W7-X Logbook REST API for processing experimental metadata and data enrichment at the Wendelstein 7-X stellarator

M. Grahl, A. Spring, T. Andreeva, T. Bluhm, S. Bozhenkov, S. Dumke, J. Geiger, O. Grulke, M. Grün, A. Holtz, U. Höfel, H. Laqua, M. Lewerentz, H. Riemann, J. Schilling, A. von Stechow, J. Svensson, A. Winter, The W7-X Team

Max-Planck-Institut für Plasmaphysik, Wendelsteinstr. 1, 17491 Greifswald, Germany

Contextual metadata becomes an important factor for the work with research data, which is created in large amounts at fusion experiments. In its first operational phases, the stellarator experiment Wendelstein 7-X (W7-X) produces about 500 TB of experimental data in only a few weeks of scientific campaign. A new central logbook software was implemented for the processing of contextual metadata at W7-X. The bulk data of W7-X experiments can be enriched with additional information, such as comments and tags for categorizing. Since its introduction, many enhancements have been implemented based on user feedback. This was possible due to a flexible software architecture. By using the logbook REST API, users can integrate logbook requests into their own software and programmatically add new content. The logbook quickly became a crucial tool for W7-X operation and is now considered a central hub for all experiment-related information. This paper describes the implementation of the W7-X logbook and the experience with the integration of metadata via REST API.

1. Introduction

The optimized stellarator Wendelstein 7-X (W7-X) successfully completed its first three operational phases (OPs). In a large cooperation with scientists and engineers from 40 institutes around the world, the W7-X team has been working on a variety of research topics. Next to commissioning of the machine and diagnostic systems, the program was already covering a wide range of scientific topics. About 40 different diagnostic systems were covering the experiments. Even in these early stages with a not-fully-equipped machine, the experiments showed very promising results [1]. Due to a tight schedule for the completion of the machine, only a small number of experiment days could be used for physics experiments. Nevertheless, up to 60 experiment programs (XPs) per day and about 1000 XPs per OP could be conducted. The diagnostics and technical systems produce about 500 terabyte of experimental data. Typical plasma durations where 10 seconds, while a small number of discharges had a length of 100 seconds. For demonstrating steady-state operation, plasma durations of up to 30 min are planned for upcoming OPs. Therefore, a large flexibility is needed to satisfy the different use cases and requirements of such an experimental environment, esp. for data processing and analysis [2]. Thereby, contextual metadata is an important factor for efficient research, as well as for the traceability of scientific findings, as large machines often have a lifespan of several decades. For this reasons, an electronic logbook software [3] was developed for the support and documentation of Wendelstein 7-X operation and a cooperative aggregation of contextual metadata.

This paper is divided into three parts. The first part introduces experimental metadata and highlights the importance of it for large fusion experiments like W7-X (sec. 2). The second part describes the implementation principles of the W7-X Logbook (Sec.3) and especially its web service interface (sec. 4). This includes examples for data enrichment for operation and analysis use cases. Finally, section 5 discusses the experiences in the first operational phases of W7-X and gives an outlook on future developments.

2. Experimental metadata at W7-X

At W7-X, contextual metadata is used for the traceability of experiment operation and therefore the exploration of mass data produced by the data acquisition systems. In the first operation phases the machine monitoring and diagnostic systems of W7-X produced about 500 TB of experimental data. These systems include video cameras and high-speed ADCs. Moreover, measurements are often done on multiple machine locations for covering the three-dimensional geometry of a stellarator. Since each W7-X component (i.e. diagnostic or machine monitoring system) is a complex system by itself, additional information is needed for the interpretation of the measured data. Examples for such metadata are system setup and preferences, calibration parameters, or information about different versions of data. Especially in plasma physics, the importance (and often the extent) of this expert knowledge can be extensive since key plasma parameters (e.g. electron / ion temperature, etc.) are determined in an indirect way by the diagnostic systems. In addition, the data analysis often needs cross-calibrations with other measurements or complex physics models. The conservation and providence of this expert knowledge as contextual metadata in a decent form is an

