



## D5.1 Workshop with research infrastructures

Marco Raciti, Yoann Moranville, Carsten Thiel

### ► To cite this version:

Marco Raciti, Yoann Moranville, Carsten Thiel. D5.1 Workshop with research infrastructures. [Research Report] DARIAH. 2019. hal-02081139

**HAL Id: hal-02081139**

**<https://hal.science/hal-02081139>**

Submitted on 12 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## D5.1 Workshop with research infrastructures

### DESIR

---

#### DARIAH ERIC Sustainability Refined

INFRADEV-03-2016-2017 - Individual support to ESFRI and other world-class research infrastructures

Grant Agreement no.: 731081

Date: 26-03-2019

Version: 1.0



DESIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731081.

Grant Agreement no.:	731081
Programme:	Horizon 2020
Project acronym:	DESIR
Project full title:	DARIAH-ERIC Sustainability Refined
Partners:	DIGITAL RESEARCH INFRASTRUCTURE FOR THE ARTS AND HUMANITIES  GEORG-AUGUST-UNIVERSITAET GOETTINGEN STIFTUNG OEFFENTLICHEN RECHTS  UNIVERSITEIT GENT  UNIWERSYTET WARSZAWSKI  FACULDADE DE CIENCIAS SOCIAIS E HUMANAS DA UNIVERSIDADE NOVA DE LISBOA  CENTAR ZA DIGITALNE HUMANISTICKE NAUKE  GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER  INSTITUT NATIONAL DE RECHERCHE ENINFORMATIQUE ET AUTOMATIQUE  KING'S COLLEGE LONDON  UNIVERSITY OF GLASGOW  KNIHOVNA AV CR V. V. I.  HELSINGIN YLIOPISTO  SIB INSTITUT SUISSE DE BIOINFORMATIQUE  UNIVERSIDAD NACIONAL DE EDUCACION A DISTANCIA  UNIVERSITY OF HAIFA  UNIVERSITY OF NEUCHATEL
Topic:	INFRADEV-03-2016-2017

**DESIR**

INFRADEV-03-2016-2017 - Individual support to ESFRI and other world-class research infrastructures, Grant Agreement no. 731081.



Project Start Date:	01-01-2017
Project Duration:	36 months
Title of the document:	Workshop with research infrastructures
Work Package title:	Robustness
Estimated delivery date:	31/12/2018
Lead Beneficiary:	DARIAH
Author(s):	Marco Raciti [marco.raciti@dariah.eu], Yoann Moranville [yoann.moranville@dariah.eu], Carsten Thiel [carsten.thiel@CESSDA.eu]
Quality Assessor(s):	Yoann Moranville [yoann.moranville@dariah.eu]
Keywords:	Software sustainability, research software, technical reference, FAIR software, research infrastructures.

## Revision History

Version	Date	Author	Beneficiary	Description
0.1	9-10/10/17	Marco Raciti, Carsten Thiel	DARIAH UGOE	Organisation of the first workshop in Berlin
0.2	30/06/18	Yoann Moranville Michelle Rodzis Carsten Thiel	DARIAH UGOE UGOE	Publication of the first version of the technical reference
0.3	12-13/03/19	Stefan Buddenbohm Yoann Moranville Marco Raciti	UGOE DARIAH DARIAH	Organisation of the second workshop in Utrecht
0.4	19/03/19	Marco Raciti	DARIAH	First version of the summary report
0.5	25/03/19	Yoann Moranville	DARIAH	Final revision
1.0	26/03/19	Marco Raciti	DARIAH	Submission of deliverable D5.1 Workshop with research infrastructures

## Table of Content

<b>Executive Summary .....</b>	<b>6</b>
<b>1. Workshop “Software Sustainability: Quality and Reusability”, 9-10 October 2017 .....</b>	<b>8</b>
1.1 Agenda.....	8
1.2 Speakers.....	9
1.3 Participants.....	11
1.4 Summary .....	12
<b>2. EURISE Network and Technical Reference.....</b>	<b>14</b>
<b>3. EURISE Workshop: Software Sustainability within Research Infrastructures, 12-13 March 2019</b>	<b>15</b>
3.1 Agenda.....	15
3.2 Speakers.....	16
3.3 Participants.....	19
3.4 Summary .....	20

## Executive Summary

This report documents the activities undertaken in the framework of task 5.2 of the project DESIR. The initial aim of the task was to organise a workshop with ESFRI Research Infrastructure networks and providers. The topic originally foreseen was mostly revolved around generic challenges for Research Infrastructures, such as the **distribution of the services, governance, funding and business models, competence profiles needed for running efficient infrastructures**. However, since the submission of the proposal, the landscape had greatly changed in terms of increased possibilities of discussing these issues with other ERICs on a regular basis. For example, the informal ERIC network initiative became in the meanwhile a more structured platform, and was rebranded as **ERIC Forum**. Recently, the ERIC Forum received funding under Horizon 2020 from the European Commission and DARIAH is one of the beneficiaries. To avoid a duplication of efforts, the Consortium decided to change the focus of the workshop to investigate **software sustainability** in Research Infrastructures.

