From Declarative to Machine-Actionable Openness: Analysis of Research Data Sharing in COVID-19 Publications with Spanish-Affiliated Authors

ROXANA CERDA-COSME

EVA MENDEZ

*Author affiliations can be found in the back matter of this article

RESEARCH PAPERS

ubiquity press

ABSTRACT

This study examines research data-sharing practices in COVID-19 publications from 2020 to 2024 in which the first author and/or corresponding author held an institutional affiliation in Spain. A data-centric framework combines the analysis of Data Availability Statements (DAS), repository verification, and automated Findable, Accessible, Interoperable, and Reusable (FAIR) assessment. The corpus comprises 7,387 articles with processable full text, including 1,156 with a DAS. Five declared modalities were identified: Repository, Upon request, Supplementary material, Not available, and Other. DAS prevalence remained stable, but Upon request increased while Repository declined. Of 342 repository declarations, 107 corresponded to verifiable deposits, 68 contained data generated by the study, and 42 were compatible with automated FAIR evaluation. Regression models showed no temporal association with DAS adoption or verifiable deposit. Declared funding was positively associated with verifiable deposit, while deposits associated with 2023 publications had lower automated evaluability than those from 2020. No consistent structural pattern in FAIR maturity was identified, and repository category and object type were not significantly associated with its scores. These findings show that declared availability does not necessarily ensure verifiable or machine-actionable data sharing.

CORRESPONDING AUTHOR:

Roxana Cerda-Cosme

Library and Information Science Department, Universidad Carlos III de Madrid, Spain

rcerda@bib.uc3m.es

KEYWORDS:

Data Availability Statements (DAS); Research Data Sharing; FAIR Principles; COVID-19; Open Science; Text Mining; Spain

TO CITE THIS ARTICLE:

Cerda-Cosme, R. and Mendez, E 2026 From Declarative to Machine-Actionable Openness: Analysis of Research Data Sharing in COVID-19 Publications with Spanish-Affiliated Authors. *Data Science Journal*, 25: 42, pp. 1–18. DOI: https://doi.org/10.5334/dsj-2026-042

1. INTRODUCTION

1.1 DATA AVAILABILITY STATEMENTS AND UNCERTAINTY ABOUT THEIR EFFECTIVENESS

Data Availability Statements (DAS) have become an important editorial mechanism for promoting transparency within open science. Through these structured sections, authors indicate whether the research data associated with an article are available, where they can be found, and under what conditions they may be accessed.

DAS are intended to support reproducibility and data reuse (Colavizza *et al.*, 2020). They may also represent an initial form of data curation because researchers must communicate the location and access conditions of their datasets (An, 2026). However, DAS do not necessarily imply immediate access, technical verifiability, or interoperability (Federer *et al.*, 2018). Determining whether these declarations correspond to accessible and reusable data therefore remains a central challenge for the scalable assessment of research data openness.

1.2 PREVIOUS RESEARCH ON DATA AVAILABILITY AND REUSE

Previous studies have documented the increasing adoption of DAS across disciplines and journals (Asmundo *et al.*, 2024; Graf *et al.*, 2020; McGuinness and Sheppard, 2021). Access to research data and clear processing instructions can support the verification of published results and facilitate new research (Harris *et al.*, 2018). Appropriate organisation and documentation also improve the potential for reuse (Samuel and König-Ries, 2021), while recognising datasets as research outputs can encourage sharing (Kratz and Strasser, 2015).

Nevertheless, the presence of a DAS does not guarantee access to the associated data. Statements may refer readers to information contained within the article or its supplementary materials, while external repository deposit remains less frequent (Jenkins *et al.*, 2022). Access upon request is particularly uncertain because authors may not respond or may decline to provide the data (Jiao *et al.*, 2024; Kim *et al.*, 2025).

Accessibility also does not ensure reusability. Insufficient documentation, unstructured metadata, and fragmented repository environments can limit the interpretation and reuse of openly shared data (Kim *et al.*, 2025; Qiu and Hu, 2025). The FAIR principles address these conditions by promoting data that are Findable, Accessible, Interoperable, and Reusable (FAIR) (Wilkinson *et al.*, 2016). Research data openness must therefore be examined beyond the formal declaration of availability.

1.3 RESEARCH GAP: TOWARDS SCALABLE AND MACHINE-ACTIONABLE MEASUREMENT

The COVID-19 pandemic provides a relevant context for examining the relationship between declared availability and effective openness. International initiatives promoted rapid access to data and scientific results to accelerate research and support the global response (CODATA, 2020; Harrison *et al.*, 2021; Wellcome Trust, 2020; Xu *et al.*, 2020).

Despite these efforts, data-sharing practices remained uneven. Reported barriers included the time required to prepare data, uncertainty regarding licences and usage rights, difficulties selecting repositories, and sharing costs (Dron *et al.*, 2022; Gao *et al.*, 2020). The urgency of the pandemic did not substantially alter the behaviour of many researchers (Zuiderwijk, 2024). Authors declaring that COVID-19 data were available upon request frequently failed to respond or declined to share them (Strcic *et al.*, 2022), while persistent identifiers and open licences did not necessarily ensure FAIR compliance or the absence of technical barriers (Sofi-Mahmudi *et al.*, 2024).

It remains unclear how textual declarations correspond to verifiable repository deposits and machine-actionable metadata conditions across a large publication corpus. Addressing this gap requires connecting DAS classification, infrastructural verification, and automated FAIR assessment within the same workflow.

1.4 SPANISH INSTITUTIONAL CONTEXT

Data-sharing practices are shaped by national regulations, infrastructures, and research evaluation systems. This study therefore focuses on publications in which the first author and/or

corresponding author held an institutional affiliation in Spain. The Spanish research system has introduced several policies to promote open science and access to research outputs (Bordons, González-Albo and Moreno-Solano, 2023; De Filippo, Lascurain-Sánchez and Sánchez, 2023).

This delimitation reduces regulatory heterogeneity and enables a focused examination of how declarative transparency translates into infrastructural availability. Open science practices in Spain are influenced by public funding, the internationalisation of scientific production, and the role of academic institutions in repository management (Bordons, González-Albo and Moreno-Solano, 2023).

During 2020–2024, Law 17/2022 on Science, Technology and Innovation strengthened the promotion of open science and the deposit of publications and associated data in institutional or specialised repositories (De Filippo and Sastrón-Toledo, 2023; Ollé *et al.*, 2023). The National Open Science Strategy 2023–2027 also promotes FAIR research data management and interoperable infrastructures connected to the European Open Science Cloud (Ministerio de Ciencia e Innovación, 2023).

1.5 CONCEPTUAL CONTRIBUTION: DECLARATIVE, VERIFIABLE, AND MACHINE-ACTIONABLE OPENNESS

This study distinguishes three analytical levels of research data openness: declarative transparency, observed through the presence and content of DAS; verifiable openness, demonstrated through identifiable repository deposits; and machine-actionable openness, assessed through FAIR maturity indicators based on structured metadata.

Connecting these levels provides a scalable framework for assessing research data sharing beyond textual declarations. DAS are treated as structured evidence that can be linked to repository verification and automated metadata assessment.

1.6 RESEARCH QUESTIONS

This study addresses the following research questions:

RQ1: How can declared modalities of research data sharing be systematically identified at scale using Data Availability Statements as structured textual evidence?

RQ2: How did declared data-sharing modalities evolve between 2020 and 2024, and to what extent did repository declarations correspond to verifiable deposits?

RQ3: Which article-level and infrastructural factors are associated with verifiable repository deposit, automated FAIR evaluability, and FAIR maturity?

The study analyses 7,387 COVID-19 articles published between 2020 and 2024 in which the first author and/or corresponding author held an institutional affiliation in Spain. It combines text mining, repository verification, and automated FAIR assessment to examine the transition from declarative transparency to machine-actionable openness.

