



Certification

F. Genova (merci à I. Dillo and H. L'Hours)

research data sharing without barriers
rd-alliance.org



Perhaps the biggest challenge in sharing data is trust: how do you create a system robust enough for scientists to trust that, if they share, their data won't be lost, garbled, stolen or misused?

The Data Harvest:

How sharing research data can yield knowledge, jobs and growth

An RDA Europe Report

December 2014

La confiance est au coeur du partage des données

- Utilisateurs
 - Ceux qui déposent leurs données
 - Ceux qui les utilisent
- Partenaires
- Agences de financement



Les éléments de la confiance

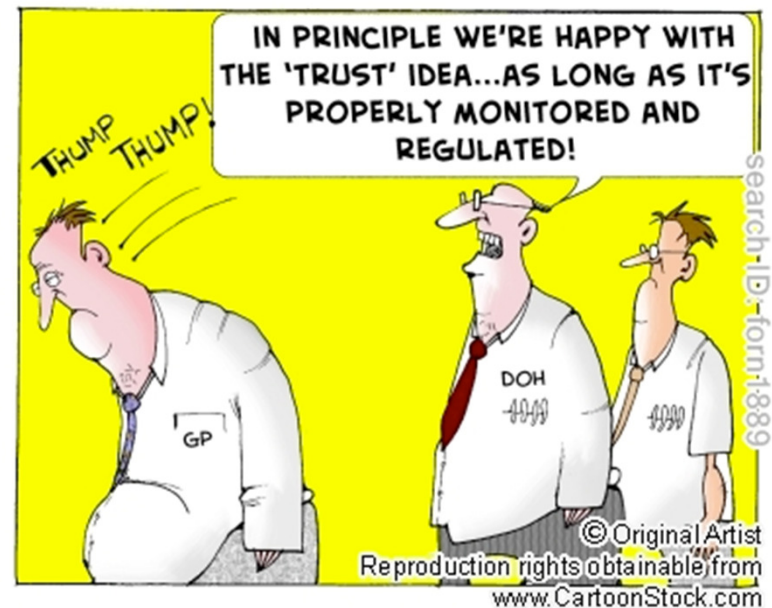
4

- Ce que vous êtes sensé faire (quelle est votre mission?)
- Ce que vous êtes (stabilité, compétence, qualité, réputation, ...)
- Transparence (certification)



"They don't trust each other to share research."

research data sharing without barriers
rd-alliance.org




RESEARCH DATA ALLIANCE

Qu'est-ce qu'un centre de données de confiance?

5

- Mission = fournir un accès fiable et sur le long terme à des ressources numériques qu'il gère pour sa communauté spécifique, maintenant et dans le futur
- Monitoring, planning et maintenance
- Connaître les risques
- **Cycle régulier d'audit et/ou certification**



Pourquoi un audit ou une certification formelle?

6

- Ou, pourquoi, puisque nos utilisateurs ont pleinement confiance en nous?

L'exemple du Centre de Données astronomique de Strasbourg (CDS)

7



- Créé en 1972
- ~ 800 000 requêtes/jour sur les services en 2014
- 664 articles publiés dans des journaux à referee en 2014 citent le mot « SIMBAD », 306 « VizieR » et 55 « Aladin »
- Nos utilisateurs ont confiance en nous!
- Les producteurs de données aussi:
 - Données envoyées par des astronomes
 - Collaboration avec les journaux pour conserver et distribuer des données attachées aux articles
 - Collaboration avec les observatoires et les projets

Première étape: devenir membre de la « communauté d'excellence » WDS

8

The screenshot shows the homepage of the Centre de Données astronomiques de Strasbourg (CDS). The page is in French and includes a navigation menu on the left with links like 'Home', 'About CDS', 'People', 'Support', 'Help and Tutorials', 'Developers corner', 'Publication support', 'myCDS', 'Virtual Observatory projects', 'Other projects', 'Authorities', 'Links', 'Internal', and 'Contact us'. The main content area features the CDS logo and name, followed by a row of service icons: 'Entry point to all services', 'Object database', 'Catalogue database', and 'Interactive My star'. Below this are search boxes for 'Object position', 'Object position/ID/name', 'Keywords, target', and 'Object position'. The 'Other services' section includes icons for Xmatch, Dictionary, Seams, and SimPlay. The 'Hosted services' section lists ADS mirror, ABA, and TPTOPbase. The 'Latest news' section contains a list of recent updates, such as 'Catalogue added between 15-Aug-2015 and 22-Aug-2015'. The 'Featured news' section highlights a '40 ans' anniversary for the CDS. The footer contains logos for CIFS INSU, University of Strasbourg, and ICSU World Data System, with the latter being circled in red.

