Certification Schemes for Research Infrastructures
-
Technology Watch Report 2

Felix Helfer, Stefan Buddenbohm, Thomas Eckart, Philipp Wieder

DARIAH-DE Working Papers Nr. 43

Helfer, F.; Buddenbohm, S.; Eckart, T., Wieder, P.: Certification Schemes for Research Infrastructures : Technology Watch Report 2
Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2021
(DARIAH-DE working papers 43)

Verfügbar:
PURL: http://resolver.sub.uni-goettingen.de/purl/?dariah-2021-3
URN: http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2021-3-1

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

Erschienen in der Reihe
DARIAH-DE working papers

ISSN: 2198-4670

Herausgeber der Reihe
DARIAH-DE, Niedersächsische Staats- und Universitätsbibliothek

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

Abstract: This working paper discusses the use and importance of various certification systems for the field of modern research infrastructures. For infrastructures such as CLARIAH-DE, reliable storage, management and dissemination of research data is an essential task. The certification of various areas, such as the technical architecture used, the work processes used or the qualification level of the staff, is an established procedure to ensure compliance with a variety of standards and quality criteria and to demonstrate the quality and reliability of an infrastructure to researchers, funders and comparable consortia. The working paper conducts this discussion based on an overview of selected certification systems that are of particular importance for CLARIAH-DE, but also for other research infrastructures. In addition to formalised certifications, the paper also addresses the areas of software-specific and self-assessment-based procedures and the different roles of the actors involved.

Keywords: Zertifizierung, Audit, Digitale Forschungsinfrastruktur, CLARIAH-DE, CLARIN, DARIAH, Standards, Qualitätskriterien

Certification, Audit, Digital research infrastructure, CLARIAH-DE, CLARIN, DARIAH, standards, quality criteria

Certification Schemes for Research Infrastructures

Technology Watch Report 2

Felix Helfer[1] Stefan Buddenbohm[2] Thomas Eckart[1]

Philipp Wieder[3]

[1] University of Leipzig
[2] Göttingen State and University Library
[3] Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen

Felix Helfer, Stefan Buddenbohm, Thomas Eckart, Philipp Wieder: "Certification Schemes for Research Infrastructures". *DARIAH-DE Working Papers* No. 43. Göttingen: DARIAH-DE, 2021.
URN: urn:nbn:de:gbv:7-dariah-2021-3-1.

The *DARIAH-DE Working Papers* are published by Mirjam Blümm, Thomas Kollatz, Stefan Schmunk and Christof Schöch.

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Abstract

This working paper discusses the use and importance of various certification systems for the field of modern research infrastructures. For infrastructures such as CLARIAH-DE, reliable storage, management and dissemination of research data is an essential task. The certification of various areas, such as the technical architecture used, the work processes used or the qualification level of the staff, is an established procedure to ensure compliance with a variety of standards and quality criteria and to demonstrate the quality and reliability of an infrastructure to researchers, funders and comparable consortia. The working paper conducts this discussion based on an overview of selected certification systems that are of particular importance for CLARIAH-DE, but also for other research infrastructures. In addition to formalised certifications, the paper also addresses the areas of software-specific and self-assessment-based procedures and the different roles of the actors involved.

Schlagwörter

Zertifizierung, Audit, Digitale Forschungsinfrastruktur, CLARIAH-DE, CLARIN, DARIAH, Standards, Qualitätskriterien

Keywords

Certification, Audit, Digital research infrastructure, CLARIAH-DE, CLARIN, DARIAH, standards, quality criteria

Contents

1 Introduction

This[1] is the second of three reports[2] composing the *technology watch of CLARIAH-DE*. The first technology watch report discusses digital repository solutions, in the context of research infrastructures[3]. Simultaneously, an evaluation of current PID solutions is published[4].

The aim of all three reports is to give an overview of technological developments relevant to the project and its partners, and to offer recommendations concerning their adaptation within CLARIAH-DE.

CLARIAH-DE is the merger of the two established German research infrastructures CLARIN-D and DARIAH-DE. An important task within this merge is the evaluation and – where possible – the integration of infrastructure components or services. Particularly the aspect of evaluation is of interest for a broader audience as a merger of research infrastructures always comes with challenges related to legacies or restrictions on various levels. Such legacies or restrictions may contain:

- Technical or infrastructural legacies: certain software stacks, standards, APIs, or reliance on components that lay in the responsibility of third parties (such as a data centre) and may not easily be subjected to change;
- Science- and user-related legacies: e.g. as simple as a well-established branding of services or more complex like the practices how to describe, present or work with research data; use of selected metadata schemas; publishing practices;

These two examples illustrate that decisions in a merger process are never to be made unattached to the past[5]. Although this is a commonplace for most informed readers it is helpful to bring this up as context. Only against this background certain decisions or deficiencies (and very often delays) can be explained. The value for a research infrastructure to acquire honest communication in this regard may not be underestimated. It adds to the level of trust and understanding among its users.

The "technologies" focused on in this report are *certification schemes* relevant for various stakeholders and participants of research infrastructures. Technology is to be perceived in a broader sense here

[1] The authors express their gratitude to Marthe Eisner (Göttingen State and University Library) for her valuable support in preparing this report.

[2] The third report is available: Eckart, Thomas, Jander, Melina, Helfer, Felix, Jegan, Robin, & Buddenbohm, Stefan. (2021, March 31). CLARIAH-DE and the European Research Infrastructure Level. Zenodo. http://doi.org/10.5281/zenodo.4650440. For information: all CLARIAH-DE related publications are available as Zotero bibliography: https://www.zotero.org/groups/2494199/clariah-de/library

[3] See: Arnold, Denis, Fisseni, Bernhard, Helfer, Felix, Buddenbohm, Stefan, & Kiraly, Peter. (2020, March 31). Repository Solutions - Technology Watch Report 1. Zenodo. http://doi.org/10.5281/zenodo.3873027

[4] Although not a TechWatch report but of interest in this context. The CLARIAH-DE report on current PID solutions with focus on CLARIN and DARIAH: Buddenbohm, Stefan & Eckart, Thomas. (2020, April 8). Persistent Identifiers in CLARIAH-DE Summary & Best Practices (AP4.1.5). Zenodo. http://doi.org/10.5281/zenodo.3744091

[5] An illustrative article on legacies in CLARIAH-DE with the example of search logics is available (to be published) as DARIAH-DE Working Paper. See: Eckart, Thomas et al. (2021): CLARIAH-DE Cross-Service Search: Prospects and Benefits of Merging Subject-specific Services. DARIAH-DE Working Paper Nr. 41, Göttingen: DARIAH-DE, 2021. http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2021-1-9. See also: Buddenbohm, Stefan. (2020, November). CLARIAH-DE - Aligning two Research Infrastructures: Experiences and Challenges. Presented at the Scholarly Primitives - DARIAH Annual Event 2020, Zagreb, Croatia: Zenodo. http://doi.org/10.5281/zenodo.4266478

as it includes aspects of standardisation, aids, resources, quality guidelines and procedures to ensure these. With research infrastructures making up the core of CLARIAH-DE, in turn, the reliable storage, management and dissemination of research data form an essential task in this environment. Formal certification of used architectures, processes, qualification status of personnel, etc. are established means to ensure compliance with a variety of criteria and to demonstrate the reliability of an infrastructure to its users, funders and other research infrastructure consortia, all of which have an interest in reliable research infrastructures based on individual motives.

In the following, the significance of audit and certification schemes in scientific environments in general is introduced, followed by a descriptive part of certification schemes which is augmented with the perspective of CLARIAH-DE. Nevertheless the collected material tries to be as applicable as possible for other research infrastructures addressing other disciplines and audiences than CLARIAH-DE does. The report is concluded by a larger discussion trying to establish analysing patterns and highlight selected aspects of relevance from a CLARIAH-DE point of view. The selection of the discussed certification schemes has to be read from the perspective of a research infrastructure like CLARIAH-DE and may turn out differently for research infrastructures in other research domains. Guiding criteria for the selection of the certification schemes below will be discussed in the following chapter but it is helpful to distinguish the impact or use of certificates along the main audiences, for example certain certificates may be legally required to be allowed to offer a service. Other certificates may be necessary for a data centre to express interoperability or adherence to scientific standards and gain access to consortial, e.g. for persistent identification or longterm preservation of research data. And of course, many certificates intend to add to the level of trust the users have in a service. For a research infrastructure such users may be researchers, research institutes or research infrastructures like libraries. The following schemes will be introduced:

- Core Trust Seal (CTS)
- DINI Certificate for Open Access Repositories and Publication Services
- Nestor/DIN 31644
- Information Technology Infrastructure Library (ITIL)
- IT Service Management FitSM
- ISO norms 9001, 16363, 20000, 27001
- DataCite Registered Service Provider Program
- DuraSpace Service Provider Program
- ePIC Quality of Service and Policies

Some self-assessment guidelines are also relevant for research infrastructures even though they are not formalised certification or audit schemes. However, they come with the clear advantage being derived from a research context, which proves for a research-oriented perspective. The following five examples have been identified as relevant from a CLARIAH-DE perspective:

- FAIR Principles
- Plan S
- TRUST Principles
- EURISE Network
- DARIAH-DE Service Life Cycle

From a formal point of view, *certification schemes*[6] (and as a consequence thereof the audits) in scientific environments resemble certifications in other environments, e.g. in commercial settings, in terms of:

- A certificate aims to ensure compliance to a certain set of criteria.
- A certificate can be a (legal) requirement to be able to operate a service or can be a visible sign for users and customers to create trust in the service.
- Usually certifications have to be renewed in defined intervals and by this are capable to take step with the overarching development in the sector.
- Repeated audits are proof of the actuality of the certificate.

