Schema Playground: A tool for authoring, extending, and using metadata schemas to improve FAIRness of biomedical data

Marco Cano, Ginger Tsueng, Xinghua Zhou, Laura D. Hughes, Julia L. Mullen, Jiwen Xin, Andrew I. Su, Chunlei Wu

Abstract

Background: Biomedical researchers are strongly encouraged to make their research outputs more Findable, Accessible, Interoperable, and Reusable (FAIR). While many biomedical research outputs are more readily accessible through open data efforts, finding relevant outputs remains a significant challenge. Schema.org is a metadata vocabulary standardization project that enables web content creators to make their content more FAIR. Leveraging schema.org could benefit biomedical research resource providers, but it can be challenging to apply schema.org standards to biomedical research outputs. We created an online browser-based tool that empowers researchers and repository developers to utilize schema.org or other biomedical schema projects.

Results: Our browser-based tool includes features which can help address many of the barriers towards schema.org-compliance such as: The ability to easily browse for relevant schema.org classes, the ability to extend and customize a class to be more suitable for biomedical research outputs, the ability to create data validation to ensure adherence of a research output to a customized class, and the ability to register a custom class to our schema registry enabling others to search and re-use it. We demonstrate the use of our tool with the creation of the Outbreak.info schema—a large multi-class schema for harmonizing various COVID-19 related resources.

Conclusions: We have created a browser-based tool to empower biomedical research resource providers to leverage schema.org classes to make their research outputs more FAIR.

Introduction

Funding agencies, international consortia, institutional policies, and publisher requirements have helped promote the adoption of the FAIR (Findability, Accessibility, Interoperability, and Reusability) guiding principles (Wilkinson MD et al, 2016)(Boeckhout M et al, 2018) for biomedical research data sharing to varying degrees of success. While it is now standard to make datasets accessible and potentially reusable via deposition of the dataset in a repository, standardization issues continue to make it challenging for researchers to make datasets findable, interoperable, and reusable. To address these issues, domain experts and data stewards have been inspecting the gap between principle and practice (Koesten L et al, 2020), extending (Jauer ML and Deserno TM, 2020) and adapting the principles (Holub P et al, 2018), creating their own metadata standards (Canham S and Ohmann C, 2016) and data schemas (Hruby GW et al, 2016)(Papadiamantis AG et al, 2020). However a large gap remains between

the communities that develop standards and the adoption of these standards by data and resource providers due to issues in communication, education/training, incentives, and the availability of supportive tools (Hollman et al, 2020).

Schema.org is a metadata vocabulary standardization project founded by the major search engine companies such as Google, Microsoft, Yahoo, and Yandex. It is an open source, collaborative initiative that develops metadata standards for improved searchability. Schema.org already includes some biomedically relevant classes like Datasets and Medical Study, and applying schema.org classes to biomedical research resources would improve interoperability, enabling researchers readily ingest existing resources and to leverage search engine-based solutions (like Google Dataset Search) to find resources of interest. Although there have been some efforts to leverage schema.org to improve findability of scientific research data (Sansone SA et al, 2017)(Papadiamantis AG, 2020)(Jones MB et al, (2021) and many generic repositories (like Figshare and Zenodo) are compliant, schema.org remains largely underutilized by the biomedical research community. Bioschemas is an open and collaborative effort that has been actively promoting the use of schema.org in the life sciences by serving as a hub for researchers to create new biomedically relevant classes with the goal of refining and proposing these classes to schema.org (Gray A et al, 2017)(Profiti G et al, 2018), and by raising awareness about the usefulness of metadata schemas. The Bioschemas community has also developed a set of tools which are useful for generating or scraping metadata based on Bioschemas classes, but do not necessarily improve the ease of using schema.org classes.

Here, we describe the Data Discovery Engine's (DDE) Schema Playground, a web-based tool that improves the ease of using any registered schema or schema.org classes. Our tool allows users to easily find and visualize relevant schema.org classes, extend them, create validations, and save/share the newly created classes for others to reuse. Our tool also includes a framework for building data registries and creating guides for data submission; however, the implementation and integration of these features on our site is restricted to partner organizations. We introduce the features of this tool, review its value to different types of users, and demonstrate its application towards the creation of a new schema for COVID-19-related resources.