WDS: « Trusted Data Services for Global Sciences »

9



- Created by ICSU in 2008, successor of the « World Data Centers » but re-accreditation required
- Mostly Earth data (also astronomy historically) but open to all
- Promoting universal and equitable access to, and long-term stewardship of, quality-assured scientific data and data services, products, and information covering a broad range of disciplines from the natural and social sciences, and humanities.
- Coordinates **trusted scientific data services** for the provision, use, and preservation of relevant datasets.
- Certification to join the WDS community

Le paysage de la certification

- 4 standards



DIN 31644



ICSU
WORLD DATA SYSTEM



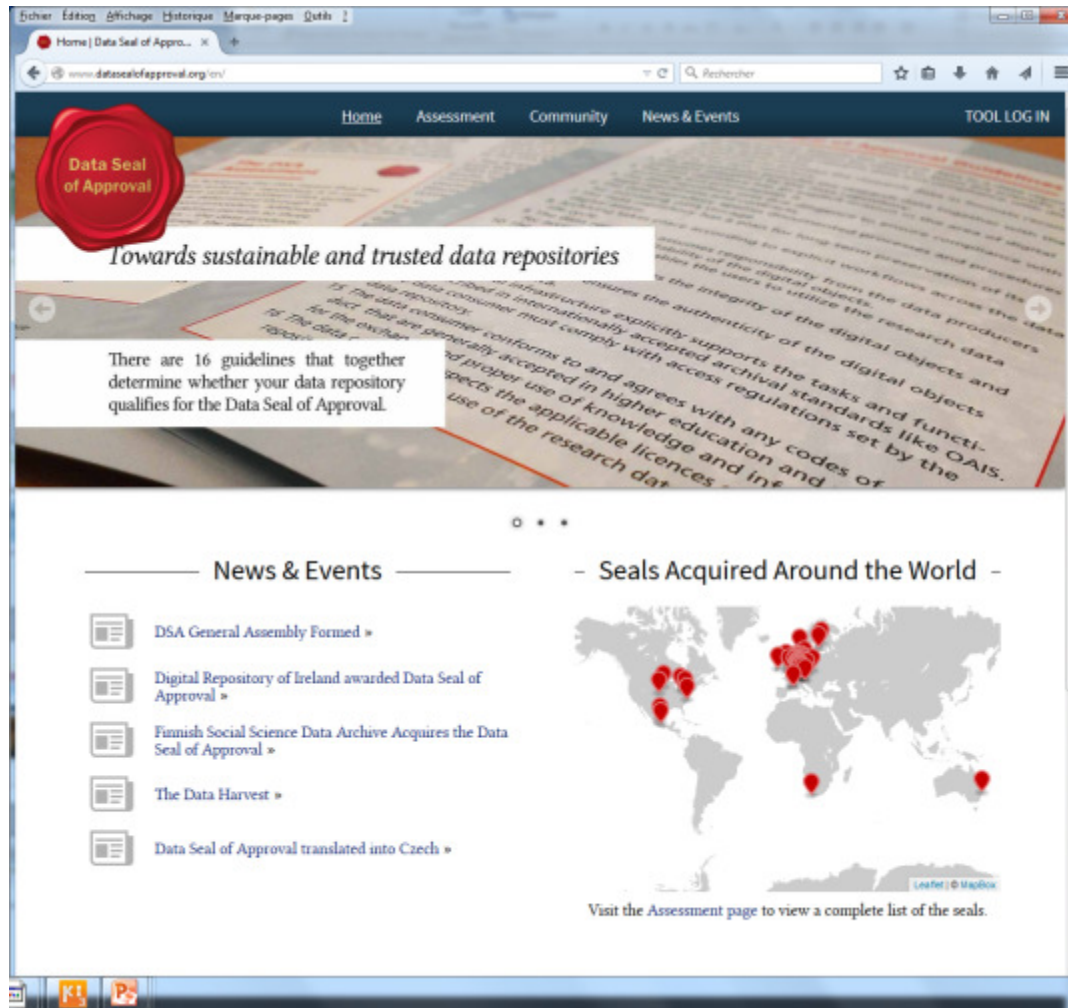
ISO 16363

Le cadre européen

- **Basic Certification** is granted to repositories which obtain DSA certification
- **Extended Certification** is granted to Basic Certification repositories which *in addition* perform a structured, externally reviewed and publicly available self-audit based on DIN 31644/nestorSeal
- **Formal Certification** is granted to repositories which *in addition to* Basic Certification obtain full external audit and certification based on ISO 16363



The Data Seal of Approval



Ajouter une certification DSA pour le CDS?

