

Review

Mapping of Data-Sharing Repositories for Paediatric Clinical Research—A Rapid Review

Mariagrazia Felisi ¹, Fedele Bonifazi ², Maddalena Toma ², Claudia Pansieri ^{1,*}, Rebecca Leary ³, Victoria Hedley ³, Ronald Cornet ^{4,5}, Giorgio Reggiardo ¹, Annalisa Landi ², Annunziata D'Ercole ², Salma Malik ⁶, Sinéad Nally ⁷, Anando Sen ³, Avril Palmeri ³, Donato Bonifazi ¹ and Adriana Ceci ²

- ¹ Consorzio per Valutazioni Biologiche e Farmacologiche (CVBF), Via Luigi Porta, 14, 27100 Pavia, Italy
 - ² Fondazione per la Ricerca Farmacologica Gianni Benzi Onlus, Via Giulio Petroni 91/D, 70124 Bari, Italy
 - ³ John Walton Muscular Dystrophy Research Centre, Translational and Clinical Research Institute, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE1 3BZ, UK
 - ⁴ Department of Medical Informatics, Amsterdam Public Health Institute, Amsterdam UMC (Academic Medical Center)—University of Amsterdam, Medical Informatics, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands
 - ⁵ Amsterdam Public Health, Methodology & Digital Health, 1081 HV Amsterdam, The Netherlands
 - ⁶ The European Clinical Research Infrastructure Network (ECRIN), 30 Bd Saint-Jacques, 75014 Paris, France
 - ⁷ Novartis Pharmaceuticals, 203 Merrion Rd, Dublin 4, D04 NN12 Dublin, Ireland
- * Correspondence: cpansieri@cvbf.net

Abstract: The reuse of paediatric individual patient data (IPD) from clinical trials (CTs) is essential to overcome specific ethical, regulatory, methodological, and economic issues that hinder the progress of paediatric research. Sharing data through repositories enables the aggregation and dissemination of clinical information, fosters collaboration between researchers, and promotes transparency. This work aims to identify and describe existing data-sharing repositories (DSRs) developed to store, share, and reuse paediatric IPD from CTs. A rapid review of platforms providing access to electronic DSRs was conducted. A two-stage process was used to characterize DSRs: a first step of identification, followed by a second step of analysis using a set of eight purpose-built indicators. From an initial set of forty-five publicly available DSRs, twenty-one DSRs were identified as meeting the eligibility criteria. Only two DSRs were found to be totally focused on the paediatric population. Despite an increased awareness of the importance of data sharing, the results of this study show that paediatrics remains an area in which targeted efforts are still needed. Promoting initiatives to raise awareness of these DSRs and creating ad hoc measures and common standards for the sharing of paediatric CT data could help to bridge this gap in paediatric research.

Keywords: data sharing; paediatric; clinical trials; repository



Citation: Felisi, M.; Bonifazi, F.; Toma, M.; Pansieri, C.; Leary, R.; Hedley, V.; Cornet, R.; Reggiardo, G.; Landi, A.; D'Ercole, A.; et al. Mapping of Data-Sharing Repositories for Paediatric Clinical Research—A Rapid Review. *Data* **2024**, *9*, 59. <https://doi.org/10.3390/data9040059>

Academic Editor: Rüdiger Pryss

Received: 18 December 2023

Revised: 13 March 2024

Accepted: 17 April 2024

Published: 20 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Conducting clinical trials (CTs) in the paediatric population can be challenging due to specific ethical, regulatory, methodological, and economic issues [1–3]. Difficulties, such as the well-documented health equity issues in paediatric clinical research, can arise from the earliest stages of patient recruitment. For example, inequitable access to clinical care, institutional barriers and equity issues in the consent process, and barriers to participation related to study procedures hinder the work of research teams worldwide [4]. Even the final stages of research (i.e., publication/sharing of results) are not without complications; the long time lag between the completion of a clinical trial and the publication of results in a peer-reviewed article can in many cases limit public awareness of research [5]. Over the past 20 years, the need for greater transparency of ongoing and completed CTs has been in the spotlight [6–11]. From an institutional point of view, since the 2000s, there have been several regulatory initiatives aimed at promoting paediatric CTs. In the US, the Pediatric Research

Equity Act (PREA) expanded upon the previously enacted Food and Drug Administration Modernization Act (FDAMA), which provided incentives for companies to test drugs in paediatric populations [12]. Shortly thereafter, the Best Pharmaceuticals for Children Act (BPCA) was enacted to encourage research to address gaps in paediatric therapeutic knowledge and to sponsor investment by pharmaceutical companies in clinical trials for off-patent drugs that need further study in children [13]. In the EU, the introduction of the Paediatric Regulation on 26 January 2007 marked a significant milestone in the regulation of medicines. This regulatory framework is based on the need to ensure the quality, ethical research, and appropriate authorization of medicines intended for use in children, with the broader goal of narrowing the gap between adult and paediatric treatments [14]. An extensive revision of the Paediatric Regulation is currently ongoing [15]. The Clinical Trials Regulation (EU) No. 536/2014, implemented in the European Union (EU) in 2022, introduces the concept of “low-intervention clinical trials” by proposing a “risk-proportional approach” to ensure a better balance between patient safety and data quality, by optimizing time and cost [16].

Despite these regulatory efforts, the lack of evidence-based research and unmet medical needs, coupled with poor investment from pharma companies in paediatric research, has resulted in a drive towards the reuse of existing data [17].

The research community has increasingly been committed to enhancing public health through responsible sharing of CT data that respects both data subjects’ privacy and intellectual property rights [16,18,19].

Several advantages have been attributed to sharing data from CTs. Representative examples include a decrease in the number of patients needed for new clinical trials thanks to the reuse of existing data, increased transparency of research results [10,20,21], and increased public trust in CTs [22,23]. To share reliable data, standardization and harmonization are essential steps in data production. The global scientific community is becoming more aware of the importance of data principles. For example, the Findable, Accessible, Interoperable, and Reusable (FAIR) data principles describe how research outputs should be organized, so they can be more easily accessed, understood, exchanged, and reused [24].

CT data can be shared through dedicated information systems called data-sharing repositories (DSRs). DSRs provide access to digital content by storing, organizing, and preserving data [25]. They are hosted on platforms (computer environments) that facilitate the interaction between data owners and data users (e.g., researchers). They are a central “safe point” where researchers can search for available data from CTs shared by public or private sponsors [26]. Two major types of data can be accessed and shared: individual patient data (IPD) and study-level data (summary data). IPD refers to information on individual patients collected during a CT (e.g., demographic data, lab results). This type of information is recorded on case report forms (CRFs). Study-level data consist of patient-level data that have been aggregated, tabulated, stratified, or otherwise organized to be used for interpreting the outcome of a clinical study.

The availability of raw data in open DSRs consisting of IPD from CTs in an electronic structured format facilitates the adequate development of IPD-meta-analyses that are placed at the highest level of the evidence pyramid. It facilitates the cumulative evaluation of evidence for specific topics, especially for high-dimensional data (such as results from genomics, transcriptomics, or epigenomics) [27].

This work is conducted in the framework of the Collaboration Network for European Clinical Trials for Children (conect4children or c4c, website: <https://conect4children.org/>, accessed on 19 April 2024) [2]. The work package “Data coordinating centre and data quality standards” within c4c attempts to create systems, tools, and standards to enhance quality, utility, reusability, and uniformity of the data collected during paediatric clinical trials. The c4c project is a collaborative initiative funded by the Innovative Medicines Initiative (IMI), a public–private partnership between the European Union (EU) and the Eu-

ropean pharmaceutical industry, represented by the European Federation of Pharmaceutical Industries and Associations (EFPIA).

The aim of this study was to describe and analyse through eight indicators existing data-sharing repositories (DSRs) that share individual patient data (IPD) from clinical trials, with a particular focus on paediatric IPD. This evaluation aimed to address barriers to data sharing, such as concerns about handling sensitive data among data managers and sponsors, and the lack of standardized policies among clinical trial funders, which can hinder the adoption of data-sharing practices [28].

This paper represents a pioneering effort to map all repositories where paediatric clinical trial data can be accessed by paediatric stakeholders. Given the well-documented challenges of conducting clinical trials in this population, it is imperative to maximize the use of data already collected. Through this effort, we have now established a clear roadmap for access to clinical trial data that will significantly benefit the paediatric research community.

