

Data Quality in Materials Science: A Quality Management Manual Approach

Thorsten Wuest, Jakub Mak-Dadanski, and Klaus-Dieter Thoben

BIBA - Bremer Institut für Produktion und Logistik an der
Universität Bremen, Bremen, Germany
{wue,dad,tho}@biba.uni-bremen.de

Abstract. Experimental research data in the materials science domain is often insufficiently described with regard to metadata and frequently displays an incoherent form of documentation. These circumstances often hinder current and future researchers significantly in reuse and comprehension of data. To support researchers during the archiving and provision of materials science research data (incl. supplementing material), a Quality Management Manual (QMM) approach as an established QM tool is proposed in this paper. Quality-assurance of experimental research data and the perpetuation of good scientific practice in provision and archiving research data are examined before the QMM approach is applied in a case study. The preliminary results indicate that QMM allows to provide practitioners basic guidelines which support integrity, availability and reusability of experimental research data in materials science for subsequent reuse.

Keywords: data quality, experimental research data, materials science, QM.

1 Introduction

Research data plays a central role in the scientific process and is the basis of scientific knowledge creation. However, without adequate documentation and corresponding metadata, research data cannot be understood and managed [1]. Neatly stored research data (incl. metadata) allows researchers to e.g., test their hypotheses, evaluate experimental results and disseminate the findings in a cohesive way that furthers the state of the art. Thus, a coherent and long-term archiving of research data supports the scientific process. Additionally, the quality assurance of research data can be considered to be essential for the competitiveness of the research location [2].

The focus of this work has been placed on the development of Quality Management Manual (QMM) guidelines, which are intended to support researchers during the archiving and provision of materials science research data and to complement the quality manuals of research institutes with the research data aspect. The usage of the above mentioned guidelines is supposed to ensure the quality-assurance of research data and simultaneously perpetuate the terms of good scientific practice in archiving of the created research data.

Research data generally represent the result of experiments and/or observations. These experiments and observations often consist of informal workflows, which may frequently exhibit differences. The so obtained data is often insufficiently described by researchers with regard to its metadata and displays an incoherent form of documentation. This condition hinders the current and future researchers significantly in the reuse and comprehension of the data [3], which in turn affects the scientific knowledge creation progress.

Following, the fundamentals of data quality as well as materials science research data will be succinctly addressed. On this basis, a framework for the model of quality assurance of materials science data will be created and presented in the following section. Thereafter, an exemplary guideline for documentation of materials science research data will be constructed. In the following discussion and limitations section, the developed solution is critically discussed. Furthermore, potential challenges concerning the implementation of quality assurance measures are illustrated. In the final section, the results are summarized and a short outlook on future work is presented.

2 State of the Art

In this section, a short overview of the basic terms and the current state of the art of Quality Management (QM) in the domain of research data management are presented. First, existing guidelines, in this case based on a German foundation, regarding the handling of research data are illustrated before the topic of research data in material science is highlighted in more detail. The last subsection is looking into data quality measures for research data.

2.1 Existing Guidelines of Research Data Management

The German Research Foundation (DFG) serves as the self-governing organization of science in Germany. On the basis of the association's issued recommendations for safeguarding good scientific practice [4] and recommendations for the secure storage and delivery of digital primary research data [5] a substantive framework for quality assurance of research data has been created. The recommendations state that securing and storing of primary (research) data should be guided by the rules of good scientific practice, as these form the basis of the scientific value chain. Research data should be kept for at least ten years in the institution where they originated. This is useful for efficiency reasons, since it allows the researchers to access the results of previous experiments or observations without having to repeat them. This takes also the often expensive (money and time efforts) creation processes into consideration. Additionally, the long storage time serves the preservation of evidence in the event the published results are ever doubted. Because each discipline is different when it comes to the needs and challenges in the area of research data handling, each domain should develop its own subject-specific approach to safeguarding research data [5]. The handling and storage of data should take place within (internationally) recognized and established standards. All information required for repeated use of the research data

should be included and stored [5]. Each research institution should, therefore, place clear rules for the storage of primary data and data carriers [4].

2.2 Research Data in Materials Science

For each scientific discipline research data are the result of an individual scientific value chain. Materials science is no exception. Research and development of materials are the main tasks in the field of materials science [6]. Experiments and tests examine the properties of materials under different conditions. The number of different test methods and the necessary testing machines make data gathering and data analysis partially demanding, since although the older machines are considered good and reliable, they often are not compatible with the latest software. Due to the high research worker turnaround of universities, it is particularly important to collect the relevant metadata and store it in a comprehensive way. Metadata provides the research data with context and describes the contents by means of controlled vocabulary [7]. It involves, in case of e.g. material delivery, information on the name of supplier, place of manufacture, chemical composition and possible prior heat treatment of the material.

