DELIVERABLE

D4.5 EIDA Documentation System

Work package	WP4 Expanding access to the European seismic monitoring infrastructure
Lead	KNMI
Authors	Reinoud Sleeman (KNMI), Javier Quinteros (GFZ), Jarek Bienkowski (KNMI), Helle Pedersen (CNRS), Christos Evangelidis (NOA), John Clinton (ETH), Constantin Ionescu (INFP)
Reviewers	Florian Haslinger, ETH
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	30.04.2020
Submission date	27.04.2020
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D4.5_EIDA_documentation_system.pdf

Table of Contents

Summary

The EIDA infrastructure is a complex, distributed system with different stakeholders in terms of maintenance, operations, management and usage. Documentation of such a complex system is an essential requirement for a sustainable future for EIDA in which expansion is foreseen in number of data, data types, extended services and increased usage. Currently, information is available in different forms and locations, often maintained by one person, and can be public or restricted. To benefit from the distributed resources in EIDA we set up a framework for writing, adding and editing documentation by multiple authors. This framework is based upon Sphinx, GitHub and Read the Docs that allows members of the various stakeholders to write, contribute and edit specific, version controlled, documentation.

1   Introduction

We give an overview of SERA WP4 and its objectives to provide the background for this report: SERA deliverable "D4.5 EIDA Documentation System". This introduction is followed by the description of the documentation system.

SERA WP4

The objectives of SERA WP4 (NA2) are focused on the strengthening and extending of the existing European infrastructure for seismological waveform data from less covered regions as well as of new types of data. Notably the objectives are to:

- foster the integration of new data from the Euro-Mediterranean region into ORFEUS EIDA with a focus on the Balkans through the existing nodes in the region (INFP, NOA).
- facilitate data integration from other regions like North-East Europe (e.g. Belarus, Scandinavia), South-West Europe (Spain, Portugal) and possibly North Africa.
- promote extension of metadata models to make the EIDA system ready to accommodate more multidisciplinary time series data.
- create and maintain up-to-date end user documentation, system documentation and training and outreach material to preserve and share knowledge and to foster sustainability of the system.
- support new EIDA nodes and data centers in preparing (meta-)data for the EIDA infrastructure and promote best practices.
- establish a robust and reliable federated European data center that provides open and transparent access to the widest amount of quality controlled seismic data with the smallest delay.

To reach these goals, SERA WP4 is structured in 4 tasks:

Task 4.1 "Expanding open access to seismic data" by conducting workshops, technical meetings and demonstrating the benefits of data integration in EIDA. Deliverable "D4.1 Outcome of regional technical EIDA meetings" presents the different activities that took place.

Task 4.2 "Expanding open access to non-seismic data" by integration of special data types from Near Fault Observatories, Volcano Observatories, induced seismicity observation networks, structural monitoring arrays, etc. Required modifications or extensions of current metadata models are reported in the impressive deliverable "D4.2 Report on metadata challenges and proposed solutions", while deliverable "D4.4 EIDA Metadata model standards" describes the challenges of and progress in the implementation of new metadata in the current metadata for seismic stations and networks.

Task 4.3 "EIDA Infrastructure Knowledge System" targets the necessity of an information system for EIDA documentation, training material and best practices. Deliverables "D4.3 Best practices guidelines" and "D4.5 EIDA Documentation System" report about the activities and the current status.

Task 4.4 "Future networks design" explores the strategies for future deployment and processing, which are reported in deliverables "D4.6 Key scientific questions for future network design" and "D4.7 Strategies for future network design".

ORFEUS

The existing European infrastructure for seismological waveform data is coordinated by ORFEUS (Observatories and Research Facilities for European Seismology; https://www.orfeus-eu.org/), a collaborative, non-profit foundation that promotes seismology in the Euro-Mediterranean area through the collection, archival and distribution of digital seismic waveform data, metadata and derived products. The ORFEUS infrastructure comprises European seismological observatories, seismic networks, research communities, waveform data archives and data services. ORFEUS is one of the three infrastructures of the Thematic Core Service for Seismology within the European Plate Observing System (EPOS Seismology). Hence, all ORFEUS services are developed in coordination with EPOS and are EPOS compliant.

Among the goals of ORFEUS are: (a) the development of waveform data products; (b) the coordination of a European federated data distribution system, and the support for seismic networks in archiving and exchanging seismic waveform data; (c) fostering standardization of services and products; (d) the adoption of best practices for seismic network operation, data quality control and data management; (e) the promotion of open access to seismic waveform data, products and services for the broader Earth science community. These goals are achieved through the development and maintenance of services targeted to a broad community of seismological data users.

Two Service Management Committees (SMCs) are established within ORFEUS with the tasks to manage, operate and develop: (a) the European Integrated waveform Data Archive (EIDA; https://www.orfeus-eu.org/data/eida/); and (b) the European Strong-Motion databases (SM; https://www.orfeus-eu.org/strong/).