2. Methods

A rapid review of existing CT DSRs was conducted through the literature database PubMed and the search engine Google in 2020 and updated in September 2023.

The PubMed search strategy consisted of five main steps: (1) identification of research questions (Supplementary Materials); (2) identification of keywords that best answer the research questions; (3) creation of search strings and application of the “within the last 10 years” and “English” search filters when using strings (Table S1); (4) selection and screening of articles, reporting keywords in the title and/or abstract, performed by two authors; (5) extrapolation, screening, and analysis of DSRs according to the criteria described below.

The keywords identified in step (2) were also used in the Google search, and the first hundred results were screened. Details of the search strategy are provided in Figure S1.

The following definition was used: “Data-sharing repository (DSR)”, an information system set up to manage, archive, and provide access (share) to datasets from CTs. To be included in the analysis, DSRs had to provide access to IPD, or at least provide clear instructions for submitting an access request. IPD data had to be publicly available in English. At least some of the data held in the DSR should be from paediatric-only trials and/or from trials on paediatric and adult patients from both industry-sponsored and academic-led trials. DSRs that did not meet the eligibility criteria or were no longer active were excluded.

DSRs that met the eligibility criteria were analysed and compared using a two-step process.

Firstly, publicly available information was searched on the DSR’s website and tabulated. The following information was extracted: general features; type of data collected and related documentation; specific guidance on data composition/structure/format, relevance, and paediatric specificity (i.e., availability of paediatric CT data filtered by age group); legal provisions for uploading and reusing data; IT security measures and protocols. Some of these features, such as relevance and paediatric specificity, access to IPD, and data privacy measures, were classified as shown below:

- Relevance was assessed as the ability of each DSR to provide access to the IPD of the subjects included in the CTs.
- Paediatric specificity was evaluated based on the ability to filter by a generic age group (e.g., 0–6, 6–12, 6–18) or through the availability of specific keywords (e.g., paediatric, neonate).
- Access to IPD was evaluated on the basis of the type of access. Three different access types were identified:
 - Direct sharing: Data are provided after a data-sharing agreement outlining the rules for data utilization has been signed by the user. No other action is required.

- Controlled access: Access is permitted after the user submits a formal application requesting data access. The data requestor may need to provide a research protocol and analysis plan, including information on data management and plans for publication of results. The data can only be accessed and analysed within the DSR’s workspace and not on the user’s computer. A data-sharing agreement must be signed by the user.
 - Open access: There is no formal process to access data. Researchers may explore but not download the data without a specific request.
 - Different de-identification measures are adopted to protect the privacy of data subjects, e.g., pseudonymization or anonymization. An adequate de-identification (or encryption) measure is key to protecting study participants from reidentification. In the de-identification process, the participant’s identifiable information is removed or replaced with a code, usually a random code number.
 - Pseudonymization processes personal data in such a way that it can no longer be attributed to a specific data subject without the use of additional information (e.g., a specifically created confidential key). Anonymization is a process that destroys any link to an identified or identifiable person via a pseudonym.
- Details on the information collected are presented in Table 1.

Table 1. Data collected.

| General features |
|--------------------------------------------------------------------------------------------------------------|
| ● Geographical location of the DSR |
| ● Funding: private or public or public/private sponsor |
| ● Funding: regular/sustained funding and business continuity measures |
| ● Source of CT data (i.e., non-commercial, commercial sources, and both sources) |
| ● Therapeutic area/s currently covered |
| ● Availability of instructions for data owners/data submitters |
| ● Availability of instructions for prospective data users |
| Data and document features |
| ● Data “in” standards (e.g., CDISC) |
| ● Availability of guidance on data composition/structure/format |
| ● Document content (annotated CRF, study protocol, statistical analysis plan, clinical study report) |
| ● Possibility to download data and analyse outside of the DSR |
| ● Standard of data downloaded |
| ● Measures to protect data privacy (anonymization, pseudonymization, encryption, aggregation, other) |
| Paediatric specificity |
| ● Availability of paediatric CT data |
| ● Possibility to filter information per specific paediatric age-cohort/generic age groups |
| Legal provisions |
| ● Data access type (direct sharing, controlled access, open access, as defined in Grant Agreement n. 777389) |
| ● Data accessibility restriction |
| ● Data protection policy available |
| ● Existing informed consent form allowing data storage and reuse |
| ITs |
| ● Security protocols |

Secondly, the DSRs were analysed through a set of purpose-built indicators: (1) relevance and paediatric specificity; (2) instructions for data owners/submitters; (3) instructions for prospective data users; (4) guidance on data composition/structure/format for data owners/submitters; (5) data protection; (6) procedures for patient-level data access; (7) IT security measures/protocols; (8) organizational and financial sustainability of the DSR.

These eight indicators were identified from the questionnaire (34 items) proposed by the CORBEL and IMPACT Observatory projects [29]. The indicators provide a general characterization of the DSRs and include aspects used to analyse them. The eight indicators were selected for relevancy and measurability. The validation process of the indicators provides internal consistency (Cronbach's alpha = 0.768, calculated from the pairwise correlations between items). The factorial analysis indicated a structure of these eight principal components that explain 58.7% of the total variance explained by the CORBEL and IMPACT questionnaire. Cronbach's alpha is a function of the number of test items and the average inter-correlation among them, and it was calculated in SPSS Statistics (IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY, USA: IBM Corp.) using the Reliability Analysis feature [29].

Two authors independently rated each DSR by classifying each indicator with a score from 0 to 2. Details about the classification system adopted by the authors are reported in Table 2.

This classification does not assess the quality of the DSR but evaluates its performance against the set of eight indicators. All cases of uncertainty, discrepancy, or missing data were resolved through discussion, searches for additional data sources, and consensus. Disagreements were resolved by consensus building with two other authors. To determine the degree of concordance between authors, we used Cohen's kappa approach [30] and an assessment of the DSRs was conducted through a performance score cluster analysis [31,32]. More specifically, the cluster analysis was performed to identify DSRs that fully meet the evaluation criteria based on the eight purpose-built indicators (total score descriptive analysis). By comparing the indicators' total score values of a model choice criterion across different clustering solutions, the procedure automatically determined the optimal number of clusters using Schwarz's Bayesian information criterion (SBIC) [33]. The likelihood distance measure assumes that variables in the cluster model are independent. Further, each categorical variable is assumed to have a multinomial distribution. Empirical internal testing indicates that the procedure is fairly robust to violations of both the assumption of independence and the distributional assumptions.

Patient and public involvement: No patients involved.

Table 2. Evaluation of the eight indicators.

| Score | 1. Relevance and Paediatric Specificity | 2. Instructions for Data Owners/Data Submitters | 3. Instructions for Prospective Data Users | 4. Guidance on Data Composition/Structure/Form for Data Owners/Submitters | 5. Data Protection | 6. Procedures for Patient-Level Data Access | 7. It Security Measures/Protocols | 8. Sustainability |
|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| 0 | If the DSR currently contains paediatric CT data but there is no chance * to filter and download/access only paediatric data. | If no instructions are available * for data owners/submitters to advise on what data are in scope and how to submit their data. | If no clear instructions for prospective data users are publicly available. * | If the DSR has no publicly available guidance or recommendations for data owners/submitters on specific models, standards, or formats for data OR for metadata. * | If there is no public mention of data protection. * | If measures or procedures to access data are not reported. * | If no security protocol or safety measures are publicly mentioned on the website. * | If funding is not currently available and the sustainability of the DSR appears uncertain. * |
| 1 | If the DSR currently contains paediatric CT data but limited filtering options are available: (e.g., filters for generic age groups only, such as paediatric, adults, geriatric). * | If only basic 'how-to upload' instructions are publicly available (and/or if detailed instructions on what is in scope and how data should be submitted DO exist but are not publicly available). * | If only basic instructions on 'how to access/download/analyse' data are available publicly, or if detailed instructions are available but are not publicly visible. * | If some guidance or recommendations are provided for data owners/submitters concerning specific models, standards, or formats for data OR for metadata but this guidance is not exhaustive. * | If data protection measures are mentioned generally, but no data protection policy is specified or made publicly available. * | If measures or procedures to access individual patient data are generally mentioned but not clearly explained. * | If the DSR makes available on the website only a summary protocol for regularly testing, assessing, and evaluating the effectiveness of technical and organizational measures to ensure the security of the processing is in place. * | If the DSR has no regular/sustained funding but has business continuity measures in place. * |