2.3 Data Quality Measures for Research Data

The DIN EN ISO 8402 standard defines quality as "the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs" [8]. According to a survey by K. C. Tan et al., many companies define quality, as the fulfillment or surpassing of customer-specific product requirements [9]. It is generally accepted that the degree with which customer requirements are fulfilled represent the most decisive quality criterion [8]. However research data, unlike a physical product, is not produced in accordance to the quality specifications of a client, but in a particular framework of the scientific research question. The performing of experiments, the evaluation and analysis of generated data are an essential part of scientific research [10]. The subsequent use of the research data defines the framework of requirements for the quality of the research data. An important requirement for the quality of research data is to enable its use outside of the original research question, which originally led to the creation of the research data. To say with certainty how a research data set was generated, all input variables have to be known. The quality of scientific research data can therefore be measured by the extent to which anyone who wants to continue to work with these data objects may comprehend the circumstances of the data generation and the information gained from the data.

3 Background for the Development of a QMM Framework for Experimental Research Data

In this section, the background of the developed QM method is described before it is applied in the next section. First, a case study which frames the application is introduced before the QM procedure of choice, QMM, is elaborated in more detail.

3.1 Case Description

The scope of the conducted case study is limited to the Department of Structural Mechanics of the “*Foundation Institute of Materials Science*” in Bremen, Germany (IWT). The mechanical experiments carried out in 2011/2012 were subjected to an analysis (according to test type and frequency). 47% of these experiments formed the tensile test. It was thus chosen for the creation of the requirements for the research data. On the basis of the evaluated test figures, and with the aid of expert interviews with scientific staff of the IWT, a picture of the state of data-handling in IWT has been created.

The projects are financed from public funds, or performed within an industrial partnership. The project defines many input variables for a series of experiments. The sample dimensions are therefore project-dependant. The tensile test testing machine generates tension values, which, through software interfaces, are evaluated to first characteristic values and converted to an Excel data set. The researcher analyzes the data on his or her workstation, PC or laptop, where the copies of the converted Excel data sets and framework data are also individually stored. Information on the chemical material composition, sample size, heat treatment conditions and surface characteristics are required for a complete analysis of the sample. However, since the data are located at different places and with different researchers, it is very difficult to access the original research data and metadata among older (> 6-8 years) publications.

3.2 QMM and Procedural Instructions

A QMM is the central document of every QM system (QMS). It describes the quality policy and its scope, and regulates the responsibilities and powers of quality assurance measures [11]. The function of the manual is that of a reference book [12]. All necessary contents of a QMM are generally defined in ISO 9001:2008 [13].

The transparency, the ability that a third party can understand and comprehend the processes and procedures in an organization is a central point of each QMS. One can transfer this perspective on to the quality of research data. Information and data necessary for repeated testing are the minimum requirements considering the quality of research data. A viable solution would be to pass all input parameters that can be captured and measured with a plausible effort (in terms of efficiency). Every relevant input parameter, which has not been passed on, reduces the reproducibility of the research data and thus reduces the data quality. According to ISO 9001:2008 clause 4.2.3 Control of Documents and 4.2.4 Control of Records [13], procedural instructions can be used as in-house regulations for implementation of quality requirements for research data. They can be used to establish clear structures of information transmission [14] and thus support linkage of relevant research data documents.

4 Exemplary Application of QMM

In this section the case study with an exemplary application of the previously described QMM method is presented. First, the scope of the method application within

the case is detailed before the specifications are elaborated briefly. Subsequently, an exemplary reference sheet illustrates the application in practice.

4.1 Scope of Method Application and Specification

The procedural instructions regulate certain cross-divisional process flows [14]. In the presented case, these are designed to ensure the processes of archiving and provision of materials science research data are based on clear criteria and contribute to quality assurance of the research data. An exemplary procedural instruction has been created, designed to support the documentation workflow of tensile tests for testing metallic materials (as illustrated in Figure 1).