2. A DATA-CENTRIC FRAMEWORK FOR MEASURING DECLARATIVE AND MACHINE-ACTIONABLE OPENNESS

2.1 DECLARATIVE TRANSPARENCY AND VERIFIABLE OPENNESS

DAS operate as article-level declarations of how research data may be accessed. Their incorporation into journal policies reflects broader efforts to operationalise transparency in scientific publishing (Klebel *et al.*, 2025). They appear across different levels of journal data policy, from recommendations to mandatory requirements (Hrynaszkiewicz *et al.*, 2020), and publishers increasingly require authors to indicate where supporting data can be found or to provide links to archived datasets (Abdul Razack, Aranjani and Mathew, 2022).

The inclusion of a DAS constitutes evidence of declarative transparency. Verifiable openness requires the declaration to correspond to an identifiable and accessible repository deposit. Data available upon request, included as supplementary material, or deposited in a repository represent different modalities that do not provide equivalent levels of independent verification.

The effectiveness of DAS therefore depends on verifying the existence and accessibility of the associated data (Gabelica, Bojčić and Puljak, 2022). This distinction is particularly relevant during scientific emergencies, when policies promoting rapid sharing do not always produce consistent repository deposit and access practices (Kamath *et al.*, 2023). A data-centric assessment must consequently connect textual declarations with evidence from repository infrastructures.

2.2 OPENNESS AS A SOCIO-TECHNICAL AND INFRASTRUCTURAL PHENOMENON

Data sharing is a socio-technical phenomenon mediated by infrastructures, metadata standards, disciplinary norms, and editorial policies. Open access, open data, and the FAIR principles jointly contribute to more transparent, accessible, and reproducible science (Umbach, 2024). Repositories support this process through metadata schemas, persistent identifiers, and preservation mechanisms.

The characteristics of these infrastructures affect how openness is implemented. Disciplinary repositories commonly use standardised schemas and controlled vocabularies, while generalist and institutional repositories may provide different levels of flexibility, metadata standardisation, curation, and technical support.

The absence of adequate infrastructure or implementation guidance can limit data-sharing policies. In health research, sharing requirements are not always accompanied by sufficient guidance across funders, ethics committees, trial registries, and journals (Tan *et al.*, 2024). During health emergencies, common standards are especially important because locating a dataset does not guarantee its quality or effective reuse (Lucas-Dominguez *et al.*, 2021). Measuring openness therefore requires attention to both repository deposits and the technical characteristics of the resulting digital objects.

2.3 FAIR AS AN INDICATOR OF MACHINE-ACTIONABLE OPENNESS

The FAIR principles provide an operational framework for evaluating whether data are FAIR through criteria including persistent identifiers, structured metadata, controlled vocabularies, and explicit licensing (Wilkinson *et al.*, 2016). In this study, these criteria represent the machine-actionable dimension of research data openness.

Machine-actionability is important when computational systems must locate, retrieve, and integrate distributed resources. During the COVID-19 pandemic, using Fast Healthcare Interoperability Resources (FHIR) as a canonical data model enabled the automated production of research data marts and reduced data-transformation and update times (Lenert *et al.*, 2021). Reproducibility in electronic health record research similarly depends on machine-readable metadata, standardised governance, code sharing, and technical support (Williams *et al.*, 2025).

F-UJI (FAIRsFAIR Research Data Object Assessment Service) assesses FAIR maturity through automated tests of metadata retrieved from digital objects and repository infrastructures. Its scores reflect structured, machine-readable metadata and technical conditions supporting discovery and reuse. FAIR maturity is operationalised in this study on a scale from 0 to 3 and does not assess the scientific quality, validity, or substantive value of the underlying data.

2.4 FROM DECLARATION TO MACHINE-ACTIONABLE OPENNESS: AN INTEGRATED FRAMEWORK

This study examines research data openness across three interconnected levels:

1. Textual level: declared modalities in DAS.
2. Infrastructural level: verification of repository deposits through identifiable access mechanisms.
3. Metadata level: FAIR maturity as an indicator of machine-actionable openness.

Each level introduces additional technical and infrastructural requirements. The transition from declared availability to machine-actionable openness is therefore characterised by progressive filtering.

Two analytical gaps are examined:

1. Declaration to verifiable deposit: the difference between textual claims of availability and identifiable, accessible datasets in repository infrastructures.
2. Verifiable deposit to machine-actionable openness: the difference between the existence of a dataset and its compatibility with automated discovery, access, and reuse under the FAIR principles.

The framework integrates text mining, repository verification, and automated FAIR assessment within a unified workflow. These levels are operationalised through DAS modalities, verifiable repository deposits, and FAIR maturity metrics, respectively.

2.5 AUTHORSHIP POSITION AND RESEARCH DATA RESPONSIBILITY

National publication corpora are commonly delimited through authors' institutional affiliations. However, considering any national affiliation may overestimate the connection between a research system and decisions concerning the management, organisation, and communication of research data, particularly in extensive international collaborations.

This study therefore applied an operational criterion based on two complementary authorship positions. First authorship was considered because it commonly reflects substantial involvement in conducting and documenting research, although it does not necessarily imply responsibility for managing or communicating the associated data. Corresponding authorship was included because the corresponding author serves as the formal contact for questions concerning the article, its methods, and its underlying data. Articles were included when the first author and/or corresponding author held an institutional affiliation in Spain.

The criterion identifies publications in which an authorship position associated with conducting or formally communicating the research was connected to the Spanish research system. Decisions concerning repositories, metadata, infrastructure, and code may nevertheless involve several team members and depend on institutional support, including secure data environments, linked datasets, software licensing, computational resources, and specialised data-management assistance (Williams *et al.*, 2025).

Authorship conventions differ across disciplines, and some fields use alphabetical or other ordering practices. Alphabetical order was not assessed separately. The combined use of first and corresponding authorship should therefore be understood as an operational criterion rather than direct evidence of responsibility for research data.

3. DATA AND METHODS

3.1 OVERALL DESIGN AND ANALYTICAL WORKFLOW

This study adopts a data-centric meta-research design to examine, in a stepwise manner, the transition from declarative transparency to verifiable infrastructural openness and machine-actionable FAIR maturity in COVID-19 publications in which the first author and/or corresponding author held an institutional affiliation in Spain.

The analytical strategy integrates three complementary components:

1. Text mining and rule-based natural language processing (NLP) to identify and classify DAS at scale;
2. Infrastructural verification of declared repository deposits, distinguishing between textual declarations and verifiable availability;
3. Automated FAIR maturity assessment using F-UJI to analyse the machine-actionable dimension of deposited digital objects.

Rather than treating these components as independent procedures, the study combines them in a unified analytical workflow structured around the transitions defined in the conceptual framework. The workflow comprises five sequential phases: construction of the analytical universe, textual normalisation and preparation, linguistic modelling, infrastructural extraction, and analytical integration. This design connects article-level declarations, repository-level verification, and digital-object-level FAIR assessment within a coherent and reproducible pipeline.

3.2 PHASE 1: CONSTRUCTION OF THE ANALYTICAL UNIVERSE

The analytical universe was defined through a systematic Scopus search for scientific publications on COVID-19 published between 2020 and 2024 in the field of medical sciences.

The query combined pandemic-related terms in the TITLE-ABS-KEY fields with structural filters that restricted the corpus to original articles (DOCTYPE(ar)), published in scientific journals (SRCTYPE(j)), in final publication stage (PUBSTAGE(final)), and classified in the MEDI subject area. The filter AFFILCOUNTRY("Spain") was also applied.