Table 2. Cont.

| Score | 1. Relevance and Paediatric Specificity | 2. Instructions for Data Owners/Data Submitters | 3. Instructions for Prospective Data Users | 4. Guidance on Data Composition/Structure/Form for Data Owners/Submitters | 5. Data Protection | 6. Procedures for Patient-Level Data Access | 7. It Security Measures/Protocols | 8. Sustainability |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| 2 | If the DSR currently contains paediatric CT data which can easily be filtered by specific paediatric age groups (e.g., below 2, between 5 and 10 years, etc). | If the DSR provides clear instructions for data owners/submitters on what data are in scope and how to submit their data (including information on any specific formats or schemas requested) and makes these instructions publicly available. | If the DSR has—and makes publicly available—clear instructions for prospective data users on how to access and/or analyse data, including. | If the DSR provides clear guidance or recommendations for data owners/submitters on specific models, standards, or formats for data OR for metadata. | If a data protection policy for the DSR is publicly available . | If procedures and materials relating to individual patient data access agreements are clearly presented and/or a data access agreement template is proposed for adoption. | If the DSR has a security protocol for regularly testing , assessing, and evaluating the effectiveness of technical and organizational measures to ensure the security of the processing in place and the protocol is available on the website. | If the DSR receives regular/sustained funding and can demonstrate business continuity measures. |

* We were not able to find, or it was impossible for us to demonstrate their presence at the time of our evaluation.

3. Results

3.1. Literature Search

The literature search identified a total of 773 articles via PubMed ($n = 743$) or other sources (Google and co-author suggestions, $n = 26$) selected by checking whether the title and abstract contained mentions of DSRs. Articles containing DSRs that did not give access to IPD, articles citing the same DSR, and articles citing no longer active DSRs were excluded. A total of 31 articles were identified as eligible for analysis ($n = 22$ from PubMed; $n = 6$ from Google; $n = 3$ suggested by co-authors) (Figure S1).

3.2. DSRs Selection

From these 31 articles, 45 publicly accessible DSRs that potentially included paediatric CT data were identified (Figure S1). Sixteen were identified in PubMed, and twenty-seven through other sources mentioned above. A preliminary screening phase was carried out to check the eligibility of each DSR. Nineteen DSRs were excluded as they did not host a real DSR or did not give access to IPD. Two DSRs overlapped with another DSR, and one no longer existed. At the end of the screening phase, twenty-one DSRs met the identified eligibility criteria and were included in the analysis (Figure 1).

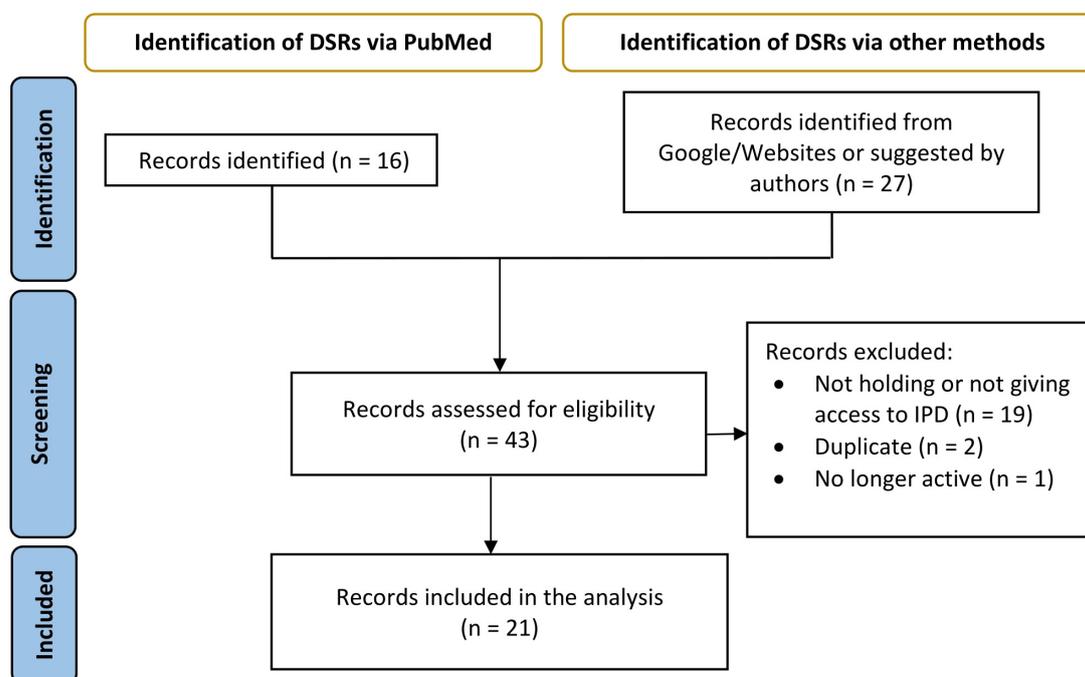


Figure 1. Flow diagram for the identification of DSRs.

Most of the DSRs identified were found through PubMed searches (sixteen out of twenty-one). Two DSRs were identified through Google searches, and three were suggested by authors with expertise in the field (Table S2).

The BioCelerate DSR only provides access to detailed information about the DSR to members of associated companies, so an in-depth analysis of search options was not possible. Nevertheless, it was agreed not to exclude it from the analysis as it represents a possible data source.

3.3. Data-Sharing Repositories' Characteristics

The overall characteristics of the DSRs as well as the URLs of the associated webpages are reported in Table 3.

Table 3. Main characteristics of the included DSRs.

| Name | Year of Establishment | Location | Funding Source | Data Access Type (Direct Sharing, Controlled Access, Open Access) | URL |
|---------------------------------------------------|-----------------------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| BioCelerate (TransCelerate) | 2018 | US | Private: TransCelerate BioPharma (not-for-profit) | Controlled access | https://www.transceleratebiopharmainc.com/initiatives/datacelerate/ (accessed on 3 October 2023) |
| BioLINCC | 2000 | US | Public: US Government's National Institutes of Health—National Heart, Lung, and Blood Institute (NHLBI) | Controlled access | https://biolincc.nhlbi.nih.gov/home/ (accessed on 3 October 2023) |
| Clinical Study Data Request (CSDR) | 2016 | US | Private: Astellas, Bayer, Bill and Melinda Gates, Boehringer Ingelheim, Cancer Research UK, Chugai/Roche, Daiichi-Sankyo, Eisai, gsk, Lilly, etc. | Controlled access | https://www.clinicalstudydatarequest.com (accessed on 3 October 2023) |
| Dryad Digital DSR | 2008 | US | Public sponsor: National Science Foundation to the National Evolutionary Synthesis Center and other partners in the US | Controlled access | https://datadryad.org/ (accessed on 3 October 2023) |
| European Genome-phenome Archive (EGA) | 2008 | EU | Public: ELIXIR infrastructure | Direct sharing | https://ega-archive.org/ (accessed on 3 October 2023) |
| Health Data Research Innovation Gateway | 2020 | EU | Public: UK Research and Innovation's (UKRI) Industrial Strategy Challenge Fund | Direct sharing | https://www.healthdatagateway.org/about/our-mission-and-purpose (accessed on 3 October 2023) |
| Infectious Diseases Data Observatory (IDDO) | 2009 | EU | Public sponsor: Oxford University | Controlled access | https://www.iddo.org/about-us/about-iddo (accessed on 3 October 2023) |
| Immunology Database and Analysis Portal (ImmPort) | 2017 | US | Public: National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Health and Human Services (HHS) | Open access | https://www.immport.org/shared/ (accessed on 3 October 2023) |

Table 3. Cont.

