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

Trust - Confiance

La confiance est au coeur du partage des données

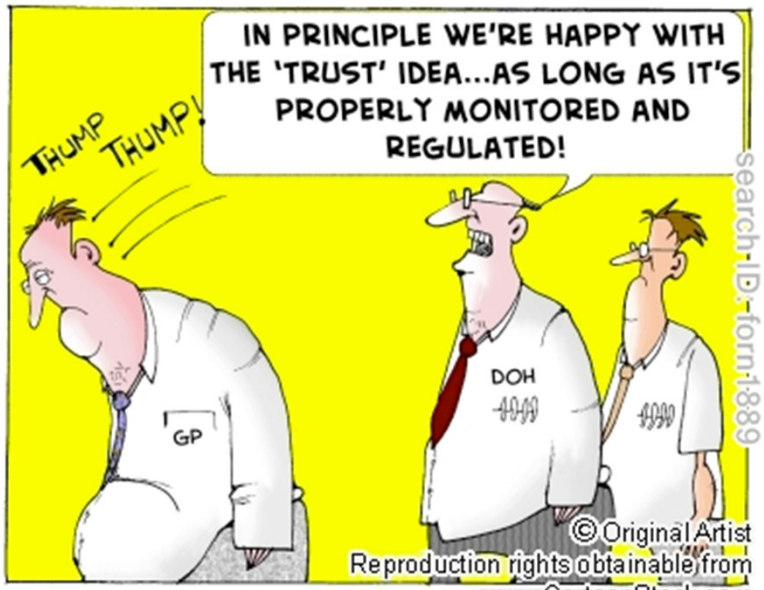
- Utilisateurs
  - Ceux qui déposent leurs données
  - Ceux qui les utilisent
- Partenaires
- Agences de financement

Les éléments de la confiance

- 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



RDA
RESEARCH DATA ALLIANCE

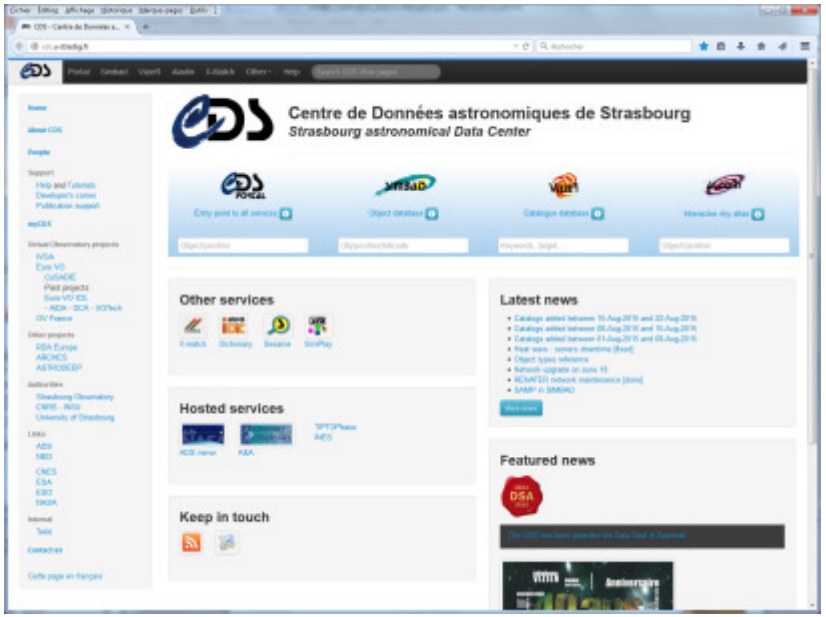
Qu'est-ce qu'un centre de données de confiance?

- 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?