The initial search returned 10,808 records. To delimit the corpus more specifically, an article was retained when the first author and/or corresponding author held an institutional affiliation in Spain. This criterion reduced the analytical universe to 8,077 publications.

Full texts were retrieved through the Elsevier application programming interface (API) and, when necessary, through Unpaywall. The process was automated using Python scripts to ensure traceability and reproducibility.

3.3 PHASE 2: NORMALISATION, VALIDATION, AND TEXTUAL PREPARATION

The bibliographic records and retrieved documents were subjected to technical validation to ensure consistency between the metadata and files. Digital object identifiers (DOIs) were harmonised, duplicates were removed, and the integrity of the PDF files was verified. The documents were then converted into plain text (TXT), applying procedures to mitigate distortions caused by editorial formats, multi-column structures, tables, and figures.

After quality control, 690 records were excluded, resulting in a final corpus of 7,387 articles with processable text.

3.4 PHASE 3: LINGUISTIC MODELLING AND OPERATIONALISATION OF DAS MODALITIES

A structured extraction scheme based on text mining and deterministic NLP was designed. The corpus included publications written in English and Spanish, and Spanish-language articles were retained and analysed alongside English-language articles. Rather than using machine-learning models, a rule-based approach was adopted because it is appropriate in contexts where editorial patterns are recurrent and methodological transparency and reproducibility are priorities.

To identify and classify DAS consistently across both languages, two bilingual English-Spanish lexical repertoires were developed:

1. To identify and extract complete DAS sections;
2. To classify the declared modalities of data availability.

The primary categories were:

- Repository
- Supplementary material
- Upon request
- Not available
- Other

During the development of the rule-based classifier, the researchers conducted manual quality-control checks on randomly selected articles to verify the identification and extraction of DAS sections and the assignment of data-sharing modalities. These checks formed part of the classifier-development process and were not conducted using an independently coded validation sample.

This operationalisation translated heterogeneous textual content into comparable categorical variables suitable for the large-scale analysis of declared data-sharing practices.

3.5 PHASE 4: STRUCTURED EXTRACTION AND INFRASTRUCTURAL CHARACTERISATION

DAS were identified and classified using lexical rules and explicit syntactic patterns, prioritising reproducibility and semantic control.

For articles classified in the Repository modality, two elements were extracted: the name of the repository and the persistent identifier. A distinction was made between generic repository mentions and deposits accompanied by technically resolvable identifiers, including DOIs, Handles, accession numbers, and other persistent identifiers.

Resolvable identifiers provided the basis for verifying repository deposits and, when technically feasible, conducting automated FAIR maturity assessment through F-UJI.

3.6 PHASE 5: ANALYTICAL INTEGRATION AND OPERATIONALISATION OF OPENNESS TRANSITIONS

The results were consolidated into structured analytical matrices aligned with the three levels defined in the conceptual framework.

First, bibliographic metadata, authorship-position and affiliation variables, DAS presence or absence, and declared modality were integrated into an article-level matrix. Second, the infrastructural dimension was operationalised by distinguishing between repository declarations, verifiable deposits with identifiable access mechanisms, and non-verifiable deposits.

Third, automated measurability was incorporated by distinguishing between deposits compatible with F-UJI evaluation through a resolvable DOI and non-evaluable deposits. The latter included objects referenced through Handles, accession numbers, restricted-access mechanisms, other identifiers, or non-operational links.

In the implemented workflow, automated F-UJI assessment was restricted to research-data deposits identified by a resolvable DOI. Deposits referenced through Handles, accession numbers, restricted-access mechanisms, or other identifiers were not submitted to F-UJI because they were not compatible with the automated procedure used in this study. These objects were classified as non-evaluable rather than non-FAIR.

This analytical sequence operationalised two central openness transitions:

1. Declaration → verifiable deposit;
2. Verifiable deposit → machine-actionable FAIR evaluability.

These transitions provide the empirical basis for the staged analysis presented in the Results section.

3.7 STATISTICAL MODELLING STRATEGY

Four linked statistical models were estimated to examine factors associated with successive stages from declarative transparency to machine-actionable openness. Models A and B operate at the article level, while Models C and D use progressively restricted subsets of research-data deposits and evaluable digital objects. The FAIR maturity score represents aggregated FAIR maturity on a scale from 0 to 3. Table 1 summarises the specification of each model.

The complete equations, model-ready analytical datasets, computational scripts, and full statistical outputs are available in the accompanying Zenodo repository (Cerda-Cosme and Méndez, 2026).

Publisher-level variables were excluded from Models C and D because their reduced samples could produce unstable dummy-variable estimates and increase the risk of overfitting. F-UJI evaluability and FAIR maturity also depend primarily on repository infrastructure and digital-object metadata rather than on the journal publisher. Model D was specified parsimoniously, and its results are interpreted as exploratory and non-causal.

All statistical analyses were conducted in Python using pandas, NumPy, SciPy, and statsmodels. Scripts and software dependencies are documented in the accompanying computational package.

4. RESULTS

4.1 DAS AND DECLARED SHARING MODALITIES

The final corpus consisted of 7,387 COVID-19 articles with processable full text in which the first author and/or corresponding author held an institutional affiliation in Spain. Of these, 1,156 articles (15.6%) included a DAS, whereas 6,231 (84.4%) did not contain a formal availability section.

MODEL	ANALYTICAL STAGE	SAMPLE	DEPENDENT VARIABLE	ESTIMATION METHOD	PREDICTORS
A	DAS adoption	7,387 articles	DAS presence: 1 = present; 0 = absent	Logistic regression	Year, publisher, open access, funding, number of authors, language
B	Repository declaration to verifiable deposit	342 repository declarations	Verifiable deposit: 1 = generated research data or identifiable pre-existing data; 0 = non-verifiable deposit	Logistic regression	Year, publisher, open access, funding, number of authors, language
C	Research-data deposit to automated evaluability	68 research-data deposits	F-UJI evaluability: 1 = evaluable through a resolvable DOI; 0 = non-evaluable	Logistic regression	Year, open access, funding, number of authors, language
D	Automated evaluability to FAIR maturity	42 F-UJI-evaluable digital objects	FAIR maturity score (0–3)	OLS with HC3 robust standard errors	Repository category, object type, year

Table 1 Summary of the staged statistical modelling strategy.

The analysis of declared data-sharing modalities was restricted to articles with a DAS (*n* = 1,156), as this level captures the declarative dimension of openness. The textual classification identified five primary modalities (Table 2).

CATEGORY	*n*	% (WITHIN DAS)
Repository	342	29.6
Upon request	502	43.4
Supplementary material	220	19.0
Not available	73	6.3
Other	19	1.6

Table 2 Distribution of declared data-sharing modalities among articles with DAS (n = 1,156).

The categories Not available and Other do not indicate a mechanism through which the associated data can be accessed. The remaining categories, Repository, Upon request, and Supplementary material, account for 1,064 articles (92.0% of articles with a DAS).

Upon request was the most frequent modality (43.4%), followed by Repository (29.6%) and Supplementary material (19.0%). This distribution indicates that, even when a formal availability statement was present, author-mediated access remained the dominant declared mechanism for obtaining research data.

4.1.1 Verifiability of repository deposits

To examine the transition from declarative transparency to verifiable openness, the subset classified in the Repository modality (*n* = 342) was analysed to determine whether the declarations corresponded to identifiable deposits with functional access mechanisms (Table 3).

Only 107 articles (31.3% of repository declarations) provided identifiable digital objects with functional access mechanisms. These comprised 68 deposits containing research data generated by the study and 39 deposits containing pre-existing data reused in the research.

When the analysis was restricted to deposits containing research data generated by the study, the number decreased to 68 cases, representing 5.9% of articles with a DAS and 0.9% of the total corpus. This progressive filtering reveals a substantial discrepancy between declared repository availability and verifiable data access.