| Name | Year of Establishment | Location | Funding Source | Data Access Type (Direct Sharing, Controlled Access, Open Access) | URL |
|--------------------------------------------------------------------------------------|-----------------------|----------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Immune Tolerance Network (ITN) TrialShare | 2017 | US | Public: NIAID, NIH | Controlled Access | Sign In: /home (itntrialshare.org) (accessed on 3 October 2023) |
| Laboratory of Neuroimaging Image & Data Archive (IDA) | 2003 | US | Public: National Institutes of Health, National Institute of Biomedical Imaging and Bioengineering | Controlled Access | https://ida.loni.usc.edu/login.jsp (accessed on 3 October 2023) |
| National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Central DSR | 2003 | US | Public | Direct sharing | https://DSR.niddk.nih.gov/home/ (accessed on 3 October 2023) |
| National Institute on Drug Abuse (NIDA) | 2014 | US | Public: National Institutes of Health (NIH) | Direct sharing | https://datashare.nida.nih.gov/ (accessed on 3 October 2023) |
| National Sleep Research Resource (NSRR) | 2014 | US | Public: National Heart, Lung, and Blood Institute (NHLBI) | Controlled access | https://sleepdata.org/ (accessed on 3 October 2023) |
| Pediatric Cancer Data Commons (PCDC) | 2015 | US | Public: Collaborative grant from the University of Chicago, philanthropic support, and contracts from The National Cancer Institute (NCI) | Controlled access | https://portal.pedscommons.org/ (accessed on 3 October 2023) |
| Paediatric Trials Network (PTN)/DASH (data and specimen hub) | 2010 | US | Public: Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) | Direct sharing | https://www.pediatrictrials.org/ (accessed on 3 October 2023) |
| Project Data Sphere (PDS) | 2014 | US | Public: CEO Roundtable on Cancer, Inc. (non-profit organization) | Controlled access | https://www.projectdatasphere.org/projectdatasphere/html/home (accessed on 3 October 2023) |
| Rare Disease Cures Accelerator–Data Analytical Platform (RDCA-DAP) | 2021 | US | Collaborative grant from the FDA | Controlled access | https://c-path.org/programs/rdca-dap/ (accessed on 3 October 2023) |

Table 3. Cont.

| Name | Year of Establishment | Location | Funding Source | Data Access Type (Direct Sharing, Controlled Access, Open Access) | URL |
|-------------------------------------------------------------|-----------------------|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| The National Cancer Institute's Genomic Data Commons (GDC) | 2016 | US | Public | Controlled access | https://gdc.cancer.gov/ (accessed on 3 October 2023) |
| The National Institute of Mental Health (NIMH) Data Archive | 2014 | US | Public: National Institute of Mental Health (NIMH); National Institute of Neurological Disorders and Stroke (NINDS); National Institute of Environmental Health Sciences (NIEHS); The Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) | Controlled access | https://nda.nih.gov/ (accessed on 3 October 2023) |
| Vivli | 2017 | US | Private non-profit: Doris Duke charitable Foundation; Lyda Hill Foundation; The Leona M. and Harry B. Helmsley charitable trust; Laura and John Arnold Foundation Private profit: PhRMA | Direct sharing | http://vivli.org/ (accessed on 3 October 2023) |
| YODA project | 2014 | US | Private: Johnson & Johnson Medtronic, Inc.; Queen Mary University of London; SI-BONE, Inc. | Direct sharing | http://yoda.yale.edu/ (accessed on 3 October 2023) |

Most DSRs are located in the US ($n = 18$) and 3 in Europe. The Biologic Specimen and DSR Information Coordinating Center (BioLINCC) was the first DSR established in 2000, while the most recent is the Rare Disease Cures Accelerator–Data Analytical Platform (RDCA-DAP) founded in the US in 2021. Details about the year of establishment are reported in Table 3. Most of the DSRs cover more than one therapeutic area ($n = 15$), and only a few focus on specific therapeutic areas: toxicology ($n = 3$), neuroscience/neurology, sleep disorders, infectious disease, or cancer ($n = 3$).

DSR performance against the set of eight identified indicators was independently evaluated by two authors. The adopted indicators were: (a) relevance and paediatric specificity; (b) instructions for data owners/data submitters; (c) instructions for prospective data users; (d) guidance on data composition/structure/format for data owners/submitters; (e) data protection; (f) procedures for patient-level data access; (g) IT security measures/protocols; (h) sustainability (Table 2).

The degree of concordance between authors in the evaluation of the eight indicators was good overall. A strong degree of concordance was obtained for the relevance

and paediatric specificity indicator, with Cohen's kappa 0.856, 95% C.I. (0.670–1.042), and for the instructions for prospective data users, with Cohen's kappa 0.878, 95% C.I. (0.666–1.090). Likewise, the procedures for patient-level data access indicators (a), the IT security measures/protocols (b), the guidance on data composition/structure/format for data owners/submitters (c), and the sustainability indicator (d) were found to have a moderate degree of concordance: (a) Cohen's kappa 0.781, 95% C.I. (0.369–1.193); (b) Cohen's kappa 0.635, 95% C.I. (0.374–0.896); (c) Cohen's kappa 0.613, 95% C.I. (0.331–0.895); (d) Cohen's kappa 0.644, 95% C.I. (0.007–1.281). Only the data protection indicator and the instructions for data owners/data submitters were found to have a weak concordance between authors: Cohen's kappa 0.573, 95% C.I. (0.271–0.875), and Cohen's kappa 0.577, 95% C.I. (0.277–0.877), respectively.

3.4. Analysis of the Eight Indicators

Details are reported in Table 4.

1. Relevance and Paediatric Specificity

In twelve out of the twenty-one DSRs, it was possible to search for paediatric CT data due to a filter for a specific or a generic age group (e.g., 0–6; 6–12; 6–18), or through the availability of specific keywords for the paediatric population (e.g., paediatric, neonates), with the exception of Biocelerate due to restricted access to detailed information about the DSR. Only the PTN and PCDC DSRs are completely dedicated to the paediatric population. Notably, PTN does not host its own DSR but shares data through the Data and Specimen Hub DSR (DASH). Five DSRs contain paediatric CT data that can be filtered by specific paediatric age groups (e.g., less than 2, between 5 and 10 years, etc.). Seven DSRs provided limited filtering options (e.g., filters for generic age groups) at the time of our evaluation, and in nine DSRs we were not able to filter, download, or easily access exclusively paediatric data.

2. Instructions for data owners/data submitters

Thirteen DSRs provide clear, easily understandable instructions for data owners/submitters on which data are in scope and how to submit data, including information on any specific formats or requested schemas. Two DSRs provide only basic, minimal, and non-exhaustive instructions about 'how to upload', or these do exist but were not publicly available at the time of our review. In six DSRs, we were not able to find instructions for data owners/submitters to advise what data are in scope and how to submit data.

3. Instructions for prospective data users

Seventeen DSRs provide clear, easily understandable instructions for prospective data users on how to access and/or analyse data. In four DSRs, we were not able to find clear instructions for prospective data users, or only basic, minimal, and non-exhaustive instructions are publicly available.

4. Guidance on data composition/structure/format for data owners/submitters

Seven DSRs provide clear, easily understandable guidance or recommendations for data owners/submitters on specific models, standards, or formats for data or metadata that can be hosted in the DSR. The most common types of file formats for the data download are SAS and CSV. Three DSRs provide basic, non-exhaustive guidance or recommendations. In eleven DSRs we were not able to find any guidance or recommendations freely available within the DSRs.

5. Data Protection

Three DSRs clearly reported a data protection policy, providing on their webpage information about measures to protect data privacy through de-identification and anonymization (or pseudonymization) processes. At the time of our research, eight DSRs reported only general information about the data protection measures adopted, but no data protection policy was specified or made publicly available. We were not able to easily find this information in ten DSRs.