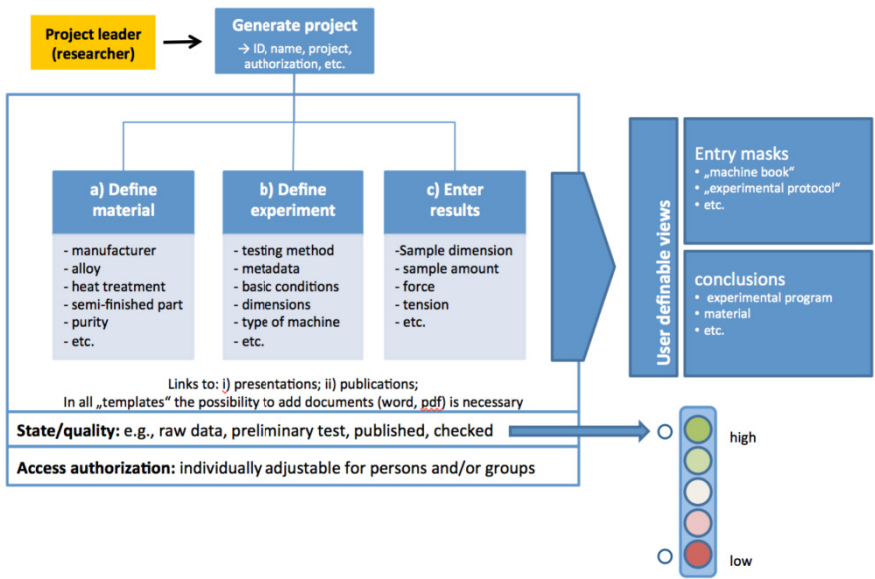


Fig. 1. Illustration of data model from user perspective

Research data management needs to include monitoring of the entire process of scientific work. This can be achieved through description of the necessary steps into the procedural instructions. Necessary for a completeness of a data set are (among many other factors) project description, the affected actors (e.g. researcher and technician), their tasks, and responsibilities along the process chain, material and experiment description. The in Figure 1 illustrated data model presents, as such, all the essential data required for a quality assured materials science data set. The researcher defines the framework of the experiment and the accompanying metadata. The technician performs the experiment, and forwards the results to the researcher. The traffic lights in Figure 1 indicate the 'completeness' of the data set and its quality.

The aim of procedural instructions in the QMM would be to define the scope and system boundaries of research data generation, as well as ensuring a link between all relevant input variables (e.g. sample size) and metadata (e.g. project number, material composition, heat treatment) necessary to recreate the test results, or at least to repeat critical processes. Following points of a procedural instruction could be determined and adapted [11; 14]: Objective of the instruction; Scope; Terminology; Description; References to documents; ID-Key; Responsibilities; Contact person for the process.

4.2 Exemplary Reference Sheet of a Procedural Instruction

Figure 2 (based on: [14; 15]) illustrates the structure of an exemplary reference sheet of a procedural instruction designed for documentation of the tensile testing of metallic materials. The top header of the pattern sheet contains general document information. The second header contains the version, the title of the instruction, and page numbers. The main area includes the previously identified elements of the instruction. The bottom line of the sheet contains information on the person responsible for the publication of the document, date of publication, information on the person who approved the document and the date of approval.

Quality Manual	Procedural Instruction		Annex
Version 0	Title: Documentation of the tensile test for metallic materials testing		Page 1 of 1
Documentation of the tensile test for metallic materials testing 1. Objective of the instruction 2. Scope 3. Terminology 4. Description 5. References to documents 6. ID-Key 7. Responsibilities 8. Contact person for the process			
Published by:	Date:	Approved by:	Date:

Fig. 2. QM, procedural instruction reference sheet

5 Discussion and Limitations

Compliance with the above described exemplary procedural instruction can contribute to the quality assurance of the research data, as laid down by the DFG recommendations. A quality assurance representative would be responsible for the implementation of the procedural instructions into the documentation of the quality system, management and distribution among the affected employees. In terms of the continuous improvement of processes and workflows in the QMS, the procedural instructions should be continuously maintained and developed. This could be done by systematic routine internal auditing, according to the ISO 19011:2011 guidelines.

To ensure the success of the procedural instructions, it is essential that the affected employees recognize and use them regularly. Employee participation is indispensable and the basis of the continuous improvement process [16]. Therefore it is important to convey to the employees the positive effects of a complete documentation, the procedural instructions, and the benefits of process transparency on the quality of the research data [17]. By knowing where research data and documents can be found, unnecessary duplication of effort can be avoided and access to information is improved [18]. The procedural instructions and included regulations, should be checked regularly and developed simultaneously with the affected employees. It is necessary to always remain critical, while examining existing processes and workflows [17].

