D2.7: Certification Implementation Report

Author(s)	Heiko Tjalsma
Status	Final
Version	v1.0
Date	22/03/2018

Abstract: In this deliverable the question is answered *how* trustworthiness by means of DSA certification can be achieved within the EUDAT infrastructure. This turned out to be a complicated issue. A partial solution has been found for making the EUDAT services trustworthy for data repositories wanting to user these services for (long-term) storage.

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

Document identifier: EUDAT2020-DEL-WP2-D2.7	
Deliverable lead	DANS
Related work package	WP2
Author(s)	Heiko Tjalsma (DANS)
Contributor(s)	
Due date	01/09/2017 - 01/03/2018
Actual submission date	22/03/2018
Reviewed by	Rob Baxter (EPCC)
Approved by	PMO
Dissemination level	PUBLIC
Website	www.eudat.eu
Call	H2020-EINFRA-2014-2
Project Number	654065
Start date of Project	01/03/2015
Duration	36 months
License	Creative Commons CC-BY 4.0
Keywords	DSA, CTS, certification, trust

TABLE OF CONTENT

LIST OF TABLES

EXECUTIVE SUMMARY

The question for this task of the EUDAT2020 project was *how* trustworthiness by means of DSA certification could be achieved within the EUDAT infrastructure. The certification task within the EUDAT2020 project was built upon the assumption that building trust into the data infrastructure itself is an essential task *"to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment. When using the services of EUDAT the existing community data repositories (i.e. those external to EUDAT) need to be able to maintain this trust towards the researchers using their repositories"*. The original idea of how to realise this was that EUDAT needs certification mechanisms. EUDAT chose the Data Seal of Approval, the DSA, as "its primary vehicle for certification".

At the beginning of the task two action lines were identified and defined in the Certification Strategic Plan [1]. Both have been pursued:

A. The sites (i.e. the service providers within the EUDAT domain) should aim for full DSA certification, if possible.
B. *The EUDAT services as such,* in particular *B2SAFE* and *B2SHARE,* should be able to keep data safe in the long term. In order to be trustworthy, therefore, these services should be DSA-compliant.

In the course of the project this task transformed itself more and more into the question of *if* rather than *how* trustworthiness by means of DSA certification could be achieved within the EUDAT infrastructure. The answer to that question is *yes*, but only in a limited way.

The conclusion for track A, individual service provider certification, is that this certification exercise has been useful, but has not always had a direct effect or use for some of the EUDAT sites. Nevertheless, eight out of twelve participating organisations or projects are or will be certified under the DSA's replacement, the CoreTrustSeal.

The conclusion for track B, service certification, is that, due to the present requirements of the CoreTrustSeal, certification of EUDAT services in the way originally envisaged is currently not possible.

1. INTRODUCTION

This document is deliverable D2.7, Certification Implementation Report, which describes the progress and challenges of implementing the "trustworthy repository certification goals" of the EUDAT2020 project as defined in D2.3, Certification Strategic Plan [1].

Reaching these goals turned out to be a long and complicated road. In particular the challenge of getting EUDAT sites certified according to the Data Seal of Approval (DSA) standards is quite a huge one for most of these organisations. Besides from that there was also the handicap of the transition from the DSA into the CoreTrustSeal (CTS) certification as elaborated in chapter 3 and annex A. This report covers these challenges; a final analysis of certification and thoughts for the future, can be found in D2.9, Certification Assessment.

1.1. The challenges of trust in EUDAT

EUDAT2020 deliverable D2.3 set out the certification plan for the years ahead. The certification task within the EUDAT2020 project was predicated upon the assumption that building trust into the data infrastructure itself is an essential task [quoting from this deliverable] *"to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment. When using the services of EUDAT the existing community data repositories (i.e. those external to EUDAT) need to be able to maintain this trust towards the researchers using their repositories"*. The plan of how to enable this trust was that EUDAT should pursue certification mechanisms from existing standards for trustworthy digital repositories. In the project plan EUDAT chose the Data Seal of Approval (DSA) as "its primary vehicle for certification". DSA, and its successor CoreTrustSeal CTS[1], is a light-weight certification process aimed at assessing the trustworthiness of a repository. It is a basic certification standard for "Trustworthy Digital Repositories", 'basic' as compared with the DIN 31644 – *NESTORseal* [3] and the ISO 16363 - *Audit and certification of trustworthy digital repositories* [2], both of which are certification systems on a higher level.

