information from ELNs and support the semantic modelling. This, however, requires highly sophisticated methods due to the usually bad textual quality of ELN protocols when compared with experimental protocols. Information extraction from the latter shown to be successful [14,26].

Following the examples of Stocker et al. [29], and Samuel and König-Ries [19], the data analysis and its provenance could semantically be represented by use of jupyter notebooks, which were previously established [22] for the use case discussed here. The combination of the semantic description of the experiment and the data analysis would result in a provenance model covering more aspects of the investigation and could help to automatically check for validity and reproducibility.

Finally, modern techniques of knowledge graph based text generation [13] could be used to automatically create parts of the methods and materials section under restrictions of word count or granularity of the description.

References

1. Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M.H., Bug, B., et al.: The ontology for biomedical investigations. PLOS ONE **11**(4), e0154556 (Apr 2016). https://doi.org/10.1371/journal.pone.0154556

2. Bartusch, F., Hanussek, M., Krüger, J.: Automatic generation of provenance metadata during execution of scientific workflows. In: Atkinson, M., Gesing, S. (eds.) Proceedings of the 10th International Workshop on Science Gateways (IWSG 2018) (Jun 2018), http://ceur-ws.org/Vol-2357/paper8.pdf
3. Bose, R., Frew, J.: Lineage retrieval for scientific data processing: A survey. ACM Comput. Surv. **37**(1), 1–28 (Mar 2005). https://doi.org/10.1145/1057977.1057978, http://doi.acm.org/10.1145/1057977.1057978
4. Brack, A., Hoppe, A., Stocker, M., Auer, S., Ewerth, R.: Requirements analysis for an open research knowledge graph. In: Digital Libraries for Open Knowledge, pp. 3–18. Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-54956-5_1
5. Davidson, S.B., Freire, J.: Provenance and scientific workflows: Challenges and opportunities. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. p. 1345–1350. SIGMOD '08, Association for Computing Machinery, New York, NY, USA (2008). https://doi.org/10.1145/1376616.1376772, https://doi.org/10.1145/1376616.1376772
6. Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., et al.: ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Research **36**(Database), D344–D350 (dec 2007). https://doi.org/10.1093/nar/gkm791
7. German Council for Scientific Information Infrastructures (RfII): The data quality challenge. recommendations for sustainable research in the digital turn (Feb 2020), http://www.rfii.de/?p=4203
8. Giraldo, O.L., Castro, A.G., Corcho, O.: Smart protocols: Semantic representation for experimental protocols. In: LISC @ ISWC. pp. 36–47 (2014)
9. Gremse, M., Chang, A., Schomburg, I., Grote, A., Scheer, M., Ebeling, C., Schomburg, D.: The BRENDA tissue ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. Nucleic Acids Research **39**(Database), D507–D513 (oct 2010). https://doi.org/10.1093/nar/gkq968
10. Herman, I., Adida, B., Sporny, M., Birbeck, M.: Rdfa 1.1 primer - third edition: Rich structured data markup for web documents (March 2015)
11. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., et al.: Knowledge graphs (2020)
12. Kim, J.M., Choi, S., Kwack, K.H., Kim, S.Y., Lee, H.W., Park, K.: G protein-coupled calcium-sensing receptor is a crucial mediator of MTA-induced biological activities. Biomaterials **127**, 107–116 (May 2017). https://doi.org/10.1016/j.biomaterials.2017.02.038
13. Koncel-Kedziorski, R., Bekal, D., Luan, Y., Lapata, M., Hajishirzi, H.: In: Proceedings of the 2019 Conference of the North. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/n19-1238
14. Kulkarni, C., Xu, W., Ritter, A., Machiraju, R.: An annotated corpus for machine reading of instructions in wet lab protocols. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). pp. 97–106. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018). https://doi.org/10.18653/v1/N18-2016, https://www.aclweb.org/anthology/N18-2016
15. Moreau, L., Groth, P., Cheney, J., Lebo, T., Miles, S.: The rationale of PROV. Journal of Web Semantics **35**, 235–257 (Dec 2015). https://doi.org/10.1016/j.websem.2015.04.001

16. Murta, L., Braganholo, V., Chirigati, F., Koop, D., Freire, J.: noworkflow: Capturing and analyzing provenance of scripts. In: Ludäscher, B., Plale, B. (eds.) Provenance and Annotation of Data and Processes. pp. 71–83. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-16462-5_6
17. Peroni, S., Shotton, D.: FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. Journal of Web Semantics **17**, 33–43 (dec 2012). https://doi.org/10.1016/j.websem.2012.08.001
18. Samuel, S., König-Ries, B.: Reproduce-me: Ontology-based data access for reproducibility of microscopy experiments. In: Blomqvist, E., Hose, K., Paulheim, H., Ławrynowicz, A., Ciravegna, F., Hartig, O. (eds.) The Semantic Web: ESWC 2017 Satellite Events. pp. 17–20. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-70407-4_4
19. Samuel, S., König-Ries, B.: Provbook: Provenance-based semantic enrichment of interactive notebooks for reproducibility. In: The 17th International Semantic Web Conference (ISWC) 2018 Demo Track (2018)
20. Sarntivijai, S., Lin, Y., Xiang, Z., Meehan, T.F., Diehl, A.D., et al.: CLO: The cell line ontology. Journal of Biomedical Semantics **5**(1), 37 (2014). https://doi.org/10.1186/2041-1480-5-37
21. Schnell, S.: Ten simple rules for a computational biologist's laboratory notebook. PLOS Computational Biology **11**(9), e1004385 (sep 2015). https://doi.org/10.1371/journal.pcbi.1004385
22. Schröder, M., Staehlke, S., Krüger, F.: Encoding reproducible and executable experiments with jupyter notebooks. In: Reference Module in Biomedical Sciences. Elsevier (2020). https://doi.org/10.1016/b978-0-12-801238-3.11686-1
23. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., et al.: The obo foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology **25**(11), 1251–1255 (2007)
24. Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., et al.: Relations in biomedical ontologies. Genome biology **6**(5), R46 (2005)
25. Smith, B., Kumar, A., Bittner, T.: Basic Formal Ontology for Bioinformatics. IFOMIS Reports (2005)
26. Soldatova, L.N., Nadis, D., King, R.D., Basu, P.S., Haddi, E., et al.: EXACT2: the semantics of biomedical protocols. BMC Bioinformatics **15**(S14) (Nov 2014). https://doi.org/10.1186/1471-2105-15-s14-s5
27. Staehlke, S., Koertge, A., Nebe, B.: Intracellular calcium dynamics dependent on defined microtopographical features of titanium. Biomaterials **46**, 48–57 (Apr 2015). https://doi.org/10.1016/j.biomaterials.2014.12.016
28. Staehlke, S., Rebl, H., Finke, B., Mueller, P., Gruening, M., Nebe, J.B.: Enhanced calcium ion mobilization in osteoblasts on amino group containing plasma polymer nanolayer. Cell & Bioscience **8**(1) (Mar 2018). https://doi.org/10.1186/s13578-018-0220-8
29. Stocker, M., Prinz, M., Rostami, F., Kempf, T.: Towards research infrastructures that curate scientific information: A use case in life sciences. In: Proc. of DILS, pp. 61–74. Springer (Dec 2018). https://doi.org/10.1007/978-3-030-06016-9_6
30. Takahashi, A., Camacho, P., Lechleiter, J.D., Herman, B.: Measurement of intracellular calcium. Physiological Reviews **79**(4), 1089–1125 (Jan 1999). https://doi.org/10.1152/physrev.1999.79.4.1089