Table 4. Analysis of the eight indicators.

| DSR | 1. Relevance and Paediatric Specificity | 2. Instructions for Data Owners/Data Submitters | 3. Instructions for Prospective Data Users | 4. Guidance on Data Composition/Structure/Format, for Data Owners/Submitters | 5. Data Protection | 6. Procedures for Patient-Level Data Access | 7. IT Security Measures/Protocols | 8. Sustainability |
|--------------------------------------------------------------------------------------|-----------------------------------------|-------------------------------------------------|--------------------------------------------|------------------------------------------------------------------------------|--------------------|---------------------------------------------|-----------------------------------|-------------------|
| Biocelerate (Transcelerate) | + | ++ | ++ | ++ | + | ++ | + | +++ |
| BioLINCC | + | +++ | + | +++ | + | +++ | + | +++ |
| ClinicalStudy DataRequest (CSDR) | + | + | +++ | + | + | +++ | ++ | +++ |
| Dryad Digital DSR | ++ | +++ | + | +++ | ++ | + | +++ | +++ |
| European Genome-phenome Archive (EGA) | + | +++ | +++ | + | ++ | +++ | +++ | +++ |
| Health Data Research Innovation Gateway | ++ | ++ | +++ | + | + | +++ | + | +++ |
| Infectious Diseases Data Observatory (IDDO) | ++ | +++ | +++ | +++ | + | +++ | + | +++ |
| Immunology Database and Analysis Portal (ImmPort) | +++ | +++ | +++ | ++ | + | +++ | ++ | +++ |
| Immune Tolerance Network (ITN) TrialShare | ++ | +++ | +++ | + | ++ | ++ | + | +++ |
| Laboratory of Neuroimaging Image & Data Archive (IDA) | + | +++ | +++ | + | + | +++ | ++ | +++ |
| National Institute on Drug Abuse (NIDA) | +++ | + | +++ | + | ++ | +++ | ++ | +++ |
| National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Central DSR | + | + | +++ | + | + | + | + | +++ |
| National Sleep Research Resource (NSRR) | + | +++ | + | + | ++ | +++ | ++ | +++ |
| Pediatric Cancer Data Commons (PCDC) | +++ | + | +++ | + | + | +++ | ++ | ++ |

Table 4. Cont.

| DSR | 1. Relevance and Paediatric Specificity | 2. Instructions for Data Owners/Data Submitters | 3. Instructions for Prospective Data Users | 4. Guidance on Data Composition/Structure/Format, for Data Owners/Submitters | 5. Data Protection | 6. Procedures for Patient-Level Data Access | 7. IT Security Measures/Protocols | 8. Sustainability |
|--------------------------------------------------------------------|-----------------------------------------|-------------------------------------------------|--------------------------------------------|------------------------------------------------------------------------------|--------------------|---------------------------------------------|-----------------------------------|-------------------|
| Paediatric Trials Network (PTN)/DASH (data and specimen hub) | +++ | +++ | +++ | +++ | ++ | +++ | + | +++ |
| Project Data Sphere (PDS) | + | + | +++ | + | + | +++ | + | +++ |
| Rare Disease Cures Accelerator–Data Analytical Platform (RDCA-DAP) | +++ | +++ | +++ | +++ | +++ | +++ | +++ | ++ |
| The National Cancer Institute’s Genomic Data Commons (GDC) | + | +++ | +++ | +++ | ++ | +++ | + | +++ |
| The National Institute of Mental Health Data Archive (NDA) | ++ | +++ | +++ | +++ | +++ | +++ | ++ | +++ |
| Vivli | ++ | +++ | +++ | ++ | +++ | +++ | ++ | +++ |
| YODA project | ++ | + | +++ | + | ++ | +++ | ++ | +++ |

Legend

| | | | |
|-------|---|----|-----|
| SCORE | 0 | 1 | 2 |
| Level | + | ++ | +++ |

6. Procedures for Patient-Level Data Access

Seventeen DSRs clearly present procedures and materials relating to IPD access agreements, and/or a data access agreement template is available for adoption. Two DSRs mentioned the procedures that should be adopted, but they were not extensively explained. In two DSRs, we were not able to identify clear, easily understandable measures/procedures to access data.

Access to IPD varies between DSRs:

- Data sharing is adopted in six DSRs.
- The controlled access model is adopted by twelve DSRs.
- Open access is adopted only by one DSR.

7. IT Security Measures/Protocols

Three DSRs had protocols available on their websites for regularly testing, assessing, and evaluating the effectiveness of technical and organizational measures to ensure the security of the processing in place. Nine DSRs reported only a summary protocol. For nine DSRs we were not able to find a security protocol or safety measures publicly available on the website.

8. Sustainability

Nineteen DSRs reported on their website that they receive regular funding or are regularly sustained and can demonstrate business continuity measures. Only two seem to have no regular/sustained funding but have business continuity measures in place. Sixteen DSRs are sustained by a public funding source, including all three European DSRs. Four DSRs are sustained by private funding, and one is based on public-private partnerships.

3.5. Cluster Analysis

A cluster analysis was performed to identify DSRs that fully meet the evaluation criteria based on the eight purpose-built indicators (total score descriptive analysis). The number of clusters to be formed was not specified in advance and was calculated using Schwarz’s Bayesian information criterion (SBIC). The cluster outcome showed two groups in terms of elements evaluated: one cluster consisting of five DSRs that meet our evaluation criteria (cluster centroid mean score = 13.40) reporting a higher performance score and one cluster consisting of the remaining sixteen DSRs (cluster centroid mean score = 8.81) with a lower performance score (Table 5).

Table 5. Total score descriptive analysis.

| Repository | Cluster |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Immunology Database and Analysis Portal (ImmPort) Paediatric Trials Network (PTN)/DASH (data and specimen hub Rare Disease Cures Accelerator – Data Analytical Platform (RDCA-DAP) The National Institute of Mental Health (NIMH) Data Archive Vivli | 1 |
| Biocelerate (Transcelerate) BioLINCC Clinical Study Data Request (CSDR) Dryad Digital Repository European Genome-phenome Archive (EGA) Health Data Resaerch Innovation Gateway IDDO Infectious Diseases Data Observatory ITN TrialShare Laboratory of Neuroimaging Image & Data Archive (IDA) National Sleep Research Resource (NSRR) NIDA NIDDK Central Repository Pediatric Cancer Data Common (PCDC) Project Data Spere (PDS) The National Cancer Institute’s Genomic Data Commons YODA project | 2 |

4. Discussion

The majority of the DSRs analysed in this research were identified through a PubMed search, Google search, or through private contacts. All the DSRs were fully active at the time the research was carried out. Only one DSR, Rapid-19 (<https://www.rapid-19.org/DSR-data> accessed on 3 October 2023) was no longer active, probably because it was built during the SARS-COVID-19 emergency. It was therefore excluded from the analysis.

Eighteen of the twenty-one identified DSRs are located in the US, and three are in Europe, highlighting the lack of eligible DSRs in the rest of the world. The origin of the IPD stored in the identified DSRs was beyond the scope of this review and was not investigated. The US led the evolution of transparency in CTs with the requirement for registration of clinical trials by ICMJE and FDAAA (2004) [34–36]. This was followed by EMA Policy 0070 (2014) [37]. Other relevant initiatives include the PhRMA/EFPIA principles for data sharing (2014) and the IOM Sharing Clinical Trial Data report (2015).

Since 2018, hundreds of ICMJE journals have started to require authors to complete a data-sharing statement describing who, what, when, where, and why IPD will be shared [38].

Different types of DSRs were identified: those specific to study data (e.g., age-specific, disease-specific, or stakeholder-specific) and generic DSRs that collect broader clinical research data. Generic DSRs represent the majority of the twenty-one analysed DSRs.

Most of the DSRs supported research on paediatric IPD. However, only the PTN and the PCDC DSRs were focussed on paediatrics. PTN is sponsored by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NIHCD) and does not have a dedicated DSR. The data available in the PTN DSR are shared through the Data and Specimen Hub (DASH) (<https://dash.nichd.nih.gov/> (accessed on 3 October 2023)), of which PTN is only a subset.

PCDC, sponsored by the University of Chicago, hosts the world's largest harmonized clinical data set for paediatric cancer research and provides data through the unified DSR for researchers, PCDC Data Portal.

Most of the identified DSRs appeared to satisfy the proposed indicators and almost all the DSRs could provide useful data for secondary use. All the DSRs identified allow data to be stored and accessed for free. Access models range from publicly accessible web-based systems with the option to download datasets to different types of request/review mechanisms that may or may not allow data to be downloaded.

Most of the DSRs provide guidance or instructions for both data owners and data users and clear, understandable information about the procedures for patient-level data access. This information is available on the platform's website. The most common method for patient-level data access requires the submission of a research proposal which is reviewed by an independent review panel (IRP). A signed data-sharing agreement is usually required before accessing data. Most of the DSRs have clear, easily understandable measures to protect data privacy and provide an environment with anonymized data upon approval of the request.

DSRs are mostly sustained by public funds, and all the European DSRs are in this category. This is a significant advantage since developing and maintaining useful DSRs for efficient data sharing tends to be expensive. Data from different sources are often collected in different formats, using different protocols and endpoints and must be quality-controlled and standardized before analysis can be performed across studies. The upfront costs of developing community standards and networks of collaboration may be high [2]. This could be impactful when the population is small, as with the paediatric population, where stakeholders are fragmented, and there are a limited number of interested parties. However, once these investments have been made, the time and effort required by potential users is relatively low, and the potential for data to be reused in ways that benefit public health is high, making the investment cost-effective [2].