author's email:michael.grahl @ipp.mpg.de

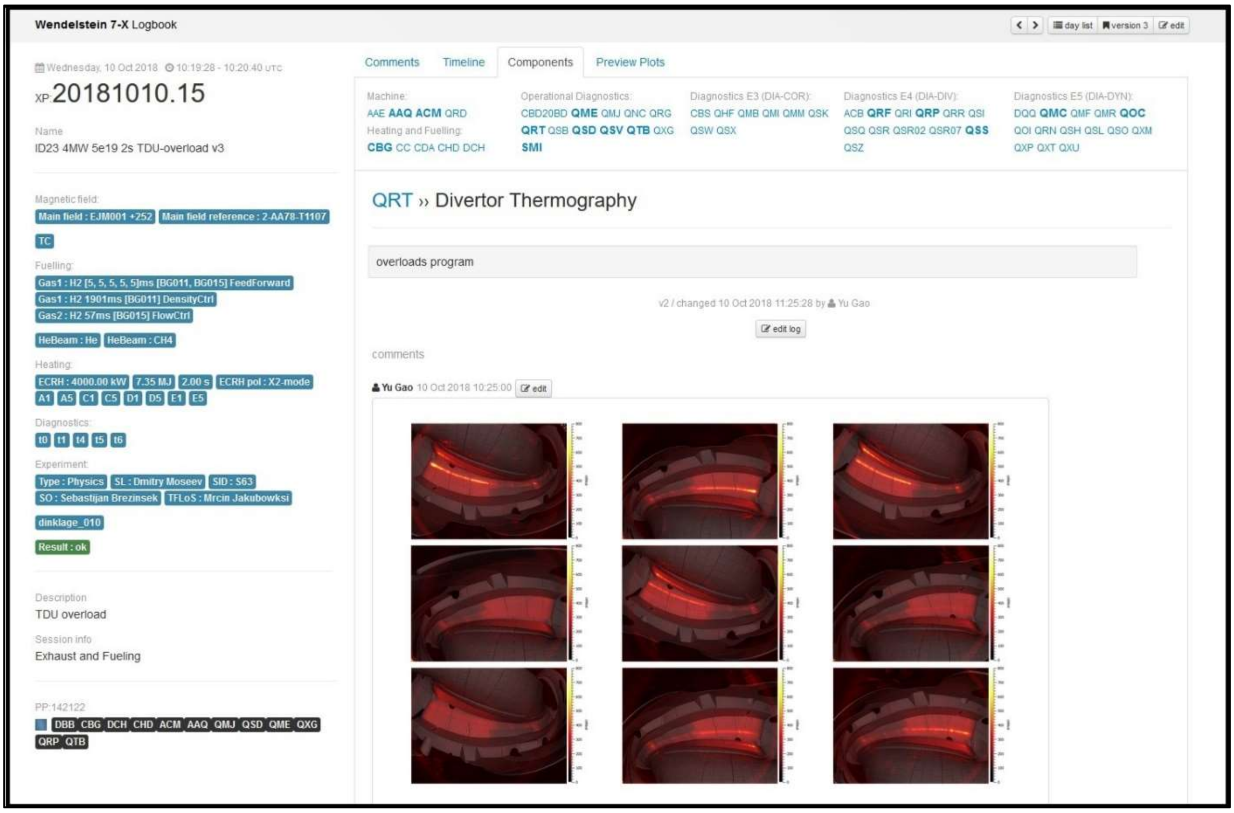

Fig. 1 - Logbook webpage for XP log showing the top-level attributes on the left, e.g. program id (XP_20181010.15), name, description, and tags. The tags are grouped by category (magnetic field, fueling, heating, diagnostics, and experiment), as well as by component (e.g., tags for gas fueling: main valve system and helium beam). The diagnostic logs for this experiment are shown on a separate tab (right).

important factor for the evaluation of measurements. This is especially important at long running, large-scale experiments with many contributors, which is typical for fusion research [4] [5]. At the same time, doing this in a consistent way is a challenging task due to the complexity of the machine and the variety of steadily evolving diagnostic systems.