13

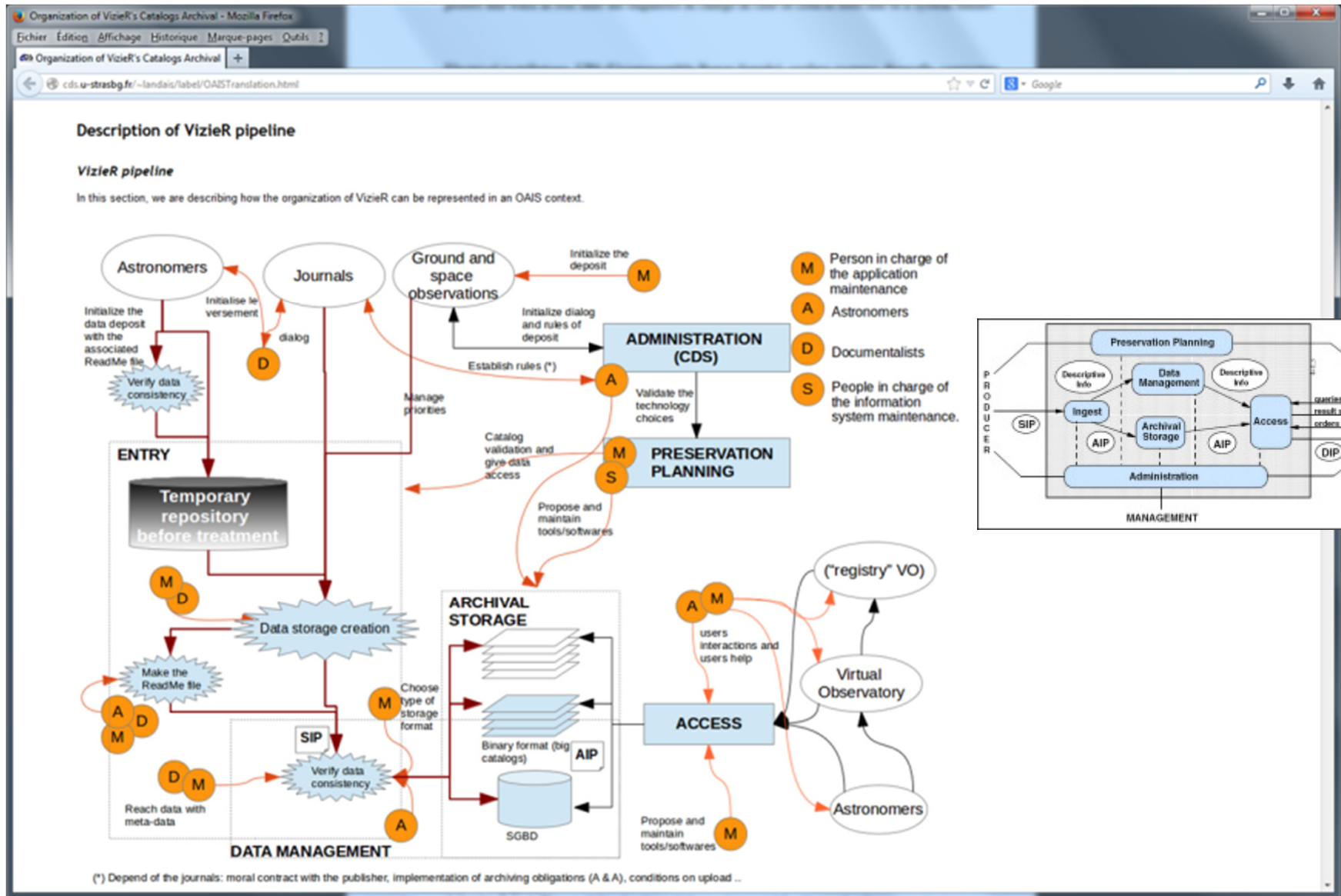
- Plus orienté « repository » alors que la mission du CDS est de fournir des services
- Les agences de financement demandent de plus en plus des plans de gestion des données – le dépôt dans un centre de données de confiance est un élément important, et pourrait devenir obligatoire à l’avenir dans certains cas
- Le DSA est le premier niveau de certification dans le contexte européen
- Très connu dans la communauté SHS
- Décision de candidater après une discussion avec un membre du DSA, et ça a marché – le premier DSA en sciences physiques, le second en France après le CINES