These aspects are relevant in scientific environments especially considering the growing importance of long-term availability/preservation, usability, and re-use of (scientific) resources.

Well-known certificates considerably increase the *level of trust* users have in a service or infrastructure[7]. This is of particular importance for research infrastructures as their use by researchers often does not fit into a conventional provider-customer pattern (i.e. they are not charged for the use of resources or don't have a SLA or contract with the provider). Research infrastructures are often provided in a distributed or non-local way (i.e. the user no longer deals with their known home institution, be it a university or a data centre). Last but not least scientific output – be it a set of research data or a publication – often comes with a close personal bonding (i.e. the user guarantees for the scientific quality with her/his name). However, the landscape of available solutions is complex and constantly evolving, making a documented comparison even more relevant for an informed overview to help selecting a viable solution.

[6] Audit is mentioned here to direct attention to the process side of certifications schemes. By audit, it is made clear that the review and decision on the application for a certification is usually exerted by an independent third party. An exception is given by the self-assessment schemes, which are described separately in this report.

[7] For the importance of trust and community involvement with the example of DARIAH-DE, see: Blümm, Mirjam/ Neuroth, Heike/ Schmunk, Stefan (2016): DARIAH-DE – Architecture of Participation. Bibliothek Forschung und Praxis | Band 40: Heft 2. DOI 10.1515/bfp-2016-0026

2 Requirements

From the perspective of CLARIAH-DE the following points are considered particularly relevant:

- Origin and organisational structure of the certification scheme,
- Thematic focus, which might be broadly ranged or focusing on very specific issues (like long term archiving, technical interoperability, security),
- General certification procedure,
- Availability of the specification, documentation material and guides, or general support,
- Effort, costs and continuance of the certification scheme,
- General spread or specific communities where the certification is in active use,
- Potential relevance for stakeholders in research infrastructures.

3 Certification Schemes

In this section, a variety of certification schemes will be presented and discussed, mostly regarding the previously stated requirements, if applicable. They are discussed in subsections each dedicated to a specific structural, technical, or organisational aspect that is relevant for a research infrastructure. The presentation of the certification schemes is followed by a discussion of the relevance, highlighting selected certification schemes. For some certifications specific examples from CLARIAH-DE can be provided, e.g. certified services. However, this is not possible for all certifications, but the relevance of a specific schema is always explained.

3.1 Repositories & Archives

3.1.1 Core Trust Seal CTS

Figure 1: Logo of Core Trust Seal

The Core Trust Seal (CTS)[8] is a certification organisation that was formed as a joint project of the World Data System of the International Science Council (ICSU-WDS[9]) and the former Data Seal of

[8] https://www.coretrustseal.org/

[9] https://www.worlddatasystem.org/

Approval (DSA) under the umbrella of the Research Data Alliance RDA[10]. The CTS is a non-profit organisation and a Dutch legal entity ("CoreTrustSeal Stichting") situated in The Hague. It is governed by 12 members of the "Standards and Certification Board" which is elected by members of the "Assembly of Reviewers."[11] The CTS certification is designed as a core level certification focusing on the trustworthiness of data repositories regarding its ability to host digital research data with a long-term perspective. It is designed as a first step of repository certification which could lead to extended certifications including Nestor/DIN 31644 or ISO 16363. It is based on a catalogue of 16 requirements[12] dealing with a wide range of organisational, technical and other issues where a certain level of compliance has to be reached by an applicant. There are five levels of compliance ranging from 0 ("not applicable") to 4 ("fully implemented"). A successful certification process requires a minimum compliance level 3 ("in the implementation phase") for all requirements except non-applicable ones which have to be justified in detail. Certification documents have to be provided in English.

The seal is awarded for a period of three years after which it has to be renewed using the then current version of the requirements catalogue. The requirements catalogue has undergone significant changes in the past. However, the declared goal of CTS is that a reapplication should be possible with minimal revisions in most cases. The catalogue in its current version is valid for the period of 2020 – 2022.

The requirements catalogue is divided in three main sections, focusing on the following major aspects:

- Organisation infrastructure, including mission scope, mid- and long-term continuity plans, compliance with disciplinary and ethical norms, details about number and qualification of staff or sufficient funding of the repository in general (six requirements).
- Details about management of digital objects, including the ability to guarantee integrity and authenticity of hosted data, a long-term preservation plan, data and metadata quality assurance procedures, existence and use of defined workflows for all relevant processes, suitable means of data discovery and identification for end users, enabling and support of data reuse (eight requirements).
- Information about the used technology stack (hardware and software), technical infrastructure development and the IT security system including a risk and threat analysis (two requirements).

All requirements and their thematic scope are explained in the catalogue by giving lists of specific questions for the respective requirement which have to be answered by the applicant. It is allowed to refer to publicly available evidence in the application form.

The process of certification relies on the self-assessment of the applicant which has to be submitted using the CTS Application Management Tool[13]. The application form is then reviewed by two auditors and – if necessary – returned with remarks or follow-up questions for a revised submission. A maximum of five revision rounds are typically allowed for a single application. To support interested data repositories,

[10] https://rd-alliance.org/

[11] All illustrations in this document are the intellectual property of the respective owners and are used by permission.

[12] CoreTrustSeal Standards and Certification Board. (2019, November 20). CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 (Version v02.00-2020-2022). Zenodo. http://doi.org/10.5281/zenodo.3638211

[13] https://amt.coretrustseal.org/

the CTS provides introductory documents including a glossary, extended guidance documents[14] and pre-recorded or live Webinars[15]. On the one hand, the CTS certification usually requires a considerable time until achievement. On the other hand, this phase is often an opportunity of learning and adaptation for the applying services, and such useful invested time.

The CTS currently tries to develop a sustainable business model to cover the cost of operation and maintenance, further development of the certification service and training material. At the moment an "administrative fee" of 1000 Euro[16] has to be paid for an application. According to CTS this is seen as "an initial step to cover administrative costs". Under certain conditions waivers or discounts are possible including cases where the repository is situated in a low- or middle-income country or a volume discount for umbrella organisations with ten or more data repositories.

As of November 2020, 101 data repositories are certified according to CTS with a geographical focus on Europe and North America. In addition, 62 repositories are certified using the predecessor certification schemes WDS or DSA[17]. Certified repositories come from a wide range of scientific disciplines, including the humanities, social sciences, geosciences, physics, climatology, astronomy and more.

In the European context, CTS is a well-established means to certify data repositories in research data infrastructures. At least in part, this is due to the requirement for centres participating in the European CLARIN infrastructure for the humanities and social sciences to be certified according to CTS[18], which applies to 24 centres at the time of writing. As a consequence, extensive knowledge and experience about the certification process is available in many German institutions that work in the context of research data infrastructures. Within CLARIAH-DE (taking CLARIN-DE and DARIAH-DE as well into account) several services are CTS certified and verifiable in the CLARIN Centre Registry[19]. For instance, repositories like the DARIAH-DE repository, the TextGrid repository, or repositories situated at the universities of Hamburg, Leipzig, and Tübingen are CTS certified. The CTS may likely become a mandatory certificate for more classes of CLARIN Centres in the future, respectively for CLARIN B-Centres this is already the case today.

The impact of the CTS on information infrastructures and its status as -de facto-standard can also be illustrated by its take in the EOSC Social Sciences and Humanities Open Cloud (SSHOC) project, which in 2020/21 devoted a whole task to the support of repositories applying for CTS certification[20].

[14] CoreTrustSeal Standards and Certification Board. (2020, May 15). Change file: CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2017–2019 to 2020–2022 (Version v01_00). Zenodo. http://doi.org/10.5281/zenodo.3828636

[15] https://www.coretrustseal.org/why-certification/requirements/ ("Extended Guidance")

[16] https://www.coretrustseal.org/apply/administrative-fee/

[17] https://www.coretrustseal.org/why-certification/certified-repositories/

[18] https://www.clarin.eu/content/assessment-procedure

[19] https://centres.clarin.eu/

[20] See: https://www.sshopencloud.eu/d82-certification-plan-sshoc-repositories and accordingly: Kleemola, Mari, Alaterä, Tuomas J., Koski, Niko, Ala-Lahti, Henri, Jerlehag, Birger, L'Hours, Hervé, … Van Horik, René. (2020). SSHOC D8.2 Certification plan for SSHOC repositories (Version v1.0). Zenodo.

