Rare Diseases Cures Accelerator Data and Analytics Platform (RDCA-DAP) best practices and recommendations for FAIR data, toward alignment with International Regulatory agencies

Alexandre Bétourné[1], Ramona L. Walls[1], Alison Bateman-House[2], Cécile Ollivier[1], Huong Huynh[1], Daniel Olson[1], Amanda Borens[1] and Jeffrey S. Barrett[1]

1. Critical Path Institute
2. NYU Langone Heath

Introduction

Rare diseases are defined in different jurisdictions as conditions or disorders which affect small populations of people: fewer than 200,000 people in the United States (1), fewer than five in 10,000 people in the European Union and Canada (2)(3), and fewer than 50,000 people in Japan (4). Although many of the over 7,000 known rare diseases affect very small numbers of people, in aggregate an estimated 350 million people worldwide are affected by such diseases. Many rare diseases result in significant disability and/or early mortality, but approximately 90% of individuals living with a rare disease have no approved treatment or therapy. By the very nature of small and often geographically dispersed populations, gathering sufficient data to inform research, identification of subjects for studies, and long-term follow up in longitudinal studies are very difficult. As a result, many potential therapies for rare diseases fail to meet statistical significance and do not demonstrate impact in clinical trials (5). Suboptimal clinical trial design, spanning endpoint selection, inclusion/exclusion criteria, and size and length of trial, is frequently blamed for these failures, and disease communities continue to debate the effectiveness of therapies after negative trial results (6). These shortcomings result in a waste of resources and most crucially, time for patients with progressive and often life-threatening conditions, including those patients who volunteered for the failed trials.

To overcome these challenges, higher quality, more informative clinical trials are required to obtain definitive evidence concerning candidate therapies' safety and efficacy. Higher quality trials depend on comprehensive characterizations of individual rare diseases, as measured by accepted outcome assessments and biomarkers, which in turn depend on standardized data on global populations of patients. Because rare disease data are so rare, scientific efforts to accelerate the development of data-driven models and tools for rare diseases must be able to integrate and reuse existing data from multiple sources. The time is now to apply FAIR data principles – that data should be Findable, Accessible, Interoperable, and Reusable (7) to improve the quality and impact of rare disease data. FAIR data principles require that datasets include rich metadata that use standards and ontologies, permanent identifiers, and a clear license, and that they are available via standardized, machine-readable protocols.

The Rare Disease Cures Accelerator – Data and Analytics Platform (RDCA-DAP), developed by Critical Path Institute (C-Path) in collaboration with National Organization for Rare Disorders (NORD), is a US Food and Drug Administration (FDA) -funded effort to help accelerate drug development for rare diseases (https://portal.rdca.c-path.org/). This initiative combines C-Path's expertise in research and data analytics and NORD's leadership in the rare disease community for over 35 years to help researchers leverage existing knowledge and analyze data to inform and optimize clinical trial design with new sources of evidence. The platform supports the use of data to improve the characterization of rare disease progression and define novel biomarkers and endpoints, and provides analytical tools to inform the design of innovative trial protocols. To help meet these goals, RDCA-DAP integrates existing datasets from various sources within the rare disease community including data from clinical trials, patient registries, preclinical data, natural history studies, and electronic health records of individual hospitals and health systems. RDCA-DAP also plans to federate with existing data aggregators and databases via shared data models and application programming interfaces (APIs) to help create an ecosystem of rare disease data and information.