Obtenir le DSA

14

- Focalisé sur les aspects « repository » aspects qui sous-tendent certains des services du CDS
- Travail de fond sur les procédures en répondant aux questions
- Travail d'équipe très intéressant
- Finalement pas de modification de nos procédures mais une description complète, clarification des aspects licences, etc
- Réaction très positives de nos autorités, des journaux avec lesquels nous travaillons, etc
- **PLUSIEURS MOIS ETP QUI ONT VALU LA PEINE!**

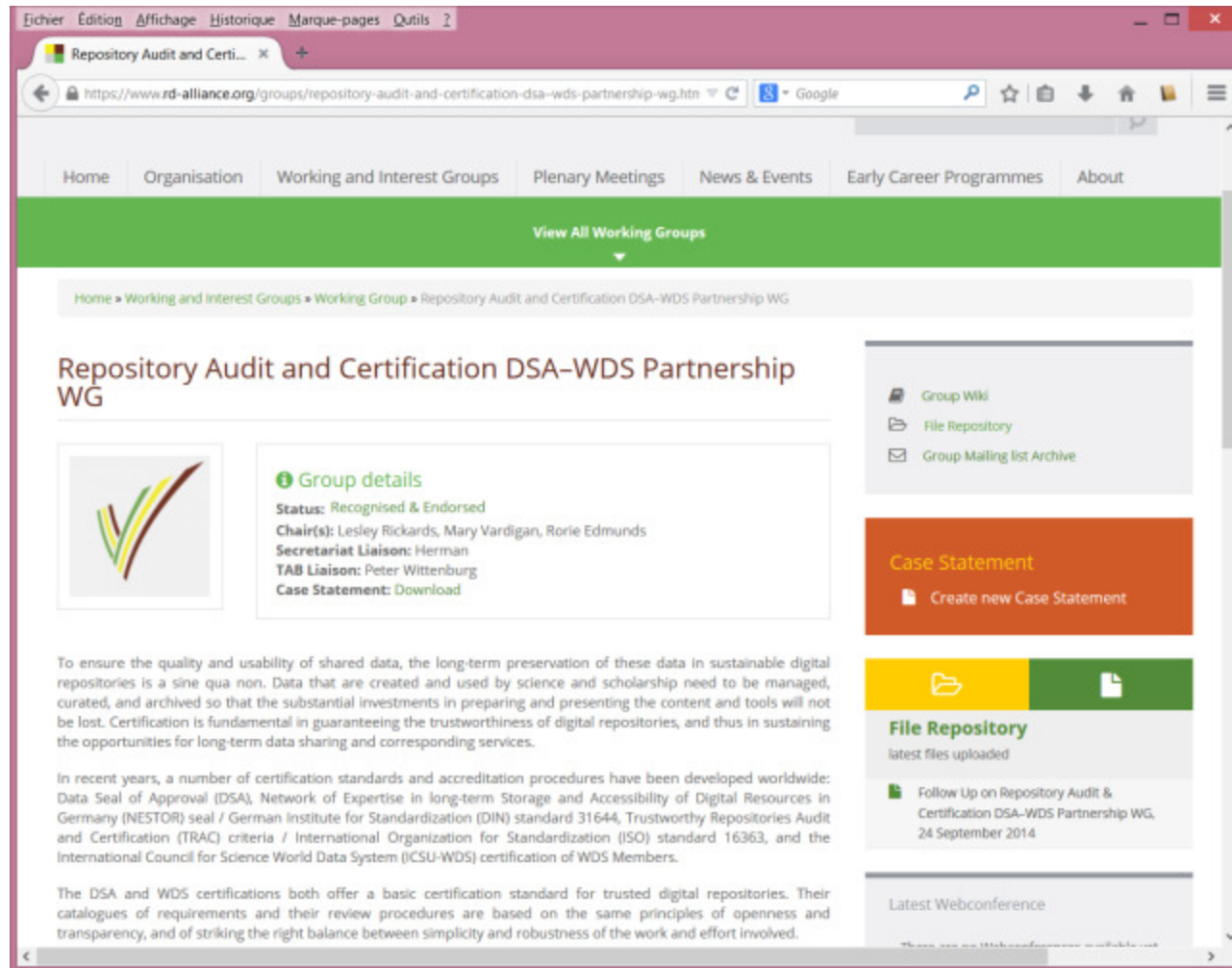
Description « à la OAIS » des services CDS