Implementation

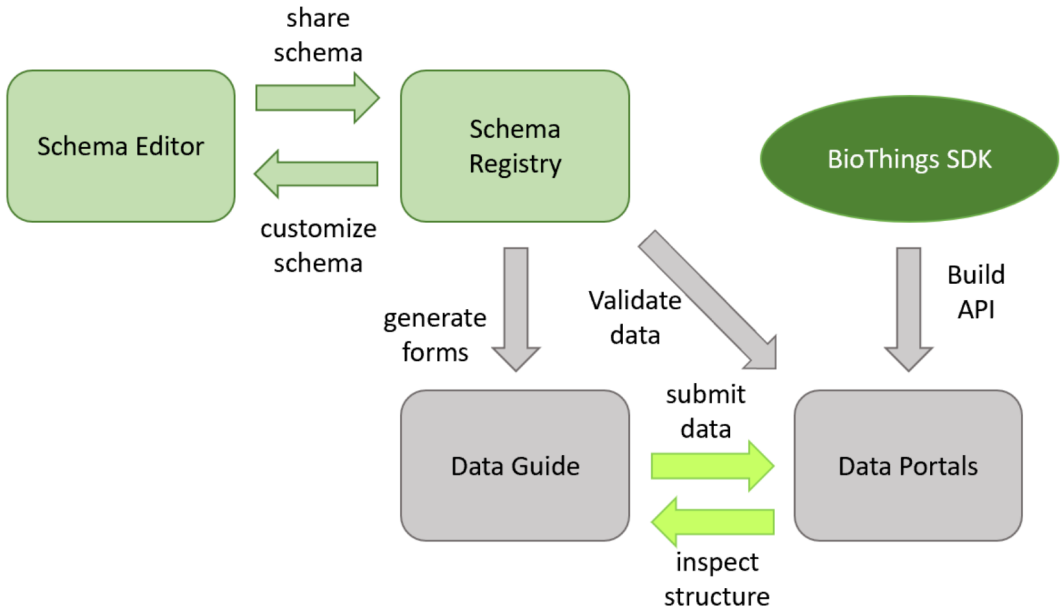
The Data Discovery Engine's Schema Playground is a browser-based tool built with Vue.js, Python/Tornado, and the BioThings Software Development Toolkit (https://docs.biothings.io/, a framework for building biomedical APIs). Schemas from schema.org and other consortia/projects are stored and made searchable using MongoDB and Elasticsearch. The code for the Schema Playground can be found at https://github.com/biothings/discovery-app and is free to use under the Apache License 2.0. The COVID-19 outbreak.info resource schemas were developed by comparing metadata properties across multiple type-specific repositories to identify properties in common. For example, metadata from LitCovid/PubMed, BioRxiv/MedRxiv, various journals like JAME, NEJM and others, and the metadata from publications found on Zenodo, Figshare and others were compared in order to identify a suitable

schema for COVID-19-related publications. Similarly, protocols from protocols.io and the BioSchemas LabProtocol class were compared to develop a schema for COVID-19-related protocols. Once the desired properties and structure for each class of COVID-19-related resource was identified, the schemas were created by extending existing schema.org classes using the DDE Schema Playground.

Results

The DDE Schema Playground consists of two standard (and fully-accessible) components and two related, custom (limited-access) components (Figure 1). The standard components improve the ease of use of schemas and classes, while the custom components help communities to reap the benefits of their use. The Schema Editor allows users to import community standard schemas like schema.org and customize them for biomedical purposes. These extended schemas can then be shared in the Schema Registry, which allows users to view the schemas and reuse them. When used in conjunction with Data Portals built with BioThings SDK, The DDE Schema Playground can automatically generate data submission forms known as Data Guides.

Figure 1 - Components of the DDE Schema Playground and how they work together



To understand how the Schema Playground might help to bridge the gap between data standardization communities and data resource providers, we identified potential utility and value of each of the DDE Schema Playground components for different types of users in our partner communities (Figure 2).