The first event took place in Berlin, hosted by the DARIAH Coordination Office on 9-10 October 2017 as a continuation of ongoing efforts to **align technology** across three ERICs in the Arts, Humanities and Social Sciences. The event saw the participation of representatives from **CESSDA, CLARIN and DARIAH** as well as developers, users, service operators and IT managers. The discussions covered a wide range of topics related to software sustainability. Speakers presented work already accomplished as part of the tasks the infrastructures have undertaken in their efforts to become operational. Among these are the **CLARIAH-NL Software Quality Guidelines, the CESSDA software maturity model and the DARIAH-DE Lifecycle**. The overall problems these approaches try to address are similar to the challenges the software industry is facing: **training, quality management**, and dealing with an ever-growing **technical debt** are challenges that need to be addressed and re-evaluated on a constant basis. A concrete common goal expressed at the workshop was to **align the existing criteria catalogues** with each other and also existing alternatives such as the ones led by the **UK Software Sustainability Institute** and the **Netherlands eScience Center**.

The event led to two important results: first of all, the constitution of the **EURISE network**, formed by CESSDA, CLARIN and DARIAH, to create an umbrella where research infrastructures meet research software engineers to discuss about **standards, best practices, education and training and sustainability**. Secondly, a **Technical Reference**, available online with an open license, was developed in the framework of D4.2 to provide a collection of guidelines and references for development and maintenance of infrastructure services. The guidelines are continuously adapted to evolving needs and discussions within and beyond the EURISE network.

After a positive evaluation of the results of the first event, the Consortium decided to organise another workshop on **12-13 March 2019 in Utrecht** on the same topic. To ensure the widest possible participation, it has been organised as a satellite meeting of the SSHOC project<sup>1</sup> kick-off meeting. Forty people from very different backgrounds attended this second workshop in which topics such as software maturity models, software sustainability from the point of view of **research infrastructures, business models, guidelines, FAIR software, security and human capital** as key for sustainability were discussed. Representatives from CESSDA, CLARIN and DARIAH showcased their individual responses towards these challenges, while the workshop was complemented with interesting presentations from the EGI foundation on the implementation of IT service management in federated infrastructures and global initiatives on security. Furthermore, a case study in the Netherlands was presented to highlight the processes towards a **FAIR software**. The connection with the **EOSC**<sup>2</sup> was also made and the ideas gathered during the event will fuel the activities carried out within the EURISE network and will enrich the technical reference, which will be a steady starting point for developing the future **SSHOC marketplace**.

Nature of the deliverable		
	R	Document, report
	DEM	Demonstrator, pilot, prototype
✓	DEC	Websites, patent fillings, videos, etc.
	OTHER	
Dissemination level		
✓	P	Public
	CO	Confidential only for members of the consortium (including the Commission Services)
	EU-RES	Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
	EU-CON	Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
	EU-SEC	Classified Information: SECRET UE (Commission Decision 2005/444/EC)

### Disclaimer

DESIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731081. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

<sup>1</sup> <https://www.sshopencloud.eu/>

<sup>2</sup> European Open Science Cloud



## 1. Workshop “Software Sustainability: Quality and Reusability”, 9-10 October 2017

The workshop took place on 9-10 October 2017 at the DARIAH-EU Coordination Office premises in Berlin. The workshop was open to anyone interested in software quality and sustainability.

### 1.1 Agenda

#### 9 October

13:00-14:00 Registration / Lunch

14:00-14:05 Welcome: **Marco Raciti** (DARIAH-EU)

14:05-14:30 Introduction: **Carsten Thiel** (SUB)

14:30-15:00 CESSDA Software Maturity Model: **John Shepherdson** (UK Data Service)

15:00-15:30 CLARIAH-NL Software Quality Guidelines: **Reinier de Valk** (DANS)

15:30-15:45 Coffee break

15:45-16:15 DARIAH-DE Service Life Cycle: **Danah Tonne** (KIT)

16:15-16:45 The infrastructure view

- DARIAH-EU: **Tibor Kalman** (GWDG)
- CESSDA: **Hossein Abroshan** (Cessda MO)
- CLARIN: **Matej Durco** (ÖAW)

16:45-17:00 Coffee break

17:00-17:30 Plenary session

17:30-18:30 “A different point of view” – external keynote: **Phillip Hagedorn** (GfK DE)

#### 10 October

09:30-10:00 Recap of the first day

10:00-11:00 First group discussion

11:00-11:30 Coffee break

11:30-12:30 Second group discussion

12:30-13:00 Group reports, results and next steps

13:00-14:00 Lunch

## 1.2 Speakers

### Carsten Thiel

Carsten is the technical head and Co-Manager of the DARIAH-DE Consortium at the Research and Development Department of the Göttingen State and University Library, where is also involved with Humanities-at-Scale and DESIR. His research interests include digital research infrastructures, distributed development processes, and the DevOps approach to infrastructure management. From 2014 to 2016 Carsten was a research associate for the CENDARI project at the department and he holds a PhD in mathematics from the University of Magdeburg, where he also worked as a teaching assistant (2010-2014)

### John Shepherdson

John is an Associate Director at the UK Data Archive where he provides strategic direction and senior management for the products and systems that enable resource and data discovery, data access, data management and preservation of social scientific data. Since its launch in 2012 he is also Director of Technical Services for the UK Data Service.

John works closely with CESSDA (the Consortium of European Social Science Data Archives), representing the UKDS at CESSDA Service Provider Forum meetings and leading its Technical Working Group (which provides guidance on technical matters, including the product development roadmap). He plays a key role in the specification and implementation of the various phases of the Technical Framework project, which defines the rules and technologies for developing, testing and deploying components for CESSDA's European Research Infrastructure.