- La préservation à long terme des données dans un centre de données pérenne est un élément clé du partage des données scientifiques
- Les centres de données doivent être de confiance
- Très intéressant donc d'obtenir une certification externe
- Mais... que faire en pratique?
- Un thème important pour la RDA
- Interest Group RDA/WDS pour discuter de la certification, qui a mis en place le Working Group RDA/WDS Certification of digital repositories

Repository Audit and Certification IG

17



The screenshot shows a web browser window displaying the RDA website. The browser's address bar shows the URL: <https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa-wds-partnership-wg.htm>. The website has a navigation menu with items: Home, Organisation, Working and Interest Groups, Plenary Meetings, News & Events, Early Career Programmes, and About. Below the navigation is a green bar with the text "View All Working Groups". The main content area is titled "Repository Audit and Certification DSA-WDS Partnership WG". It features a logo on the left and "Group details" on the right, including: Status: Recognised & Endorsed; Chair(s): Lesley Rickards, Mary Vardigan, Rorie Edmunds; Secretariat Liaison: Herman; TAB Liaison: Peter Wittenburg; Case Statement: Download. Below this is a paragraph of text: "To ensure the quality and usability of shared data, the long-term preservation of these data in sustainable digital repositories is a sine qua non. Data that are created and used by science and scholarship need to be managed, curated, and archived so that the substantial investments in preparing and presenting the content and tools will not be lost. Certification is fundamental in guaranteeing the trustworthiness of digital repositories, and thus in sustaining the opportunities for long-term data sharing and corresponding services." This is followed by another paragraph: "In recent years, a number of certification standards and accreditation procedures have been developed worldwide: Data Seal of Approval (DSA), Network of Expertise in long-term Storage and Accessibility of Digital Resources in Germany (NESTOR) seal / German Institute for Standardization (DIN) standard 31644, Trustworthy Repositories Audit and Certification (TRAC) criteria / International Organization for Standardization (ISO) standard 16363, and the International Council for Science World Data System (ICSU-WDS) certification of WDS Members." The final paragraph states: "The DSA and WDS certifications both offer a basic certification standard for trusted digital repositories. Their catalogues of requirements and their review procedures are based on the same principles of openness and transparency, and of striking the right balance between simplicity and robustness of the work and effort involved." On the right side of the page, there are several widgets: "Group Wikis" with links to "Group Wikis", "File Repository", and "Group Mailing list Archive"; a "Case Statement" widget with a "Create new Case Statement" button; a "File Repository" widget showing "latest files uploaded" and a "Follow Up on Repository Audit & Certification DSA-WDS Partnership WG, 24 September 2014" link; and a "Latest Webconference" widget.

The Certification Working Group

18

The screenshot shows a web browser window displaying the RDA/WDS Certification of Digital Repositories Interest Group page. The browser's address bar shows the URL: <https://www.rd-alliance.org/groups/rdawds-certification-digital-repositories-ig.html>. The website header includes the RDA logo and navigation links: Home, Organisation, Working and Interest Groups, Plenary Meetings, News & Events, Early Career Programmes, and About. A green bar below the navigation contains the text "View All Interest Groups". The main content area features a breadcrumb trail: Home » Working and Interest Groups » Interest Group » RDA/WDS Certification of Digital Repositories IG. The page title is "RDA/WDS Certification of Digital Repositories IG". On the left, there is a logo consisting of three stylized triangles in green, yellow, and red. To the right of the logo is a "Group details" section with the following information: Status: Recognised & Endorsed; Chair(s): Micheal Diepenbroek, Ingrid Dillo, Mustapha Mokrane; Case Statement: Download. Below this is a paragraph of text: "Status: Recognised & Endorsed Joint RDA/WDS IG. In order to guarantee data sharing, the long-term preservation of these data in sustainable digital repositories is a sine qua non. Data that are created and used by science and scholarship need to be managed, curated and archived, making sure that the substantial investments in preparing and presenting the content and tools will not be lost. Researchers need to be sure that the resources the repositories offer remain meaningful and usable over time. Moreover, the repositories themselves need to have sustainable business models. Preservation and sustainability raise challenges in many areas. The main issues related to long term preservation and sustainability remain basically unresolved, as many organizational, technical, financial and legal aspects remain open. Certification is therefore fundamental in guaranteeing the trustworthiness of digital repositories and thus in sustaining the opportunities for long-term data sharing. The Interest Group will build on previous work in the area of certification. It will deliver the global overview and the necessary recommendations and requirements that allow the effective implementation of certification of digital". On the right side of the page, there is a sidebar with several sections: "Group Wiki", "File Repository", and "Group Mailing list Archive"; a "Case Statement" section with a "Create new Case Statement" button; a "File Repository" section with the text "Latest file attachment in this group repository empty"; and a "Latest Webconference" section with a "Webconference Repository Audit and Certification" link.