TYPE	*n*	% (WITHIN REPOSITORY)
Research data generated	68	19.9
Pre-existing data reused	39	11.4
No verifiable link	221	64.6
Broken link or restricted access	14	4.1

Table 3 Verification status of repository declarations (*n* = 342).

Starting from 7,387 articles, the analytical sequence was reduced to 1,156 articles with a DAS and 342 repository declarations. Of these declarations, 107 corresponded to verifiable deposits, 68 contained research data generated by the study, and 42 were compatible with automated FAIR evaluation. This reduction empirically illustrates the two structural gaps identified in the conceptual framework.

4.2 LONGITUDINAL DYNAMICS OF SELECTED DECLARED DATA-SHARING MODALITIES (2020–2024)

The longitudinal analysis focuses on three declared modalities: Repository, Upon request, and Supplementary material. Percentages were calculated within each publication year using the total number of articles assigned to these three modalities as the denominator (Figure 1).

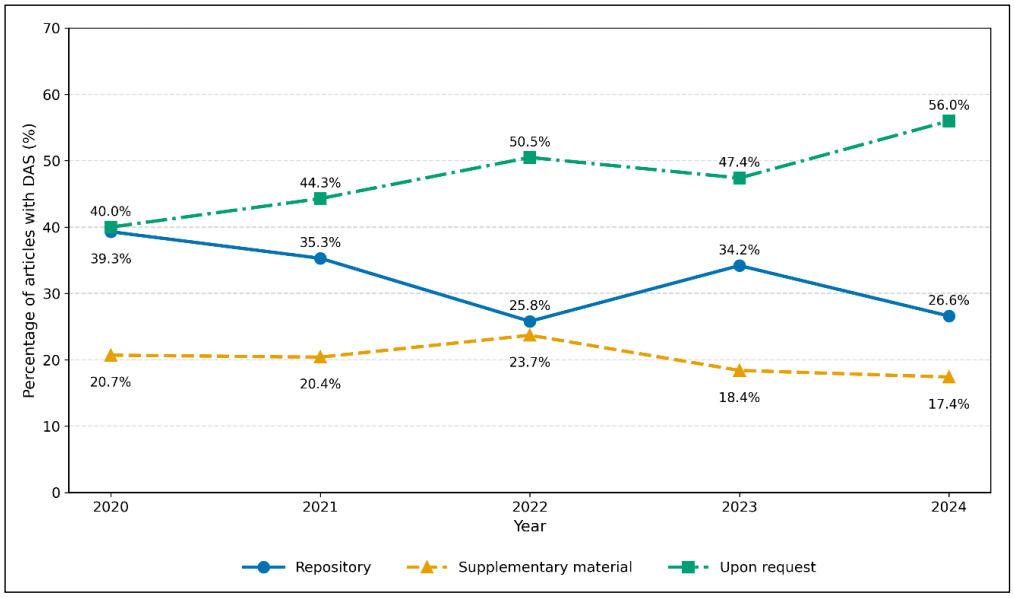

Figure 1 Annual distribution of selected declared data-sharing modalities (2020–2024).

In 2020, Upon request was the most frequent modality (40.0%), closely followed by Repository (39.3%). Upon request remained the predominant modality throughout the study period, increasing to 44.3% in 2021 and 50.5% in 2022. After a temporary decline to 47.4% in 2023, it reached its highest value in 2024 (56.0%).

In contrast, Repository showed an overall downward trajectory, decreasing from 39.3% in 2020 to 26.6% in 2024, despite a temporary increase from 25.8% in 2022 to 34.2% in 2023. Supplementary material also declined overall, from 20.7% in 2020 to 17.4% in 2024, with values ranging between 17.4% and 23.7% during the study period.

4.2.1 Cumulative variation

Between 2020 and 2024:

- Repository decreased by 12.7 percentage points, corresponding to a relative decline of 32.3%.
- Upon request increased by 16.0 percentage points, corresponding to a relative increase of 40.0%.
- Supplementary material decreased by 3.3 percentage points.

The difference between request-based and repository-based access increased from only 0.7 percentage points in 2020 to 29.4 percentage points in 2024.

4.2.2 Structural interpretation

Rather than indicating a consolidation of infrastructural openness, the observed trend reflects an increasing predominance of conditional and author-mediated access. This pattern reinforces the distinction between declarative transparency and operational openness.

4.3 FACTORS ASSOCIATED WITH DAS ADOPTION (MODEL A)

A binary logistic regression model was estimated to examine the factors associated with the presence of a DAS. The model included the full corpus of COVID-19 publications with processable text in which the first author and/or corresponding author held an institutional affiliation in Spain (n = 7,387). Table 4 presents the model results.

VARIABLE	ODDS RATIO	95% CI	*p*-VALUE
Year (reference = 2020)			
2021	1.02	0.83–1.26	0.839
2022	0.99	0.80–1.22	0.915
2023	1.02	0.81–1.28	0.880
2024	0.98	0.76–1.28	0.903
Publisher (reference = Other)			
Elsevier	1.01	0.84–1.22	0.877
MDPI	1.02	0.83–1.25	0.841
Frontiers	1.31	1.00–1.71	0.050
Springer	1.12	0.82–1.52	0.480
Wiley	1.40	1.04–1.89	0.028
BioMed Central	1.83	1.36–2.47	<0.001
Oxford University Press	0.74	0.49–1.13	0.166
SAGE Publications	1.13	0.69–1.86	0.629
Lippincott Williams and Wilkins	0.83	0.47–1.48	0.530
BMJ Publishing Group	0.49	0.22–1.07	0.074
Article characteristics			
Open access	1.05	0.81–1.36	0.702
Funding declared	0.88	0.76–1.01	0.079
Log(Number of authors)	0.99	0.91–1.07	0.773
Non-English language	0.80	0.57–1.12	0.188

Table 4 Factors associated with Data Availability Statement adoption (Model A).

Note: Binary logistic regression estimating the likelihood of DAS inclusion among COVID-19 publications in which the first author and/or corresponding author held an institutional affiliation in Spain (n = 7,387). Odds ratios and 95% confidence intervals are based on robust standard errors using the sandwich estimator. Reference categories: Year = 2020; Publisher = Other. McFadden pseudo-R^2 = 0.006.

No statistically significant temporal association was observed after adjusting for publisher and article characteristics. The odds of DAS inclusion remained similar across the five publication years. Open access status, declared funding, number of authors, and publication language were also not significantly associated with DAS presence.

Publisher-level differences were observed. Compared with publishers included in the residual category, articles published by BioMed Central and Wiley had significantly higher odds of including a DAS. The association for Frontiers was marginal, with the confidence interval reaching 1.00 after rounding. The remaining publisher contrasts were not statistically significant.

The model had limited explanatory capacity, as indicated by its McFadden pseudo-R^2 of 0.006. The results therefore suggest that the variables included in the model explain only a small proportion of the variation in DAS adoption within the analysed corpus.

4.4 FACTORS ASSOCIATED WITH VERIFIABLE REPOSITORY DEPOSIT (MODEL B)

A binary logistic regression model was estimated to examine the factors associated with the transition from repository declaration to verifiable deposit. The analysis included the 342 articles classified in the Repository modality. A deposit was classified as verifiable when the repository declaration corresponded to research data generated by the study or to identifiable pre-existing data. This resulted in 107 verifiable and 235 non-verifiable repository declarations. Table 5 presents the model results.