There is also heterogeneity among the DSRs in terms of data upload, data handling, and DSR access. The lack of commonly acknowledged guidelines on the structure of DSRs

for the sharing of CT data inevitably leads to inconsistency in the data available. This may affect the availability and quality of paediatric data for secondary use and represents a barrier to data availability that could be mitigated by adopting common international standards [39]. This heterogeneity limits the interoperability within DRSs and, consequently, the ability to have a representative homogeneous IPD sample.

Information about IT security measures is only reported on a few websites, despite the importance of making this information publicly available. An adequate and transparent data protection policy and ad hoc IT measures may guarantee better quality of data, prevent data breaches, and increase the confidence of users of the DSR.

4.1. Strengths

To our knowledge, this is the first report addressing the availability of paediatric IPD within DSRs of CT data. Only four studies have previously addressed similar topics: the first was carried out by N. Anthony et al., the second by Ohmann et al., the third by Banzi et al., and the fourth was published by the Clinical Research Data Sharing Alliance (CRDSA). N. Anthony and co-workers have recently analysed the digital impact of published reuse of clinical data in terms of media attention and citation rates on three DSRs (CSDR, YODA, and Vivli). They did not find a substantial difference between reusing data from DSRs and using a sample of equivalent studies published in the same journals [40]. The study by Banzi et al. is not focused on this specific population [26]. On the other hand, Ohmann and colleagues provide an overview of the status and use of the sharing of IPD and make recommendations to address common barriers, such as structuring data and metadata using recognized standards, managing DSR data, and accessing and monitoring data sharing, but do not specifically address the paediatric field [41]. Last but not least, the “Review of Biopharma Sponsor Data Sharing Policies and Protection Methodologies”, recently published by the CRDSA, provides an overview of key policy elements that impact the value of research benefits to end users, such as what data are shared and how data are transformed to protect patient privacy across only three data-sharing platforms: Vivli, CSDR, and YODA [42].

This study also highlights the importance of the work being carried out by c4c (<https://conect4children.org/> Accessed on 17 April 2024) that aims to facilitate the development of new drugs and other therapies for the entire paediatric population through the creation of systems, tools, and standards to enhance the quality, utility, reusability, and uniformity of the data collected during paediatric clinical trials.

4.2. Limitations

None of the twenty-one DSRs are designed specifically for the paediatric population and its specific characteristics. This inevitably impacts the ability to carry out paediatric research due to the huge differences in the paediatric population, ranging from neonates to adolescents. To carry out effective research on a specific cohort of paediatric patients, a dedicated DSR is needed that is tailored to their characteristics (e.g., that allows for the selection of preterm subjects under 1 kg).

The PTN alone does not support this level of specificity.

This study has some potential weaknesses and limitations. Not all the available DSRs have been identified, mainly due to the dynamic nature of the topic and the time needed for a publication. Despite this, we attempted to use the most rigorous and extensive search strategy to identify as many DSRs as possible. The search strategy adopted was not intended to be systematic, but it can be considered the most appropriate to provide a descriptive overview of the available DSRs. Since we included only DSRs with information available in English, it is likely that some DSRs, mainly from non-English speaking countries, have been missed. Using only PubMed and Google for literature and DSR searches may lead to bias, as relevant studies or repositories not indexed in these databases may be missed. Investigating additional databases or sources could provide a more comprehensive overview. Limiting

the Google search to the first hundred results may not capture all relevant information, as other repositories or studies may appear beyond the first hundred results.

4.3. Future Perspectives

The number of DSRs is expected to grow over time, mainly due to the policies and initiatives implemented in the last two decades [10], and standard instruments (e.g., checklists) for assessment of the suitability of DSRs could be beneficial. Further efforts are needed to raise awareness about these DSRs as a central, safe point for researchers to find data from CTs shared by public or private sponsors, enhancing their value and creating ad hoc methods or procedures to reuse data in a responsible and standardized way [26]. This is especially important in the paediatric context in which nonreporting/nonpublication of findings remains common [8].

Moreover, the process of harmonization and standardization of data (e.g., CDISC standards <https://cdisc.org/> accessed on 3 October 2023) is time-consuming and costly but is an essential step in making data more FAIR. It would be beneficial to address the economic hurdles associated with making DSRs more FAIR across different specialities and countries [24,43].

Achieving interoperability between different DSRs means ensuring that these repositories can exchange data and work together seamlessly by implementing mechanisms for mapping and translating data between different formats and schemas.

Collaborative initiatives and data-sharing networks can promote data sharing by establishing common practices, protocols, and governance frameworks. They support semantic interoperability to standardize the semantics of data elements and facilitate accurate interpretation and integration of data across repositories. Data interoperability has several advantages, such as greater statistical power, poolable for post hoc analysis, pragmatic clinical trials, and analysis of under-represented subgroups [43–47].

Several additional challenges must be addressed, particularly in emerging economies. These challenges include legal and policy issues, scarcity of coordination between research groups, lack of a culture for data sharing, ethical/privacy considerations, insufficiency of proper infrastructure (including high-speed Internet connectivity), deficiencies in the interoperability of DSRs, shortage of data managers and data scientists, and a scarcity of open data DSRs to facilitate data sharing [27].

Data sharing can also help to develop and validate artificial intelligence (AI) models in the medical field in areas such as electronic medical records, medical imaging technology, medical big data, intelligent drug design, and smart health management systems. AI solutions can potentially improve the standardization and accuracy of clinical decision making while providing more dimensions of data accumulation for medical knowledge-based systems. These developments can also support physicians and researchers in the optimization of treatment plans and decision making about optimal treatment options [7].

5. Conclusions

Although data sharing is widely recognized as a fundamental requirement of scientific research and strongly encouraged, only a few CT DSRs exist in the paediatric field. To the best of our knowledge, this work is the first report addressing the availability of paediatric IPD within DSRs of CT data.

This work provides an inventory of the main DSRs containing paediatric clinical trial data, describing their main characteristics to disseminate and encourage the knowledge and subsequent use of DSRs. The latter may facilitate a clear and transparent sharing of paediatric CT information in the scientific community and relieve researchers, data managers, and sponsors of the ethical, regulatory, and economic burdens, shortening the time to respond to paediatric therapeutic needs.

Eight criteria were identified and used to assess the comprehensive suitability of DSRs.

The overall result shows heterogeneity between DSRs in terms of data upload, data handling and access to the DSR, instructions for data submitters and users, procedures for

patient-level data access, and privacy and security protocols (which are often lacking), as well as paediatric specificity. To the best of our knowledge, only two DSRs are exclusively dedicated to paediatrics (PTN and PCDC). We can hypothesize that this inconsistency in the available data may be due to the lack of generally accepted guidelines on the structure of DSRs for sharing CT data.

Lessons learned from existing paediatric DSRs highlight the importance of developing dedicated infrastructure, standardizing protocols, and fostering collaboration to support paediatric data-sharing initiatives and advancing paediatric research. Despite the growing awareness that data sharing can contribute to the successful development and validation of AI models for optimizing treatment plans and making decisions about optimal treatment options in the paediatric population, the results of this study show that paediatrics remains an area where focused efforts are still needed.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/data9040059/s1>, Research questions; Google search keywords list; Table S1: Research strings; Table S2: List of initiatives and sources from which DSRs have been identified; and Figure S1: Flow diagram for the selection of articles from literature.

Author Contributions: Conceptualization: A.C., M.F., and F.B.; Formal analysis: C.P., G.R., and M.T.; Investigation: A.D., A.L., C.P., M.T., V.H., and R.L.; Methodology: A.D., G.R., M.F., and F.B.; Project administration: M.F.; Supervision: M.F. and F.B.; Visualisation: A.D., A.S., C.P., S.M., and S.N.; Writing—Original Draft Preparation: A.D. and C.P.; Writing—Review and Editing: ALL. All authors have read and agreed to the published version of the manuscript.

Funding: The conect4children (c4c) project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 777389. The Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Data Availability Statement: The following items have been reported in the Supplementary Materials. No further data are available to share.

Acknowledgments: All the DSRs/platforms' reference persons who participated.