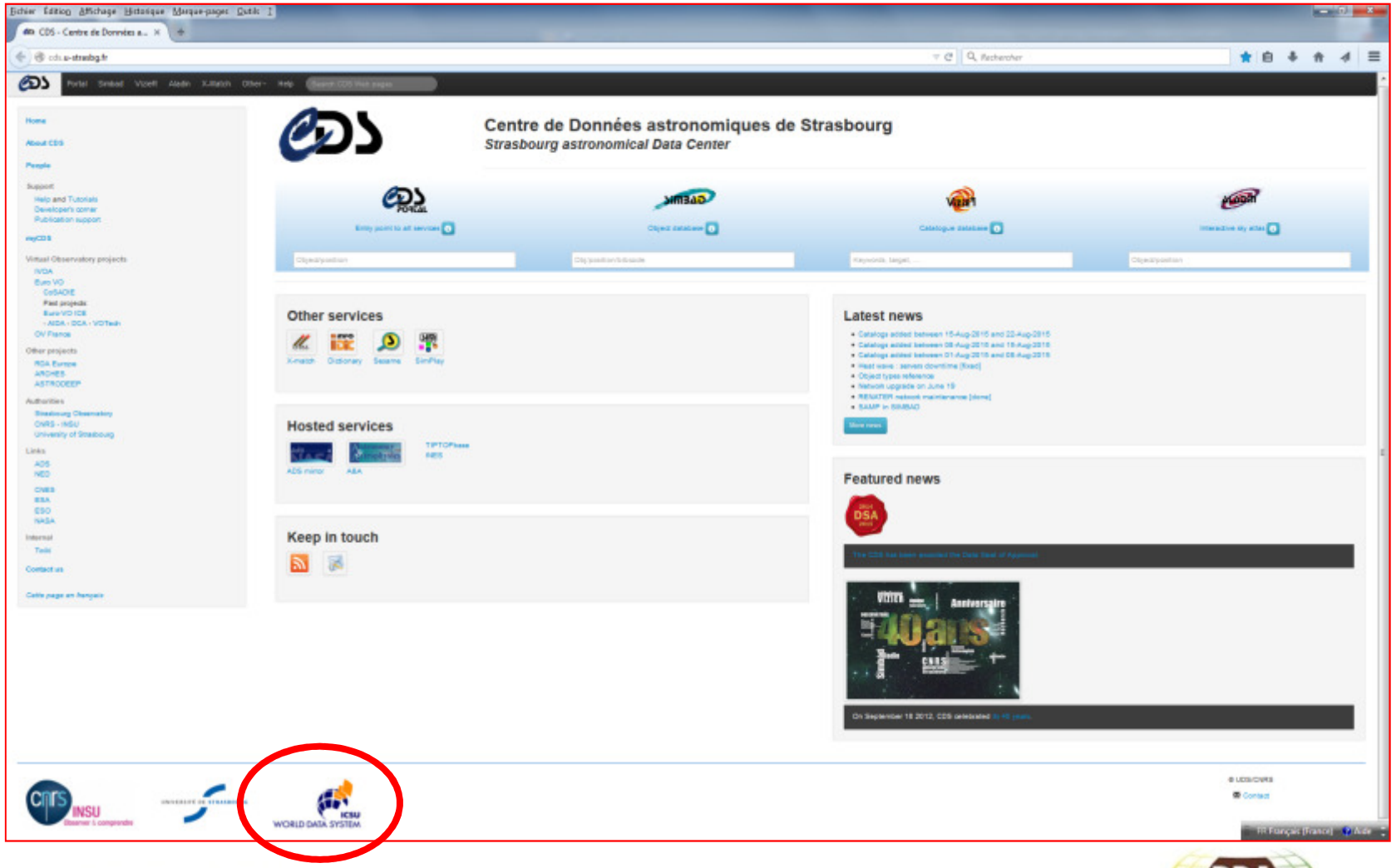
- Ou, pourquoi, puisque nos utilisateurs ont pleinement confiance en nous?