3.1.2 DINI Certificate for Open Access Publication Services

The *German Initiative for Network Information (DINI)*[21] has been for many years an *advocacy initiative* addressing challenges related to the transformation of information and communication technologies in the scientific environment. DINI serves not only as discussion forum for stakeholders from research organisations and information infrastructure institutions, it also (among other things) strives to promote best practices and exemplary solutions for re-use; to promote the adaptation and further development of standards; to register competence centres (such as the Göttingen based Competence Centre for Interoperable Metadata). Several working groups are devoted to subjects such as long-term preservation (see also: Nestor); research information systems; metadata interoperability; E-learning, or electronic publishing. All of this is pursued under the principle of community involvement[22].

Figure 2: Logo of DINI

The working group on electronic publishing is authoring the DINI Certificate for Open Access Repositories and Publication Services which will be discussed in the following. Along with the authoring responsibility the working group also maintains the certification process. The subjects of the working group evolve over time according to the general developments in the scientific system: *infrastructures and services for electronic publishing in open access* (e.g. repositories); electronic open access journal platforms; platforms for open access books; adaption of principles of open science for these services; coupling of such publication services to the research data life cycle and research data infrastructures.

The Certificate for Open Access Repositories and Publication Services is DINI's main instrument to promote *standardisation for open access publication services*. In this abstract category fall the majority of publication repositories operated by German research institutions, be they institutional or subject-specific/disciplinary repositories. The certificate has now (last edition in 2019) reached its sixth generation and is by far the most known and reputable catalogue of criteria for publication services in the European information infrastructure landscape. The catalogue of criteria[23] covers all relevant areas for publication services such as organisational, technical, and legal aspects. The DINI Certificate describes itself as the de-facto standard for scientific publication services and has gained international recognition over the years.

[21] https://dini.de/

[22] Müller, Uwe, Scholze, Frank, Arning, Ursula, Beucke, Daniel, Deinzer, Gernot, Höhner, Kathrin, … Wolf, Stefan. (2019). DINI-Zertifikat für Open-Access-Publikationsdienste 2019 - Request for Comments. Zenodo. http://doi.org/10.5281/zenodo.2575346

[23] Müller, U., Scholze, F., Arning, U., Beucke, D., Blumtritt, U., Braun, K., Deppe, A., Deinzer, G., Fenner, M., Klotz-Berendes, B., Meinecke, I., Pampel, H., Schirrwagen, J., Bove, K., Severiens, T., Summann, F., Steinke, T., Tullney, M., Vierkant, P., Voigt, M., Walger, N., Weiland, J. B., Weimar, A., Wolf, S.(2020): DINI Certificate for Open Access Repositories and Publication Services 2019, Berlin : Humboldt-Universität zu Berlin, 45 p. https://doi.org/10.18452/21759

The DINI certificate pursues two strategic goals closely related to the emergence of the open access paradigm in around 2000: (1) Improvement of the open access publishing infrastructure in Germany and (2) strengthening of the open access-based publication formats. Although there is a considerable improvement of both aspects, particularly the second goal prevails.

The criteria are described with look at the core processes for open access publishing services:

- Service provision and consultation for authors and editors,
- Ingestions, curation and long-term preservation of publications, consisting of the electronic document and the according metadata,
- Public provision of publications, ensuring findability and accessibility for human and machine users.

These core processes are usually enforced by the following components, which are subject to the certification process:

- Organisational structure (not subject to certification),
- Technical operation system,
- Interfaces for human users, particularly the web frontend,
- Technical interfaces, particularly OAI-PMH.

To address the dynamic nature of the scientific publication system and to have a means to promote best practices to evolve into de-facto standards, the certificate differentiates between obligatory minimum requirements and optional criteria. It is a common practice that optional criteria become obligatory minimum requirements.

The assessment of an open access publication service, in many cases a repository, follows these steps:

- Preparatory phase by the applicant, i.e. documentation of the current state along the criteria of the certificate. Usually, some open issues remain but this mustn't thwart the submission of the application.
- Selection of reviewers: any application is reviewed by two reviewers (technical review; information science review), who ideally share some experience with the specific application, e.g. with its disciplinary scope or technological base.
- Review process, which is partly based on the documentation submitted by the applicant and auditing of the actual service along the set of criteria. Examples: does the publication provide the necessary documentation for its users in a legally appropriate way? Does the metadata and the OAI-PMH API fulfil the minimum requirements? Is the technical operation of the services meeting established professional standards, e.g. in terms of data security or long term preservation? Usually, this phase of the review includes a communication with the applicant to clarify open questions or to consult on how to meet a certain criterion.

- Decision on the application: Usually the application gets approved as the DINI reviewers see themselves in a consulting role. The certification process is used as an opportunity to share knowledge and experience and to improve open access publication services on a specific level.
- Seal for a DINI certified open access publication service is awarded: this adds to the level of trust users have in the individual service but it is also practically a quality seal, enabling the service to be included in an overall added-value level. This aspect is described below in more detail.

With growing differentiation of roles and the distribution of responsibility regarding the provision of open access publication services, the DINI-ready seal has been introduced. This seal is intended as alleviation for the applying service as it now is spared from documenting the basic technical infrastructure. The basic technical infrastructure is often being provided by data centres, which may receive the DINI-ready status and is from then on a certified hosting service for open access publication services.

In 2021, the DINI certificate can without question be seen as a well-networked and established *quality and trust seal for open access publication services*. This is not only proven by the number of certified services (62 publication services in 2021[24]), but also by the constant revisions and language-versions of the certificate and the involvement of the community for feedback, revisions and absorption of new trends in information infrastructures. This level of trust is considerably contributing to the uptake and use of the services by the researchers and research institutions and one may designate the DINI certificate indeed as a community-driven initiative and stands on the other side of the spectrum as, for instance, commercially-driven certification schemes such as ISO.

On the other hand, the certification comes with a progressing level of standardisation which allows for added value services. A simple but nevertheless impressive example is given by the Bielefeld Academic Search Engine[25], which functions as a harvesting instance and creates a large search space of OA publications including the DINI certified services via OAI-PMH. Due to the standardised use of metadata schemas the publications from the DINI certified services can be better included in such search spaces and allow for a convenient user experience. Also, the DINI certificate proves as a promoter of standardisation. Examples may be seen in the (persistent) author, institution, funder or document identifiers.

This pull to standardisation comes as a mandatory criterion for services which want to certify but exerts also impact on other services as well. Even if a publication service is not opting for the certificate, the guidelines may function as helpful and inspirational for the setup of the service.

From a CLARIAH-DE point of view the DINI certificate gains importance with a look at the publication practices, for instance the DARIAH-DE Working Papers, which are published in a DINI-certified service. Currently CLARIAH-DE doesn't offer a publication service applicable for the DINI certificate but this may change.

[24] https://dini.de/dienste-projekte/publikationsdienste/

[25] https://www.base-search.net/about/en/index.php

3.1.3 Nestor / DIN 31644

Nestor[26], the *network of expertise in long-term storage of digital resources* in Germany, was originally set up as a BMBF (German Federal Ministry for Education and Research) sponsored project between 2003 and 2009. Since July 2009 it has been continued on an independent basis by the former project partners and other organisations. Nestor brings together a "disparate array of institutions"[27] concerned with digital preservation. It is itself not restricted to Germany-based activities and has actively been involved in a number of national and international initiatives and projects.

Figure 3: Logo of Nestor

The network partners with organisations from different fields, connected in some way with the subject of *digital preservation*. There are currently 22 partnerships with German organisations and institutions (e.g. the Bundesarchiv[28]), a small number of international agreements, and a few associated partnerships with German institutions, for example the Computerspiele Museum Berlin[29].

Interested parties from private and public domains can contribute to Nestor working groups, providing expert knowledge on selected individual topics (e.g. Personal Digital Archiving[30]). International developments are monitored, evaluated and factored into future planning and activities. A total of 13 active and five inactive working groups constitute the so-called *nestor competence network*.

Nestor offers a certification in the form of the *nestor Seal for Trustworthy Digital Archives*[31]. It serves as a supervised extended self-evaluation based on the "criteria for trust-worthy digital long-term archives" defined in DIN 31644. The current fee is 500 euros and it is possible to acquire extended certification as part of an European certification process. According to the "Memorandum of understanding" the Nestor seal qualifies as "extended certification": It requires more effort than a simple self-evaluation ("basic certification"), but less effort than an intensive examination by field experts ("formal certification"). Besides DIN 31644, the evaluation criteria can also be based on ISO 16363[32].

According to Nestor, a digital repository is "an organisation (consisting of people and technical systems) which has assumed responsibility for the long-term preservation and long-term availability of digital data and its provision for a specified designated community" (Nestor 2009). The evaluation is thus based on organisational as well as technical aspects, not solely on software or hardware solutions, and does not assess the archive's contents.