The question right at the start of the EUDAT2020 project was how trustworthiness through DSA certification could be achieved within the EUDAT infrastructure. After reflection, discussion and consultation in the first phase of the task two action lines were distinguished. Both have been pursued separately in the remainder of this task:

A. The sites (i.e. the service providers within the EUDAT domain) should aim for full DSA certification, if possible.
B. *The EUDAT services themselves,* in particular *B2SAFE* and *B2SHARE,* aim to keep data safe in the long term. These services should therefore, in order to be trustworthy, be DSA-compliant. (See Deliverable D2.3, page 5.)

These two lines both created far more difficulties in implementing than foreseen at the start of the project for varying reasons. In the following sections we describe why that was the case.

1.2. Outline

We cover action lines A and B separately in, respectively, chapters 2 and 3, and we offer general conclusions for this stage of the certification activity in chapter 4.

[1] From September 11th 2017, the ICSU World Data System (ICSU-WDS) and the Data Seal of Approval (DSA) have been combined in a new certification organisation: *CoreTrustSeal*. In this report, this seal is often referred to under its old name DSA, as this was in use during most of the time of the EUDAT2020 project. The term CTS is, however, also used.

2. DSA/CTS – COMPLIANCE OF THE EUDAT SITES

The underpinning idea of the "site compliance" line of action was that external community repositories (i.e. outside the EUDAT CDI infrastructure), whether DSA-compliant or not, can be offered DSA-compliant long-term archive storage by service-provider sites within EUDAT *if the latter are DSA-compliant themselves*. This is in fact rather straightforward. It means that the sites participating in the EUDAT infrastructure who are in a position to do this, should be able to reach DSA-compliance.

From the start it proved necessary to make most participants acquainted with the concept of certification in general and the Data Seal of Approval in particular. In order to achieve this, two workshops and a number of webinars were held. In addition there was a lot of contact between the task leader and the participants in this task on a bilateral level. It was decided in the first workshop held in this task, 21–22 April 2015 in The Hague, that the participating sites within the EUDAT domain principally should aim at full DSA certification.

2.1. Workshops and webinars: training and work

2.1.1. 21–22 April 2015 in The Hague

This workshop was aimed at familiarising all participating EUDAT CDI partners with the DSA and its underlying principles. Additionally, it was agreed that in this task priority should be given to the first action line. The highest urgency was to get the participants in this task to know what involvement with DSA certification would mean in practice for them. It was proposed that all these partners – the service provider sites – should carry out a DSA self-assessment as a test. This could give an indication how far these sites are away from acquiring DSA certification and what problems or hurdles could arise.

2.1.2. 4–5 November 2015 in Montpellier

Eleven of the twelve partners actually produced a self-assessment in October/November as agreed at the The Hague workshop. Two of these self-assessments were from CINES and DANS – the only two partners already DSA-compliant at that moment. Before the workshop all the self-assessments were reviewed both by CINES and DANS, in strict confidence. These two leading partners have both considerable experience in reviewing DSA self-assessments.

The results were discussed in this workshop amongst all the participants. Several points could be observed.

Although most participants seemed to have good knowledge of the DSA and its guidelines many gaps were observed in the self-assessments of most partners in terms of having a high-enough level for actually getting certified. The analysis showed that most of the gaps were in the field of explicit documentation and procedures. One of the main issues here was that most partners were applying *implicit* procedures, regarding for example internal workflows or back-up procedures, but *explicit* documentation, being a prerequisite for any certification, but in particular for DSA/CTS, was missing. Knowledge and observance of legal obligations as well as national codes of conduct was also reason for concern, but this was not *seen* as such by most partners. This, most likely, arises from the type of data managed by most of the sites, much of which are natural-scientific experimental or simulation data with no significant disclosure risk.

DSA guideline 13 on Technical Infrastructure Standards scored particularly badly. This guideline is in essence about the application of the OAIS reference model[2] [4]. It recommends that, for a DSA-compliant repository, "*the technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS*". Initially, there was much misunderstanding on this requirement as the OAIS was primarily seen as a technical standard (which it is not) rather than a reference model focused on the actual *organisation* of a data repository.