### Reinier de Valk

Reinier holds a PhD in computer science (Music Informatics Research Group, City University London, 2015), and has a background in musicology and music. His interests lie in the fields of computational modeling, machine learning, music information retrieval and computational music analysis, digital humanities, data science, and semantic web technology. Professionally, he seeks to blend these fields together. Reinier is currently employed as an information scientist at Data Archiving and Networked Services (DANS). He is involved in several (inter)national research infrastructure projects including EHRI ([ehriproject.eu](http://ehriproject.eu)), PARTHENOS ([parthenos-project.eu](http://parthenos-project.eu)), and CLARIAH ([clariah.nl/en](http://clariah.nl/en)), where topics such as data and software quality and sustainability, format standardisation, data linking and enrichment, and virtual research environments are addressed.

---

#### DESIR

INFRADEV-03-2016-2017 - Individual support to ESFRI and other world-class research infrastructures, Grant Agreement no. 731081.



## **Danah Tonne**

Danah received her PhD in computer science for modeling and quantifying the reliability in bit preservation architectures. Her fields of interest include research data management, preservation and the digital humanities.

Since 2011 Danah has been working for the Karlsruhe Institute of Technology (KIT) and within the DARIAH-DE consortium. She is now co-leading the working group "Service Life Cycle" which developed a formal process for service integration into the DARIAH infrastructure. Further involvements include "eCodicology" - Algorithms for the Automatic Tagging of Medieval Manuscripts and building up the information infrastructure of the Collaborative Research Center 980 "Episteme in Motion".

## **Tibor Kalman**

Tibor is vice-chair of the Joint Research Committee of the DARIAH-ERIC, co-head of DARIAH's Virtual Competency Center 'eInfrastructure', as well as Chief Technical Officer of the Persistent Identifier Consortium for eResearch. He is also member of the Steering Group of DARIAH-DE, member of the Senior Management Team of DARIAH-ERIC, and member of the Joint Research Committee of DARIAH-ERIC.

Tibor has contributed at the GWDG to several European and national funded projects with the aim to build up research infrastructures for various disciplines. His teaching activities include amongst others the theory and practice of distributed systems, parallel computing, and data management. His current research focuses on (1) distributed computing paradigms, (2) automated system deployment, and (3) research data management, including persistent identifiers. He holds a PhD in Computer Science and his thesis focused on achieving interoperability between information and monitoring systems of distributed research infrastructures, by information exchange, resource discovery, and service quality monitoring across virtual organizations.

He also has 15+ years experience as an independent IT consultant specialising in providing software solutions for small and medium-sized enterprises.

## **Hossein Abroshan**

Hossein is managing technical direction and maintaining IT governance. He manages the technology strategy, systems implementations, and IT operations. Hossein leads CESSDA's IT services and coordinates technical activities between the CESSDA ERIC Service Providers.

## **Matej Durco**

Matej is head of the ACDH-ÖAW's technical working group "Tools, Services & Systems" and was one of the key figures in founding the institute. Since 2009, he has been part of the

Austrian research infrastructures core group, contributing substantially to the development of key technical components of the CLARIN infrastructure such as the Component Metadata Infrastructure (CMDI), the Federated Content Search (FCS) and the Vocabulary Repository, both on the local and on the European level. Currently, he acts as coordinator of the Metadata Curation Task Force in CLARIN and is co-head of the Virtual Competence Centre I (e-Infrastructure) in DARIAH. In his role as head of the ACDH-ÖAW's technical group, he has been coordinating the development of applications and the provision of services for the numerous projects of the institute and its cooperation partners. His focus lies on building and providing a diverse but at the same time consolidated portfolio of technical solutions, striving to offer the most appropriate tools for specific requirements of individual projects while maintaining the highest technical standards.

### **Phillip Hagedorn**

Phillip is IT-Manager for DevOps teams within the GfK (Gesellschaft für Konsumforschung) as well as head of the cloud transformation steering group and head of the Community of Practice for DevOps and Cloud Best practice. In his former positions he worked as Director of Application Management and Automation Development for GfK Nurago and as Senior Consultant at Namics AG for several E-Commerce and Enterprise CMS Customers.

## **1.3 Participants**

The registration for the workshop was available on Eventbrite<sup>3</sup>. There were 38 registrations for the event. The first day 25 people attended the meeting while 21 were present on the second day. Participants came from different backgrounds: developers, users, service operators and IT managers within and beyond the established research infrastructures' communities. The institutions represented were the following:

- CESSDA
- SUB-University of Göttingen
- Max Planck Institute for the History of Science
- Berlin-Brandenburg Academy of Sciences and Humanities
- DANS-KNAW
- Humboldt University of Berlin
- Austrian Academy of Sciences
- University of Trier
- GWDG
- TU Braunschweig
- DARIAH

---

<sup>3</sup> <https://www.eventbrite.co.uk/e/software-sustainability-quality-and-re-usability-tickets-35911777092>

- Free University Berlin
- Hermann von Helmholtz Center for Cultural Techniques
- Karlsruhe Institute of Technology
- CLARIN
- Software Sustainability Institute

## 1.4 Summary

As part of ongoing efforts to **align technology** across the three Pan-European infrastructures for the Social Sciences, Arts, and Humanities, representatives from **CESSDA**, **CLARIN**, and **DARIAH** held a workshop on “Software Sustainability: Quality and Re-usability<sup>4</sup>”, on 9-10 October 2017 at Centre Marc Bloch in Berlin. With participants from Austria, Germany, Great Britain, Norway and The Netherlands representing, talks and discussions covered a **wide range of topics related to software sustainability**. Speakers presented work already accomplished as part of the tasks the infrastructures have undertaken in their efforts to become operational.