- Deux systèmes de certifications reconnus mais “légers”
 - Auto-évaluation, pas de visite sur site
 - Evaluation de la candidature par les pairs
- DSA : point de départ en SHS, WDS : sciences physiques, mais leurs champs d'intervention s'étendent
- Des membres en commun
- Etablir un catalogue de critères commun et des procédures d'évaluation communes
- Des critères semblables mais des différences, en particulier DSA plus centré sur la gestion des données, WDS sur la stabilité et l'adéquation de l'organisation, DSA plus soucieux de confidentialité et éthique

The starting point: Two different « lightweight » frameworks

20

■ DSA

- 16 guidelines for Thrustworthy Digital Repositories (data producers/repositories/consumers)
- DSA granted for a period of 2 years
- 45 seals acquired, some 40 underway

■ WDS

- Assessment to allow membership
- 17 criteria
- Review every 3-5 years
- ~50 members

Les critères sont harmonisés

21

- Point clé: prendre en compte la mission du centre candidat pour l'évaluer
- Prise en compte par le DSA et le WDS en cours



<https://rd-alliance.org/group/repository-audit-and-certification-dsa%E2%80%93partnership-wg/outcomes/dsa-wds-partnership>

Les critères communs

22

- 16 critères qui viennent avec des informations et des conseils sur les réponses possibles
- Contexte – **la clé de l'évaluation est la mission du centre candidat!**
- Trois thèmes
 - Infrastructure organisationnelle
 - Gestion des données
 - Technologie

Organisational infrastructure

23

- Mission/scope
- Licenses
- Continuity of access
- Confidentiality/Ethics
- Organisational infrastructure
- Expert guidance

Digital object management

24

- Data Integrity and authenticity
- Appraisal
- Documented storage procedure
- Preservation plan
- Data quality
- Workflows
- Data discovery and identification
- Data reuse

Technology

25

- Technical infrastructure
- Security

Un message fort

26

- Extrêmement utile pour une auto-évaluation même si on ne va pas jusqu'à soumettre la candidature pour une évaluation extérieure!
- Tous les critères ont un guide de réponse et une auto-évaluation du niveau de conformité
- On peut tester sa propre progression en comparant l'auto-évaluation à une évaluation antérieure.

- R0 Context

Please provide context for your repository

- R1 Mission/Scope Organizational Infrastructure

The repository has an explicit mission to provide access to and preserve data in its domain

- R2 Licenses Organizational Infrastructure

The repository maintains all applicable licenses covering data access and use and monitors compliance

- R3 Continuity of access Organizational infrastructure

The repository has a continuity plan to ensure ongoing access to and preservation of its holdings

- R4 Confidentiality/ethics Organizational Infrastructure

The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms

- R5 Organizational infrastructure Organizational Infrastructure

The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission

- R6 Expert guidance Organizational Infrastructure

The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant)

- R7 Data integrity and authenticity Digital Object Management

The repository guarantees the integrity and authenticity of the data

- R8 Appraisal Digital Object Management

The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users

- R9 Documented storage procedures Digital Object Management

The repository applies documented processes and procedures in managing archival storage of the data

- R10 Preservation plan Digital Object Management

The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way

- R11 Data quality Digital Object Management

The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations

- R12 Workflows Digital Object Management

Archiving takes place according to defined workflows from ingest to dissemination

- R13 Data discovery and identification Digital Object Management

The repository enables users to discover the data and refer to them in a persistent way through proper citation

- R14 Data reuse Digital Object Management

The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data

- R15 Technical infrastructure Technology

The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community

- R16 Security Technology

The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users

- WDS definition

"Scientific data services assist organizations in the capture, storage, curation, long-term preservation, discovery, access, retrieval, aggregation, analysis, and/or visualization of scientific data, as well as in the associated legal frameworks, to support disciplinary and multidisciplinary scientific research."

- No clear divide between “repositories” and “services”, and the situation is evolving...