An organisation interested in acquiring the Nestor seal registers its interest and two contact persons. It also provides an accurate specification of the evaluation subject. The registration is confirmed by

[26] https://www.langzeitarchivierung.de/

[27] https://www.langzeitarchivierung.de/Webs/nestor/EN/nestor/nestor_node.html

[28] https://www.langzeitarchivierung.de/Webs/nestor/EN/nestor/Partner/bundesarchiv.html

[29] https://www.langzeitarchivierung.de/Webs/nestor/DE/nestor/Partner/CSM.html

[30] https://www.langzeitarchivierung.de/Webs/nestor/EN/Arbeitsgruppen/AG_Personal_Digital_Archiving/ag_personal_digital_archiving.html

[31] https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html

[32] http://nbn-resolving.de/urn:nbn:de:0008-2019062507

Nestor and a single contact person responsible for the remaining process is appointed. The applicant then starts a detailed self-evaluation based on a form provided by Nestor. Different criteria need to be evaluated based on their degree of fulfillment, within a system of 0, 3, 6, or 10 points. When a 6 or 10 is given, public documents verifying the claims must be provided with the form. To pass, 12 criteria need to have 10 points, with the rest above an average of 7. The completed form and additional documentation in German and English is sent to Nestor's contact person and reviewed for plausibility and quality by two reviewers. Based on their report, the seal will be awarded or refused. The seal includes the year it was awarded and does not expire.

For the Technische Informationsbibliothek Hannover (TIB), the certification process took 12 months[33], with a total of 193.5 working days between 16 employees. Between 2016 and 2017, another three archives have been certified: the ZBW Leibniz-Informationszentrum Wirtschaft, the Data Archiving and Networked Services (DANS), and the Deutsche Nationalbibliothek (DNB).

Currently no CLARIAH-DE service or component is Nestor-certified but as the abovementioned examples make clear, the involved data centres and libraries in CLARIAH-DE may follow up on the Nestor seal in the future with stabilising sustainability of its offerings as long-term preservation is an important aspect for any service archiving research data or research publications. However, the aspect of long-term preservation is covered partly already by the Cores Trust Seal. The CTS is a quasi-standard among the CLARIN Data Centres and seems to be more compatible on a European scale.

Nestor has gained a reputation as important focal point in the German information science and infrastructure community for standards, information and consultation and is in this regard important for CLARIAH-DE although no service may have received the Nestor seal. For example the Nestor summer schools[34] are know as important multiplicator and knowlegde sharing events and may illustrate the indirect but considerable impact of Nestor.

3.1.4 ISO 16363[35]

The "Audit and Certification for Trustworthy Digital Repositories"[36] is a framework focusing on the organisational structure of a repository, its digital object management, and infrastructure and security risk management. The standard defines more than 100 metrics which can be used for self-auditing but also for an external certification. A similar certification scheme established in the research environment is available with Core Trust Seal (CTS) described above. The CTS currently marks a common quality level for research data repositories and is the successor of the Data Seal of Approval (DSA).

[33] https://www.langzeitarchivierung.de/Webs/nestor/SharedDocs/Downloads/DE/praesentationen/2019VertrauenSieIhremArchivSchwab.pdf

[34] http://nestor.sub.uni-goettingen.de/education/index.php

[35] https://www.din.de/de/mitwirken/normenausschuesse/nl/veroeffentlichungen/wdc-beuth:din21:151263066

[36] https://public.ccsds.org/pubs/652x0m1.pdf

[37] Illustration by the German National Library: https://www.dnb.de/DE/Professionell/Erhalten/Zertifizierung/zertifizierung.html

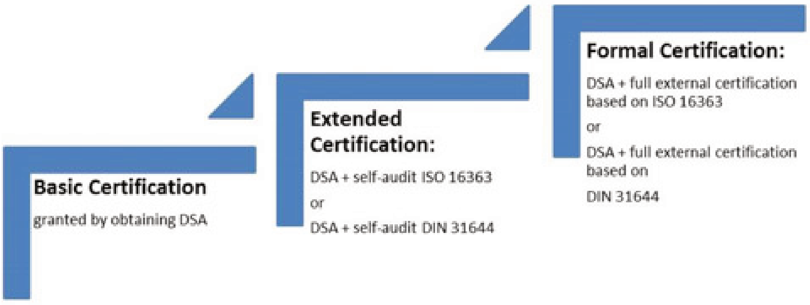
Basic Certification
granted by obtaining DSA
Extended Certification:
DSA + self-audit ISO 16363
or
DSA + self-audit DIN 31644
Formal Certification:
DSA + full external certification based on ISO 16363
or
DSA + full external certification based on
DIN 31644

Figure 4: The German National Library (DNB) described in 2019 ISO 16363 as a possible extension of DSA/CTS[37]

With regard to CLARIAH-DE and possibly all research data repositories the attention for the CTS as successor of the DSA seems to prevail compared to the ISO norm. For this reason ISO 16363 is not discussed in detail in this document.

3.2 General Quality Management

3.2.1 ISO 9001

ISO 9001[38] is a wide-spread norm and, as a foundation for *quality management systems*, basically applicable to any kind of organisation with the aim of ensuring certain levels of quality for services or products. The norm is not specifically intended or designed for research infrastructures or data centres.

The norm came into being in 1987, is currently available in the fifth revision (ISO 9001:2015) and is authored and revised by the *International organisation for Standardization (ISO)*. It has to be noted that the ISO itself is not the auditing body, but the particular auditing rights are granted to specific organisations like e.g. TÜV or DQS, to name some examples in Germany. ISO 9001 is available in various national derivatives, which are similarly structured along ten categories[39]:

- Section 1: Scope
- Section 2: Normative references
- Section 3: Terms and definitions
- Section 4: Context of the organisation
- Section 5: Leadership
- Section 6: Planning

[38] https://www.din.de/de/mitwirken/normenausschuesse/nqsz/veroeffentlichungen/wdc-beuth:din21:242367583

[39] A number of other ISO norms, eg. 20000 and 27001, follow a similar structure. This helps to integrate different management systems into one.

- Section 7: Support
- Section 8: Operation
- Section 9: Performance evaluation
- Section 10: Continual Improvement

Figure 5: Logo of ISO 9001

Along these categories, the fee-based audit is exercised by an external auditor, who certifies the compliance of the applicant to the criteria according to the above mentioned categories. This process is executed yearly to check whether the audited organisation follows the documented processes within the organisation's scope and according to the requirements of customers, users and other interested parties.

The definition, documentation, and introduction of a quality management system according to ISO 9001 is a perennial endeavour that requires substantial resources, motivation and staff training within the respective organisation. But as ISO 9001 allows to maintain any kind of processes as part of a quality management system, it is well-suited to be individually applied to IT service environments with a self-defined scope. Examples exist where academic data centres solely applied ISO 9001 to their service desk compared to others, where all relevant processes (referred to as value-adding processes within the norm) are maintained and improved according to ISO 9001.

As norms like ISO 9001, ISO 20000 and ISO 27001 follow the same structure and share quite a substantial amount of common requirements, ISO 9001 is a good foundation to start with quality management in an organisation and later fulfil the requirements of other, selected norms within the already existing framework.

With looking at CLARIAH-DE - or in this regard any research infrastructure - certificates aimed at quality management like the ISO families mentioned above, become important for data centres[40], libraries or other research infrastructures entities. Such certificates may not be associated directly with the research infrastructure by its users - the researchers - but an informed observer will take note of it.

The relevance for research infrastructure comes with the internalisation of best practices and quality standards in general, e.g. for data centres or information infrastructure institutions. Larger funding schemes such as the German National Research Data Infrastructure (NFDI)[41] or the European Open Science Cloud (EOSC)[42] may expect such quality declarations from beneficiaries.

[40] For instance the ISO 9001 certificate of the GWDG as data centre: https://www.gwdg.de/documents/20182/62973/Zertifikat_ISO+9001.pdf

[41] https://www.nfdi.de/

[42] https://eosc-portal.eu/

3.3 IT Service Management

Information Technology Service Management (ITSM) are activities and processes to organise, operate, and develop IT services provided by an institution. They are in part focused on the qualification and certification of personnel (not of institutions as a whole) or specific parts of their infrastructure. In the contexts of research infrastructures like CLARIAH-DE, many services with a public visibility and high requirements for their resilience and performance are provided to the research community. Well-structured processes are therefore required to ensure adequate performance on all organisational and technical levels. This includes a wide range of services and is supported by infrastructure components like central technical monitoring, a helpdesk, quality criteria, processes for developing and using research software, and more.

3.3.1 ITIL

The *Information Technology Infrastructure Library (ITIL[43])* contains best practices and guidelines for providing IT services in a structured way on the basis of defined processes, procedures and tasks. It was originally developed at the UK's Central Computer and Telecommunications Agency (CCTA) in the 1980s as recommendations for standardising IT management practices. Since 2013 it has been a trademark of AXELOS which is a joint venture by the Government of the United Kingdom and the Capita plc.

ITIL structures the delivery of an IT service into six key activities (Plan, Improve, Engage, Design and Transition, Obtain/Build, Deliver and Support) and defines 34 management practices (structured in „General management practices", „Service management practices" and „Technical management practices").

The ITIL certification is person-related. Organisations cannot, contrary to e.g. ISO/IEC 20000, be certified according to ITIL. As of ITIL 4, certification offers two separate paths which both start with the „Foundation" level. „Managing Professional" (MP) and „Strategic Leader" (SL) are the follow-up levels, focusing on the technical operation of an IT service (MP) or cross-relationships between IT and business strategy (SL). All these certifications consist of a set of dedicated certification modules. A certification level „ITIL Master" can be reached after a minimum of five years working in ITSM which is not based on a fixed set of exam modules.

There is a broad range of training providers[44] with widely varying prices. Costs for the certification exam also varies heavily between countries, typically ranging from 150€ to 450€.