To date, RCDA-DAP has ingested 74 datasets from industry, academic, and registry contributors. Furthermore, C-Path's Data Collaboration Center has 15 years of experience standardizing and integrating diverse data sources for dozens of projects. We have observed many common issues that make data integration more challenging and limit the extent to which data are FAIR. These issues (detailed in Table 1) include poor practices in data collection leading to low quality data, issues in data management, lack of or gaps in standardization, disparate privacy laws and international regulations, and the shortage of shared international regulatory endorsements of best practices for data collection, management, and sharing. These problems can occur in any type of data, including registry, preclinical, natural history, and, to a lesser extent, clinical trial data. Issues in quality, completeness, and relevance of source data inevitably lead to uncertainty in downstream findings, which limits their utility in applications for regulatory decision making. In the following section, we propose solutions to the data integration challenges we have encountered while building RDCA-DAP. We are making these solutions and recommendations freely accessible in the hope of engaging with the rare disease data ecosystem. In particular, we aim to foster collaborations with regulatory agencies internationally, in the hopes of gathering global alignment on best practices and standards, toward greater quality of rare disease data.

Table 1. Common issues and challenges with rare disease data. Many of these issues apply to data outside rare diseases.

Category	**Specific issues**
Data quality	• missing data, missing fields • no or incomplete data dictionaries (e.g., missing definition of scoring values, data derivation formulas, or units) • lack of longitudinality
Data management	• lack of globally unique and persistent identifiers for patients, datasets, biosamples, etc. • data edits not traceable • difficult to determine duplicate data points or link patients across studies • non-FAIR data
Standards and ontologies	• no existing standard or common data model for registries • diverse data types and formats • SDTM is not easy to integrate with other data formats
Interoperability of international platforms.	• difficult to harmonize data across languages • lack of globally unique, persistent identifiers, variable data models.
Ethical and regulatory issues	• globally inconsistent laws, regulations, and ethical norms concerning what is required for patient protection; who has access to the data; who may withdraw patient data; who vets proposed uses of the data, by what criteria, and with what consequences if the use is deemed objectionable • relevant laws, regulations, and ethical norms may not only vary across national or state borders but also by funder, where the datasets are housed, the stated intention of data collection, and what

	patients were guaranteed with regard to the storage and use of their data • Unique sensitivities may arise given instances of historical or contemporary misuse of data

RDCA-DAP's best practices and collaboration proposal

Data quality

A non-exhaustive list of the most common data quality problems is detailed in Table 1, above. Clinical trial data are often standardized to the CDISC Study Data Tabulation Model (SDTM) (8), and therefore are generally of better quality than registry data. However, as clinical trials are required to report all data, they have issues such as outliers or missing data that can inhibit reuse. We propose solutions to common data quality issues in Table 2 below. All these data quality recommendations apply to registries and other non-trial data, but many of them will be relevant to selected clinical trials as well.

Table 2. Recommendations for improving data quality in rare disease datasets.

Data quality issue	**Proposed solution**
Missing data	Improve study design. Standardize data collection protocols. Provide reasons for missing data.
Incomplete data (e.g., missing units, drug amount and frequency)	Include complete data with units and frequency, not just values, in the dataset itself. Do not rely on data dictionaries for this information. In some instances, missing values may be inferred using rigorous data imputation mathematical methods, but imputed data must be marked as such.
Uncertainty around dates	When recording clinical measures, record the measurement date, not just the date it was entered into the record.
Number and frequency of repeat measures	Choose intervals of repeated measures to optimize longitudinal variation discovery without over-surveying.
Highly heterogenous questions and answers across datasets; uncertainty about what survey questions and answers mean	Use standardized instruments and protocols where they exist (e.g., from PROMIS question bank), as well as common data elements and standardized terminologies (discussed more below). Leverage standardized Clinical Outcomes Assessments (COA) and Questionnaires, Ratings, and Scales (QRS) documents. Limit response options to predefined answers rather than free text.
Inability to track patients across studies leading to duplication and loss of longitudinality	Develop a system that uniquely identifies patients while preserving their privacy. Ensure that all parties use the same GUIDs (no re-inventing IDs).
Measures outside of expected range	Provide standard ranges. Use tooling that prohibits data entry outside plausible ranges. Build in QC to look for unit or typing errors that often lead to unusually large or small values.
Ability to verify data (e.g., from genetic report, EHR, pharmacy records)	Include medical documentation that substantiates self-reported results.
Poorly formatted or uninterpretable data	Use standard data structures (e.g., single CSV tables). Make data tidy (9). If using Excel, only have a single table per worksheet, do not merge columns. Do not rely on color