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



- Créé en1972
- ~ 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



WDS: « Trusted Data Services for Global Sciences »

- 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.
- Cooordinates **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

2014 DSA 2015

nestor Seal 2015

DIN 31644

ISO

ISO 16363

ICSU WORLD DATA SYSTEM

research data sharing without barriers
rd-alliance.org

RDA RESEARCH DATA ALLIANCE

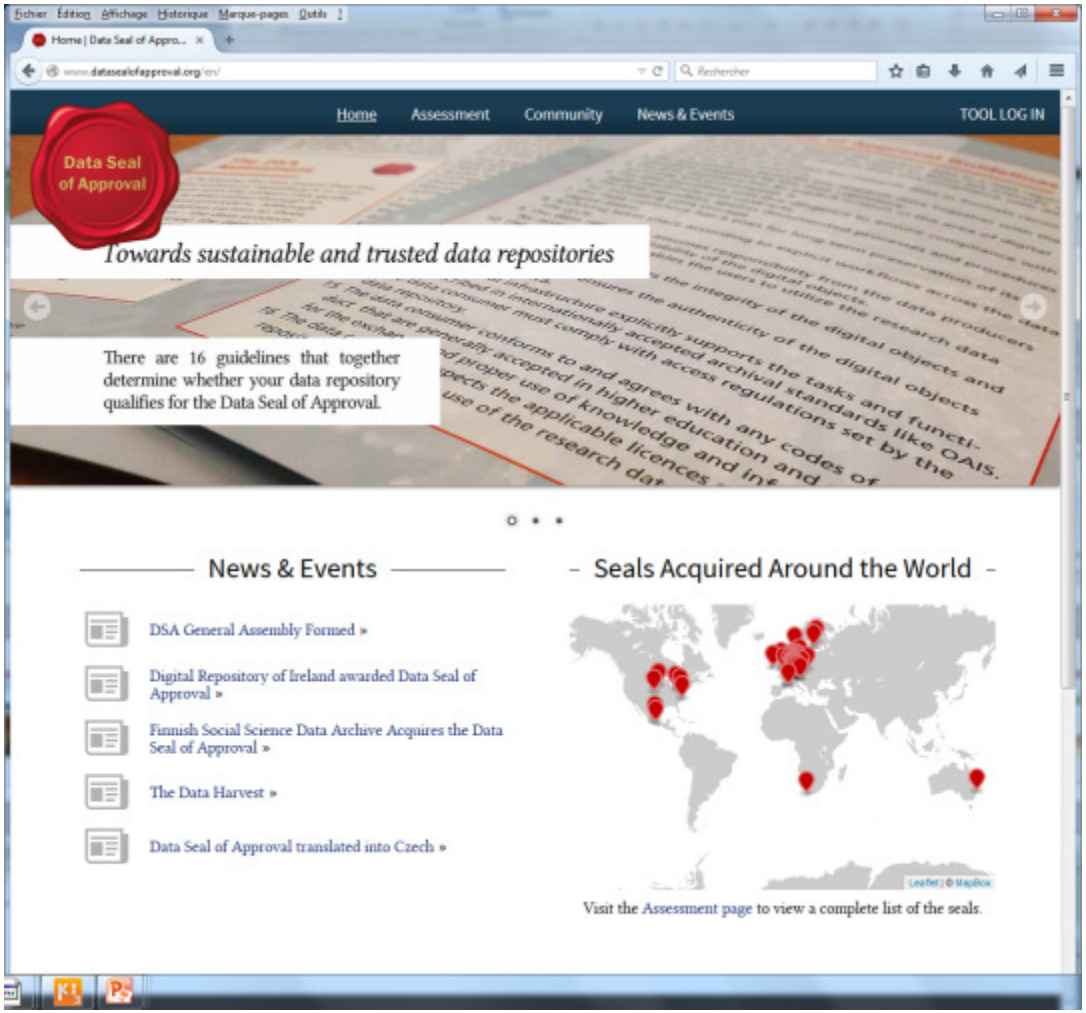
Le cadre européen

- **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 t*o Basic Certification obtain full external audit and certification based on ISO 16363

The Data Seal of Approval



Ajouter une certification DSA pour le CDS?