3. The W7-X Logbook

The W7-X Logbook was developed for the processing of Wendelstein 7-X metadata. It was designed as central knowledge-sharing software for the W7-X team and it can be used interactively via a web browser (Fig 1.) or integrated into other software via a web service interface. The quantity of components and the complexity of the experiment makes it challenging to provide a user interface with the right level of detail for different user roles and use cases. Therefore, the logbook provides data presentations for different abstraction levels. A general experiment view containing top-level information (incl. magnetic coils, heating, fueling) and a component-specific view for details of each machine subsystem or diagnostic. The majority of experiment-specific metadata is represented in tags, which can be chosen from a catalog. This allows the consistent usage of tags by the whole team.

3.1 Domain model

Considering the experimental environment of W7-X, a leading requirement of the logbook software is to allow flexible processing of experimental metadata in an extensible way. Therefore, the domain model of the application was designed as shown in Figure 2. The main objects of the logbook domain model are *Logs* and *Comments*. *Comments* are used for free-form user remarks, which are associated to a log by the log id. Comments can include formatted text (HTML) and multimedia elements, such as images or videos. On the other hand, *Log* objects are used for structured metadata with a pre-defined schema. This makes them more suitable for automated processing and analysis. Therefore, the log objects hold many significant values in an array of *Tag* objects. In principle, tags are key/value pairs, which have several additional attributes. Next to a description and a unit for numeric values, further attributes are used for classification.

In the logbook domain model, different log objects represent different uses cases of W7-X operation. Experiment program (XP) logs representing a top-level overview of W7-X segment programs, including plasma discharges, as well as central tests and commissioning of the whole machine. In addition, separate component logs for these programs contain information for each