ITIL is considered to be one of the most popular references and certification schemes in the context of ITSM. However, its complexity and required effort is sometimes criticised as being too extensive. This might also be the case from the perspective of research Infrastructures like CLARIAH-DE, where these person-specific certifications could carry less weight, as with possible fluctuations in personnel, a focus on the services themselves could prove more beneficial. Thus, an extensive process like ITIL might not be worthwhile comparing costs and gains.

[43] https://axelos.com/best-practice-solutions/itil

[44] https://www.axelos.com/find-a-training-provider

3.3.2 FitSM

FitSM[45] is a standards family focusing on a „lightweight“ approach to ITSM. It is based on the results of the FedSM project that aimed to improve service management in “federated e-Infrastructures” while ensuring compatibility with other ITSM schemes like ITIL. FitSM is managed and developed by a working group of the *IT Education Management Organisation* (ITEMO[46]).

Figure 6: Logo of FitSM

FitSM is based on 14 processes as well as 16 general and 69 process-specific requirements for an ITSM system. The standard is structured in seven parts (including parts explaining vocabulary, requirements, the FitSM role model or implementation guides). The project emphasises open availability of all material, including extensive training material[47] which is provided in different languages (English, German, partially Spanish). Training courses are provided by several organisations[48]. FitSM certification is organised in three levels (Foundation, Advanced and Expert). Exams containing 20 – 30 multiple choice questions can be taken at several partner organisations with costs ranging from 80 – 160€ “exam/certification fee” varying between the different levels. FitSM certifications have gained popularity in contexts where more complex and expensive certifications might not be sustainable. It is used in the European Open Science Cloud (EOSC[49]) in both EOSCpilot[50] and EOSC-Hub[51] as a reference framework to qualify staff and to structure processes. With its more lightweight approach compared to ITIL, this certification might also be a more feasible option for research infrastructures like CLARIAH-DE and the abovementioned EOSC example may serve in the future as blueprint for large national schemes like the Nationale Forschungsdateninfrastruktur[52] (NFDI) as well.

3.3.3 ISO/IEC 20000

ISO/IEC 20000[53] is a norm for IT service management and defines minimum requirements for IT processes to ensure a defined level of quality. These minimum requirements have to be implemented and documented by the applying organisation – which can come from a broad range, including academic data centres – which is subject to the audit. By these processes the proper management of IT services by the organisation is ensured.

[45] https://www.fitsm.eu

[46] https://www.itemo.org/

[47] https://www.fitsm.eu/downloads

[48] https://www.fitsm.eu/training-organisations/

[49] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[50] https://eoscpilot.eu/

[51] https://www.eosc-hub.eu/

[52] https://www.nfdi.de/en-gb

[53] https://www.din.de/de/mitwirken/normenausschuesse/nia/veroeffentlichungen/wdc-beuth:din21:296602063

Figure 7: Logo of ISO

ISO/IEC 2000 defines the following requirements and process, which are obviously relevant for the development and deployment of IT services in the academic environment:

- Requirements for the management system
- Conceptualisation and implementation of a service management
- Conceptualisation and implementation of new or changed services
- Service level management
- Service reporting
- Availability and service continuity management
- Planning of budget and accounting for IT services
- Capacity management
- Information security management
- Business relationship management
- Supplier management
- Incident management
- Problem management
- Configuration management
- Change management
- Release and deployment management

As ISO/IEC 2000 is focused more on oranisations than personnel, it has a lot more significance for research infrastructures compared to, for example, ITIL. Its postulated requirements are also very much in line with the goal of modern academic digital services, though with all ISO norms its public facing signaling power may be restrained by the restricted access of its content.

3.4 Information Security

3.4.1 ISO/IEC 27001

ISO/IEC 27001[54] "*Information technology – Security techniques – Information security management systems – Requirements*" documents by its successful audit the presence of an information security management system (ISMS) in an organisation. The norm includes requirements for the assessment and handling of information security risks. ISO 27001 is part of the larger family of norms ISO27k which contains specific derivatives for various organisation types.

The purpose of the norm is to define "*the requirements for establishing, implementing, maintaining and continually improving an information security management system*". Whether or not these requirements are fulfilled is subject to regular audits executed by an external, certified entity. The norm is similarly organised as the ISO/IEC 9001 norm and requires organisations to clearly define the following: context of the organisation, leadership, planning, support, operation, performance evaluation, and improvement. Annex A of the ISO/IEC 27001 norm contains a list of so-called controls and their objectives, which are in details described in ISO/IEC 27002 (see below).

The ISO 27001 norm, which describes the actual requirements for an ISMS, is complemented by a number of so-called guideline standards, which cover topics like risk management or how to implement an integrated management system covering ISO 27001 und ISO 20000. With respect to the scope of this document and the envisaged focus of certification interests of CLARIAH-DE partners and similar organisations, these guideline standards are in general not in the core focus. There is one particular exception, which is ISO/IEC 27002[55] "Information technology — Security techniques — Code of practice for information security controls". This guideline standard contains 114 controls in 35 main security categories, which are further subsumed in 14 security control clauses. These controls can be seen as guiding principles for the management of information security that are applicable to most organisations. They also help to design and implement the ISMS processes and to evaluate the actual compliance of an organisation with respect to ISO 27001. Examples for controls are "Inventory of assets", which provides guidance on how "Assets associated with information and information processing facilities should be identified and an inventory of these assets should be drawn up and maintained.[56]" or "Secure log-on procedures", which describes that "Where required by the access control policy, access to systems and applications should be controlled by a secure log-on procedure.[57]"

Evaluating and implementing the ISO/IEC including the 114 controls requires substantial know-how and effort. It is therefore necessary to properly plan an ISO 27001 certification and support it with sufficient resources. Furthermore, it should be carefully evaluated whether the operation of an ISMS is something an organisation wants to do and, evenly important, whether it has the resources and the intention to maintain and evolve it. From an information security perspective, ISO/IEC 27001 is a norm with substantial impact on the processes of an organisation, which helps to improve the security standards, raise awareness, and get a clear view on risks and contingency measures. As such, it complements

[54] https://www.din.de/de/mitwirken/normenausschuesse/nia/veroeffentlichungen/wdc-beuth:din21:194462684

[55] https://www.iso.org/standard/54533.html

[56] ISO/IEC 27002, second edition, 2013-10-01, Section 8.1.1. "Inventory of assets"

[57] ISO/IEC 27002, second edition, 2013-10-01, Section 9.4.2. "Secure log-on procedures"

repository or archive-specific norms and provides the necessary grounding for "service-specific end-to-end information security". However, it is generally too much effort for organisations operating a domain-specific repository or a smaller archive to get and maintain an ISO/IEC 27001 certificate. Although one can limit the scope of the certification to this one repository, still the controls have to be implemented. We therefore see ISO 27001 certificates mainly at large IT, telecom, or cloud providers or larger data centres.

3.5 Tool-specific Certifications

Besides the more general certifications and standards previously discussed, a growing number of organisations and companies offer certifications specifically related to their own software or services. This is of particular interest for research infrastructures as they are usually not constructed as monolithic blocks but can be better described as a framework or array of resources, which can function and be used separately from one another.

These tool-specific certifications are only feasible if the underlying software or service is in use by the party seeking certification, but can, in that case, signify competency and expertise in the respective system, while often also allowing for a closer and more influential communication with the maintainer of the technology in question. However, they often entail significant requirements, not least including financial contributions. Three of these certification programs will be exemplarily discussed, as their topics relate to the context of research infrastructures. This should give some insight into tool-specific certifications and their costs and benefits without going beyond the scope of this report.

3.5.1 DataCite Registered Service Provider Program

Figure 8: Logo of DataCite

DataCite[58] is a non-profit organisation providing digital object identifiers (DOIs) for research data and other research outputs and developing services for DOI management. Their API can be integrated to allow other organisations the registration of DataCite DOIs. The membership entails an annual fee (different tiers starting at around 500€). To join the Registered Service Provider Program, an application can be submitted via an online form and is expected to take up to two weeks to review[59]. A re-registration through self-assessment is to be performed every January.

[58] https://datacite.org/service-provider-program.html

[59] A good overview from the perspective of the applying institution is available by: Kümmet, Sonja, Lücke, Stephan, Schulz, Julian, Spenger, Martin, & Weber, Tobias. (2019, November 15). DataCite Best Practice Guide (Version Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.3559800

There are a number of requirements for becoming a *registered service provider with DataCite*: Most obviously, the applicants need to have an existing integration with the REST API for their DOI registration services and be able to demonstrate that findable DOIs have been registered. Furthermore, metadata submitted to DataCite is expected to be compliant with the DataCite Metadata Schema 4 (or more recent), the applicant is expected to provide a "secure means for their users to submit DataCite member credentials" and user support, both however not specified further on the DataCite website. The applicant should also follow DataCite's best practices for service providers[60], which details practices for DOI assignment and handling. Lastly, a designated contact for communication with DataCite and some basic metadata about the provided service are required.

Service providers that registered successfully for the program will then be listed on the DataCite website. They receive a badge to display on their website and are supposed to be held in close communication with DataCite staff and other registered service providers.