Conflicts of Interest: The authors declare that they have no competing interests.

List of Abbreviations

Analysis Data Model (ADaM)
 Artificial Intelligence (AI)
 Best Pharmaceuticals for Children Act (BPCA)
 Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC)
 Case Report Forms (CRFs)
 Clinical Data Acquisition Standards Harmonization (CDASH)
 Clinical Data Interchange Standards Consortium (CDISC)
 Clinical Research Data Sharing Alliance (CRDSA)
 Clinical Study Data Request (CSDR)
 Clinical trials (CTs)
 Coordinated Research Infrastructures Building Enduring Life-science (CORBEL)
 Comma-Separated Values (CSV)
 connect4children (c4c)
 Data and Specimen Hub (DASH)
 Data-sharing repositories (DSRs)
 Data-sharing repository (DSR)
 European Federation of Pharmaceutical Industries and Associations (EFPIA)
 European Genome-phenome Archive (EGA)
 European Medicines Agency (EMA)
 European Union (EU)
 Findable, Accessible, Interoperable, Reusable (FAIR)
 Food and Drug Administration Amendments Act (FDAAA)

Food and Drug Administration Modernization Act (FDAMA)
 Genomic Data Commons (GDC)
 Health and Human Services (HHS)
 International Business Machines Corporation Statistical Package for the Social Sciences (IBM SPSS)
 International Committee of Medical Journal Editors (ICMJE)
 Image & Data Archive (IDA)
 Infectious Diseases Data Observatory (IDDO)
 Immune Tolerance Network (ITN)
 Immunology Database and Analysis Portal (ImmPort)
 IMProving Access to Clinical Trial data (IMPACT)
 Independent Review Panel (IRP)
 Institute of Medicine (IOM)
 Innovative Medicines Initiative 2 (IMI2)
 International Business Machines Corporation (IBM)
 Individual Paediatric patient data (IPD)
 National Cancer Institute (NCI)
 National Institutes of Health (NIH)
 National Heart, Lung, and Blood Institute (NHLBI)
 National Institute of Allergy and Infectious Diseases (NIAID)
 National Institute of Child Health and Human Development (NICHD)
 National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)
 National Institute of Environmental Health Sciences (NIEHS)
 National Institute of Mental Health (NIMH)
 National Institute of Neurological Disorders and Stroke (NINDS)
 National Institute on Drug Abuse (NIDA)
 National Sleep Research Resource (NSRR)
 Pediatric Cancer Data Commons (PCDC)
 Paediatric Trials Network (PTN)
 Pharmaceutical Research and Manufacturers of America (PhRMA)
 Project Data Sphere (PDS)
 Rare Disease Cures Accelerator–Data Analytical Platform (RDCA-DAP)
 Statistical Analysis System (SAS)
 Schwarz's Bayesian Information Criterion (SBIC)
 Study Data Tabulation Model (SDTM)
 United Kingdom Research and Innovation (UKRI)
 Yale University Open Data Access (YODA)

References

1. Wan, M.; Alessandrini, E.; Brogan, P.; Eleftheriou, D.; Warris, A.; Brüggemann, R.; Turner, M. Risk-Proportionate Approach to Paediatric Clinical Trials: The Legal Requirements, Challenges, and the Way Forward under the European Union Clinical Trials Regulation. *Clin. Trials* **2022**, *19*, 573–578. [[CrossRef](#)] [[PubMed](#)]
2. Turner, M.A.; Hildebrand, H.; Fernandes, R.M.; De Wildt, S.N.; Mahler, F.; Hankard, R.; Leary, R.; Bonifazi, F.; Nobels, P.; Cheng, K.; et al. The conect4children (c4c) Consortium: Potential for Improving European Clinical Research into Medicines for Children. *Pharm. Med.* **2021**, *35*, 71–79. [[CrossRef](#)] [[PubMed](#)]
3. Chiaruttini, G.; Felisi, M.; Bonifazi, D. Challenges in Paediatric Clinical Trials: How to Make It Feasible. In *The Management of Clinical Trials*; Abdeldayem, H., Ed.; InTech: London, UK, 2018; p. 90. ISBN 978-1-78923-238-7.
4. Weiss, E.M.; Porter, K.M.; Sullivan, T.R.; Sotelo Guerra, L.J.; Anderson, E.E.; Garrison, N.A.; Baker, L.; Smith, J.M.; Kraft, S.A. Equity Concerns Across Pediatric Research Recruitment: An Analysis of Research Staff Interviews. *Acad. Pediatr.* **2023**, *24*, 318–329. [[CrossRef](#)] [[PubMed](#)]
5. Brewster, R.; Wong, M.; Magnani, C.J.; Gunningham, H.; Hoffer, M.; Showalter, S.; Tran, K.; Steinberg, J.R.; Turner, B.E.; Goodman, S.N.; et al. Early Discontinuation, Results Reporting, and Publication of Pediatric Clinical Trials. *Pediatrics* **2022**, *149*, e2021052557. [[CrossRef](#)] [[PubMed](#)]
6. Goldacre, B.; Lane, S.; Mahtani, K.R.; Heneghan, C.; Onakpoya, I.; Bushfield, I.; Smeeth, L. Pharmaceutical Companies' Policies on Access to Trial Data, Results, and Methods: Audit Study. *BMJ* **2017**, *358*, j3334. [[CrossRef](#)]
7. Bertagnolli, M.M.; Sartor, O.; Chabner, B.A.; Rothenberg, M.L.; Khozin, S.; Hugh-Jones, C.; Reese, D.M.; Murphy, M.J. Advantages of a Truly Open-Access Data-Sharing Model. *N. Engl. J. Med.* **2017**, *376*, 1178–1181. [[CrossRef](#)] [[PubMed](#)]
8. Taichman, D.B.; Sahni, P.; Pinborg, A.; Peiperl, L.; Laine, C.; James, A.; Hong, S.-T.; Haileamlak, A.; Gollogly, L.; Godlee, F.; et al. Data Sharing Statements for Clinical Trials. *BMJ* **2017**, *357*, j2372. [[CrossRef](#)] [[PubMed](#)]