Among these are the **CLARIAH-NL Software Quality Guidelines**<sup>5</sup> and the **CESSDA software maturity model**, which both define evaluation criteria for software products. Their approaches differ, in that the former focuses on explicit implementation guidelines, while the latter, modelled on NASA’s Reuse Readiness Levels, describes a generalised framework for evaluating a given software product. While criteria are also an important part of the **DARIAH-DE Service Lifecycle**<sup>6</sup>, its focus is on describing processes and necessary considerations when taking software from initial design through development and testing to production use.

The overall problems these approaches try to address are similar to the challenges the software industry is facing: **training**, **quality management**, and dealing with an ever-growing **technical debt** are challenges that need to be addressed and re-evaluated on a constant basis.

The specific requirements and expectations are very different from the perspectives of the developer, the user, and the maintainer or service provider. This multifaceted understanding of sustainability is similar to that of data, for which FAIR principles<sup>7</sup> have become the de-facto definitions of in this respect. A similar approach, the “Recommendations to encourage

---

<sup>4</sup> <https://calenda.org/411096>

<sup>5</sup> <https://github.com/CLARIAH/software-quality-guidelines>

<sup>6</sup> <https://wiki.de.dariah.eu/display/publicde/DARIAH-DE+Service+Life+Cycle>

<sup>7</sup> <https://www.nature.com/articles/sdata201618>

best practices in research software<sup>8</sup>” aims at improving the overall quality and **re-usability of research software**.

Agreement reigned among workshop participants on the need to foster and spread awareness of these different needs. At various levels, these can possibly be addressed through **guidelines, recommendations, and criteria catalogues** at varying levels of detail. Elementary guidelines for software developers just starting out, which typically only involve modest scripting and ad-hoc solutions implemented for highly specialised problems, can support the process from the start. In general, not all software development can or must aim at reaching production level maturity. On the other hand, when software is re-used or applied by third parties, **external evaluation can improve trust**. This is particularly relevant to infrastructures which not only have to rely on the code that runs their services, but are facilitators in providing a wider research community and thus end users access to the tools concerned.

Improving on the status quo involves a **cultural change** in the academic acceptance of software as a form of relevant research output as it is currently the case for data. In this light, the adoption of requirements by funders plays as significant a role as any cultural shift among researchers. Defining similar general requirements for research software by funders may be an incentive to invest in quality. But as research thrives on new and innovative technology and ideas, adhering to formal procedures and requirements in the IT world with its fast-paced rate of change can potentially also limit progress. Ultimately, **balancing the interests of the different stakeholders** remains a core challenge when considering and defining the meaning of sustainability for a given piece of software.

A concrete common goal expressed at the workshop was to **align the existing criteria catalogues** with each other and also existing alternatives such as the ones led by the **UK Software Sustainability Institute** and the **Netherlands eScience Center**. The plan to continue the exchange among the three infrastructures should involve these institutions as well as the Research Software Engineering community at large. Further discussions on the topic are open to interested contributors within the DH-Tech Group’s Slack channel #sustainability. The plan to continue the exchange among the three infrastructures should involve these institutions as well as the Research Software Engineering community at large.

---

<sup>8</sup> <https://softdev4research.github.io/recommendations/>

## 2. EURISE Network and Technical Reference

The first workshop in Berlin produced two concrete outcomes. The **EURISE network**<sup>9</sup> have been formed by CESSDA, CLARIN and DARIAH to create an umbrella where research infrastructures meet research software engineers. Currently it is coordinated by Carsten Thiel for CESSDA, Dieter Van Uytvanck for CLARIN and Tibor Kálmán and Yoann Moranville for DARIAH. The EURISE network is an informal platform to discuss standards, best practises, sustainable technology, software engineering, infrastructure sustainability, reusability of research software and to promote software as research output.

Furthermore, a **Technical Reference**<sup>10</sup> was developed as a collection of best practices and basic software development guidelines as well as quality checklists for developers and operators of infrastructure components. It lists general requirements and considerations, but does not force the use of specific licenses, technology stacks or hosting services. This work relies on the Netherlands eScience Center Guide<sup>11</sup> and the CLARIAH-NL Software Quality Guidelines<sup>12</sup>. The Technical Reference has been delivered as part of D4.2 in DESIR, however, the work is continuously adapted to evolving needs and discussions within and beyond the EURISE network.

---

<sup>9</sup> <https://eurise-network.github.io/>

<sup>10</sup> <https://technical-reference.readthedocs.io/en/latest/#>

<sup>11</sup> <https://guide.esciencecenter.nl/>

<sup>12</sup> <https://github.com/CLARIAH/software-quality-guidelines>

### 3. EURISE Workshop: Software Sustainability within Research Infrastructures, 12-13 March 2019

The EURISE workshop took place on 12-13 March 2019 in Utrecht at Surf Offices. The workshop focused on various aspects of software sustainability and quality within research infrastructures from different research realms. The workshop on March 12-13 2019 is following up a first workshop on 9-10 October 2017 in Berlin which led to the establishment of the EURISE network. The topics discussed were developer guidelines and best practices, software quality measures and assessment, infrastructure service management, infrastructure security, business models and sustainability and the role played by Research Infrastructures in Research Software Engineering.