This led to the conclusion that one or maybe two partners might reach DSA certification soon, while other partners might reach DSA certification in a timeframe of 6 to 12 months *if they were prepared to put some*

[2] For abbreviations see Glossary - Annex B.

effort by themselves in it and with some guidance from within the task. For a number of partners it was questionable whether they would be able to reach DSA-compliance either soon or at all.

One overall conclusion was that more training on DSA and OAIS was desired!

The workshop also discussed whether it should be possible at all to certify the *EUDAT-services themselves* and how that could that be accomplished? This touches on action line B, and we return to this in chapter 3.

2.1.3. Webinars

Following the wish expressed in the workshops, a number of well-attended webinars were held:

- Webinar *on OAIS: 29 June 2015.* This webinar was concentrated on the OAIS as there were many questions on the principles and use of the OAIS.
- *Webinar on Trust and Certification: 18 April 2016.* A general webinar on the basics of Trust, specially certification. This was still based on the DSA, requested by many.
- *Webinar on Trust: The New Requirements DSA/WDS: 30 January 2017.* Webinar on the changes from DSA Guidelines into CoreTrustSeal Requirements.

All these webinars were given by Heiko Tjalsma (DANS) as task leader.

In the course of this task there have been many contacts on an individual level between the task leader and the various participants.

2.2. Outcome

The final results of the implementation phase of the certification task are shown in table 1. As foreseen in deliverable D2.3 (page 9) only a few organisations have submitted or will submit their self-assessment in 2018. Of the twelve participants, three had acquired the DSA earlier and had embarked on the process of acquiring the CTS seal after submitting their self-assessments for the new scheme. Five participants indicated intentions to apply for CTS certification in the future, two of them very soon, two later in 2018/2019. Four indicated either that they are not going to do this at all or, possibly, only much later.

Table 1: Status CTS certification participants in task 2.2 EUDAT2020

Status February 2018	Number
CTS acquired or submitted	3
CTS to be submitted in the (near) future	5
NO CTS application planned	4
Total	**12**

2.3. Discussion

Why have more partners in the task not reached certification? There several reasons for this.

One factor is the change of the DSA into the CTS during the course of the EUDAT2020 project, in late 2016/early 2017 (see Annex A). This aside, a far more important reason lies in the mission and background of the participating organisations. During the test assessments it turned out that some of the DSA/CTS requirements were difficult to comply with. This arises chiefly from the fact that a number of these sites are computing centres. They have no difficulty to comply with the technical requirements, but they are not set up as data repositories as defined in the CTS Requirements (in following, for example, a logical organisational model like OAIS). Only two of the sixteen CTS requirements are concerned with technical infrastructure. All the other requirements are in the field of the “organisational infrastructure” and “digital object management”. In particular the former of these two main areas, organisational infrastructure, is certainly not a natural pattern for computing centres. Organisational infrastructure in CTS-terms is about subjects like mission, data licences, governance, guaranteeing long-term access and preservation, handing of legal and ethical requirements. Digital object management covers questions around the guaranteeing of the

authenticity and integrity of data, including provenance of the data, around appraisal and data quality. These issues are not traditionally part of the activities of computing centres, or, more precisely a number of these centres.

To summarise: the organisational framework that requires having an explicit mission, complying with legal regulations and long-term preservation planning for data, is not in place for these sites. Some of these sites are, by their background, not institutions who act as repositories. In particular it emerged very clearly that using the OAIS for many of these sites is something which was not immediately in their scope. They do not work in the same way as digital repositories do and, generally speaking, functioning as a fully qualified repository is simply too far out of scope

Participants gave a variety of reasons for not being able to reach certification yet or not at all. Some of these have to do with the fact that they are not (yet) sites but projects for which getting CTS-certification is not currently their highest priority. This applies in particular for data projects like ICOS or LTER. That said, these data projects remain committed to seeking CTS status in the not too distant future, typically after transformation onto a more sustainable footing.