	coding or fonts for interpretability. Include data dictionaries and appendices (e.g., questionnaires, protocols, survey instruments).
Poor interpretability of questionnaires	Validate questions before conducting survey.
Inability to or uncertainty about sharing data	Use of consent forms that allow sharing. This is especially important internationally.
Unethical data collection	Follow all relevant laws concerns human subjects research, data storage, use, and sharing. Have formal agreements with involved parties, especially funders and researchers, about who has authority to do what, under what circumstances. While not all use of registry data constitutes human subjects research, researchers globally should be aware of the International Committee of Medical Journal Editors (ICMJE) statement on protection of research participants (to which attestation of adherence must often be made when submitting articles for publication) (10)

Data Management

Good data collection practices cannot guarantee high quality data, which requires management practices throughout the data life cycle (11) plus attention to FAIR data principles (7) and why they are important. Having a solid data management plan at the beginning of data collection is important for any project, but particularly crucial for groups collecting longitudinal natural history data. Biopharmaceutical companies and large academic labs conducting clinical trials generally have staff trained in data management, whereas smaller labs and patient registry groups are much less likely to have such expertise in-house. A key element of data management that is missing from rare disease data (and in many other areas) is the use of globally unique, permanent, resolvable identifiers (GUIDs). The need to uniquely identify patients is well known, and resources exist to mint identifiers (12). GUIDs are a key part of the solution to duplication of patient data (not knowing that two records represent the same information for the same person) and loss of longitudinal data (not being able to track patients across time when their data are spread across multiple datasets). Mathematical methods allow this data to still be used (13), but consistent use of GUIDs for patients would add value to the data and save patients precious time in not re-reporting. Unfortunately, with rare diseases, GUIDs may not provide anonymity, and we encourage investment in developing a solution to this challenge. GUIDs are also important for other data elements, including specimens, data, and variables, and we encourage their use in place of free text wherever appropriate. Additional infrastructure to mint and resolve GUIDs and host and serve their metadata is also needed.

Standards and Ontologies

Standardizing data across so many rare diseases around the globe may seem daunting, but it is technically within reach. Data standards encompass data structures, data elements, ontologies, and data exchange. Resources exist to help standardize material in each of these categories, although additional work to adapt them to rare diseases will be necessary. An important component of any standardization process is that all data manipulations and conversions must be traceable. Therefore, not only are the standards themselves important, but also tooling to work with the standards and record their use. We encourage the development of open-source tooling for working with standards of all types. We propose below a list of suggestions and recommendations for data standards best practices. We welcome future discussions with international regulators, and ultimately hope to partner with them to refine our recommendations and drive consensus and adoption from the rare disease community.

Data structure: Different structures are required for different purposes (e.g., analysis, submission to regulatory authorities, different data types). However, for multiple structures to be interoperable, standard data structures need to readily convert among each other without loss of information required for each use case. SDTM is required for submission of clinical trial data to FDA, Pharmaceuticals and Medical Devices

Agency (PMDA), and National Medical Products Administration (NMPA), a preferred standard for European Medicines Agency (EMA) and Health Canada, and may be a good candidate for a single international standard for clinical trial data. SDTM, however, has two serious shortcomings for data reuse: it is complex and difficult for non-experts to understand, and the format does not lend itself well to typical analysis tools and integration with other data sources. Therefore, if SDTM or a similar model is endorsed for clinical trial data, the community must support efficient and lossless translators to other formats that will encourage reuse and integration of trial data with other data types such as registry data and electronic health records. There is no single widely used Common Data Model (CDM) for registry data, and the heterogeneity of formats across registries is a huge barrier to their integration. Rather than develop yet another CDM, we recommend extending an existing model to accommodate registry data. C-Path is working internally to extend the Observational Medical Outcomes Partnership (OMOP) CDM to work with registry data, as are other groups (14). OMOP has a large and growing user/developer community, is easily extensible, and works well for data analysis.