VARIABLE	ODDS RATIO	95% CI	*p*-VALUE
Year (reference = 2020)			
2021	0.75	0.37–1.54	0.438
2022	0.88	0.41–1.92	0.754
2023	0.75	0.34–1.66	0.476
2024	0.95	0.37–2.42	0.909
Publisher (reference = Other)			
Elsevier	0.91	0.47–1.74	0.768
MDPI	0.49	0.23–1.08	0.078
Frontiers	0.33	0.10–1.12	0.076
Springer	0.48	0.12–1.89	0.292
Wiley	0.81	0.28–2.39	0.705
BioMed Central	1.03	0.26–4.04	0.968
Oxford University Press	1.67	0.34–8.11	0.527
SAGE Publications	0.43	0.05–4.03	0.462
Lippincott Williams and Wilkins	1.73	0.09–34.04	0.720
BMJ Publishing Group	1.12	0.12–10.34	0.920
Article characteristics			
Open access	0.56	0.23–1.35	0.197
Funding declared	1.84	1.05–3.20	0.032
Log(Number of authors)	0.81	0.58–1.13	0.219
Non-English language	1.31	0.48–3.60	0.601

Table 5 Factors associated with verifiable repository deposit (Model B).

Note: Binary logistic regression estimating the likelihood of a verifiable deposit among articles classified in the Repository modality (n = 342). Verifiable deposits include research data generated by the study and identifiable pre-existing data (n = 107). Odds ratios and 95% confidence intervals are based on robust standard errors using the sandwich estimator. Reference categories: Year = 2020; Publisher = Other. McFadden pseudo-R^2 = 0.038.

No statistically significant temporal association was observed. The likelihood that a repository declaration corresponded to a verifiable deposit did not differ significantly between 2020 and the subsequent publication years. Publisher, open access status, number of authors, and publication language were also not significantly associated with verifiable deposit.

Declared funding was the only statistically significant predictor. Articles reporting funding had 1.84 times the odds of providing a verifiable deposit compared with articles without declared funding, after adjustment for the remaining variables.

The model had limited explanatory capacity, with a McFadden pseudo-R^2 of 0.038. The results indicate that the included bibliographic and publication-level characteristics explain only a small proportion of the distinction between repository declarations with and without verifiable deposits.

4.5 FACTORS ASSOCIATED WITH AUTOMATED FAIR EVALUABILITY (MODEL C)

A binary logistic regression model was estimated to examine the factors associated with automated FAIR evaluability among deposits containing research data generated by the study. The analysis included 68 research-data deposits, of which 42 had a resolvable DOI compatible with the automated F-UJI procedure and 26 were not evaluable through the implemented workflow. Table 6 presents the model results.

Compared with deposits published in 2020, those published in 2023 had significantly lower odds of being evaluable through the automated F-UJI procedure. No statistically significant differences were observed for 2021, 2022, or 2024. The confidence intervals for the annual estimates were wide, reflecting the reduced number of observations available at this stage of the analysis.

Open access status, declared funding, and publication language were not significantly associated with automated evaluability. The number of authors showed a positive association that approached, but did not reach, the conventional threshold for statistical significance.

The model had a McFadden pseudo-R^2 of 0.157, indicating that the included variables accounted for a modest proportion of the variation in automated evaluability. Given the reduced sample size and wide confidence intervals, these results should be interpreted as exploratory and non-causal.

VARIABLE	ODDS RATIO	95% CI	*p*-VALUE
Year (reference = 2020)			
2021	1.10	0.22–5.52	0.904
2022	1.65	0.26–10.25	0.592
2023	0.19	0.04–0.91	0.038
2024	0.20	0.03–1.47	0.114
Article characteristics			
Open access	0.22	0.04–1.39	0.107
Funding declared	1.30	0.37–4.54	0.679
Log(Number of authors)	1.89	0.98–3.62	0.057
Non-English language	0.81	0.15–4.50	0.814

Table 6 Factors associated with automated FAIR evaluability (Model C).

Note: Binary logistic regression estimating the likelihood of automated FAIR evaluability among deposits containing research data generated by the study (n = 68). Evaluable deposits were those with a resolvable DOI compatible with the automated F-UJI procedure (n = 42). Odds ratios and 95% confidence intervals are based on robust standard errors using the sandwich estimator. Reference category: Year = 2020. McFadden pseudo-R^2 = 0.157.

4.6 STRUCTURAL CORRELATES OF FAIR MATURITY (MODEL D)

An ordinary least squares regression model with HC3 robust standard errors was estimated to examine the structural factors associated with FAIR maturity among the 42 digital objects evaluated through F-UJI. The dependent variable was the aggregated FAIR maturity score on a scale from 0 to 3. The predictors included publication year, repository category, and digital object type. Table 7 presents the model results.

VARIABLE	β	95% CI	*p*-VALUE
Year (reference = 2020)			
2021	−0.017	−0.351–0.318	0.923
2022	−0.086	−0.491–0.319	0.677
2023	−0.681	−1.269–−0.093	0.023
2024	0.006	−0.583–0.594	0.985
Repository category (reference = Generalist)			
Disciplinary	−0.217	−0.912–0.479	0.542
Institutional	−0.021	−0.410–0.368	0.915
Object type (reference = Dataset)			
Scholarly article	−0.258	−0.711–0.195	0.264
Software	0.035	−0.385–0.454	0.872
Unknown	−0.226	−0.818–0.366	0.454

Table 7 Structural correlates of FAIR maturity (Model D).

Note: Ordinary least squares regression estimating the structural correlates of overall FAIR maturity among F-UJI-evaluable digital objects (n = 42). The dependent variable is the aggregated FAIR maturity score on a scale from 0 to 3. HC3 robust standard errors are reported. Reference categories: Year = 2020; Repository category = Generalist; Object type = Dataset. Model fit: R^2 = 0.337; adjusted R^2 = 0.151.

Digital objects published in 2023 had FAIR maturity scores that were, on average, 0.681 points lower than those published in 2020, after adjusting for repository category and object type. No statistically significant differences were observed for 2021, 2022, or 2024.

Repository category was not significantly associated with FAIR maturity. Digital objects deposited in disciplinary or institutional repositories did not show statistically significant differences from those deposited in generalist repositories. Similarly, scholarly articles, software objects, and objects without a specified type did not differ significantly from datasets.

The model explained 33.7% of the observed variation in FAIR maturity, although the adjusted R^2 was 15.1%. Given the reduced sample size and the small number of observations in some repository and object-type categories, the results should be interpreted as exploratory and non-causal.

5. DISCUSSION

The findings reveal a disconnection between the formal presence of DAS and the conditions required for verifiable and machine-actionable data sharing. Although DAS prevalence remained stable, declared modalities shifted towards access upon request and away from repository deposit. The discussion examines this divergence, the discontinuities across the openness process, and their policy implications.

5.1 FROM DECLARATIVE TRANSPARENCY TO VERIFIABLE OPENNESS

DAS provide a scalable source for identifying declared data-sharing modalities, but their presence does not necessarily indicate effective availability. The limited progression from formal statements to verifiable deposits shows that declarative transparency is necessary but insufficient for infrastructural openness.

Previous research similarly shows that shared data often comprise derived or processed datasets, while raw data remain less frequently available (Klungthanaboon, 2026). Declarations also do not ensure that data are adequately organised, documented, or reusable (Deeb *et al.*, 2025). DAS can therefore be understood as instruments of editorial standardisation that communicate expectations of transparency without enforcing their technical implementation (Montague-Hellen and Montague-Hellen, 2023).

The longitudinal shift from repository declarations towards access upon request indicates growing reliance on author-mediated rather than independently verifiable access. Formal expectations for data sharing were therefore not matched by comparable development in repository-based practices. Barriers such as insufficient metadata standardisation, limited dataset visibility, repository selection, licensing, and code deposit may contribute to this divergence (Yehudi *et al.*, 2025). During the pandemic, inconsistent recording practices, storage systems, and data infrastructures also affected the reliability and completeness of epidemiological data (Starnini *et al.*, 2021). The gap between declared and verifiable openness consequently reflects both author practices and institutional conditions.