EIDA

EIDA is organized and managed by the EIDA Management Board (EMB), which is in effect an ORFEUS working group. The EMB is responsible for the management of EIDA. Each EIDA node is represented in the EMB. The ETC (EIDA Technical Commission) operates under the EMB and guarantees EIDA operations. The EMB and ETC continually address challenges such as efficient data management, content metadata, quality maintenance, provenance, and access services in a distributed network of archives.

To ensure sustainability and efficient maintainability of the complex EIDA infrastructure the setup of a documentation system that holds up-to-date collections of documents to describe the EIDA infrastructure (e.g. technical requirements, installation, limitations, operations, maintenance, management, procedures, user tools) is a key element.

FDSN – StationXML

ORFEUS and IRIS are currently working with technical writers from ISTI to provide good documentation on StationXML. Currently, the SEED manual was the most comprehensive document to describe seismic station metadata, e.g. instrumentation characteristics, basic site information. Since the adaption of StationXML by the FDSN as standard format for metadata, proper

documentation is lacking. This document is now in preparation and will most likely become available through 'Read the Docs'.

2   Documentation

The EIDA infrastructure is a complex, distributed system with different stakeholders in terms of maintenance, operations, management and usage. This is reflected in the current organization of available documentation. Each group of stakeholders has its own preferred platform to archive and expose information, from public website to restricted areas, depending on various aspects, e.g. public vs. restricted, technical vs. management, developers vs. users, static vs. dynamic, differences in technical skills, distributed knowledge, shared ownership, etc. We can roughly group the information into 3 domains: public (e.g. users, non-EIDA archives), internal (EIDA Management and Technical Commission) and software (developers, testers). The goal of this SERA task is to set up a platform that will bring together, and homogenize, the different pieces of documentation into a sustainable system. At the same time such a framework for documentation has to take advantage of the distributed organization of the EIDA infrastructure and allows the different groups to work collaborative on the documentation in a version controlled system.

Public

The general information on EIDA is maintained at a specific landing page: www.orfeus-eu.org/eida. This page and the links herein provide up-to-date information on the organization and structure of EIDA, details on the different nodes that build up EIDA, contributing networks in EIDA, EIDA data policy, EIDA authentication system, user feedback and the different services that are available to request data across EIDA: webinterface, webservices, data quality and StationBook.

The information on these website pages is relatively static and relatively easy to maintain, however the workload is centralized at ORFEUS Data Center. The dynamic parts of the general information, like the list of networks in EIDA or the corresponding DOIs, are generated on-the-fly using EIDA webservices.

For reporting issues concerning the use of EIDA (technical difficulties, questions, suggestions) we set up the EIDA User Feedback Repository hosted on github.com/EIDA/userfeedback. In this way we foster the communication with our user community by providing official channels to report problems or questions. Ultimately, encountered problems and their solutions can be taken up in the user documentation. Currently, 52 issues were reported of which 46 are already closed.

Public information is open accessible without login credentials. The general public information targets the general public and (new) seismologists with interest in ORFEUS and EIDA, as well as researchers/ scientists who are using the EIDA infrastructure for collecting data. The EIDA User Feedback Repository in particular will build up very specific – technical – information on issues encountered during the use of services and corresponding solutions, and is also open accessible.

Internal

Internal information on EIDA for both EMB and ETC is archived and available at Redmine at: https://dev.knmi.nl/projects/eida/wiki/Wiki

Both EMB and ETC have access to these areas and maintain their documentation with shared responsibilities, for example: meeting minutes, general documents, technical documents, etc. These areas are accessible after registration and thus provide open access to users working for EIDA.

Software / Developers

For software development and documentation versioning Git (version control system) and GitHub (web based version for version control) are commonly used. Git allows developers to work together, for example within the ETC, on the same project at the same time without disturbing each other's files. EIDA created a number of repositories on GitHub (https://github.com/EIDA) to share the development of a number of tools and services (e.g. routing, stationbook, logging, wfcatalog etc.).

Writing good documentation is difficult and time consuming. Also there is no 'definition' of what characterizes 'good' documentation. However, a number of best practices can be identified, e.g.: up to date, easy to find, extensive and comprehensive, easy to contribute to, etc. Platforms that have been explored to host such documentation are 'Confluence' and 'Read the Docs'. Confluence software, developed by Atlassian, is an effective team collaboration software suite which provides a common platform for teams to work together and share information efficiently. It is a tool to centralize a knowledge repository like a wiki with advanced content creation tools. We initially started to explore Confluence but decided not to continue with this platform as this software may be too comprehensive for our purpose to set up the EIDA documentation. We therefore started to build our documentation environment using Read the Docs (which actually is used also by Confluence).

3   Read the Docs

Read the Docs is one of the largest open source documentation hosting sites in the world, a software documentation hosting platform with free source code and free usage of services. It is a platform for both developers and technical writers. Read the Docs can be seen as a documentation platform that builds, runs and tests the documentation on each commit of an edit of the code. We use Read the Docs together with Sphinx and GitHub. Sphinx is a document generator using the reStructuredText (rst) markup language, while GitHub is used as a repository to host versions of the Sphinx reStructuredText files.