Data elements: The heterogeneity of data elements (i.e. variables) used across rare disease studies is another serious barrier to data reuse. In 2021, the main issues preventing the standardization of data are not unique to rare diseases, such as description of demographics, drugs, or pedigree. They are social, not technical. Standard vocabularies and reporting methods exist for many common variables (e.g., (15), see also list from (16)), and we encourage the adoption of one or a few standards for these fields. An advantage of OMOP CDM is that it already includes concept mappings to several widely used vocabularies such as the Unified Medical Language System (UMLS, (17)). Non-standard laboratory tests for rare diseases that are not included in LOINC (Logical Observation Identifiers Names and Codes, a database and universal standard for identifying medical laboratory observations, (18)) or existing ontologies should be added (see next section). Variable domains such as family history, phenotype and symptom descriptions, and physical and mental tests may be harder to standardize but making them interoperable is possible through common data elements (CDEs) and ontologies. We suggest the use and expansion of these resources for rare diseases. Where CDEs are truly not possible, data providers should be required to include data dictionaries explaining the variables and use resources like protocols.io or osf.io.

Ontologies: An ontology is a machine interpretable representation of the knowledge in a domain, structured as concepts, instances of concepts, and the relationships among them. Ontologies are widely used in biomedicine to help structure and standardize data and to logically infer additional facts. Due to the sparse and heterogeneous nature of rare disease data, ontologies are critical for combining data across diseases and study types, because they allow related, but not identical, concepts to be grouped together. The UMLS provides a comprehensive thesaurus of biomedical terminology that can be used as an ontology, but it provides limited logical expressivity. To achieve full value of ontologies and provide full international operability, we recommend using the open-source ontologies recommended by the Global Alliance for Genomes and Health's (GA4GH) Phenopacket standard (19). These ontologies are particularly important for describing rare diseases, because there are many diseases with similar common names, and single diseases with different names in different countries. Where possible, using the Human Phenotype Ontology (HPO, (20) and (21)) for phenotypes/symptoms of diseases will greatly increase the value of data for reuse and integration.

Data exchange: The Phenopackets standard was developed to standardize phenotypic data exchange within the medical and scientific settings and allow phenotypic data to flow between clinics, databases, clinical labs, journals, and patient registries in ways currently only feasible for more quantifiable data. A Phenopacket is a standard file format that contains a set of mandatory and optional fields to share information about a patient phenotype (e.g., clinical diagnosis, age of onset, lab tests results or disease severity). It is also able to link to a separate file containing a patient's genetic sequence data, if available. Fast Healthcare Interoperability Resources (FHIR) was developed by HL7 International as a standard for exchanging healthcare information electronically (22), specifically for electronic health records. FHIR can be used as a stand-alone data exchange standard but can also be used in partnership with existing widely used standards. We recommend the adoption of FHIR and Phenopackets for the sharing data but recognize that those two standards may only cover a portion of rare disease data. Before an exchange standard for rare disease registry data can be accomplished, there must be standard data structures and vocabularies, as described above.

Interoperability of International platforms

The standardization practices recommended in the previous paragraphs, if encouraged by regulators in all jurisdictions, would achieve improved international interoperability. Nonetheless, several challenges remain. One obvious challenge is harmonizing data collected in different languages. English is the lingua franca of science, and while some types of data could be standardized by mapping them to English language CDMs, the burden of mapping data such as questionnaires and natural history reports to English will be large. Automated translations and natural language processing tools can help, but still require effort. We support the translation of standards to multiple languages where appropriate and suggest that starting with CDEs and ontologies may be a valuable first step, in order to encourage their use outside English-speaking countries. Our colleagues from multi-lingual areas such as Canada and the EU may have additional insights into solutions to this challenge. The adoption of shared GUID systems for patients, specimens, and datasets (as mentioned above under Data Management) is also key to the interoperability of international data platforms.