5.2 STRUCTURAL DISCONTINUITIES IN THE OPENNESS PROCESS

The first discontinuity occurs between repository declaration and verifiable deposit. The absence of temporal improvement suggests that formal repository statements did not increasingly translate into identifiable data objects. Declared funding was positively associated with verifiable deposit, whereas open access was not, confirming that publication access and research data openness are related but distinct dimensions. This finding is consistent with evidence that data-sharing declarations may provide formal compliance without ensuring effective access (Elghzali *et al.*, 2025).

The second discontinuity concerns automated FAIR evaluability. A substantial proportion of research-data deposits could not be evaluated through the implemented F-UJI procedure, and no consistent temporal improvement was identified. Heterogeneous identifiers, fragmented governance, and limited interoperability may restrict automated assessment even when data have been deposited (Gerber *et al.*, 2025).

The third discontinuity concerns FAIR maturity. No consistent pattern was observed across publication years, repository categories, or object types. Automated evaluability therefore does not ensure higher FAIR maturity, which also depends on documentation, licensing, data preparation, technical expertise, and institutional incentives (Bajraktari *et al.*, 2025).

Research data openness should consequently be understood as a staged rather than binary process. Each transition requires additional technical and institutional conditions, while FAIR compliance does not itself guarantee effective reuse (Ghose, 2024; Tai *et al.*, 2025). Persistent identifiers, metadata standards, interoperable formats, and semantic structures are necessary to connect these stages (Behan *et al.*, 2023; Parks *et al.*, 2023).

5.3 IMPLICATIONS FOR OPEN SCIENCE POLICIES

Policies requiring DAS may increase the visibility of data-sharing practices without ensuring effective openness. Editorial and funder mandates should connect these statements with

expectations concerning repository deposit, resolvable identifiers, structured metadata, licensing, and documentation. Compliance monitoring should distinguish between the presence of a statement and the existence of accessible and machine-actionable data.

The limited compatibility of deposits with automated assessment also highlights the need for greater interoperability among repository and identifier infrastructures. Investment in repositories, metadata creation, persistent identifiers, and specialised data-management support is necessary to reduce dependence on the technical characteristics of individual infrastructures.

Open access to publications should not be treated as a proxy for research data openness. Policies should evaluate declared availability, verifiable deposit, automated evaluability, and FAIR maturity separately to identify where the process breaks down and where institutional support is required.

5.4 LIMITATIONS

The study focuses on COVID-19 publications in the medical field in which the first author and/or corresponding author held an institutional affiliation in Spain. These positions were used as complementary operational criteria and do not constitute direct evidence of scientific leadership or personal responsibility for data management. Authorship conventions vary across disciplines, and alphabetical order was not assessed separately.

The findings may reflect the Spanish research and repository environment, medical publishing practices, and the exceptional conditions of the pandemic. They should not be directly generalised to other countries, globally distributed teams, disciplines, or non-emergency contexts.

The rule-based classification relied on bilingual lexical repertoires and explicit linguistic patterns. Manual quality-control checks were conducted on randomly selected articles, but the classifier was not evaluated against an independently coded sample, and precision, recall, and F1 scores were not calculated. Future studies could use an independent reference sample, multiple coders, inter-coder agreement, and formal performance metrics. The analysis was also limited to processable full texts, while repository links and identifiers were verified at a specific point in time and may change.

F-UJI evaluates automatically retrievable metadata, and its scores may be affected by repository-specific metadata exposure, identifier systems, and interoperability mechanisms. Assessment was restricted to deposits with resolvable DOIs. Objects using other identifiers or access mechanisms were classified as non-evaluable rather than non-FAIR. F-UJI compatibility should therefore not be equated with FAIRness or interpreted as evidence of scientific quality or practical reusability.

The statistical models were observational and do not establish causality. Models A and B had limited explanatory capacity, while the reduced samples in Models C and D produced wide confidence intervals and limited statistical power. Their results should be interpreted as exploratory associations within this corpus.

Despite these contextual limitations, the analytical framework can be tested across other countries, disciplines, research settings, identifier infrastructures, and FAIR assessment tools. Comparative applications could distinguish corpus-specific findings from broader discontinuities between declared availability, verifiable deposit, automated FAIR evaluability, and FAIR maturity.

6. CONCLUSIONS

This study examined research data-sharing practices in COVID-19 publications from 2020 to 2024 in which the first author and/or corresponding author held an institutional affiliation in Spain. It combined text mining, repository verification, and automated FAIR assessment to analyse the transition from declared availability to machine-actionable openness.

Regarding RQ1, DAS provided a scalable source for identifying declared sharing modalities, although their presence did not guarantee accessible or independently verifiable data.

Regarding RQ2, DAS prevalence remained stable, while declared modalities shifted towards access upon request and away from repository-based sharing. This trend indicates increasing reliance on conditional, author-mediated access rather than stronger infrastructural openness.

Regarding RQ3, no consistent temporal improvement was observed in verifiable deposit, automated FAIR evaluability, or FAIR maturity. Declared funding was positively associated with verifiable deposit, while repository category and object type were not significantly associated with FAIR maturity. Results from the restricted analytical samples should be interpreted as exploratory.

Overall, declared availability, verifiable deposit, automated evaluability, and FAIR maturity represent distinct conditions that cannot be captured by a single binary indicator. Requirements for DAS should therefore be linked to repository deposit, resolvable identifiers, structured metadata, licensing, documentation, and institutional support.

DATA ACCESSIBILITY STATEMENT

The model-ready analytical datasets, computational scripts, complete model specifications, and full statistical outputs for Models A to D are openly available through Zenodo: https://doi.org/10.5281/zenodo.23001642. The source data and computational workflow used for corpus construction, full-text processing, DAS extraction and classification, repository verification, and automated FAIR assessment form part of the reproducibility package of an ongoing doctoral thesis. These materials remain under embargo and will be made publicly available upon publication of the thesis.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Roxana Cerda-Cosme orcid.org/0000-0002-8078-0267
Library and Information Science Department, Universidad Carlos III de Madrid, Spain
Eva Mendez orcid.org/0000-0002-5337-4722
Library and Information Science Department, Universidad Carlos III de Madrid, Spain

REFERENCES

Abdul Razack, H.I., Aranjani, J.M. and **Mathew, S.T.** (2022) 'Clinical trial transparency regulations: implications to various scholarly publishing stakeholders', *Science and Public Policy*, 49(6), pp. 951–961. Available at: https://doi.org/10.1093/scipol/scac041

An, B.-G. (2026) 'The effects of research data identifiers in data availability statements on scholarly output diffusion', *Journal of Information Science*. Available at: https://doi.org/10.1177/01655515251397638

Asmundo, M.G., Durukan, E., Russo, G.I., Jensen, C.F.S., Østergren, P.B., Cimino, S. and **Fode, M.** (2024) 'Data availability statements and data sharing in urology: a false promise?', *European Urology Focus*, 10(6), pp. 999–1002. Available at: https://doi.org/10.1016/j.euf.2024.05.019

Bajraktari, R., Booth, F., Munafò, M. and **Beazley-Long, N.** (2025) 'The most prevalent perceived barriers to sharing research data at the point of manuscript acceptance in health and life sciences: a survey at the University of Bristol', *F1000Research*, 14, 482. Available at: https://doi.org/10.12688/f1000research.161819.2

Behan, B., Jeanson, F., Cheema, H., Eng, D., Khimji, F., Vaccarino, A.L., Gee, T., Evans, S.G., MacPhee, F.C., Dong, F., Shahnazari, S., Sparks, A., Martens, E., Lasalandra, B., Arnott, S.R., Strother, S.C., Javadi, M., Dharsee, M., Evans, K.R. and **Mikkelsen, T.** (2023) 'FAIR in action: Brain-CODE – A neuroscience data sharing platform to accelerate brain research', *Frontiers in Neuroinformatics*, 17, p. 1158378. Available at: https://doi.org/10.3389/fninf.2023.1158378