#### 3.1 Agenda

##### 12 March

15:00-15:30 Arrival: Tea & Coffee

15:30-16:00 Welcome: Introduction - **Carsten Thiel** (CESSDA)

16:00-16:30 CESSDA's Software Maturity Model - **John Shepherdson** (CESSDA)

16:30-17:00 CLARIN's first steps on the long path to software sustainability - **Dieter van Uytvanck** (CLARIN)

17:00-17:30 Business Models for Digital Research Infrastructures using the example of DARIAH - **Frank Fischer** (DARIAH)

18:00-21:00 Dinner: De Utrechter Stadsbrasserie

##### 13 March

09:00-09:15 Coffee break

09:15-09:45 EURISE Technical Reference - **Michelle Weidling** (SUB Göttingen)

09:45-10:15 Implementing a Service Management System in a Federated Multi-Supply Environment - **Yannick Legré** (EGI)

10:15-10:45 Infrastructure Security - **David Kelsey** (UKRI STFC)

10:45-11:00 Coffee break

11:00-11:30 A FAIR Software Route - the Dutch Way - **Mustapha Mokrane** (DANS) / **Carlos Martinez Ortiz** (NLeSC)



11:30-12:00 Sustainability beyond guidelines - How our perception of research infrastructures shapes our perception of the sustainability issue? - **Andrea Scharnhorst / Francesca Morselli** (DANS)

12:00-13:00 Breakout Session: Discussion

13:00-14:00 Lunch

14:00-15:00 Final Discussion and Closing of the Workshop

## 3.2 Speakers

### Carsten Thiel

Carsten is Chief Technical Officer at CESSDA. He is in charge of CESSDA's technical roadmap and strategy, including the design of all technical aspects of the infrastructure as well as a sustainable model for its long-term operation by Main Office and Service Providers.

### John Shepherdson

**John is Platform Delivery Director at CESSDA.** He deals with matters relating to the specification, delivery and enhancement of the cloud-based CESSDA technical infrastructure. He also leads the Technical Working Group and liaises with other relevant platform providers to explore possible synergies

### Dieter van Uytvanck

Dieter Van Uytvanck is the Technical Director of CLARIN ERIC. He graduated in Informatics (2002, Ghent University) and Language and Speech Technology (2007, Radboud University Nijmegen) and has been involved in technical infrastructure building since CLARIN's preparatory phase in 2008.

Between 2004 and 2008 he has worked in the context of The Language Archive at the Max Planck Institute for Psycholinguistics, as part-time technical assistant and corpus manager, supporting researchers to annotate, to enrich and to deposit language data into the institutional repository.

After graduating in 2007 he continued working at the MPI for Psycholinguistics, first in the CLARIN preparatory ESFRI project (until 2011), then in the technical management of the CLARIN-D infrastructure (until 2016). Since early 2016 he is working full-time as Technical Director of CLARIN ERIC, which is hosted at Utrecht University. He is the chair of the Standing Committee on CLARIN technical centres since 2012.

## **Frank Fischer**

Frank Fischer is currently Associate Professor for Digital Humanities at the Higher School of Economics, Moscow, and co-director of DARIAH-EU. He studied Computer Science, German Literature, and Spanish Philology in Leipzig and London and is an Ancien Pensionnaire de l'École Normale Supérieure in Paris. He received his PhD from the University of Jena for his study on the dramatic works of Joachim Wilhelm von Brawe and their contemporary translations into Russian, Danish, and French.

## **Michelle Weidling**

Michelle works at the Digital Humanities department of the Göttingen State and University Library. She is involved in the projects Library of Neology, TextGrid III and Theodor Fontane Notizbücher II.

## **Yannick Legré**

Yannick is the director of the EGI Foundation since February 2014. He was formerly a senior research engineer at the French National Scientific Research Centre – Grid and Cloud Institute (CNRS-IdGC). He holds a Master of Science in Information Technology (MScIT) and a degree in Law (LL.L). Over the last 15 years, Yannick has been involved in more than 30 projects in the areas of healthcare and biomedical research, as well as biodiversity and environmental research, such as the ENVRI & CReATIVE-B projects. Yannick has also been a co-founder and the president of the international HealthGrid association, and the director for International Relations of a French SME.

## **David Kelsey**

Dr David Kelsey is head of particle physics computing at STFC Rutherford Appleton Laboratory. He has been leading many Grid Security activities since 2001 including EU DataGrid, EGEE, EGI-InSPIRE, EGI-Engage, GridPP and WLCG. He has an equally long track record in federated identity management as the founder and leader of the EU Certificate Authority Coordination Group in 2000, a group that became EUGridPMA in 2004 and led to the formation of IGTF. Kelsey has been an active member of the FIM4R group collecting and documenting the research community requirements for federated identity management and has striven to establish trust between e-Infrastructure operational security teams, the identity federations and with IT infrastructures worldwide, as the founder and leader of the Security for Collaborating Infrastructures (SCI) activity. Kelsey is chair of the SCIV2 working group of WISE. Kelsey has a PhD in experimental particle physics.

---

## **DESIR**

INFRADEV-03-2016-2017 - Individual support to ESFRI and other world-class research infrastructures, Grant Agreement no. 731081.