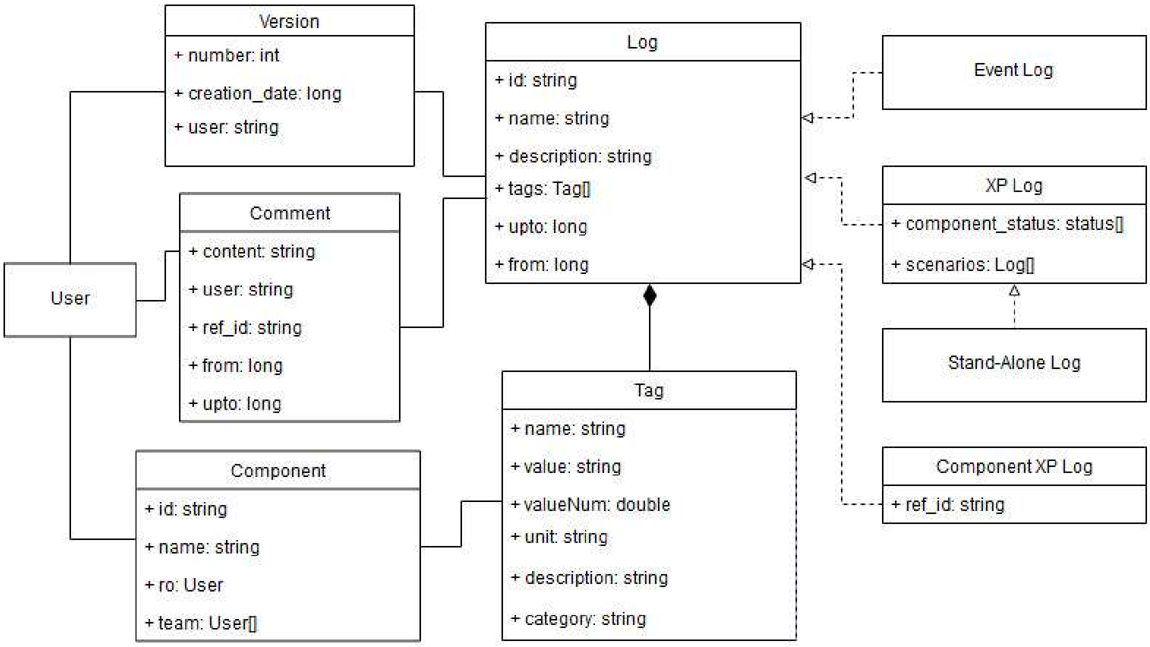

Fig. 2: Logbook domain model

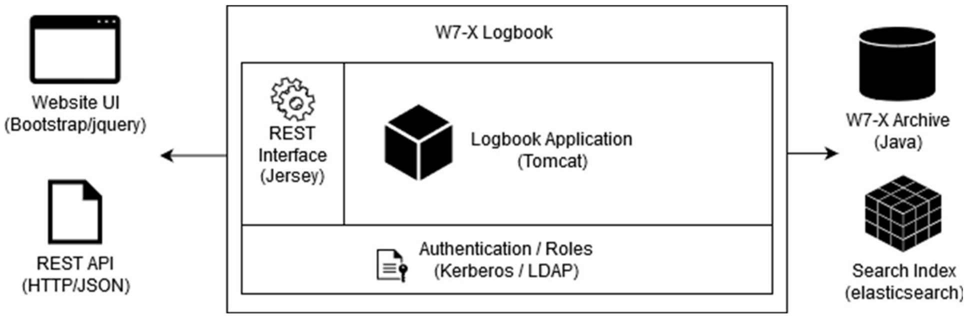

Fig. 3 - Logbook architecture top-level view

component. Every log object contains standard attributes: name, description and a start and end date. The derived XP log contains additional attributes provided by the W7-X segment control [6] e.g., component statuses and segment specific information. Next to XP logs, stand-alone logs are generated by the W7-X segment control for autonomous programs of single W7-X components. Finally, machine- or diagnostic operators can create logs for other events. This includes test measurements and calibration runs, as well as system failures or downtimes. Such event logs are often used for daily overviews, containing the diagnostic settings for an experiment day. Version control is implemented for all log types to keep track of changes, as well as for comparing planned programs with the actual results.

Relations between logs and comments are defined by identifiers and timestamps. For timestamps, the logbook uses 64-bit integer values, which are also used in the W7-X archive [7]. For the primary key of a component, the logbook uses identifiers, which were assigned by the W7-X project, e.g. *QME* for the electron cyclotron emission diagnostic. Component IDs are used as a secondary key in logs and comments (attribute *ref_id*), as well as for association in tags. User information (id, name) and roles (e.g., component team, session leader) are handled by the central user database of the institute, which is integrated by Kerberos and the lightweight directory access protocol (Fig. 3). For persistence, the different comment types (experiment and for each component) and logs types (W7-X experiments, W7-X operation events, and for each component) are stored in separate datatypes (i.e. equivalents to *tables* in relational databases) in the back-end databases. The front end is implemented as a RESTful web service, where users can request representations of these comment and log objects.

4. REST API

Several essential functions for W7-X research are implemented as web services for providing them in a convenient and consistent manner [8]. Web services can be easily integrated into other software due to the usage of standard protocols and web techniques, such as representational state transfer [9] application programming interfaces (REST APIs). REST describes a

Method	Path	Action
POST	/api/comment/XP	Create a new experiment log comment
GET	/api/comment/XP/Abcd3f	Read the comment with the id *Abcd3f*
POST	/api/comment/XP/Abcd3f	Update comment with the id *Abcd3f*
DELETE	/api/comment/XP/Abcd3f	Delete comment with the id *Abcd3f*

Fig. 4 – Listing of HTTP request methods for CRUD (create, read, update, delete) operations on experiment comments.

client/server architecture with resources located on a server and clients interacting with these resources by using a limited set of operations. Nowadays, REST via hypertext transfer protocol (HTTP) is the most common implementation for distributed software architectures. Thereby, HTTP operations are performed on a uniform resource locator (URL) for the interaction with resources, which are represented in a common message format (mostly JavaScript Object Notation/JSON).

The W7-X Logbook provides a REST API for accessing the JSON representation of logs and comments via HTTP operations (Fig. 4). Reading of an experiment log can be done via a *GET* request on the URL built from a base path and the corresponding log ID (Fig. 5). In the same way, *POST* requests are used to create new content and *DELETE* is used for deleting entities. In the logbook API, any read operation can be done without authentication. For creating or change content via POST or DELETE operations, the user needs to provide credentials in the request header via basic authentication or API key. API keys are useful for the integration of logbook requests in automatic analysis, as users do not have to put their passwords in scripts. The workflow for logbook requests with API keys is shown in Fig. 6. The user requests an access token from an API key endpoint and use it for a limited time for all requests, which need authentication. The logbook API key implementation is based on JSON Web Token [10] standard.

In addition to read operation by ID, advanced search requests are implemented with the help of an integrated search index and a query domain specific language (DSL).

