RESEARCH DATA ALLIANCE

ICSU
WORLD DATA SYSTEM

DSA-WDS Partnership on Repository Certification Working Group

research data sharing without barriers

rd-alliance.org

Working Group Members

- Lesley Rickards (UK, PSMSL, WDS-SC) [Co-chair]
- Mary Vardigan (USA, ICPSR, DSA Board) [Co-chair]
- Kevin Ashley (UK, Digital Curation Centre)
- Michael Diepenbroek (Germany, Pangaea, WDS-SC)
- Ingrid Dillo (The Netherlands, DANS, DSA Board)
- Rorie Edmunds (WDS Programme Officer)
- Françoise Genova (France, CDS, WDS-SC)
- Hervé L'Hours (UK, UK Data Archive, DSA Board)
- Guoqing Li (China, CEODE, WDS-SC)
- Jean-Bernard Minster (USA, UCSD, Chair of WDS Scientific Committee)
- Mustapha Mokrane (WDS Executive Director)
- Paul Trilsbeek (The Netherlands, MPI for Psycholinguistics, DSA Board)
- Eleni Panagou, Ph.D. Candidate in Web Engineering, Democritus University of Thrace, Greece [RDA Early Career Researcher]

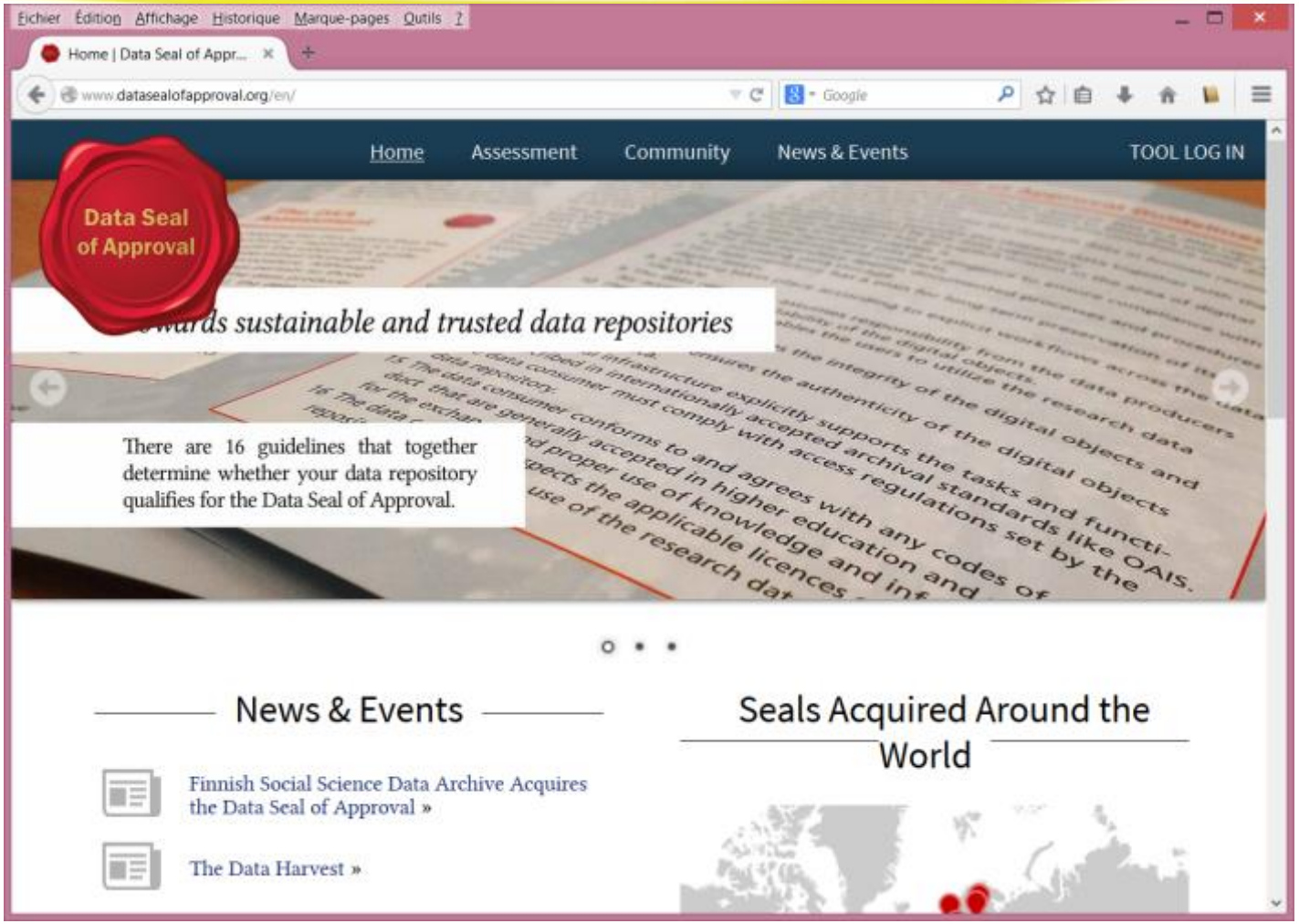
Context

- Long term preservation of data in sustainable digital repositories is a key element of scientific data sharing
- The repositories should be TRUSTED
  - By scientists who deposit their data in a repository
  - By data users
  - By funding agencies who require that project results are shared
- An important topic to tackle within the RDA
- ICSU-WDS builds a community of quality-assured scientific data and data services, products, and information
- RDA/WDS partnership, hence the RDA/WDS Certification of Digital Repositories IG