- 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

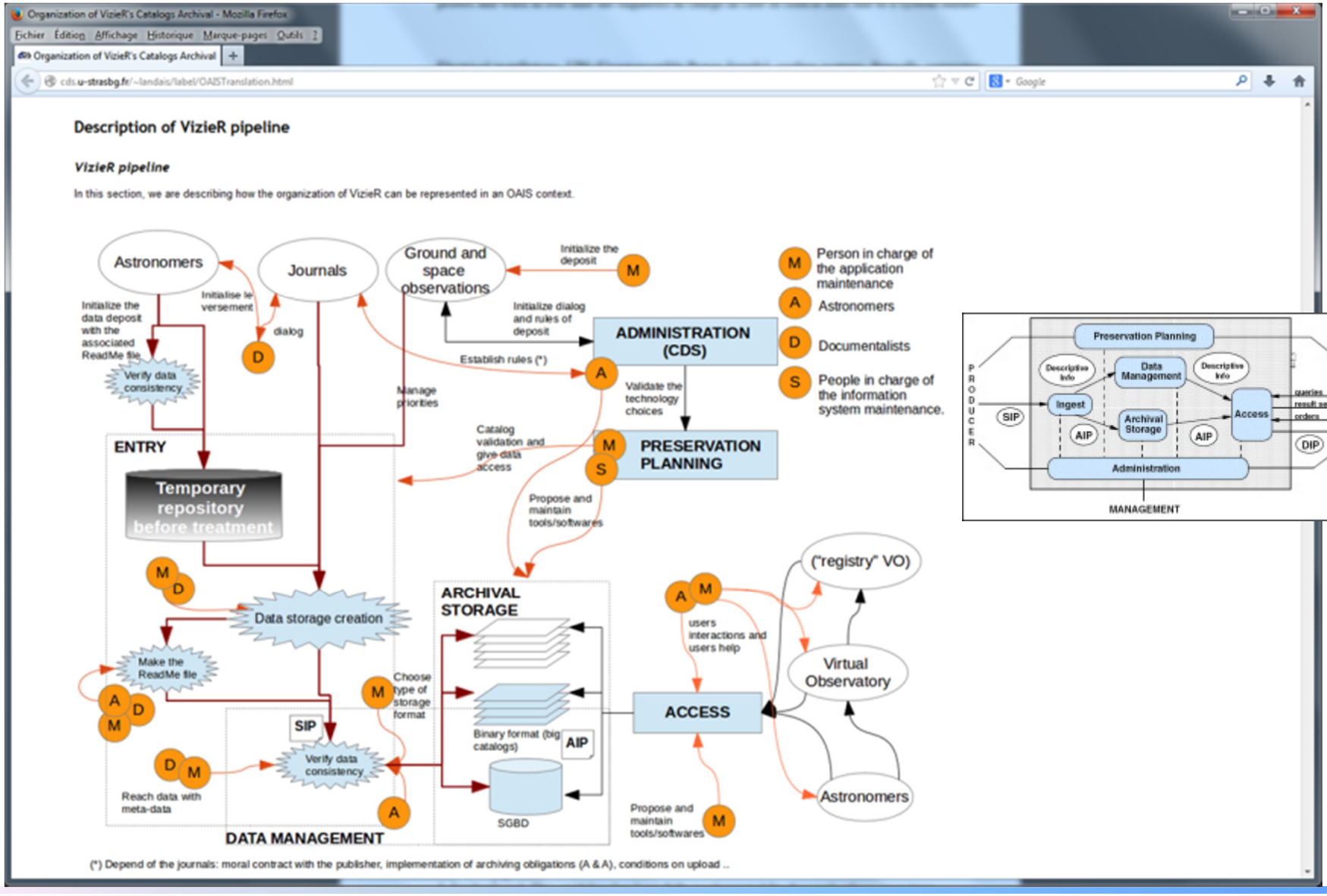
- Décision de candidater après une discussion avec un membre du DSA, et ça a marché – le premier DSA en sciences physiques, le second en France après le CINES