Other participants, on the other hand, decided that, after initial work in this task, they will not pursue DSA/CTS certification. To cite one partner, "*we are principally a high-performance computing centre rather than a digital repository; it was always tricky to map the DSA requirements onto our strategy and operational procedures. We fall down, for instance, on R1, 'The repository has an explicit mission…' "*. Another partner noted that, although it was very useful to get to know the DSA/CTS certification issues, it was not in their organisational scope and also not deemed immediately useful as their site was not connected to a EUDAT service (more on this in the next chapter).

The conclusion for this part of the task, based on the reactions of most participants, is that this certification exercise was useful, but did not in some cases have a direct effect or use for them. That said, eight out of twelve participating organisations or projects have been, or will in the near future be, CTS certified through EUDAT2020. This outcome – certification being overall a long road to travel – is very much comparable with experiences from other infrastructures (ERICs, for example) where CTS-certification has become mandatory.

3. CERTIFICATION OF EUDAT SERVICES?

As indicated in section 1.2, action line B identified the question of whether the EUDAT services in themselves, in particular B2SAFE and B2SHARE, can offer trustworthiness. Based partly on the knowledge acquired in these self-assessments the second action line was discussed intensively during the second workshop, 4-5 November 2015 in Montpellier. The question was, more specified: is it possible at all to certify the *EUDAT-services themselves,* in particular B2SHARE and B2SAFE, and how should that be accomplished?

This has always been a far more difficult track as it was difficult to conceptualise. From the beginning it was clear that it would not be possible for any of the EUDAT services per se to be able to reach CTS-compliancy. This was suspected at the time of writing of deliverable D2.3 but this sense became much stronger as the project proceeded. This assessment covers the two services which would be first in line for potential certification, B2SAFE and B2SHARE, but the fundamental principles of the CoreTrustSeal preclude this. The intent of CTS is to ensure that a *repository* is trustworthy; its data management should be aimed at guaranteeing data preservation for the long run. For this *all the aspects* of the organisational and legal governance, digital object management and technical infrastructure must be taken into account together.

Most of the EUDAT CDI sites that operate services like B2SHARE and B2SAFE are not able to achieve CTS certification because, as noted in the previous chapter, they do not fulfil all these tasks. It would mean that these EUDAT sites should work as a full data repository using archiving terminology and defining and documenting the functions of a data repository, as CTS requires, perhaps with explicit reference to the OAIS Reference Model. According to the OAIS model datasets should be handled, and if possible stored, in the form of *Archival Information Packages* (AIPs), containing both the data and the metadata. (Note that these packages are conceptual, so metadata and data need not be stored in the same physical or virtual structure.) Whether many service provider sites are geared to receive and handle AIPs (or comparable standardised concepts), as well as to be interoperable between themselves, has never been fully established: could AIPs be moved successfully between the sites via the B2SAFE service, for instance?

The consequence of this is the following. Those EUDAT sites which have become CTS-compliant, as described in chapter 2, would have to offer CTS-certified trustworthiness as part of an EUDAT service. CTS-certified sites could offer this trustworthiness on their own; they do not need the EUDAT infrastructure for that. But that is not the point here: what is needed is trustworthiness of EUDAT services. Then, however, there needs to be a standard within EUDAT ascertaining that EUDAT services such as B2SAFE offered by these sites will be “CTS-proof”. This means that the B2SAFE-services these partners are offering should be CTS-compliant on its own and should not be different from the services they are offering as individual site. If the EUDAT site X has become CTS-compliant, the B2SAFE service of X should offer the same level of services X is offering to its non-EUDAT customers. This model will not work in practice, because as indicated in the previous chapter, a number of EUDAT sites will not reach CTS-certification. As mentioned in chapter 2, it is obvious that some sites will not even apply for this. This leads to a distinction between EUDAT sites (certified ←→ non-certified) being unworkable for the EUDAT services in practice.

4. CONCLUSIONS

In planning, the concept of the EUDAT CDI as a network of trustworthy repositories and supporting data centres made a lot of sense. However, the practicalities of implementation have thrown up a number of largely intractable issues that, at base, reflect the complex nature of many of the organisations involved.

The CTS guidelines focus around the notion of "the repository". For EUDAT generic service providers, a "repository service" could indeed be one of the offerings they have – B2SHARE as a front-office repository service, or B2SAFE as a back-office component, for instance – but the CTS guidelines in practice propagate more widely than the notion of repository service, touching on broader aspects of the hosting organisation in ways that can, and do, come into conflict with other parts of the service providers' missions. Essentially, it's hard to be certified as a repository if you're not a repository.