Ethical Considerations for Rare Disease Data

There are numerous ethical issues concerning the appropriate collection, storage, use, and sharing of patient-generated data. Those involved in these activities must follow all relevant laws, including those about consent and privacy, concerning human subjects research and data storage, use, and sharing. While not all use of patient constitutes human subjects research, those contributing data to registries and using data should be aware of the International Committee of Medical Journal Editors (ICMJE) statement on protection of research participants (to which attestation of adherence must often be made when submitting articles for publication) ([10]).

Globally inconsistent laws, regulations, and ethical norms concerning patient data; who has access to the data; who may withdraw patient data; who vets proposed uses of the data, by what criteria, and with what consequences if the use is deemed objectionable will be a vexing issue for integrating rare disease data which may often be obtained from patients across wide geographic areas. These laws, regulations, and ethical norms may not only vary across national or state borders but also by funder, where the datasets are housed, the stated intention of the data collection, and what patients were guaranteed with regard to the storage and use of their data. Regulators have a role to play, if only through non-binding guidance, in both guiding entities using rare disease data to do so in accordance with high ethical standards, and in helping to facilitate the development of harmonized approaches.

Privacy, consent, and international laws

Navigating international laws and regulations around privacy and consent adds a layer of complexity to rare disease data integration. By default, the most stringent jurisdiction sets the limits for international data sharing. This currently puts the onus on locations such as the European Union and, within the US, California, to clarify the General Data Protection Regulation (GDPR) and California Consumer Protection Act (CCPA) so that other jurisdictions can comply with those regulations. However, we should not forget that people with rare diseases often want to share their data, and due to rarity, this usually means international sharing ([23]). It is the responsibility of the entire international community to develop methods, policies, and tools for data sharing that support patient privacy while not inhibiting scientific progress. We recommend that regulators encourage the development of novel consenting tools (e.g., blockchain and ontologies) that will allow the consent to travel with the data and be updated when patients' needs change. Regulators and other stakeholders should work to develop and implement international policies specifically focused on the privacy and data sharing needs of rare diseases, including incentives to share data.

Discussion

RDCA-DAP was launched to break down barriers among rare diseases data silos and establish an integrated platform able to accommodate multiple sources of patient-level data. This document provides a list of the most common issues we have encountered around making data FAIR and regulatory compliant, together with our proposals for best practices. RDCA-DAP aims at improving the data sharing ecosystem, and as such, provides feedback to its data custodians to encourage best practices in data collection and standardization. To advance rare disease drug development that is global in nature, international stakeholders should collaborate toward reaching consensus on best practices around data management, trial design, and regulatory science. The practices described above, if adopted by data contributors and endorsed globally by regulatory or other agencies, will make future data integration efforts more productive, potentially speeding the time to development of new treatments for rare diseases. Our recommendations apply to patient registries as well as all other sources of rare disease data. If stakeholders implement the suggestions offered, we anticipate over time an improvement in the quality of data available for drug development purposes and an increased ability to compare, contrast, and combine datasets.

The most frequent issues leading to poor data quality stem from the lack of standardized methods, protocols, questionnaires, and instruments used for data collection. Insufficient planning and the absence of conduct validation are also detrimental to data quality. We proposed a series of solutions that, if implemented, would drastically improve data quality. Well-implemented and controlled data management plans and infrastructure are essential but costly and harder to achieve in some smaller academic or non-profit registries with limited funding. As noted, guidance is available but challenging for rare disease communities to incorporate, and additional support might be required for starting registries. We thus support the establishment of international infrastructure that would help patient groups to develop well-managed and well-maintained registries.