Figure 2 - The value and utility of DDE Schema Playground components to different types of users

Types of Users	Standard Components: Schema Registry	Standard Components: Schema Editor	Custom Components: Data Portals	Custom Components: Data Guides
Data Providers	Find schemas for new classes of data Register in-use schema for FAIRness	Extend schemas to ensure maximum compatibility and minimize duplicate effort	Collaboratively collect and share data	Submit compliant data via forms
Data Consumers	Download schemas to understand how to consume and validate data	Easily create schema validation for structured data if schema validation not available	Explore and reuse standardized data	Understand structure of available data
Data Standardizers	Explore community needs for new classes & properties Share schemas for community uptake	Easily create schema from existing classes and validation once consensus on structure has been established	Identify data structuring needs and potential structuring improvements	Understand Structure Burden

Any data portals and guides can be used by anyone with sufficient access rights, but the creation of a data portal or data guide requires partnerships with our team to actualize. For the outbreak portal, data submission via the guide is open to all and utilizes github for authentication. For other portals, access may be restricted as required by the responsible partner organization. The data portal and data guides allow data providers and data consumers to collect, share, and use data. Since the data guide converts a custom schema into a web-based data submission form, it enables data consumers and data standardizers to visually inspect and understand the burden of structure.

The schema registry and editor allows data providers and/or standardizers to find, customize, and share schemas. Sharing schemas via the registry will make it easier for data consumers to understand how to consume data from a data provider and to create data validation if one is not available from the data provider. Having a central location for schemas submitted by data providers will also make it easier for data standardizers to evaluate the needs of the biomedical research community. To further illustrate the value of the schema registry and editor, we compare and detail the features of the DDE Schema Playground with available tools for creating, applying, and consuming other major schemas such as Schema.org and Bioschemas.

Schema.org, Bioschemas and other data standardization efforts have built strong communities to generate consensus on data modeling for the creation of new schemas or the improvement of existing schemas. Hence, there are extensive processes in place (but few tools) for the creation of a new schema based on schema.org or any other schemas. Because of its widespread adoption, there are third party tools available for utilizing and consuming markup from schema.org. The Bioschemas community also has a process in place for defining new classes

and has a set of tools which cover both the creation of a new schema (google spreadsheet conversion), utilization of a schema (markup generation), and evaluation of use (markup validation, scraping), but these tools vary in usability based on the users programming experience. In contrast to schema.org, Bioschemas also defines cardinality (allowable number of values per property) and marginality (optional vs required value) of its properties as these are important to the life sciences research community. As members of the Bioschemas community, we sought to provide complimentary schema tools to facilitate the biomedical data schema development and adoption. To do this, we identified schema tools and features available from schema.org and the Bioschemas community which would be of interest to the biomedical research community. Although tools were available for many desired features, many tools were only available as source code and required basic programming experience. We focused our efforts on features for which user-friendly tools could not easily be found resulting in a web-based application that empowers individual data resource providers to utilize and customize existing schemas from schema.org and other similar efforts. As seen in Table 1, these features include:

Table 1- Comparison of Schema.org, Bioschemas, and DDE Schema Playground

	schema.org	Bioschemas	DDE schema playground
Search for classes from own community	✓	✓	✓
Search for classes from other communities			✓
Visualize any schema in json			✓
Compare schemas			✓
Create a schema	✓*	✓*	✓
Extend a schema			✓
Create schema validation			✓
Export/Save a schema		🛠	✓
View example markup using a schema	✓	✓	
Create markup using a schema	🛠	🛠	✓**
Account for cardinality		✓	✓
Account for marginality		✓	✓
Google doc integration		🛠	
github integration	✓	✓	✓
markup validation	🛠	🛠	🛠
markup scraper	🛠	🛠	✓**

✓available, 🛠 separate tools available, *process in place, **feature exists, but not generally available

1. Searching and viewing schemas from schema.org and other metadata standardization efforts

The DDE Schema Playground allows for the visualization of json schemas hosted online either on github or elsewhere (Supplemental Figure 1A). This allows users who are familiar

with schema.org to review their compliant schema in a more human readable format. The DDE Schema Playground also has a searchable registry of classes from schema.org, BioLink (Bruskiewich et al, 2021), BioThings, and others. Users may browse and visualize the schemas for various classes from these sources to identify the classes of most interest to them (Supplemental Figure 1B). If a community like Bioschemas or consortia like N3C is interested in making their schema available for searching and viewing, they can import and register their json schema. The DDE Schema Playground also enables users to compare up to four schemas. For example, there are multiple Dataset schemas available in the registry, and users can compare them to see what properties are unique to each and what properties they share (Supplemental Figure 1C).