3.5.2 DuraSpace Service Provider Program

Figure 9: Logo of DuraSpace

In the *DuraSpace Service Provider Program*[61] from DuraSpace (which, in turn, is part of the LYRASIS non-profit organisation as of July 2019), an organisation can become a certified partner for one of the DuraSpace technologies, namely DSpace, Fedora (and Samvera/Islandora), VIVO, DuraCloud, and ArchivesDirect.

The application is submitted online and, if promising, followed up by an interview. Payment involves a percentage of gross revenue (minimum $2,500 USD). Apart from the financial contribution, the submission of two client references concerning collaboration satisfaction per year is required for a continued certification. Furthermore, the organisation is expected to be an active contributor in the associated community, either by technical, educational, or other involvement (minimum 250h annually). Contribution and revenue reports are also mandatory.

A certified partner can then participate in the project leadership committee handling "priorities and strategic direction" of the technology in question, will be included in promotional efforts (including display on program websites) and can use the corresponding logo.

[60] https://datacite.org/documents/DataCite_BestPractices_ServiceProviders_v1.pdf

[61] https://duraspace.org/community/service-providers/

3.5.3 ePIC Quality of Service and Policies

Figure 10: Logo of ePIC

Similar to DataCite, ePIC is relevant in the context of persistent identification. Identifiers are in general more and more common not only for publications but for a variety of other subjects or entities. Accordingly, the demands for quality assurance, reliability and transparency are growing as well and the ePIC Quality of Services and Policies[62] as an internal certification scheme can be mentioned as an example. All members of the ePIC consortium and – by this – the PID services have to pass through the set of organisational and technical criteria (e.g. for mirroring or replication), which aim at the transparency about the reliability of ePIC PIDs. The relevance of this quality initiative breaks down to research infrastructures such as CLARIAH-DE, which relies on PID services and make use of them conspicuously or inconspicuously.

3.6 Self-assessment Principles

There also exist various so-called *principles* concerning data or repository management for which no third-party organisation will assess their implementations and issue certifications, but which, if adhered to, can nonetheless signal a certain reliability concerning their specific scope. These might also present alternatives to actual certifications if, for example, the financial fees of a certification process prove to be an issue. Thus, five of these principles for self-assessment that are relevant to research infrastructures will be briefly presented here.

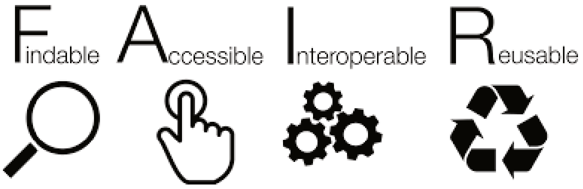

Figure 11: FAIR principles

The *FAIR principles* were designed to guide scientific data management (Wilkinson et al., 2016). For this, they postulate a list of requirements for the data, metadata and protocol of the management system. These requirements are grouped around the four main principles of findability, accessibility, interoperability and reusability. An example for a declared compliance with the FAIR principles can be seen on the Zenodo website[63]. Wherever applicable CLARIAH-DE advocates the FAIRification of services and resources provided by the research infrastructure. In this regard the FAIR principles may

[62] https://www.pidconsortium.net/?page_id=904
[63] https://about.zenodo.org/principles/

be considered similarly to the Open Access principles. Both are (or at least should be) nowadays inherent parts of research infrastructure practioning. It would be desirable for CLARIAH-DE to publish a FAIR declaration to break down the abstract FAIR principles to its individual services, tools and resources. Currently CLARIAH-DE details its support of the FAIR principles in the section "Principles and Standards"[64].

Plan S[65] is an initiative for open-access publishing of scientific data and proposes, among other things, principles for open access repositories, which are grouped into mandatory and recommended additional criteria. The criteria include requirements for metadata, PIDs, availability and more. A self-assessment example against these requirements can also be seen on the Zenodo website mentioned above.

Figure 12: TRUST principles

The *TRUST principles*[66] by the Research Data Alliance, first published in the Scientific Data journal, aim to offer guidance maintaining trustworthiness for a digital repository (Lin et al., 2020). For this they propose five principles regarding the stewardship of repositories for digital research data. These principles are differentiated further via several core requirements necessary to fulfill the principle. The TRUST principles are endorsed by a multitude of organisations, including CoreTrustSeal and Nestor, both of which were discussed in previous sections. The *EURISE Network*[67] is the European Research Infrastructure Software Engineers' Network and has been founded by the ERICs[68] CESSDA, CLARIN and DARIAH to establish a platform for discussion and exchange on matters of software quality. EURISE wants to introduce common quality standards, best practices and education in the area of software development for research.

EURISE Network
European Research Infrastructure Software Engineers' Network

Figure 13: Logo of EURISE Network

EURISE can be understood against the background of the manifold software landscape in research infrastructures in the social sciences and humanities. Often, a research-driven approach in the software

[64] https://www.clariah.de/en/consulting-training/principles-and-standards

[65] https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/

[66] https://www.rd-alliance.org/rda-community-effort-trust-principles-digital-repositories

[67] https://eurise-network.github.io/

[68] European Research Infrastructure Consortium

development is prevalent. This comes with many benefits and often the impetus to develop a software solution for tasks in the research process can only be formulated by the researchers. But the drawbacks often occur as soon as the specific development context loses its interest in the software: Aspects like re-usability and sustainability require compliance to a certain level of quality in development and documentation. These questions have been taken up by EURISE which is open beyond the ERICs CESSDA, CLARIN and DARIAH.

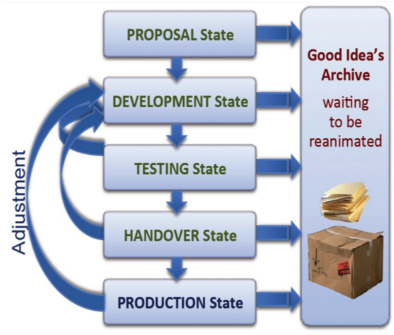

Figure 14: The DARIAH-DE Service Life Cycle as of 2015

The *DARIAH-DE Service Life Cycle*[69]/[70] provides a structured process to integrate new services in the DARIAH-DE research infrastructure. The process was conceptualised against the background of DARIAH-DE's growth over the various funding phases. As the figure depicts it is basically a simple five-step-process which culminates in the "production state". A service which wants to be integrated in the DARIAH-DE research infrastructure has to pass through the process and must successfully reach the production state. The concept as such is very similar to the well-known agile development approach and is insofar DARIAH-DE specific by the criteria assigned to each phase, translating the cycle into a digital humanities focus and applying the DARIAH-DE governance. A service leaving the handover state is evaluated as mature and standardised enough to be "owned" by a data centre for instance.

The topics concerning the service life cycle are now partly discussed within the EURISE Network, or services are even subject to established certification schemes like CTS. The DARIAH-DE Service Life Cycle is at the time of this report's writing no longer actively in use and succeeded by the abovementioned, more formalised certification schemes or, on a European scale, by the EURISE Network.

With look at CLARIAH-DE and the upcoming contributions to the NFDI - or in general the sustainable provision of services and resources - the DARIAH-DE Service Life Cycle may become important again and serve as blueprint to establish a quality and maturity management in CLARIAH-DE.

[69] https://wiki.de.dariah.eu/display/publicde/DARIAH-DE+Service+Life+Cycle

[70] See Puhl, Johanna/ Andorfer, Peter/ Höckendorff, Mareike/ Schmunk, Stefan/ Stiller, Juliane/ Thoden, Klaus: "Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften". DARIAH-DE Working Papers Nr. 11. Göttingen: DARIAH-DE, 2015 URN: http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-4-4

4 Discussion

This report gathered certification schemes which the authors consider relevant from the perspective of a research infrastructure like CLARIAH-DE. The discussed list of certification schemes is not exhaustive due to the broad range of potentially relevant frameworks and the dynamic development of schemes in general, which keeps pace with overall technological trends and leads to emergence of new schemes or the vanishing of old ones. All of the above listed certification schemes can be categorised along certain criteria, depending on the interest of the reader. This paper takes the perspective of a research infrastructure, particularly CLARIAH-DE.

CLARIAH-DE identified six categories for the certification schemes listed above, defined from the topic or area of interest of the certification scheme and the stakeholder/agent role:

Table 1: Certification categories from a CLARIAH-DE point of view

Certification category	***Area of interest and stakeholder role***
Repository/archive-focused	Focus on quality standards for metadata and content, long term preservation, and interoperability between services, so especially relevant for data providers of all kinds.
General quality management	Less topical focus than other categories. Particularly of interest for established institutions partnering in a research infrastructure undertaking, e.g. data centres and libraries.
IT service management	Focus on the sustainable provision of software and services and the orientation on quality standards in software development, therefore high relevance for any research infrastructure partner involved in such activities.
Information security	With its focus on security, important for any partner institution taking part in a research infrastructure and likely obligatory for institutions like data centres to be allowed to offer services.
Tool-specific	Of growing interest for research infrastructures that are becoming more modular, as individual components may come into question for this, particularly to be become interoperable and quality-proven for other infrastructures as well.
Self-assessment principles	Not formalised certification per se but can facilitate the development of standards and future adoption of certification schemas.