In addition to support for general good data management practices, we aim to collaborate and benefit from regulatory agencies' leadership in several areas. A pivotal element to data management, GUIDs present several challenges in the context of rare disease, as GUID may not always provide anonymity, and research and development are needed to overcome this challenge. Global adoption of shared GUID systems (through federation and shared infrastructure components) will be transformative for rare disease research. Data structures, data elements, ontologies, and data exchange should be standardized internationally across all rare diseases. This is technically within reach, and we described a series of tools available to allow standardization and traceability of all data manipulations necessary for such processes. To date, CDISC's STDM has been adopted by most regulatory agencies, but it is not well suited for integration with other data types. If international consensus is reached on the CDISC standard for submission to regulators, the development of lossless translators should be encouraged to facilitate the interchangeability to other data formats and facilitate the integration of data from multiple sources. We recommended the endorsement of an expansion of the OMOP CDM for registry data, to break down the heterogeneity of data formats currently observed across registries. Similarly, a restricted number of standard vocabularies and reporting instruments should be promoted and expanded to include rare disease variables, using CDEs and ontologies. The Phenopackets Standard and its recommended ontologies should be promoted to facilitate sharing and communication of disease phenotypic information across international platforms. However, achieving greater international interoperability will require help and guidance from funders and regulators to support the translation of standards to multiple languages, which could start with CDEs and ontologies. Finally, international sharing of rare disease data will require innovation to support patient privacy while encouraging sharing. In the context of increased stringency of patient protection regulations, novel consenting tools and new international policies should be implemented.

Our recommendations are driven primarily by the FAIR data challenges we have encountered in RDCA-DAP and C-Path more generally, but additional stakeholder input is required to build a global consensus on data standards, ontologies, and best practices. For example, additional exchanges with international regulatory agencies to understand the specific challenges they are facing around FAIR data, and – perhaps more importantly – evaluate how those challenges differ among agencies. For example, the FDA's Center for Drug Evaluation and Research (CDER) approach to data standards consists of four strategic goals: Support open, consensus-based data standards development; Maintain and promote a well-defined data standards governance function; Promote the electronic submission of regulatory data using established standards; and Optimize the regulatory review process to fully leverage data conformed to standards (24). It would be useful to engage with other agencies to understand how their approaches may differ, and how these differences may lead to gaps in information needed by regulators, globally.

We believe that the best way to achieve global data sharing and collaboration, especially for multi-national studies, is through shared standards and practices, not only for submitted clinical trial data, but for other data types in the chain leading up to trials. Especially for rare diseases, where clinical trials are so difficult, other types of data must inform medical product development, such as registries, pre-clinical research and real-world data (including patient registries). The use of RDCA-DAP, combined with appropriate methodologies, can help develop more targeted generation of evidence to ensure that patients only participate in clinical trials with specific objectives that further the scientific understanding of a medicinal product for its use in the target population. Evidence for efficacy and benefit-risk generated with RDCA-DAP, in combination with clinical trials in rare diseases, should result in the same quality of regulatory decision-making as that based on self-standing clinical trials. Consequently, the same quality and rigor used in clinical trial data needs to be applied to these other sources of data. Our experience has shown that the standards and formats used for regulatory submission are not the best for pre-clinical and basic research. We therefore encourage regulators and other stakeholders to consider the medical product development life cycle in their endorsement or encouragement of best practices. Just like the data life cycle that does not begin and end with a single study, the medical product development life cycle must encompass both the "rough" data collected before clinical trials and the reuse of data after a trial. Only by taking a life cycle approach to data management can the value of the rare and precious data associated with rare diseases be put to full use.