The workflow is: a) edit/create rst files with Sphinx, b) add, commit and push these to the GitHub repository and c) let Read the Docs build the up-to-date webpages. The use of GitHub allows multiple authors to work on the documentation easily.

An example of the Sphinx reStructuredText markup language files is:

```
.. ORFEUS documentation documentation master file, created by
   sphinx-quickstart on Fri Apr  3 07:41:10 2020.
   You can adapt this file completely to your liking, but it should at least
   contain the root `toctree` directive.

.. figure:: _static/orfeuslogo.jpg

Welcome to the ORFEUS documentation repository
==============================================
```

ORFEUS (Observatories and Research Facilities for European Seismology) is a collaborative non-profit foundation that promotes seismology in the Euro-Mediterranean area through the collection, archival and distribution of digital seismic waveform data, metadata and derived products.
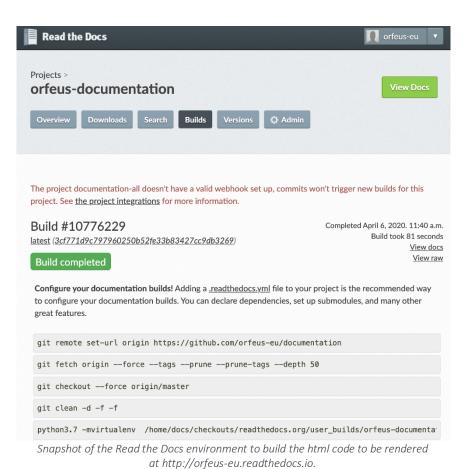
Our website is: http://www.orfeus-eu.org.

```
.. toctree::
  :maxdepth: 2
  :caption: ORFEUS:

  general

.. toctree::
  :maxdepth: 2
  :caption: EIDA:

  eida_intro
  eida_user
  eida_eas
  eida_routing
  stationbook
```
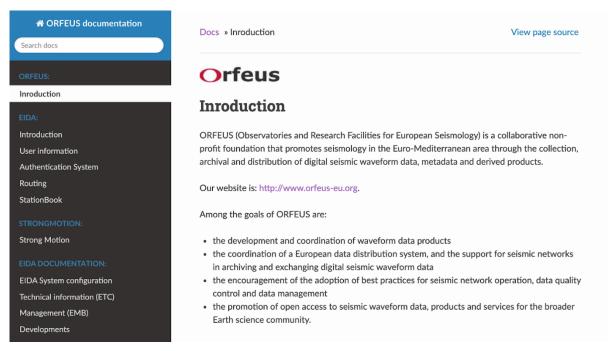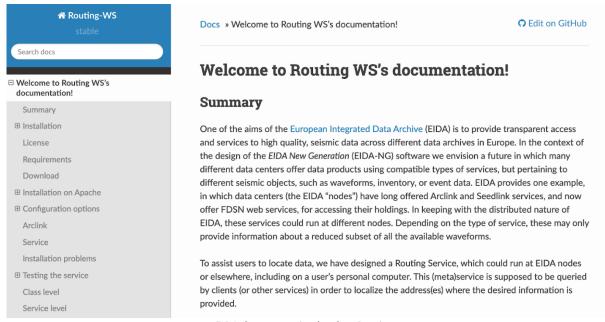
Uploading the Sphinx created rst files to GitHub allows Read the Docs to build the documentation. Based on the rst files in GitHub Read the Docs creates corresponding html files - which can be done locally by Sphinx as well for testing purposes - with a predefined layout, called theme. As theme for the ORFEUS documentation rendered by Read the Docs we have chosen to use 'sphinx_rtd_theme' (https://github.com/snide/sphinx_rtd_theme).

Snapshot of the Read the Docs environment to build the html code to be rendered at http://orfeus-eu.readthedocs.io.

ORFEUS / EIDA documentation html page created by Read the Docs.

EIDA documentation html on Routing.

The URL to the ORFEUS documentation on Read the Docs is: https://orfeus-eu.readthedocs.io

4   Conclusions

We successfully set up a platform for the generation and maintenance of ORFEUS and EIDA documentation in a shared responsibility environment. The use of Sphinx, GitHub and Read the Docs allows multiple users to work on the documentation - reflecting the distributed and federated character of EIDA. The framework is flexible to contain documentation for the different stakeholders and can be extended/modified in time using open source tools.

Contact

Project lead	ETH Zürich
Project coordinator	Prof. Dr. Domenico Giardini
Project manager	Dr. Kauzar Saleh
Project office	ETH Department of Earth Sciences
	Sonneggstrasse 5, NO H-floor, CH-8092 Zürich
	sera_office@erdw.ethz.ch
	+41 44 632 9690
Project website	www.sera-eu.org

Liability claim

The European Commission is not responsible for any use that may be made of the information contained in this document. Also, responsibility for the information and views expressed in this document lies entirely with the author(s).