2. **Extending and customizing a pre-existing schema for a particular use**
   The ability to browse and inspect pre-existing schemas makes it easier for a user to customize or extend the schema to suit their own purpose. All the properties from the pre-existing schema will be inherited in the extended schema; however, the user may select properties for which validation is desirable. The user can also create new properties to be included in the extended schema. For example, the Dataset schema from schema.org serves as a useful foundation, but a schema focused on COVID-19-related datasets may need additional fields (e.g., infectiousAgent). To tailor the Dataset schema, we find and extend it from the registry (Supplemental Figure 2A). After we create a name for our schema (the namespace) and the class, we can customize it. We can select to include any property that is available from the schema we are extending (Supplemental Figure 2B), and we can create new properties (eg- curatedBy) that are tailored to our needs (Supplemental Figure 2C).

3. **Creating validation for the schema for data quality enforcement**
   Marginality (whether a property is required or not) and cardinality (whether a property can have one or multiple values) are two aspects of schema properties that are not expressed well by schema.org but are desirable to biomedical researchers (Supplemental Figure 3A and 3C). In the DDE Schema Playground, this is handled via the creation of schema validation, and schema validation creator provides a simple drag and drop mechanism to create straightforward validations (Supplemental Figure 3B). For slightly more complex validations, the user can edit the type of validation they are trying to include before dragging and dropping it into the property of interest. In our example Dataset schema, an Organization is a potential type for our new property (curatedBy). We edit the example object validation for Person to create an Organization object validation (Supplemental Figure 3D).

4. **Exporting and saving a schema generated by the Schema Playground editor**
   The DDE Schema Playground allows you to export/download your newly created schema locally and it is also integrated with GitHub, allowing users to save to their GitHub repository (Supplemental Figure 4A-C). The integration with GitHub allows the edits to the schema to be made by multiple parties and provides the schema owner the option of pulling changes to the schema. Additionally, the schema can be forked and edited/customized allowing for

re-use of the schemas which in turn improves findability and reusability of resources which follow the schemas.

5. **Registering a newly created schema in the DDE schema registry to facilitate its extension and re-use**

   Once saved in GitHub, users can review their schema with the schema viewer and add it to the registry to enable others to easily re-use it (Supplemental Figure 1A). This provides a user-friendly interface for editing, customizing, and re-using schemas for those who prefer not to manually edit text and format in json.

The DDE Schema Playground offers any user the ability to reuse and extend existing schemas. This tool is primarily to assist in the authoring of schemas for use in other applications. In addition, we have converted three Dataset schemas into "guides", which are web-based forms for annotating resources using schemas authored in the DDE Schema Playground. Annotations created using these guides are stored within a resource registry hosted within the DDE. There are currently three public guides based on the Dataset schemas for the outbreak.info web application (Research Library, 2020), the N3C initiative (Haendel et al., 2020), and the CD2H consortium (Center for Data to Health, 2021). While the creation of guides from schemas is not a fully-automated feature that is available to all users, most of the underlying components are reusable, additional guides can be constructed and hosted within the DDE through collaboration.

Creating the COVID-19 Outbreak schema using the Schema Playground

Outbreak.info is a project from the Su, Wu, and Andersen labs at Scripps Research to unify COVID-19 and SARS-CoV-2 epidemiology and genomic data, published research, and other resources. The standardization of published research and other resources was accomplished by creating a single, multiclass schema to harmonize the metadata: The COVID-19 Outbreak schema. This schema can be found in the DDE registry at https://discovery.biothings.io/view/outbreak/ and was built via the DDE Schema Playground with some manual editing (for merging all the classes into a single schema). There are five principal classes in the Outbreak schema (Analysis, Dataset, ClinicalTrial, Protocol, Publication) and many subclasses to support the principal classes. As seen in Table 2, the classes in the Outbreak schema were extended from related schema.org classes (whenever available) and were created based on metadata comparisons from a variety of related sources. For example, the Protocol class in the Outbreak schema was extended from the HowTo class in schema.org and was based on properties identified from available metadata in protocols.io and the LabProtocol profile from Bioschemas. This schema is currently used to harmonize and improve FAIRness of metadata from over 200,000 resource entries in the outbreak.info research library at https://outbreak.info/resources.