9. Welsh, J.; Lu, Y.; Dhruva, S.S.; Bikdeli, B.; Desai, N.R.; Benchetrit, L.; Zimmerman, C.O.; Mu, L.; Ross, J.S.; Krumholz, H.M. Age of Data at the Time of Publication of Contemporary Clinical Trials. *JAMA Netw. Open* **2018**, *1*, e181065. [CrossRef]
10. Committee on Strategies for Responsible Sharing of Clinical Trial Data; Board on Health Sciences Policy; Institute of Medicine Guiding Principles for Sharing Clinical Trial Data. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*; National Academies Press: Washington, DC, USA, 2015; p. 290. ISBN 978-0-309-31629-3.
11. International Coalition of Medicines Regulatory Authorities (ICMRA) and the World Health Organization (WHO) Joint Statement on Transparency and Data Integrity. Available online: https://www.icmra.info/Drupal/En/Covid-19/Joint_statement_on_transparency_and_data_integrity (accessed on 3 October 2023).
12. U.S. Food and Drug Administration Pediatric Research Equity Act | PREA. Available online: <https://www.fda.gov/Drugs/Development-Resources/Pediatric-Research-Equity-Act-Prea> (accessed on 3 October 2023).
13. U.S. Food and Drug Administration Best Pharmaceuticals for Children Act (BPCA). Available online: <https://www.fda.gov/Drugs/Development-Resources/Best-Pharmaceuticals-Children-Act-Bpca> (accessed on 3 October 2023).
14. European Parliament, the Council of the European Union Regulation (EC) No 1901/2006 of The European Parliament and of the Council of 12 December 2006 on Medicinal Products for Paediatric Use, and Regulation (EC) No 1902/2006 Amending Regulation in Which Changes to the Original Text Were Introduced Relating to Decision Procedures for the European Commission. 2006. Available online: <https://www.ema.europa.eu/en/human-regulatory-overview/paediatric-medicines-overview/paediatric-regulation> (accessed on 11 March 2024).
15. European Commission. Evaluation of Medicines for Rare Diseases and Children Legislation. European Commission. 2020. Available online: https://health.ec.europa.eu/medicinal-products/medicines-children/evaluation-medicines-rare-diseases-and-children-legislation_en (accessed on 11 March 2024).
16. European Parliament, the Council of the European Union REGULATION (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on Clinical Trials on Medicinal Products for Human Use, and Repealing Directive 2001/20/EC. 2020. Available online: <https://eur-lex.europa.eu/eli/reg/2014/536/oj> (accessed on 3 October 2023).
17. Pisani, E.; Aaby, P.; Breugelmans, J.G.; Carr, D.; Groves, T.; Helinski, M.; Kamuya, D.; Kern, S.; Littler, K.; Marsh, V.; et al. Beyond Open Data: Realising the Health Benefits of Sharing Data: Table 1. *BMJ* **2016**, *355*, i5295. [CrossRef]
18. Ohmann, C.; Banzi, R.; Canham, S.; Battaglia, S.; Matei, M.; Ariyo, C.; Becnel, L.; Bierer, B.; Bowers, S.; Clivio, L.; et al. Sharing and Reuse of Individual Participant Data from Clinical Trials: Principles and Recommendations. *BMJ Open* **2017**, *7*, e018647. [CrossRef]
19. Kiley, R.; Peatfield, T.; Hansen, J.; Reddington, F. Data Sharing from Clinical Trials—A Research Funder’s Perspective. *N. Engl. J. Med.* **2017**, *377*, 1990–1992. [CrossRef] [PubMed]
20. Antman, E. Data Sharing in Research: Benefits and Risks for Clinicians. *BMJ* **2014**, *348*, g237. [CrossRef] [PubMed]
21. Sardanelli, F.; Ali, M.; Hunink, M.G.; Houssami, N.; Sconfienza, L.M.; Di Leo, G. To Share or Not to Share? Expected Pros and Cons of Data Sharing in Radiological Research. *Eur. Radiol.* **2018**, *28*, 2328–2335. [CrossRef] [PubMed]
22. Hrynaszkiewicz, I.; Norton, M.L.; Vickers, A.J.; Altman, D.G. Preparing Raw Clinical Data for Publication: Guidance for Journal Editors, Authors, and Peer Reviewers. *Trials* **2010**, *11*, 9. [CrossRef] [PubMed]
23. Mello, M.M.; Lieou, V.; Goodman, S.N. Clinical Trial Participants’ Views of the Risks and Benefits of Data Sharing. *N. Engl. J. Med.* **2018**, *378*, 2202–2211. [CrossRef] [PubMed]
24. 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.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
25. InfoScipedia-IGI Global Publishing House. Available online: <https://www.igi-global.com/Dictionary/Digital-Repositories/7693> (accessed on 3 October 2023).
26. Banzi, R.; Canham, S.; Kuchinke, W.; Krleza-Jeric, K.; Demotes-Mainard, J.; Ohmann, C. Evaluation of Repositories for Sharing Individual-Participant Data from Clinical Studies. *Trials* **2019**, *20*, 169. [CrossRef] [PubMed]
27. Forero, D.A.; Curioso, W.H.; Patrinos, G.P. The Importance of Adherence to International Standards for Depositing Open Data in Public Repositories. *BMC Res. Notes* **2021**, *14*, 405. [CrossRef] [PubMed]
28. Speer, E.M.; Lee, L.K.; Bourgeois, F.T.; Gitterman, D.; Hay, W.W.; Davis, J.M.; Javier, J.R. The State and Future of Pediatric Research—An Introductory Overview: The State and Future of Pediatric Research Series. *Pediatr. Res.* **2023**, *1*–5. [CrossRef]
29. Cook, D.A.; Beckman, T.J. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *Am. J. Med.* **2006**, *119*, 166.e7–166.e16. [CrossRef]
30. McHugh, M.L. Interrater Reliability: The Kappa Statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]
31. Morey, L.C.; Blashfield, R.K.; Skinner, H.A. A Comparison of Cluster Analysis Techniques Withing a Sequential Validation Framework. *Multivar. Behav. Res.* **1983**, *18*, 309–329. [CrossRef] [PubMed]
32. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A K-Means Clustering Algorithm. *Appl. Stat.* **1979**, *28*, 100–108. [CrossRef]
33. Schwarz, G. Estimating the Dimension of a Model. *Ann. Statist.* **1978**, *6*, 461–464. [CrossRef]
34. International Committee of Medical Journal Editors (ICMJE) International Committee of Medical Journal Editors (ICMJE): Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication. *Haematologica* **2004**, *89*, 264.

35. United States Congress. Senate Bill S. 558 PUBLIC LAW 110–85—SEPT. 27, 2007. Available online: <https://www.govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf#page=82> (accessed on 3 October 2023).
36. Summaries of Clinical Trial Results for Laypersons. Available online: https://health.ec.europa.eu/system/files/2020-02/2017_01_26_summaries_of_ct_results_for_laypersons_0.pdf (accessed on 3 October 2023).
37. European Medicines Agency External Guidance on the Implementation of the European Medicines Agency Policy on the Publication of Clinical Data for Medicinal Products for Human Use. Available online: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-1.pdf (accessed on 3 October 2023).
38. Siebert, M.; Gaba, J.F.; Caquelin, L.; Gouraud, H.; Dupuy, A.; Moher, D.; Naudet, F. Data-Sharing Recommendations in Biomedical Journals and Randomised Controlled Trials: An Audit of Journals Following the ICMJE Recommendations. *BMJ Open* **2020**, *10*, e038887. [[CrossRef](#)] [[PubMed](#)]
39. Nosek, B.A.; Alter, G.; Banks, G.C.; Borsboom, D.; Bowman, S.D.; Breckler, S.J.; Buck, S.; Chambers, C.D.; Chin, G.; Christensen, G.; et al. Promoting an Open Research Culture. *Science* **2015**, *348*, 1422–1425. [[CrossRef](#)] [[PubMed](#)]
40. Anthony, N.; Pellen, C.; Ohmann, C.; Moher, D.; Naudet, F. Social Media Attention and Citations of Published Outputs from Re-Use of Clinical Trial Data: A Matched Comparison with Articles Published in the Same Journals. *BMC Med. Res. Methodol.* **2021**, *21*, 119. [[CrossRef](#)] [[PubMed](#)]
41. Ohmann, C.; Moher, D.; Siebert, M.; Motschall, E.; Naudet, F. Status, Use and Impact of Sharing Individual Participant Data from Clinical Trials: A Scoping Review. *BMJ Open* **2021**, *11*, e049228. [[CrossRef](#)] [[PubMed](#)]
42. Clinical Research Data Sharing Alliance: CRDSA, Work Group: Data Protection A Review of Biopharma Sponsor Data Sharing Policies and Protection Methodologies 2023—Version: 2.0 2023. Available online: https://crdsalliance.org/?smd_process_download=1&download_id=469 (accessed on 3 October 2023).
43. McMurry, A.J.; Murphy, S.N.; MacFadden, D.; Weber, G.; Simons, W.W.; Orechia, J.; Bickel, J.; Wattanasin, N.; Gilbert, C.; Trevvett, P.; et al. SHRINE: Enabling nationally scalable multi-site disease studies. *PLoS ONE* **2013**, *8*, e55811. [[CrossRef](#)]
44. Nabbout, R.; Zanello, G.; Baker, D.; Black, L.; Brambilla, I.; Buske, O.J.; Conklin, L.S.; Davies, E.H.; Julkowska, D.; Kim, Y.; et al. Towards the international interoperability of clinical research networks for rare diseases: Recommendations from the IRDiRC Task Force. *Orphanet J. Rare Dis.* **2023**, *18*, 109. [[CrossRef](#)]
45. Loudon, K.; Treweek, S.; Sullivan, F.; Donnan, P.; Thorpe, K.E.; Zwarenstein, M. The PRECIS-2 tool: Designing trials that are fit for purpose. *BMJ* **2015**, *350*, h2147. [[CrossRef](#)] [[PubMed](#)]
46. Lehne, M.; Sass, J.; Essenwanger, A.; Schepers, J.; Thun, S. Why digital medicine depends on interoperability. *NPJ Digit. Med.* **2019**, *2*, 79. [[CrossRef](#)] [[PubMed](#)]
47. Hodson, S.; Jones, S.; Collins, S.; Genova, F.; Harrower, N.; Mietchen, D.; Petrauskaitė, R.; Wittenburg, P. *FAIR Data Action Plan: Interim Recommendations and Actions from the European Commission Expert Group on FAIR Data*; Zenodo: City, Country, 2018. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.