



HAL
open science

Data management terminologies: from challenges to community recommendations

Anne-Sophie Bage, Romain David, Pedro Luiz Pizzigatti Corrêa, Françoise Genova, Giovanni Guerrieri, Liise Lehtsalu, Nektarios Liasko, Bénédicte Madon, Paul Millar, Melanie Nentwich, et al.

► To cite this version:

Anne-Sophie Bage, Romain David, Pedro Luiz Pizzigatti Corrêa, Françoise Genova, Giovanni Guerrieri, et al.. Data management terminologies: from challenges to community recommendations. Plenary RDA P25, Oct 2025, Brisbane (AU), Australia. , 1 p., 2024. <hal-05400140>

HAL Id: hal-05400140

<https://hal.inrae.fr/hal-05400140v1>

Submitted on 5 Dec 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



**INTERNATIONAL
DATA WEEK**
13-16 October 2025
Brisbane, Australia



Authors & Contact

Romain David, Anne-Sophie Bage, Pedro Correa, Françoise Genova, Yvan Le Bras, Liise Lehtsalu, Nektarios Liaskos, Bénédicte Madon, Paul Millar, Melanie Nentwich, Alison Specht.
(RDA Professionalising Data Stewardship IG / Sensitive Data IG)

Contact: romain.david@erinha.eu

**Data management terminologies:
from challenges to community recommendations**

JOIN THE RDA SESSION!

Contribution to the Paper Sprint & Paper Validation
Assessing the Severity of Terminological Obstacles for a better Quality of Data Management in your context
Thursday, 16 October 2025 13:30 – 15:00 AEST
03:30 – 05:00 UTC | 05:30 – 07:00 CET | RDA 25th Plenary Meeting **Breakout Session 5 | Venue: Plaza Terrace Room**

Session Objectives

This session is structured around **three interactive components**:

- 1. Survey (20 – 25 min, Mentimeter/Slido)**
 - Participant input on terminological obstacles they have experienced.
 - Scores of the severity of obstacles (Levels 1 – 5) in participants' own contexts..
- 2. Round Table (15 min)**
 - Produce comparable insights across disciplines while capturing local nuance.
 - Discuss how obstacles appear across different domains and communities.
 - Compare severity ratings across contexts and examples.
- 3. Discussion (30 min)**
 - Identify stakeholders needed to address specific obstacles.
 - Connect challenges to possible mitigation strategies.
 - Shape next steps for the collaborative paper and checklist development.

Wrap-up (5 min): Summarise findings and explain how they will feed into the P25 paper sprint and future RDA recommendations (all participants are invited to be co-authors).

Despite an increase in the availability of proposed definitions for many data ecosystem tasks, there remains an urgent need to harmonise data management terminologies that support FAIR principles and actively improve the quality of data management. We use a **five-level severity scale** (from minor disruption to research termination), inspired by established risk assessment frameworks (*likelihood x impact*). The approach **aims to link terminological obstacles** directly to **consequences on projects** such as time, cost, quality, compliance, and reproducibility.

5 Levels of Severity

In your specific context (disciplinary / regional / ...)

A severity level is assigned to each obstacle

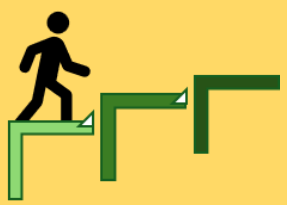
9 sources of confusion

Negligible



- **Definition:** The obstacle causes only minor inconvenience (e.g. slight misunderstanding, quickly clarified).
- **Impact:** Minimal disruption; workflow continues almost unaffected.
- **Analogy:** Like a typo in a document that does not change its meaning.

Low



- **Definition:** The obstacle introduces some ambiguity or inefficiency, requiring clarification.
- **Impact:** Small delays, but work progresses without major rework.
- **Analogy:** A mistranslation in a term that needs checking but doesn't derail a meeting.

Moderate



- **Definition:** The obstacle generates significant confusion or divergent interpretations across roles, disciplines, or languages.
- **Impact:** Requires discussion and harmonisation; project progress slowed.
- **Analogy:** Different groups use "archiving" to mean storage vs. preservation, leading to rework.

High



- **Definition:** The obstacle blocks alignment on responsibilities or processes in a critical part of the data lifecycle.
- **Impact:** Substantial delay, additional cost, or risk to data quality and compliance.
- **Analogy:** Unclear distinction between "data steward" and "data owner" leads to gaps in accountability for FAIR compliance.

Critical



- **Definition:** The obstacle halts workflows or makes collaboration impossible until resolved.
- **Impact:** Severe disruption, project risks failure, or results become non-reproducible.
- **Analogy:** Lexical gaps or conflicting role definitions across infrastructures prevent agreement on data governance.

Polysemy

one term or concept, multiple meanings → ambiguity.

Disciplinary divergence

same term, different meanings across fields

Linguistic confusion

multiple terms for the same meaning (synonyms)

Geographical variation

same term, different meanings across countries/regions

Lexical gaps

concepts without equivalents in the target language.

Idiomatic usage

imported terms with context-specific meanings → ambiguity.

Fuzziness

nuanced or evolving terminology → ambiguity.

Adoption barriers

new terms not yet endorsed or widely used by the community.

Vagueness

unclear meanings (e.g. administrative units differ in name/nature across regions)

Building on Specht et al. (2024)
<https://dx.plos.org/10.1371/journal.pone.0311967>