4.1. Search API

Central part of the logbook is a search index, implemented with the help of elasticsearch [11]. The search index offers near real-time responses times, so most search requests are answered within a couple of milliseconds. The search engine allows indexing of new content on insert. Therefore, new entries are indexed as a part of the write request and the corresponding index is updated within a second. Search requests are done to a separate endpoint with the criteria formulated in a simple query DSL as a part of the URL parameters. Search options cover free-text search over all objects and fields, as well as searches in specific fields. Range searches can be used for numeric fields, for example to get all experiment programs with a

```python
import requests, json, getpass

base = 'https://w7x-logbook.ipp-hgw.mpg.de/api/'
log_id = 'XP_20181010.15'

url = base + 'log/' + log_id
log = requests.get(url).json()

print('program name: ', log['_source']['name'])
# program name: ID23_4MW_5e19_2s_TDU-overload_v3

print('number of tags: ', len(log['_source']['tags']))
# number of tags: 32

my_comment = {'content' : 'Hello world!','ref_id' : log_id,\
              'from' : log['_source']['from']}
d = json.dumps(my_comment)

url = base + 'comment/XP'
creds = (getpass.getuser(), getpass.getpass())
res = requests.post(url, json=d, auth=creds)

print('Status code: ', res.status_code)
# Status code: 201
```

Fig. 5 – Logbook API requests in Python.

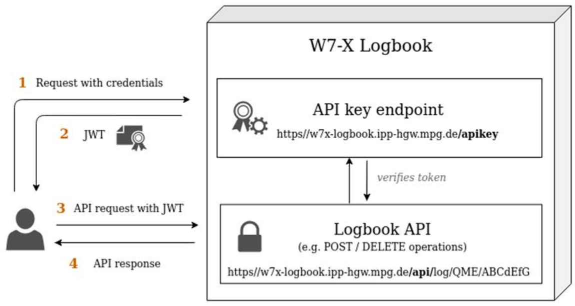

Fig. 6 – Logbook API key workflow.

specific duration or heating power. The query DSL allows complex search requests over all comments and logs with an easy-to-use syntax, which includes wildcards and logical operators. For example search criteria can be combined via *AND* / *OR* and excluded via *NOT*. The results array contains the IDs of matching logs and comments, but can also be customized (number of results, included fields, etc.) by using additional query parameters.

4.2. Integration into W7-X software

The REST API together with its fast search index allows any W7-X software to enrich its workflow with logbook data in a convenient way. Therefore, the logbook became a central component in the W7-X CoDaC software tool chain. The CoDaC software components jointly support the scientific workflow, starting from experiment proposals, to the design of the experiment sequence, program execution and monitoring, as well as processing of the results. For the main content of the logbook, the central W7-X control software automatically extracts metadata from the planned programs and logs it during experiment execution. This reduces the need of manual input on a large scale, as these logs already contain many characteristic information of W7-X experiments. It describes general experiment sequences (e.g. timing of fueling/heating systems), as well as organizational information and physics expectations. As the REST API allows the usage of logbook content in flexible way, it is integrated into CoDaC software in various ways. The search functions of several CoDaC applications are implemented by using the logbook API. The logbook websites contains links to the experiment proposals, while the proposals databank links to the logbook for showing related executed programs. In addition, the central tag catalog can be read as JSON for programmatically usage in client software. For data analysis, the processing of the experimental data from the W7-X archive is possible via the Archive Web API [12]. As the logbook uses the same timestamp format, a consistent integration of data from both systems can be done via time-aware hyperlinks. Therefore, the logbook website shows on-the-fly-generated plots from archived data on the log page of every experiment and the logbook component pages provide related archive links for relevant times (e.g. calibrations). In the first operational phases, the analysis of W7-X experimental data is implemented in various ways and many of the used software highly benefit from the integration of logbook data. Analysis are implemented as individual codes (e.g., in Python, Matlab, C++), or as part of the Minerva [13] framework (Java). Minerva is a Bayesian modelling framework, which simplifies the combined analysis of data from different diagnostics while ensuring consistency. In OP1.2b it was used to analyze data e.g. from Thomson scattering, electron cyclotron emission (ECE), or magnetic equilibrium diagnostics. Typical calculations can include several Minerva nodes or data from other W7-X web services such as logbook and archive API or physics models (e.g. plasma equilibria or the magnetic coil geometry). Thereby, several Minerva nodes use logbook data for parametrization of their analysis, e.g. the power and polarization of gyrotrons for ECE calculations. In addition, the logbook API was often used for the triggering of analysis calculations, as the first version of a log is available immediately after the start of an experiment. In this way, preliminary results were already available during or shortly after an experiment. The resulting data is displayed on the corresponding logbook pages as part of the integrated plotting feature or as images with overview plots, which were added in the comments.