ICSU World Data System

Repository Audit and Certification IG

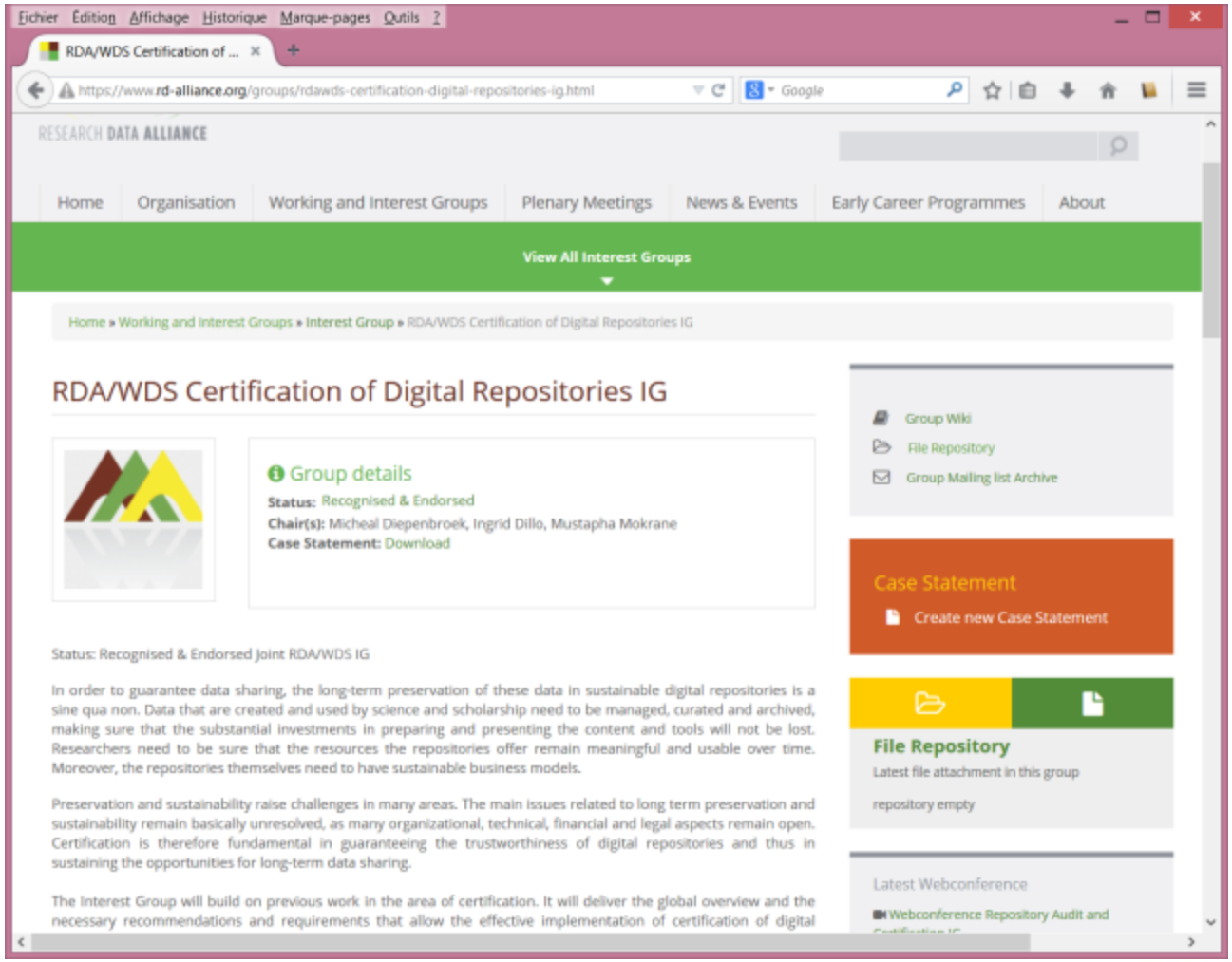
The Data Seal of Approval



Certification WG background

- Data Seal of Approval and World Data System both lightweight mechanisms for repository assessment
  - Self-assessment, no on-site visit
  - Peer-reviewed assessment supervised by the DSA Board and the WDS Scientific Committee
- DSA began in social science and humanities, WDS in natural and physical sciences but both expanding in scope
- Over past years, both groups began to see synergies
- Common members!
- When the RDA/WDS IG established, exploring a partnership seemed natural

The Certification Working Group



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

Goals

- Develop common catalog of criteria for basic repository assessment
- Develop common procedures for assessment
- Implement a shared testbed for assessment
- i.e. *alignment*

- Ultimately, create a shared framework for certification that includes other standards as well, including Nestor and ISO 16363/TRAC

Harmonization of Requirements

- Mapped the DSA to the WDS, and the WDS to the DSA
- Found that lists have similarities and differences
  - DSA guidelines more concise; WDS has multi-part criteria
  - DSA focus on data management, not organizational stability
  - WDS certification includes membership in the WDS and certification of services, not in scope for the DSA

Common Requirements

- Context
- Appraisal
- Mission/scope
- Documented storage procedures
- Preservation plan
- Workflows
- Data discovery and identification
- Data reuse
- Data integrity and authenticity

Common Requirements, continued

- Technical infrastructure
- Security
- Licenses
- Continuity of access
- Data quality
- Confidentiality/Ethics
- Organizational infrastructure
- Expert guidance

Challenges Faced

- There were significant gaps between the DSA and the WDS in some areas:
  - Technical and organizational infrastructure
  - Confidentiality and ethics
- The group had to compromise to come to good solutions

Harmonization of Procedures

- The group works on aligning procedures and is addressing:
  - Appeals process
  - Compliance ratings
  - Path to improvement
  - URLs required in evidence
  - Language
  - Renewal frequency
  - Transparency
  - Procedures for reviewers
  - Branding of new common requirements
  - How the DSA and WDS relate to each other

Next Steps

- Create mapping to Nestor and ISO standards