Formally speaking, a service or institution demonstrates with a certificate compliance with a certain set of criteria. Ideally these criteria are transparent and publicly available[71] and allow the informed reader to draw conclusions on the service or institution. An example for this could be a CTS certified research data repository or a DINI certified publication repository. If funders require to archive research data or publications according to these schemes, this is an important signal for users to consider the utilisation of these certified services. Under the headlines of open access and FAIR principles, it is an important topic for researchers or research projects as a means of ensuring compliance with their funder's requirements and also documenting this compliance. It is important to distinguish between "hard certificates" such as the DIN norms, which may be a professional or legal requirement for a data centre, and the researcher-nearer initiatives related to Open Access (DINI certificate) or the FAIR principles and self-assessment principles (CTS, EURISE Network). These may be found on the other side of the spectrum but are nevertheless important. The impact of Open Access, FAIR principles or self-assessment principles may be described as a slow diffusion towards "pro certificate attitude" influencing the actors of research infrastructures firstly, and only secondly may manifest themselves in visible certificates or seals. Nestor is a good example in this regard. Although there aren't many services bearing the Nestor seal, there is clearly a successful outreach impact by the information material and summer schools of Nestor.

Beyond the documentation of compliance, the actuality of a certificate is important as well. Most certificates are only valid for a limited period of time and have to be renewed. This conveys the message to the user that the service or institution keeps pace with the actual trends and developments. If no periodic renewal is required, there are some examples of certificates stating the date of acquisition. As an example, the DINI certificate is bestowed with a versioning information, which may be partly seen as accommodation to the applying institutions which do not always have the capacity or will for the regular effort of renewal procedures.

The decision for a specific form of certification depends on a number of factors. These include the organisational structure of the applicant, the type and scope of the services provided, the intended target group and their expectations, right through to specific details of the technical implementation. The question of costs (e.g. measured in terms of financial or personnel expenditure) must also be clarified in this context. The type and size of an institution to be certified influences the specific choice and will, for example, lead to different decisions for large data centers than for small and highly specialised participants of the infrastructure. This consideration is also taken into account when deciding on the central requirements of large infrastructure consortia and in some cases excludes the use of certification measures that are considered too costly or time-consuming[72].

With this proviso, why are certifications and audit schemes important for a research infrastructure like CLARIAH-DE? Basically the same reasons apply as for any other entity or organisation which wants to – or has to – perform a certification procedure. The following four main aspects are discussed in detail below from the angle of CLARIAH-DE:

[71] Which is not the case with the ISO norms.

[72] See for example the "network of centres" concept used in CLARIN ERIC where this consideration has to be made for an environment with many and often small participating institutions.

1. Trust
2. Legal/ethical obligations
3. Quality/maturity
4. Standardisation/interoperability

Trust is possibly the most important aspect from a research infrastructure's perspective. Its mission is to support research using resources and services and vice versa provide infrastructure components for hosting of publications, research data or – in the case of CLARIAH-DE – even complete research projects. Obviously such a level of use and reliance requires trust in the research infrastructure. The researchers have to be sure that, for instance, their research data is safely stored in compliance with legal or funder-related requirements.

Usually the relationship between the researcher and the research infrastructure does not fit in the pattern of a provider-customer relationship which is often documented by formal contracts. However, it is still a kind of customer relationship and trust becomes an important immaterial assurance for the researcher that the "contract will be honored".

Legal/ethical obligations can make it mandatory for a service or institution to conduct certifications. It may only be allowed to offer certain services or products with a valid certificate. This aspect applies in the research environment prominently to academic data centres as they have – for instance – to ensure the safety and legal compliance of hosted or processed data. The discussed ISO norms are relevant in this regard and often used as a common quality standard.

As data centres are important partners constituting research infrastructures, it is of growing importance when considering the evolution of European research funding procedures in the last decade. The legal and ethical aspects are by now a separate chapter in funding applications and grant agreements and the European Commission as a funder makes sure that the importance of these aspects is well received by the beneficiaries.

A certain *quality/maturity* of offered services and products can be displayed prominently by certificates. If the criteria are transparent this allows the user (or customer in other contexts) to evaluate the service offering. In the research environment a significant number of services are developed in a research- or community-driven manner. With a look at the sustainability and maturity of such research-driven services this becomes a problem when the status is not clear for potential external users or – very often after a project terminates – the service persists but isn't maintained anymore.

The DARIAH-DE Service Life Cycle is a good example for self-assessment guidelines emerging from research infrastructures. It has been developed against the background of maturity concepts like the NASA Readiness scale and applies this principle to the research infrastructure in a pragmatic manner. It is also a useful tool for the service providers themselves to assess the quality and maturity of their services and to compare it with other resources. Although currently (2021) not actively used, the service life cycle may become important for CLARIAH-DE in the future.

Apart from the DARIAH-DE Service Life Cycle, the EURISE Network may be cited as another very research-specific example in the software or infrastructure quality context. EURISE first and foremost functions as a structured discussion arena for software engineers involved in research infrastructures. Although topics such as software quality, sustainability, and interoperability are often promoted, adherence to these principles is not taken for granted. To promote the importance and internalisation of such norms, the EURISE Network offers an informal stage for discussion amongst the relevant research infrastructures, which is documented by the fact that it was founded by the three SSH-relevant ERICs CESSDA, CLARIN, and DARIAH. This argument relates to the formerly mentioned distinguishing of "hard certificates" (such as a DIN norm) and the "soft initiatives" (like OA or FAIR principles), which could also be seen as a scale of maturity. If the abovementioned ERICs internalise self-assessment principles discussed within EURISE Network, a "hard certificate" may be the end result (or an adaptation by other certificates such as CTS).

This connexion is also important against the background of service provision in research infrastructures. Usually the services and infrastructure components – if service maturity has been achieved – are "owned" by data centres or other infrastructure institutions such as libraries; for these institutions certifications are a natural part of their everyday work and function as quality and security impetus; this may be particularly applicable for ISO norms. The DARIAH-DE Service Life Cycle even devoted an individual phase to this aspect: the handover state.

Certificates are also drivers for *standardisation/interoperability* and in this regard play an important role for research infrastructures. The DINI certificate may serve as an example, as a substantial part of its guidelines revolve around the quality of publication metadata. Repositories adhering to the standards of the certificate can enter a common search space, e.g. via the Bielefeld Academic Search Engine, which integrates DINI certified repositories and makes use of the standardised (and thus high quality) metadata of the publications. CLARIN(-D)'s "network of centres"[73] is another example where interoperability is ensured by certification of technical and organisational key aspects including requirements for providing specific technical interfaces or formats. This is a consequence of the fact that large parts of a modern research infrastructure rely immediately on the conformance with such central standards. Taken on the European level, this general aspect is even more obvious, particularly with a look at CLARIAH-DE, the related research consortia CLARIN ERIC and DARIAH ERIC, and beyond this, the European Open Science Cloud (EOSC). A discovery service for research publications such as OpenAIRE is only feasible with a certain degree of standardisation in the partaking repositories.

Standardisation and interoperability are basic prerequisites for integrating infrastructures or individual services with one another. When looking at the German research landscape, the National Research Data Initiative (NFDI) is crucial for CLARIAH-DE. Within the NFDI several – usually discipline-oriented – infrastructure consortia try to cater for the requirements of their user communities. Using a common framework can improve their interoperability significantly. CLARIAH-DE as a long-established research infrastructure for the arts and humanities undoubtedly raises expectations in this regard but has also a long standing tradition of discussion and cooperation, institutionalised before the CLARIAH-DE project in the Technical Advisory Board of CLARIN-D and DARIAH-DE or in European committees like CLARIN's Standing Committee for Technical Centres[74] (SCCTC).

[73] https://centres.clarin.eu / https://www.clarin-d.net/en/about

[74] https://www.clarin.eu/governance/standing-committee-clarin-technical-centres

The *ISO norms family* is suitable to shed lights on the different roles within research infrastructures. In this context, the relevance of ISO norms derives from the way the service portfolio is provided to its users. Usually data centres – which are important IT service providers for universities and academic institutions anyway – have a responsible role here as they host infrastructure components ranging from basic infrastructures to very research-specific services. Data centres are partly obliged to process audits for certain ISO norms on a regular basis. For instance, ISO/IEC 20000 has to be renewed every three years and is thereby a good proof that the IT service management of a data centre is in step with the professional standards which are constantly evolving.

Even if a data centre or library is not involved from the beginning in a research infrastructure it may obtain a role as owner. Many services are developed in a research-driven approach and not within the authority of – for example – a data centre, but usually the ownership is transferred for the sake of sustainability in the case that the service was successful in terms of maturity or user uptake. The inheriting institutions are often data centres, libraries, or other infrastructure entities in the academic environment and their decision to take over from a research infrastructure may be adjusted with the requirements from, for instance, a certain ISO norm, which the data centre has to fulfill.