Obtenir le DSA

- 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

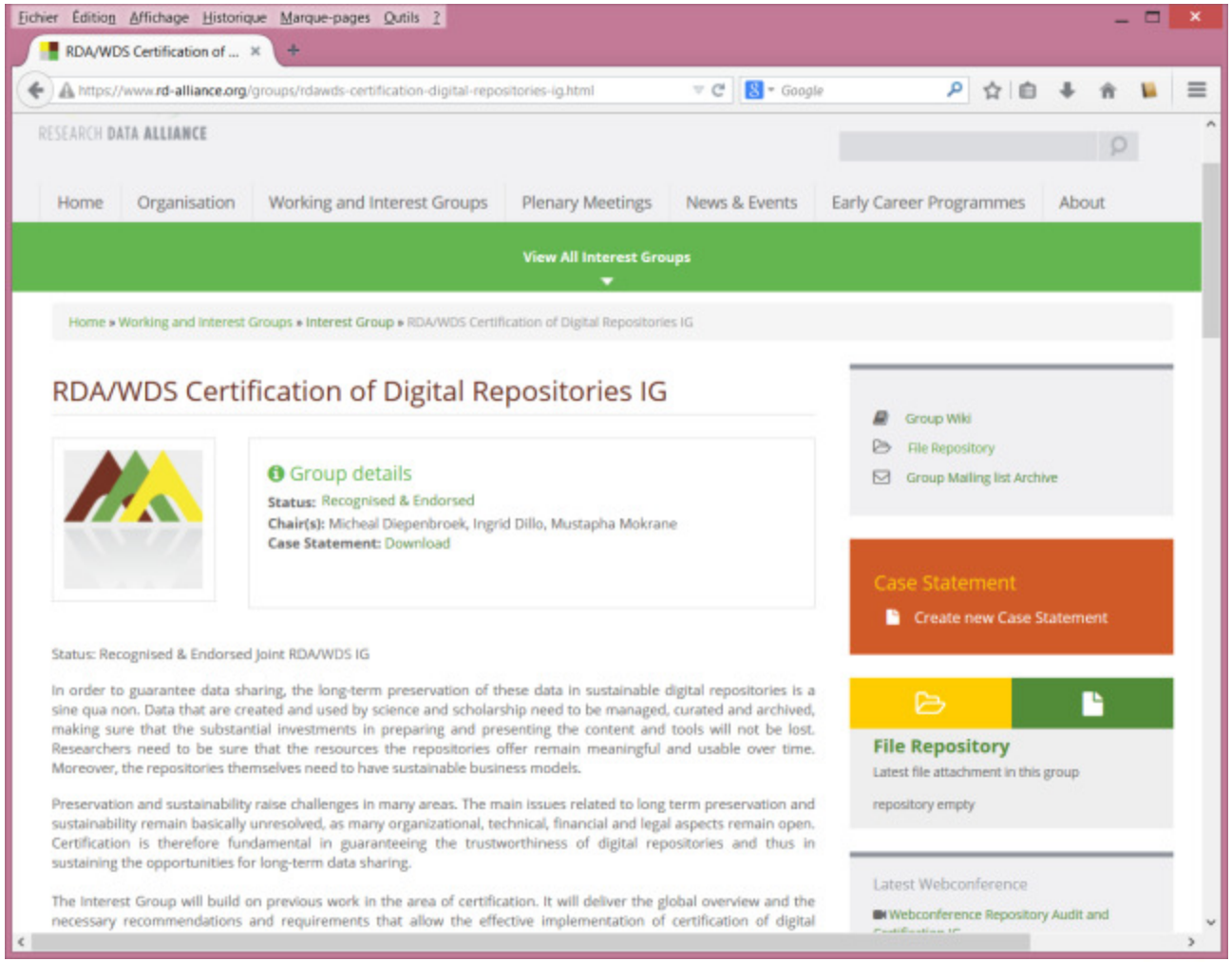


Retour au contexte

- 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

The Certification Working Group



DSA et WDS

- 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

- 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

- 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%93wds-partnership-wg/outcomes/dsa-wds-partership

Les critères communs

- 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

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

Digital object management

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

Technology

- Technical infrastructure
- Security

Un message fort

- 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.

Common catalogue

- 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

Common catalogue

- 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

Common catalogue

- 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

- R13Data 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

- R14Data 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

« Data services »

- 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…