REFERENCES

1. Office of the Federal Register NA and RA. Orphan Drug Act, 1983 [Internet]. govinfo.gov. U.S. Government Printing Office; 1982 [cited 2021 Nov 24]. Available from: https://www.govinfo.gov/content/pkg/STATUTE-96/pdf/STATUTE-96-Pg2049.pdf
2. Orphan medicine [Internet]. European Medicines Agency. [cited 2021 Nov 24]. Available from: https://www.ema.europa.eu/en/glossary/orphan-medicine
3. Canada H. Building a National Strategy for High-Cost Drugs for Rare Diseases: A Discussion Paper for Engaging Canadians [Internet]. 2021 [cited 2021 Nov 24]. Available from: https://www.canada.ca/en/health-canada/programs/consultation-national-strategy-high-cost-drugs-rare-diseases-online-engagement/discussion-paper.html
4. Overview of Orphan Drug/Medical Device Designation System [Internet]. Ministry of Health, Labour and Welfare: Pharmaceuticals and Medical Devices. [cited 2021 Nov 24]. Available from: https://www.mhlw.go.jp/english/policy/health-medical/pharmaceuticals/orphan_drug.html
5. Tambuyzer E, Vandendriessche B, Austin CP, Brooks PJ, Larsson K, Miller Needleman KI, et al. Therapies for rare diseases: therapeutic modalities, progress and challenges ahead. Nat Rev Drug Discov. 2020 Feb;19(2):93–111.
6. Morel T, Cano SJ. Measuring what matters to rare disease patients – reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J Rare Dis. 2017 Nov 2;12:171.
7. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018.
8. SDTM | CDISC [Internet]. [cited 2021 Nov 24]. Available from: https://www.cdisc.org/standards/foundational/sdtm
9. Wickham H. Tidy Data. J Stat Softw. 2014 Sep 12;59:1–23.
10. ICMJE | Recommendations | Protection of Research Participants [Internet]. [cited 2021 Nov 24]. Available from: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/protection-of-research-participants.html
11. Data Lifecycle [Internet]. NIH National Library of Medicine. [cited 2021 Dec 10]. Available from: https://nnlm.gov/guides/data-thesaurus/data-lifecycle
12. Global Unique Identifier (GUID) [Internet]. RaDaR - Rare Diseases Registry Program. [cited 2021 Dec 8]. Available from: https://rarediseases.info.nih.gov/radar/global-unique-identifier-generator
13. Data Users FAQ [Internet]. Available from: https://c-path.org/wp-content/uploads/2021/02/202011-RDCADAP_FAQ_Data_Users.pdf
14. OMOP CDM to SDTM - General [Internet]. OHDSI Forums. 2020 [cited 2021 Nov 24]. Available from: https://forums.ohdsi.org/t/omop-cdm-to-sdtm/11267
15. NIH Common Data Elements (CDE) Repository [Internet]. [cited 2021 Dec 8]. Available from: https://cde.nlm.nih.gov/home
16. Zentzis B. Common Data Element (CDE) [Internet]. OHSU Clinical Informatics Wiki. [cited 2021 Nov 24]. Available from: https://clinfowiki.org/wiki/index.php/Common_Data_Element_(CDE)
17. Unified Medical Language System (UMLS) [Internet]. U.S. National Library of Medicine; [cited 2021 Nov 24]. Available from: https://www.nlm.nih.gov/research/umls/index.html
18. Home - LOINC [Internet]. LOINC. [cited 2021 Nov 24]. Available from: https://loinc.org/
19. Recommended Ontologies — phenopacket-schema 2.0 documentation [Internet]. [cited 2021 Nov 24]. Available from: https://phenopacket-schema.readthedocs.io/en/v2/recommended-ontologies.html
20. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J-P, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019 Jan 8;47(D1):D1018–27.
21. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res. 2017 Jan 4;45(D1):D865–76.

22. FHIR Overview [Internet]. [cited 2021 Nov 24]. Available from: https://www.hl7.org/fhir/overview.html
23. Rare Disease Patient Registries [Internet]. EURORDIS - The Voice of Rare Disease Patients in Europe. [cited 2021 Nov 24]. Available from: https://www.eurordis.org/publication/rare-disease-patient-registries
24. Research C for DE and. Data Standards Program Strategic Plan and Board [Internet]. FDA. FDA; 2021 [cited 2021 Nov 24]. Available from: https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/data-standards-program-strategic-plan-and-board