5. Experiences and outlook

During the first three operational phases, the logbook became a central tool for W7-X operation. Especially the possibility to share information in an instant manner was appreciated by team members within the control room, as well as by remote participants. The users were able to integrate the logbook into their individual workflows: e.g. by adding comments and plots manually via web browser or programmatically as part of their own codes. This led to a collaborative creation of comprehensive overviews of W7-X experiment runs, which were considered as very helpful for the orientation during the campaign and afterwards. The development of the software was done in close collaboration with the members of the W7-X team and updates to the software could be added in short iterations, even during the operational phase. By the time of writing, the W7-X team wrote over 21000 comments and 37000 logs for the 4430 experiment programs of the first 3 operational phases. The Logbook is considered as the central hub for W7-X experiment related information and starting point for analysis activities. The combination of web service techniques and a flexible database schema has proven to be very beneficial, especially when working in such a heterogeneous system landscape. It endorses evolutionary development and a large flexibility, which are key factors in an experimental environment.

For coming operational phases, several improvements and new features are planned for the logbook. This includes updates of the tag catalog and the introduction of new tags. Beyond tags, further extensions of the logbook domain model are planned, which will allow custom data structures for each component, e.g., for complex metadata from video diagnostic systems [14]. In addition, several updates for the logbook website are planned. For example, a new comparison view for experiment logs, which highlights the difference between intended program and the actual results.

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

References

[1] T. Klinger, et al., Overview of first Wendelstein 7-X high-performance operation, Nucl. Fusion 59 (2019), 112004

[2] A. Spring, et al., Establishing the Wendelstein 7-X steady state plasma control and data acquisition system during the first operation phase, Fusion Eng. Des. 123 (2017), pp. 579-583.

[3] M. Grahl, et al., The new W7-X logbook – A software for effective experiment documentation and collaborative research at Wendelstein 7-X, Fusion Eng. Des. (2019), https://doi.org/10.1016/j.fusengdes.2019.02.052

[4] J. Stillerman, et al., Scientific data management with navigational metadata, Fusion Eng. Des. 128 (2018), pp. 113 – 116.

[5] M. Greenwald, et al., A metadata catalog for organization and systemization of fusion simulation data, Fusion Eng. Des. 87 (2012), pp. 2205-2208

[6] A. Spring, et al., A Metadata Framework for Assisting Experiment Planning and Evaluation at W7-X, in *IEEE Trans. Plasma Sci.*, doi: 10.1109/TPS.2019.2958034

[7] Ch. Hennig, et al., ArchiveDB—Scientific and technical data archive for Wendelstein 7-X, Fusion Eng. Des. 112 (2016), pp. 984 - 990.

[8] M. Grahl, et al., Web Services for 3D MHD Equilibrium Data at Wendelstein 7-X, IEEE Transactions on Plasma Science, 46 (5) (2018), pp. 1114-1119

[9] R. Fielding, Architectural Styles and the Design of Network-based Software Architectures, PhD thesis, University of California - Irvine (2000)

[10] JSON Web Token: https://tools.ietf.org/html/rfc7519

[11] Elasticsearch: https://www.elastic.co/de/products/elasticsearch

[12] M. Grahl, et al., Archive WEB API: A web service for the experiment data archive of Wendelstein 7-X, Fusion Eng. Des. 123 (2017), pp. 1015-1019.

[13] J. Svensson and A. Werner, Large scale Bayesian data analysis for nuclear fusion experiments, IEEE Int. Symp. on Intelligent Signal Processing pp 1–6 (2007)

[14] A. Puig Sitjes, et al., Wendelstein 7-X near real-time image diagnostic system for plasma-facing components protection, Fusion Science and Technology 74 (1-2) (2018), pp. 116-124