Nevertheless, at the time of writing eight of the 23 partner organisations in the EUDAT CDI have achieved, or are on the road to achieve, CTS certification. The reasons why more have not have been well understood. The final report in this series, D2.9 Certification Assessment, will take this forward and look at different ways forward based on the lessons learned.

5. REFERENCES

[1] H. Tjalsma, *Certification Strategic Plan*, EUDAT2020 D2.3, December 2015.

[2] ISO 16363, *Space data and information transfer systems – audit and certification of trustworthy digital repositories*, 2012, http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

[3] DIN 31644, *Information and documentation – Criteria for trustworthy digital archives*, 2012, https://www.din.de/en/getting-involved/standards-committees/nid/standards/wdc-beuth:din21:147058907

[4] CCSDS, *Reference Model for an Open Archival Information System (OAIS)*, 2012, https://public.ccsds.org/pubs/650x0m2.pdf

ANNEX A. BACKGROUND OF THE CHANGE FROM DSA TO CORETRUSTSEAL CERTIFICATION

As mentioned in the text the DSA certification has changed during the course of the EUDAT2020 project into the CTS CoreTrustSeal certification standard. This was the result of the DSA entering into a partnership with ICSU, the World Data System. The aim of the partnership was to create a set of harmonized common procedures for certification of repositories at the basic level. The evolution of standards and processes to apply trustworthy digital repository (TDR) status has met an acknowledged need for standardization and formalization in the area of data repository practice. The ISO 16363 - *Audit and certification of trustworthy digital repositories*, DIN 31644 - Criteria for trustworthy digital archives, NESTORseal, and the Data Seal of Approval (now CoreTrustSeal) have provided a number of OAIS-related approaches to certify trustworthiness.

By 2014 it was considered that the provision of a single 'core' TDR certification approach was better for the data management community as a whole and between 2014 and 2016 the DSA and WDS reviewed their goals, processes and procedures in the context of a Research Data Alliance (RDA) working group. Group members came primarily from the DSA Board and WDS Membership Committee but participation was open and progress reports were provided to the wide-ranging membership of the Repository Certification Interest Group of the RDA. The group released a set of common requirements and common procedures for public comment and undertook internal testing (by both WDS and DSA members) of the revised requirements. After a long open comment period, the new criteria were adopted by both DSA and WDS in 2016.

In the CTS, there are still 16 requirements as in the DSA, though their structure and content had changed to provide clearer language and a greater alignment with the other mentioned TDR Framework standards. There were changes to the structure of the requirements: six on organisational infrastructure, eight on digital object management and two on technology. Extensive care was taken to ensure consistency between the old and new approaches such that prior WDS or DSA recipients would be able to renew to the new requirements. However, the requirements are not identical, with more detail provided on documented public evidence requirements and a greater focus on technical infrastructure and appropriate information security.

During the self-assessment, each applicant indicates a compliance level for each of the requirements:

0 – Not applicable

1 – The repository has not considered this yet

2 – The repository has a theoretical concept

3 – The repository is in the implementation phase

4 – The guideline has been fully implemented in the repository

The compliance levels provided by the applicants will be judged against the given evidence by the reviewers.

Since July 2017, repositories applying for DSA or WDS certification are certified by the CoreTrustSeal. More information can be found at https://www.coretrustseal.org/about/. In August 2017, the two bodies constituted an interim DSA-WDS Board to manage the process and requirements. The CTS governance will be formalised in a legal structure early 2018.

A new common tool and a new CoreTrustSeal brand to support a new independent board will replace the interim board when statues and business model considerations are agreed.

ANNEX B. GLOSSARY

Term	Explanation
DSA	Data Seal of Approval
CTS	CoreTrustSeal
OAIS	Open Archival Information System
SLA	Service Level Agreement
OLA	Operational Level Agreement
ToU	Terms of Use
AIP	Archival Information Package (OAIS)
SIP	Submission Information Package (OAIS)
DIP	Dissemination Information Package (OAIS)
WDS	World Data System