Bordons, M., González-Albo, B. and **Moreno-Solano, L.** (2023) 'Improving our understanding of open access: how it relates to funding, internationality of research and scientific leadership', *Scientometrics*, 128(8), pp. 4651–4676. Available at: https://doi.org/10.1007/s11192-023-04726-1

Cerda-Cosme, R. and **Méndez, E.** (2026) *Data and code underlying the manuscript "From declarative to machine-actionable openness: Analysis of research data sharing in COVID-19 publications with Spanish-affiliated authors" (2020–2024)* [Dataset and code]. Zenodo. Available at: https://doi.org/10.5281/zenodo.23001642

CODATA. (2020) *Data Together COVID-19 appeal and actions*. Available at: https://codata.org/data-together-covid-19-appeal-and-actions/

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K. and **McGillivray, B.** (2020) 'The citation advantage of linking publications to research data', *PLOS ONE,* 15(4), p. e0230416. Available at: https://doi.org/10.1371/journal.pone.0230416

De Filippo, D. and **Sastrón-Toledo, P.** (2023) 'Influence of research on open science in the public policy sphere', *Scientometrics,* 128(3), pp. 1995–2017. Available at: https://doi.org/10.1007/s11192-023-04645-1

De Filippo, D., Lascurain-Sánchez, M.-L. and **Sánchez, F.** (2023) 'Mapping open science at Spanish universities: analysis of higher education systems', *El Profesional de la Información,* 32(4), p. e320406. Available at: https://doi.org/10.3145/epi.2023.jul.06

Deeb, H., Creasey, S., De Ugarte, D.L., Strevens, G., Usman, T., Wong, H.Y., Kutzer, M.A.M., Wilson, E., Zieliński, T. and **Millar, A.J.** (2025) 'The rise of open data practices among bioscientists at the University of Edinburgh', *PLOS ONE,* 20(7), p. e0328065. Available at: https://doi.org/10.1371/journal.pone.0328065

Dron, L., Kalatharan, V., Gupta, A., Haggstrom, J., Zariffa, N., Morris, A.D., Arora, P. and **Park, J.** (2022) 'Data capture and sharing in the COVID-19 pandemic: a cause for concern', *The Lancet Digital Health,* 4(10), pp. e748–e756. Available at: https://doi.org/10.1016/S2589-7500(22)00147-9

Elghzali, A., Molina, D., Chaudhry, M., Duncan, J., Elfar, A., Oldham, E., Dennis, B., Paul, E., Ford, A.I. and **Vassar, M.** (2025) 'Data sharing practices in high-impact rehabilitation journals', *BMC Medical Research Methodology,* 25(1), p. 129. Available at: https://doi.org/10.1186/s12874-025-02587-1

Federer, L.M., Belter, C.W., Joubert, D.J., Livinski, A., Lu, Y.-L., Snyders, L.N. and **Thompson, H.** (2018) 'Data sharing in PLOS ONE: An analysis of data availability statements', *PLOS ONE,* 13(5), p. e0194768. Available at: https://doi.org/10.1371/journal.pone.0194768

Gabelica, M., Bojčić, R. and **Puljak, L.** (2022) 'Many researchers were not compliant with their published data sharing statement: a mixed-methods study', *Journal of Clinical Epidemiology,* 150, pp. 33–41. Available at: https://doi.org/10.1016/j.jclinepi.2022.05.019

Gao, F., Tao, L., Huang, Y. and **Shu, Z.** (2020) 'Management and data sharing of COVID-19 pandemic information', *Biopreservation and Biobanking,* 18(6), pp. 570–580. Available at: https://doi.org/10.1089/bio.2020.0134

Gerber, B., Currie, G.R., Mosoiu, A., Bernier, A., Bernier, F.P., Boycott, K.M., Fiebelkorn, G., Hens, K., Knoppers, B.M., LeBlanc, C., Scherer, S.W., Shaw, D., Viney, C., Virtanen, C., Benseler, S.M., Yeung, R. S.M. and **Marshall, D.A.** (2025) 'A roadmap for navigating child health research data sharing across Canada and beyond – building on UCAN CAN-DU', *Pediatric Rheumatology,* 23(1), p. 89. Available at: https://doi.org/10.1186/s12969-025-01139-7

Ghose, A. (2024) 'Can LCA be FAIR? Assessing the status quo and opportunities for FAIR data sharing', *The International Journal of Life Cycle Assessment,* 29(4), pp. 733–744. Available at: https://doi.org/10.1007/s11367-024-02280-3

Graf, C., Flanagan, D., Wylie, L. and **Silver, D.** (2020) 'The open data challenge: an analysis of 124,000 data availability statements and an ironic lesson about data management plans', *Data Intelligence,* 2(4), pp. 554–568. Available at: https://doi.org/10.1162/dint_a_00061

Harris, J.K., Johnson, K.J., Carothers, B.J., Combs, T.B., Luke, D.A. and **Wang, X.** (2018) 'Use of reproducible research practices in public health: a survey of public health analysts', *PLOS ONE,* 13(9), p. e0202447. Available at: https://doi.org/10.1371/journal.pone.0202447

Harrison, P.W., Lopez, R., Rahman, N., Allen, S.G., Aslam, R., Buso, N., Cummins, C., Fathy, Y., Felix, E., Glont, M., Jayathilaka, S., Kadam, S., Kumar, M., Lauer, K.B., Malhotra, G., Mosaku, A., Edbali, O., Park, Y.M., Parton, A. and **Apweiler, R.** (2021) 'The COVID-19 data portal: accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing', *Nucleic Acids Research,* 49(W1), pp. W619–W623. Available at: https://doi.org/10.1093/nar/gkab417

Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R. and **Goudie, S.** (2020) 'Developing a research data policy framework for all journals and publishers', *Data Science Journal,* 19, p. 5. Available at: https://doi.org/10.5334/dsj-2020-005

Jenkins, T., Persaud, B.D., Cowger, W., Szigeti, K., Roche, D.G., Clary, E., Slowinski, S., Lei, B., Abeynayaka, A., Nyadjro, E.S., Maes, T., Thornton Hampton, L., Bergmann, M., Aherne, J., Mason, S.A., Honek, J.F., Rezanezhad, F., Lusher, A.L., Booth, A.M. and **Van Cappellen, P.** (2022) 'Current state of microplastic pollution research data: trends in availability and sources of open data', *Frontiers in Environmental Science,* 10, p. 912107. Available at: https://doi.org/10.3389/fenvs.2022.912107

Jiao, C., Li, K. and **Fang, Z.** (2024) 'Data sharing practices across knowledge domains: a dynamic examination of data availability statements in PLOS ONE publications', *Journal of Information Science, 50*(3), pp. 673–689. Available at: https://doi.org/10.1177/01655515221101830

Kamath, P., Gudi, N., Staunton, C., Jacob, A.G. and **John, O.** (2023) 'Individual participant data sharing intentions and practices during the coronavirus disease-2019 pandemic: a rapid review', *Data & Policy,* 5, p. e29. Available at: https://doi.org/10.1017/dap.2023.26

Kim, L., Lavrinienko, A., Sebechlebska, Z., Stoltenberg, S. and **Bokulich, N.A.** (2025) 'Tier-based standards for FAIR sequence data and metadata sharing in microbiome research', *Nucleic Acids Research,* 53(15), p. gkaf777. Available at: https://doi.org/10.1093/nar/gkaf777