Apart from this, data centres are by nature qualified for cooperation with other data centres, which gives them the competence to contribute to infrastructure frameworks such as AAI federations or to topics like data security, long term preservation, or persistent identification. CLARIAH-DE, deriving from the established research infrastructures CLARIN-D and DARIAH-DE, can rely on a long experience of trust and cooperation with various data centres and would not be able to provide its service portfolio without data centres and university libraries. A simple example for this can be given with the replication of services and resources along various data centres, which contributes considerably to the security and availability of CLARIAH-DE. Another aspect of growing importance is, that only with such proven cooperations a research infrastructure is capable to take part on the European level of research infrastructures, for instance in EOSC (European Open Science Cloud). In most cases not a research infrastructure as such – e.g. CLARIAH-DE – will apply for ISO certification but either components or partners, like a data centre.

Tool-specific certifications are most likely of high relevance to parties with a very strong focus on the specific underlying technology and who would highly value a more direct line of communication to its developers. Otherwise, cost and resources seem to be better directed towards the more "general-purpose" certifications, as those might signify more broadly the competencies of the relevant party, instead of focusing on a singular, technical aspect.

Self-assessment principles, although they appear not as formalised as the other examples in this report, can serve as promising first steps to signal one's competencies, especially if the financial cost of a third-party certification is too high. They could also serve as a stepping stone on the road to a certification process, as many of their guidelines and values overlap with most of the certification criteria mentioned.

This holds true from the perspective of research infrastructures as well: A certification process for all services in an infrastructure might be too resource-intensive, especially during an initial development phase. A set of preliminary principles however could help guide partners and further work towards a more standardised environment, which also could as a result, then be easier to certify at a later point in time. Taking this into account the spectrum described in this paper - reaching from DIN norms to self-assessment principles like the EURISE Network or FAIR principles - may be mapped against the maturity of a research infrastructure. Although it can't be expected from a short-term funded infrastructure project to certificate components or services as long as its occupied with building the components, it is advisable to have the relevant certification schemes in mind from the very beginning.

Self-assessment of infrastructure components or services is an important milestone towards sustainability and illustrates a drive for interoperability and cooperation. A research tool being developed in a silo-like environment for a very particular research question might not consider interoperability or standards but such a view is hardly appropriate for participating in modern research infrastructures. The formalisation of standards may then be the next natural step with one of the abovementioned certification schemes.

This report, published as a technology watch report within the CLARIAH-DE project, gathered a spectrum of relevant certification schemes from the perspective of a humanities research infrastructure. Although not comprehensive, the report claims to collect the most relevant and common schemes for CLARIAH-DE. The proposed categories of schemes:

- repository/archive-focused
- general quality management
- IT service management
- information security
- tool-specific
- self-assessment principles

may serve as pattern for other research infrastructures to adapt the list according to their requirements.

The report also discussed topics such as formalisation of schemes, the importance of timing and available resources – particularly for temporary funded projects – and the importance of "soft" or not yet formalised schemes such as the EURISE Network or the FAIR principles.

CLARIAH-DE as research infrastructure consists of the predecessor projects CLARIN-D and DARIAH-DE, both established research infrastructures for over a decade. Against this background, CLARIAH-DE wants to share its experience in a public way through this technology watch report. CLARIAH-DE numbers among its portfolio components, resources, institutions and actors certification schemes from all of the abovementioned categories. Not all of them may be attributed directly to CLARIAH-DE such as the general quality or IT service management certificates, which only the informed observer may find at the websites of a data centre. Other schemes or principles or even community-based discussion forums such as EURISE Network or the FAIR principles may be even harder to associate in a traceable way with CLARIAH-DE but clearly, they have an impact. This includes the networking with other research infrastructures (such as EURISE Network which gathers the SSH ERICs) or a "backcoupling" to its users, as may be the case with the FAIR principles. CLARIAH-DE demonstrates that it advocates, for instance, Open Access and the FAIR principles in science, which adds to the trust experienced by its users. Last but not least, one also finds individual services or components, which are certified. For example, CLARIAH-DE pursues to certify its repositories using CTS; the TextGrid repository and the DARIAH-DE repository are only the latest services in this regard.

As a conclusion, it is important to have this diverse spectrum in mind when discussing certification schemes for research infrastructures. Certificates are not an end in itself but serve the purpose to professionalise the research infrastructure (and its providers), to add to the level of trust users have in the infrastructure, and in general to contribute to the internalisation of good scientific practices.

Bibliography

Arnold, Denis, Fisseni, Bernhard, Helfer, Felix, Buddenbohm, Stefan, & Kiraly, Peter (2020): Repository Solutions - Technology Watch Report 1. Zenodo. http://doi.org/10.5281/zenodo.3873027.

Blümm, Mirjam, Neuroth, Heike, Schmunk, Stefan (2016): DARIAH-DE – Architecture of Participation. Bibliothek Forschung und Praxis | Band 40: Heft 2. https://doi.org/10.1515/bfp-2016-0026.

Buddenbohm, Stefan (2020): CLARIAH-DE - Aligning two Research Infrastructures: Experiences and Challenges. Presented at the Scholarly Primitives - DARIAH Annual Event 2020, Zagreb, Croatia: Zenodo. http://doi.org/10.5281/zenodo.4266478.

Buddenbohm, Stefan and Eckart, Thomas (2020): Persistent Identifiers in CLARIAH-DE Summary & Best Practices (AP4.1.5). Zenodo. http://doi.org/10.5281/zenodo.3744091.

CoreTrustSeal Standards and Certification Board (2019): CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 (Version v02.00-2020-2022). Zenodo. http://doi.org/10.5281/zenodo.3638211.

CoreTrustSeal Standards and Certification Board (2020): Change file: CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2017–2019 to 2020–2022 (Version v01_00). Zenodo. http://doi.org/10.5281/zenodo.3828636.

DARIAH-DE (2017): The DARIAH-DE Service Life Cycle.
https://wiki.de.dariah.eu/display/publicde/DARIAH-DE+Service+Life+Cycle.

DIN 31644:2012-04 (2012): Information und Dokumentation - Kriterien für vertrauenswürdige digitale Langzeitarchive.

Eckart, Thomas, Gradl, Tobias, Jegan, Robin, Margaretha, Eliza, Werthmann, Antonina, Helfer, Felix, Buddenbohm, Stefan (2021): CLARIAH-DE Cross-Service Search: Prospects and Benefits of Merging Subject-specific Services. DARIAH-DE Working Paper Nr. 42, Göttingen: DARIAH-DE, 2021. URN: http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2021-1-9.

Gradl, Tobias and Henrich, Andreas (2016): Die DARIAH-DE-Föderationsarchitektur – Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen. Bibliothek Forschung und Praxis | Band 40: Heft 2. https://doi.org/10.1515/bfp-2016-0027.

ISO 9001:2015 (2015): Quality management systems — Requirements.

ISO 16363:2012 (2012): Space data and information transfer systems — Audit and certification of trustworthy digital repositories.

ISO/IEC 20000-1:2018 (2018): Information technology — Service management — Part 1: Service management system requirements.

ISO/IEC 27001:2013 (2013): Information technology — Security techniques — Information security management systems — Requirements.

Kleemola, Mari, Alaterä, Tuomas J., Koski, Niko, Ala-Lahti, Henri, Jerlehag, Birger, L'Hours, Hervé, Van Horik, René (2020): SSHOC D8.2 Certification plan for SSHOC repositories (Version v1.0). Zenodo.

Kümmet, Sonja, Lücke, Stephan, Schulz, Julian, Spenger, Martin, & Weber, Tobias (2019): DataCite Best Practice Guide (Version Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.3559800.

Lin et al. (2020) Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7.

Müller, Uwe, Scholze, Frank, Arning, Ursula, Beucke, Daniel, Deinzer, Gernot, Höhner, Kathrin, Wolf, Stefan (2019): DINI-Zertifikat für Open-Access-Publikationsdienste 2019 - Request for Comments. Zenodo. http://doi.org/10.5281/zenodo.2575346.

Müller, U., Scholze, F., Arning, U., Beucke, D., Blumtritt, U., Braun, K., Deppe, A., Deinzer, G., Fenner, M., Klotz-Berendes, B., Meinecke, I., Pampel, H., Schirrwagen, J., Bove, K., Severiens, T., Summann, F., Steinke, T., Tullney, M., Vierkant, P., Voigt, M., Walger, N., Weiland, J. B., Weimar, A., Wolf, S. (2020): DINI Certificate for Open Access Repositories and Publication Services 2019, Berlin : Humboldt-Universität zu Berlin, 45 p. https://doi.org/10.18452/21759.

Nestor (2009): nestor materials 8 - Network of Expertise in long-term Storage and Accessibility of Digital Resources in Germany / Working group Trusted Repositories – Certification: nestor criteria: Catalogue of Criteria for Trusted Digital Repositories, Version 2, 2009, Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2010030806.

Nestor (2019): nestor materials 17 - Network of Expertise in long-term Storage and Accessibility of Digital Resources in Germany: Erläuterungen zum nestor-Siegel für vertrauenswürdige digitale Langzeitarchive, Version 2.1, 2019, urn:nbn:de:0008-2019062507.

Puhl, Johanna, Andorfer, Peter, Höckendorff, Mareike, Schmunk, Stefan, Stiller, Juliane, Thoden, Klaus (2015): Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften. DARIAH-DE Working Papers Nr. 11. Göttingen: DARIAH-DE, 2015. URN: urn:nbn:de:gbv:7-dariah-2015-4-4.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18.