## **Mustapha Mokrane**

Mustapha Mokrane is currently a Project leader at DANS (Data Archiving and Networked Services), the Institute for Permanent Access to Digital Research Resources in The Netherlands. He was until June 2018 the Executive Director of the World Data System International Programme Office. Previously, he worked between 2009 and 2012 at the International Council for Science (ICSU) as Science and Information Technology Officer in charge of the coordination of ICSU's Scientific Data and Information activities and the liaison with its partners. He was also responsible for the information technology related activities within ICSU before 2003 and 2009. After moving from Algeria, his home country, he trained as molecular biologist in Marseille, France. Using transgenesis and microarrays technologies, he developed a drosophila model to elucidate the role of a gene involved in human cardiomyopathies. His scientific background covers genetics and bioinformatics and he developed a strong interest for scientific data and information challenges. He holds a Ph.D. in Molecular Biology from the Aix-Marseille University.

## **Carlos Martinez Ortiz**

Carlos has a PhD in computer science and is an eScience research engineer at the Netherlands eScience Center. He is currently working in projects related to the Digital Humanities. He is responsible for development of DIVE+.

## **Andrea Scharnhorst**

Dr. Andrea Scharnhorst is Head of e-Research at the Data Archiving and Networked Services (DANS) institution in the Netherlands - a large digital archive for research data primarily from the social sciences and humanities. She is also member of the e-humanities group at the Royal Netherlands Academy of Arts and Sciences (KNAW) in Amsterdam, where she coordinates the computational humanities programme. In DARIAH, Andrea chairs the Joint Research Committee and coordinates the collection of national in-kind contributions. Her research fields are: Science dynamics, Scientometrics and Complex Systems, Evolution of Knowledge Organisation Systems, Evolution of Scholarly Communication and the Future of Libraries and Archives.

## **Francesca Morselli**

Francesca's academic background is in Media Studies (with majors in contemporary history and audiovisual semiotics) and audiovisual archiving. Before joining the DARIAH CIO team (in the context of DANS-KNAW- Data Archiving and Networked Services in The Hague) she worked and conducted research for European institutions, projects and infrastructures (Europeana; CEU, CENDARI) for the accessibility of cultural heritage, both directed to

general users and academics. At DANS-KNAW Francesca is a member of the research team and the data science group; she is also involved in DARIAH-EU as part of the Chief Integration Office and the European Project PARTHENOS as co-coordinator of the WP on Common Policies and Implementation Strategies.

### 3.3 Participants

The registration for the workshop was available on [sciencesconf.org](https://sciencesconf.org)<sup>13</sup>. There were 40 participants attending the 2 days event organised alongside the kick-off meeting of the project SSHOC. The following 24 institutions were represented in the workshop:

- Instruct-ERIC
- DARIAH ERIC
- Göttingen State and University Library
- KNAW/HuC
- CESSDA ERIC
- Ghent University
- Austrian Academy of Sciences
- Berlin-Brandenburg Academy of Sciences and Humanities
- OpenEdition
- CLARIN ERIC
- Finnish Social Science Data Archive/Tampere University
- GWDG
- AUSSDA – The Austrian Social Science Data Archive
- STFC-UKRI
- GESIS – Leibniz Institute for the Social Sciences
- FORS
- AGR
- Huma-Num/CNRS
- EGI
- King's College London
- Netherlands eScience Center
- KNAW Humanities Cluster
- DANS-KNAW
- Poznan Supercomputing and Networking Center

---

<sup>13</sup> <https://euriseworkshop.sciencesconf.org/>

### 3.4 Summary

#### a) Welcome: Introduction - Carsten Thiel (CESSDA)



Image 1 – Carsten Thiel introduces the workshop

Carsten Thiel introduced the **challenges surrounding the sustainability of research software**, which are developed within academia and used for the purpose of research. One of the main issues is often related to the nature of software built within the academia. Often research projects produce **demonstrators or prototypes** which are, in most cases, not supported anymore after the end of project funding. Other challenges are constituted by people turnover, the varying degrees of development experience, a constantly evolving IT landscape and no longer supported source codes (legacy code).

He then introduced the topic from the point of view of research infrastructure, which aim, ideally, is to **building bridges** for developers, researchers/scholars and projects on the one hand and data centers and libraries/archives on the other hand.

Carsten Thiel also showed the importance of **ensuring the software quality of services**, providing evidence for technical maturity and operating the services in a reliable manner.

The **importance of training** has also been highlighted. During the presentation, two initiatives in this respect have been mentioned: Software Carpentry<sup>14</sup>, a volunteer project

---

<sup>14</sup> <https://software-carpentry.org/>

dedicated to teaching basic computing skills to researchers, and Code Refinery<sup>15</sup>, an e-infrastructure set up to advance FAIRness of Software management and development practices.

### **b) CESSDA's Software Maturity Model - John Shepherdson (CESSDA)**

John Shepherdson presented the CESSDA's Software Maturity Model. Over the last couple of years, CESSDA has built up a comprehensive infrastructure that provides a development and test environment available to all CESSDA Service Providers, which make use of technologies such as Bitbucket, Jenkins, Selenium, SonarQube, JMeter, etc. However, such a powerful environment should be supported by a **set of quality standards** to ensure that software developed fulfil acceptance criteria. The **CESSDA Software Maturity Levels** which is based on a light version of the NASA's 9 Reuse Readiness Levels<sup>16</sup>, contains 12 criteria to be used as a standard to check whether a software can be accepted and deployed.