- Finalize the harmonized requirements with guidance and put them out to the community as Version 1
- Complete work on aligning procedures, determining the relationship of DSA and WDS to each other as organizations
- Create a testbed for certification
- Investigate shared pool of reviewers

Conclusions to date, Questions and Lessons learnt

- The common framework should be a first step in a larger continuum of certification schemes. How do we achieve integration?
- *Global geographic reach.* Basic certification must account for cultural and linguistic differences. The notion of repository certification is currently dominated by Europe and the USA. How do we reach out to others?

- The framework can also be used for self-assessment! An efficient pathway for repository improvement

Common Catalogue

- C1 Context — Org Infrastructure / Digital Object Management

Please provide context for your repository

- C2 Appraisal — Digital Object Management

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

- C3 Mission/Scope — Organizational Infrastructure

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

- C4 Documented storage procedures Digital Object Management

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

- C5 Preservation plan Digital Object Management

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

- C6 Workflows Digital Object Management

Archiving takes place according to defined workflows from ingest to dissemination

- C7 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

- C8 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

- C9 Data integrity and authenticity Digital Object Management

The repository guarantees the integrity and authenticity of the data

- C10 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(ies)

- C11 Security Organizational Infrastructure, Digital Object Management, Technology

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

- C12 Licenses Organizational Infrastructure

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

- C13 Continuity of access Org Infrastructure / Digital Object Management

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

- C14 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

- C15 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

- C16 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

- C17 Expert guidance Organizational Infrastructure

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