Klebel, T., Traag, V., Grypari, I., Stoy, L. and **Ross-Hellauer, T.** (2025) 'The academic impact of open science: a scoping review', *Royal Society Open Science,* 12(3), p. 241248. Available at: https://doi.org/10.1098/rsos.241248

Klungthanaboon, W. (2026) 'To share or not to share research data: perspectives and practices of researchers in the digital humanities community in Thailand', *Journal of Librarianship and Information Science,* 58(1), pp. 525–543. Available at: https://doi.org/10.1177/09610006251319029

Kratz, J.E. and **Strasser, C.** (2015) 'Researcher perspectives on publication and peer review of data', *PLOS ONE,* 10(2), p. e0117619. Available at: https://doi.org/10.1371/journal.pone.0117619

Lenert, L.A., Ilatovskiy, A.V., Agnew, J., Rudisill, P., Jacobs, J., Weatherston, D. and **Deans, K.R.** (2021) 'Automated production of research data marts from a canonical fast healthcare interoperability resource data repository: applications to COVID-19 research', *Journal of the American Medical Informatics Association,* 28(8), pp. 1605–1611. Available at: https://doi.org/10.1093/jamia/ocab108

Lucas-Dominguez, R., Alonso-Arroyo, A., Vidal-Infer, A. and **Aleixandre-Benavent, R.** (2021) 'The sharing of research data facing the COVID-19 pandemic', *Scientometrics,* 126(6), pp. 4975–4990. Available at: https://doi.org/10.1007/s11192-021-03971-6

Ministerio de Ciencia e Innovación. (2023) *Estrategia Nacional de Ciencia Abierta 2023–2027*. Available at: https://www.ciencia.gob.es/Estrategias-y-Planes/Estrategias/ENCA.html

McGuinness, L.A. and **Sheppard, A.L.** (2021) 'A descriptive analysis of the data availability statements accompanying medRxiv preprints and a comparison with their published counterparts', *PLOS ONE,* 16(5), p. e0250887. Available at: https://doi.org/10.1371/journal.pone.0250887

Montague-Hellen, B. and **Montague-Hellen, K.** (2023) 'Publishers, funders and institutions: who is supporting UKRI-funded researchers to share data?', *Insights: The UKSG Journal,* 36, 4. Available at: https://doi.org/10.1629/uksg.602

Ollé, C., López-Borrull, A., Melero, R., Boté-Vericad, J.-J., Rodríguez-Gairín, J.-M. and **Abadal, E.** (2023) 'Habits and perceptions regarding open science by researchers from Spanish institutions', *PLOS ONE,* 18(7), p. e0288313. Available at: https://doi.org/10.1371/journal.pone.0288313

Parks, N.A., Fischer, T.G., Blankenburg, C., Scalfani, V.F., McEwen, L.R., Herres-Pawlis, S. and **Neumann, S.** (2023) 'The current landscape of author guidelines in chemistry through the lens of research data sharing', *Pure and Applied Chemistry,* 95(4), pp. 439–450. Available at: https://doi.org/10.1515/pac-2022-1001

Qiu, Y. and **Hu, Z.** (2025) 'Data governance and open sharing in the fields of life sciences and medicine: a bibliometric analysis', *Digital Health,* 11. Available at: https://doi.org/10.1177/20552076251320302

Samuel, S. and **König-Ries, B.** (2021) 'Understanding experiments and research practices for reproducibility: an exploratory study', *PeerJ,* 9, p. e11140. Available at: https://doi.org/10.7717/peerj.11140

Sofi-Mahmudi, A., Raittio, E., Khazaei, Y., Ashraf, J., Schwendicke, F., Uribe, S.E. and **Moher, D.** (2024) 'COVID-19-related research data availability and quality according to the FAIR principles: a meta-research study', *PLOS ONE,* 19(11), p. e0313991. Available at: https://doi.org/10.1371/journal.pone.0313991

Starnini, M., Aleta, A., Tizzoni, M. and **Moreno, Y.** (2021) 'Impact of data accuracy on the evaluation of COVID-19 mitigation policies', *Data & Policy,* 3, p. e28. Available at: https://doi.org/10.1017/dap.2021.25

Strcic, J., Civljak, A., Glozinic, T., Pacheco, R.L., Brkovic, T. and **Puljak, L.** (2022) 'Open data and data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv', *Scientometrics,* 127(5), pp. 2791–2802. Available at: https://doi.org/10.1007/s11192-022-04346-1

Tai, K.H., Müller, M., Mansmann, U., Vieira Armond, A.C., Decullier, E., Le Louarn, A., Munung, N.S., Naudet, F., Prasser, F. and **Sax, U.** (2025) 'FAIRification of biomedical research data', *Journal of Clinical Epidemiology,* 187, p. 111920. Available at: https://doi.org/10.1016/j.jclinepi.2025.111920

Tan, A.C., Webster, A.C., Libesman, S., Yang, Z., Chand, R.R., Liu, W., Palacios, T., Hunter, K.E. and **Seidler, A.L.** (2024) 'Data sharing policies across health research globally: cross-sectional meta-research study', *Research Synthesis Methods,* 15(6), pp. 1060–1071. Available at: https://doi.org/10.1002/jrsm.1757

Umbach, G. (2024) 'Open science and the impact of open access, open data, and FAIR publishing principles on data-driven academic research', *Statistical Journal of the IAOS,* 40(1), pp. 59–70. Available at: https://doi.org/10.3233/SJI-240021

Wellcome Trust. (2020, January 31) *Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak*. Available at: https://wellcome.org/node/1912

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R. and **Mons, B.** (2016) 'The FAIR guiding principles for scientific data management and stewardship', *Scientific Data,* 3, p. 160018. Available at: https://doi.org/10.1038/sdata.2016.18

Williams, R., Bolton, T., Jenkins, D., Mizani, M.A., Sperrin, M., Sudlow, C., Wood, A., Heald, A., Peek, N. and **CVD-COVID-UK/COVID-IMPACT Consortium.** (2025) 'The challenges of replication: a worked example of methods reproducibility using electronic health record data', *PLOS ONE,* 20(7), p. e0326335. Available at: https://doi.org/10.1371/journal.pone.0326335

Xu, B., Kraemer, M.U.G., Gutierrez, B., Mekaru, S., Sewalk, K., Loskill, A., Wang, L., Cohn, E., Hill, S., Zarebski, A., Li, S., Wu, C.-H., Hulland, E., Morgan, J., Scarpino, S., Brownstein, J., Pybus, O., Pigott, D. and **Kraemer, M.** (2020) 'Open access epidemiological data from the COVID-19 outbreak', *The Lancet Infectious Diseases,* 20(5), p. 534. Available at: https://doi.org/10.1016/S1473-3099(20)30119-5

Yehudi, Y., Hughes-Noehrer, L., Goble, C. and **Jay, C.** (2025) 'COVID-19: an exploration of consecutive systemic barriers to pathogen-related data sharing during a pandemic', *Data & Policy,* 7, p. e4. Available at: https://doi.org/10.1017/dap.2024.79

Zuiderwijk, A. (2024) 'Researchers' willingness and ability to openly share their research data: a survey of COVID-19 pandemic-related factors', *SAGE Open,* 14(1). Available at: https://doi.org/10.1177/21582440241234985

TO CITE THIS ARTICLE:
Cerda-Cosme, R. and Mendez, E 2026 From Declarative to Machine-Actionable Openness: Analysis of Research Data Sharing in COVID-19 Publications with Spanish-Affiliated Authors. *Data Science Journal*, 25: 42, pp. 1–18. DOI: https://doi.org/10.5334/dsj-2026-042

Submitted: 19 March 2026
Accepted: 04 November 2026
Published: 23 November 2026

COPYRIGHT:
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0/.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.