The 12 criteria (each criterion contains 5 levels, from minimum to excellent) are the following:

1. Documentation (end user, operational, developer);
2. Intellectual property issues;
3. Extensibility;
4. Modularity;
5. Packaging;
6. Portability;
7. Standards Compliance;
8. Support;
9. Verification and testing;
10. Security (by design);
11. Internationalisation and Localisation;
12. Authentication and Authorisation.

### **c) CLARIN's first steps on the long path to software sustainability - Dieter van Uytvanck (CLARIN)**

Dieter Van Uytanvick, Technical Director of CLARIN ERIC, introduced first steps that CLARIN is making towards software sustainability. He briefly presented **CLARIN**, an ERIC which

---

<sup>15</sup> <https://coderefinery.org/>

<sup>16</sup> [https://cdn.earthdata.nasa.gov/conduit/upload/2004/RRLs\\_v1.0.pdf](https://cdn.earthdata.nasa.gov/conduit/upload/2004/RRLs_v1.0.pdf)

**makes digital language resources available** to scholars, researchers, students and citizen-scientists from all disciplines, especially in the Social Sciences and Humanities. Among the services offered by CLARIN we can cite the Virtual Language Observatory, the Language Resource Switchboard and the Language Resource Inventory. CLARIN deals with three types of services, which can be offered **centrally** (e.g. the Virtual Language Observatory), can be **distributed** (e.g. several validators like OAI-PMH-validation, Federated Content Search) or **centre-specific** (the majority of services fall into this category). At the beginning, CLARIN offered no central services, however, over the time, infrastructural crucial services were moved to the central level.

The issues presented, in terms of software sustainability in CLARIN, are the following:

- lack of awareness;
- living with and surviving hypes and trends;
- simplification on the side of the management, leading to a lack of resources;
- organisational factor, like people turnover, time pressure, bad or absent documentation.

Dieter Van Uytenhove explained that there are **no golden rules** to solve these issues. However, he presented some techniques to help software development in the infrastructure. For example, a **minimalistic approach** is recommended, as well as a free software approach (with the motto “publish soon, publish often”), a redundant know-how and testing. Furthermore, standardised deployments and early code peer-review and feedback are also solutions suggested for software sustainability.

#### **d) Business Models for Digital Research Infrastructures using the example of DARIAH - Frank Fischer (DARIAH)**

Frank Fischer is co-Director of DARIAH since 2017. He started his presentation on Business Models by introducing the Digital Humanities community, which is characterised by work predominantly done in teams rather than solitary, with projects as basic units of research, emphasis on building and crafting, alternative teaching formats like hackathons, workshops, summer schools, and the use VREs (Virtual Research Environments) and social network platforms for networking.

He then discussed **why the Humanities need Research Infrastructure**, answering with a citation coming from the PARTHENOS project: *“Modern definitions recognise that all disciplines have infrastructural needs, for data, for specialist knowledge, for methodological*



*support, for storage and processing power. As such, the arts and humanities need shared research infrastructure as much as any other discipline”<sup>17</sup>.*

The mission and vision of DARIAH were also introduced. DARIAH ERIC exists to **enable excellent research in the Arts and Humanities** and its vision is that these are anchored at the centre of a technologically evolving knowledge society, while the mission is to **empower research communities with digital methods** to create, connect and share knowledge about culture and society.



**Image 2 – Frank Fischer presents Business Models for Digital Research Infrastructures**

DARIAH’s vision and mission are then translated over four main pillars which are:

1. building the **Marketplace** as a humanist-friendly component of the European Open Science Cloud;
2. approaching **training and education** strategically and in a coordinated fashion;
3. deepening the **connection** to Arts and Humanities communities;
4. strengthening DARIAH’s voice in **policy and advocacy**.

Frank Fischer presented some DARIAH driven services, such as the single sign-on, the DH Course Registry (in collaboration with CLARIN), the Geo-Browser, Text-Grid, #dariahTeach.

<sup>17</sup> <http://www.parthenos-project.eu/why-humanities-need-a-research-infrastructure>



Furthermore, the income streams of the infrastructure have been discussed, in particular the original model which combines membership fees, competitive grants income and non-cash contributions from members. After the presentation, the crowd discussed how to approach the questions of **Key Performance Indicators for the Social Sciences and Humanities infrastructures** and the need to advocate different impact measurements for such infrastructures.

#### e) EURISE Technical Reference - Michelle Weidling (SUB Göttingen)

Michelle Weidling, developer at the Göttingen State and University Library presented the objectives of the **EURISE Technical Reference** (TR) which development started under the umbrella of the DESIR project. It is a collection of **basic guidelines and references** for development and maintenance of infrastructure services within DARIAH and beyond. **Maintainability and usability** are the two key words for software development. Accordingly, the Technical Reference was developed by taking into account these two priorities.

The current list of the Technical Reference contains four main parts. The developer guidelines takes into account the README, the documentation, the "tooling", interoperability and the changelog. The operational guidelines address the infrastructure documentation, documentation for (virtual) servers, security, incidents and post-mortems. Policies recommendations are also given for code hosting and release policy. Finally, the developer can answer a software **quality checklist** to make sure that the product meets certain quality standards.

The suggestion given to those who want to use the Technical Reference is to adjust the focus according to the particular needs of a specific project or institution by extending the basic guidelines found in the TR.