6 Conclusion and Outlook

Missing format requirements lead to inadequate description of the research data with metadata. This condition complicates the re-use of research data and thus reduces their quality. The quality of research data results from the ability to use the data under conditions different from the original scientific question. Only when this condition is reached in the documentation of the data, is the research data ready for sustainable use. Using the example of the Research Institute IWT a procedural instruction has been identified, which supports quality-assured data archiving, and -provision of materials science research data in a QMM. During the case study, it was found that the employees have no internal regulations for quality assurance of research data. Due to the inconsistencies in the storage of research data (incl. metadata), the retrievability of the original research data for publications (after 6 to 8 years) may be considerably hindered. There is a need for improvement, since it is apparent from the DFG recommendations that research data should be stored for at least 10 years.

The presented template of a procedural instruction can be added to the annex of an existing QMM and, in form of an in-house regulation, support it. The solution described here has been specially tailored to data generation of tensile testing in department of structural mechanics of IWT Bremen. Future studies may address the transfer to further processes and test methods.

The implementation of the Dublin Core standard for the integration of the metadata in the project documentation and a solution for potential national and free provision of research data, in terms of Open Access principles, would further benefit the documentation and quality of the research data. The creation of a documented process for updating the procedural instruction is another open issue. The procedural instructions should be maintained and further developed, within the terms of continuous improvement of processes and workflows of a QM system, e.g. through systematic routine internal auditing, according to the ISO 19011:2011 guidelines.

Acknowledgement. This work was funded by the “Deutsche Forschungsgemeinschaft” via the project “InfoSys”. The authors would like to extend their gratitude to project partners and funding party for their generous support.

References

1. Enke, H., Fiedler, N., Fischer, T., Gnadt, T., Ketzan, E., Ludwig, J., Rathmann, T.: Leitfaden zum Forschungsdaten-Management, WissGrid Deliverable 2.3.1, Version 0.6, WissGrid (2011)
2. Winkler-Nees, S.: Der Umgang mit Forschungsdaten in Wissenschaft und Lehre. DFG, Bad Honnef (2010)
3. Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S.: nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, M. Jehn (Hrsg.). Verlag Werner Hülsbusch, Boizenburg (2009)
4. DFG: Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission 'Selbstkontrolle in der Wissenschaft, Deutsche Forschungsgemeinschaft, Weinheim, Wiley-VCH (1998)
5. DFG: Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten, Deutsche Forschungsgemeinschaft, Bonn (2009)
6. Acatech: Materialwissenschaft und Werkstofftechnik in Deutschland, Empfehlungen zur Profilierung, Lehre und Forschung, acatech – Deutsche Akademie der Technikwissenschaften, Fraunhofer IRB Verlag, Stuttgart (2008)
7. Severiens, T., Hilf, E.R.: nestor: Langzeitarchivierung von Rohdaten”, nestor Materialien 6, nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland (2006)
8. Spur, G., Stöferle, T.: Handbuch der Fertigungstechnik, Band 6 Fabrikbetrieb. Carl Hanser Verlag, München (1994)
9. Tan, K.C., Kannan, V.R., Handfield, R.B., Ghosh, S.: Quality, manufacturing strategy, and global competition: An empirical analysis. *Benchmarking: An International Journal* 7(3), 174–182 (2000)
10. Blumbergs, I., Kleinhofs, M., Chatys, R.: Experimental Research Data Series Quality Analysis, XIX Intl. Scientific-Technical Conference, pp. 96–98 (2011)
11. Linß, G.: Qualitätsmanagement für Ingenieure: mit Handbuch Qualitätsmanagement, 2. Auflage, Carl Hanser, München Wien (2005)
12. Rothery, B.: Der Leitfaden zu ISO 9000; mit QM-Musterhandbuch und Erläuterungen, Carl Hanser Verlag, München (1994)
13. DIN EN ISO 9001:2008: Quality management systems – Requirements (2008)
14. Schmidt, S.: Das QM-Handbuch: Qualitätsmanagement für die ambulante Pflege, 2. Auflage. Springer, Heidelberg (2010)
15. Glaap, W.: ISO 9000 leichtgemacht: Praktische Hinweise und Hilfen zur Entwicklung und Einführung von QS-Systemen. Carl Hanser, München (1993)
16. Torre, P.G., Adenso-Díaz, B., González, B.A.: Empirical evidence about managerial issues of ISO certification. *The TQM Magazine* 13(5), 355–360 (2001)
17. Petzoldt, J., Schorcht, H., Haaßengier, C.: “Qualitätsmanagement für Lehre und Forschung: Erfahrungen der Technischen Universität Ilmenau. *Beiträge zur Hochschulforschung* 30.1, 74–93 (2008)
18. Jones, S., Ball, A., Ekmekcioglu, C.: The Data Audit Framework: A First Step in Data Management Challenge. *The International Journal of Digital Curation* 3(2), 112–120 (2008)