In conclusion, Michelle Weidling shared some good practices when developing such as asking colleagues to do a **short usability test**, dedicating at least one hour per week to **write or update the documentation**, and **using the technical reference from the start** when beginning a new project.

#### f) Implementing a Service Management System in a Federated Multi-Supply Environment - Yannick Legré (EGI)

Yannick Legré, director of the EGI Foundation, presented EGI, an e-infrastructure that provides advanced computing services for research. EGI counts more than 260 data and computing centres, serving 31 large scale ESFRI projects.

The presentation focused on the **IT Service Management** in EGI and in particular on the implementation of FitSM, which is a free standard for lightweight service management suitable for IT service providers. FitSM is operated by ITEMO (IT Education Management Organization), a German non-profit organisation, and is freely released under a CC-BY license.

FitSM should be implemented following a 7–step approach:

1. define the rationale and scope for implementing service management and get top management commitment and support;
2. Identify/assign roles and responsibilities for planning/implementation;
3. ensure training and awareness;
4. perform an initial organisation maturity assessment;
5. define a service management plan with overall goals and milestones;
6. start defining policies, activities and procedures for each process;
7. re-assess progress through formal reviews or audits (e.g. annually).

Yannick Legré also explained the benefits of having an **Integrated Management System** (IMT), which helps to put in place standard processes, procedures and agreements for managing the infrastructure efficiently and effectively, increases clarity on expectations between EGI partners and customers, makes decision-making clearer and gives knowledge and experience that can be reused by other e-infrastructures through consultancy, training, audits and workshops.

### g) Infrastructure Security - David Kelsey (UKRI STFC)

David Kelsey developed the main topic from the point of view of **security**. David is Head of the particle physics computing group at STFC UK and collaborates with EGI. The aim of the EGI operational security is to maintain confidentiality, integrity and availability of service and data and manage security risks in an environment where threats are constantly changing.

**EGI CSIRT**<sup>18</sup> team coordinates operational security activities within the EGI Infrastructure with the objectives of preventing security incidents, enabling a security monitoring, offering a proper policy framework for operational security and providing the necessary skills through training and dissemination activities.

---

<sup>18</sup> <https://www.egi.eu/wp-content/uploads/2017/07/EGI-CSIRT-report-July-2017.pdf>

The speaker also introduced the activities of the **WISE community**<sup>19</sup>, an initiative that aims to enhance best practice in information security for IT infrastructures for research. The community represents an excellent way to strengthen collaboration with other infrastructures and helps to improve security policies and procedures. At the end of the AARC project<sup>20</sup>, its policy templates will become part of WISE

#### **h) A FAIR Software Route - the Dutch Way - Mustapha Mokrane (DANS) / Carlos Martinez Ortiz (NLeSC)**

Mustapha Mokrane and Carlos Martinez Ortiz introduced the initiatives in the Netherlands to apply the **FAIR principles to software**. The main challenges particularly encompass the reusability and interoperability of software.

They suggest the following best practices to get to the FAIR software route:

- different levels of FAIR Compliance for different types of software (similar to curation levels for datasets)
- initial (minimum) focus on Findability and Accessibility
- provide recommendations for good (FAIR) coding practices
- centralised versus decentralised route/approach (distributed chain of responsibilities or single point of failure).

The Netherlands eScience Center encourages the reuse of research software by developing the **Research Software Directory**<sup>21</sup>, a platform that offers tools to promote the exchange and re-use of research software.

In conclusion, a first step in order to make the software route happen is to bring in the researchers/developers and repositories together. However, a reflection on incentives (internal and external) should also be done, as well as developing skills through training is crucial to support the whole construction.

---

<sup>19</sup> <https://wise-community.org/>

<sup>20</sup> <https://aarc-project.eu/about/>

<sup>21</sup> <https://www.research-software.nl/>

i) Sustainability beyond guidelines - How our perception of research infrastructures shapes our perception of the sustainability issue? - Andrea Scharnhorst / Francesca Morselli (DANS)

Andrea Scharnhorst and Francesca Morselli brought into the discussion a concept of sustainability that identifies the human capital as a key factor for sustainability. The project CENDARI<sup>22</sup> was presented as use-case. One of the conclusions of the project was to consider **sustainability as a process rather than a state**. Also, **human capital is decisive** in terms of sustainability of knowledge capital and sustainability of user communities.

The definition of sustainability depends on the viewpoint one chooses and ideally they are all represented in a sustainable analysis (engineering, technical...). But the accent on people is crucial and the turnover and the loss of know-how have a serious effect on sustainability.

As a summary of the discussion, sustainability is presented as a process rather than a tickbox and needs early setting of goals. It requires proper documentation, stable institutions but also the human capital plays an important role. The main threats are ephemeral institutional settings and the diverse roles, responsibilities, skills set of the human capital involved. Both can affect the transfer of knowledge which is at the centre of sustainability. In conclusion, **stable institutions, profound and agile education, defined career paths and functions are preconditions for sustainability**.



Image 3 – Andrea Scharnhorst summarises her presentation

<sup>22</sup> <http://www.cendari.eu/>

## j) Breakout Session: Discussion

Over lunch, breakout sessions were formed to discuss the following topics:

- technical reference/developer guidelines
- security
- service management
- FAIR software
- software preservation
- software discovery
- roles of humans

The outcomes of the discussions were summarised in the plenary during the last session of the workshop, right before the conclusion of the event.