Table 2- Classes in the Outbreak schema and how they were created and used

Outbreak schema class	extended from schema.org class	inspired by / derived from	used in Outbreak schema class
Analysis	CreativeWork	Outbreak Protocol, Dataset Class, various COVID-19 Analyses	--
ClinicalTrial	MedicalStudy	NCT Clinical Trials PRS, WHO ClinicalTrials	--
Dataset	Dataset	NIAID schema	--
Protocol	HowTo	Bioschemas LabProtocol, protocols.io	--
Publication	MedicalScholarlyArticle	NLM Medline, biorxiv, various journals	--
ArmGroup	Thing	NCT Clinical Trials PRS, WHO ClinicalTrials	ClinicalTrial
CitationObject	CreativeWork	Outbreak primary classes	all
Correction	CreativeWork	CitationObject	Publication
DataDownload	DataDownload	NIAID schema	Dataset
Eligibility	Thing	NCT Clinical Trials PRS, WHO ClinicalTrials	ClinicalTrial
Instrument	Product	Bioschemas LabProtocol, protocols.io	Protocol
Intervention	Thing	NCT Clinical Trials PRS, WHO ClinicalTrials	ClinicalTrial
MonetaryGrant	MonetaryGrant	NIAID schema	all
Organization	Organization	NIAID schema	all
Outcome	Thing	NCT Clinical Trials PRS, WHO ClinicalTrials	ClinicalTrial
Person	Person	NIAID schema	all
Product	Product	Bioschemas LabProtocol, protocols.io	Protocol
StudyEvent	Thing	NCT Clinical Trials PRS, WHO ClinicalTrials	ClinicalTrial
StudyStatus	Thing	NCT Clinical Trials PRS, WHO ClinicalTrials	ClinicalTrial
StudyDesign	Thing	NCT Clinical Trials PRS, WHO ClinicalTrials	ClinicalTrial
StudyLocation	Place	NCT Clinical Trials PRS, WHO ClinicalTrials	ClinicalTrial

Discussion

In an effort to make scientific resources more FAIR, communities in the biological sciences (bioschemas), earth sciences (science on schema), and more are working diligently to align and influence schema.org to suit the needs of the scientific research community. These communities play an important role in introducing schema.org to the scientific research resource providers and creating tailored schemas more suitable for the research community. Although these communities have helped to create more relevant classes or improve existing classes, it is difficult to push these suggestions to schema.org without compelling use cases or widespread adoption of these tailored classes. The availability of user-friendly tools can improve the adoption of schema.org and other community-driven schema classes, and empower data providers and researchers to engage in schema authoring and sharing.

Most tools for utilizing existing schema.org classes focus on the utilization of an existing schema (such as markup generation) and lack the ability to customize the schema in a schema-compliant way. Tools that do allow customizing/creating a schema (eg- Bioschemas GoWeb) often require some degree of programming. The DDE Schema Playground is a

browser-based tool that enables members of the research community to easily adapt schemas to suit their need and to enable community re-use of their schemas through the DDE schema registry. This encourages and empowers researchers to structure their data in a schema-compliant fashion earlier on in the scientific research process rather than as an afterthought. The schema authoring by the research community, for the research community will encourage the creation and adoption of new classes and properties, which may have previously been neglected due to the absence of representation (ie- expert subject matter volunteers) in data standardization communities. In this fashion, the DDE Schema Playground allows for researchers to express and share their data structuring needs with the data standardization community without diverting attention away from their primary research efforts. Data standardization communities also benefit because their volunteer time can be concentrated on classes already in use by researchers (but could benefit from some standardization), and diverted away from classes which lack interest/support from the research community at large.

We tested the use of the DDE Schema Playground to create customized schema.org-compliant classes that could be used to normalize metadata between multiple types (datasets, clinical trials, publications, etc.) of COVID-19-related resources and applied these schemas towards a searchable resource site (https://outbreak.info). The Outbreak resource schema is available in the DDE schema registry which is also includes schemas from schema.org, BioLink, the National COVID Cohort Collaborative (N3C), the National Institute of Allergy and Infectious Diseases (NIAID) and more. We hope others will join us in making their open data more interpretable, interoperable, and reusable by adding their schemas to the schema registry.

Conclusion

We have created a user-friendly browser-based tool which facilitates the application of schema.org towards biomedical research outputs. We demonstrate its use with the creation of the Outbreak.info schema and encourage others to register and reuse schema.org-compliant schemas.

Availability and requirements

Project name: Data Discovery Engine Schema Playground
Project home page: https://discovery.biothings.io/schema-playground
Operating system(s): Web-based, Platform independent
Programming language: Python
Other requirements: github account for schema editing
License: Creative Commons Attribution 4.0 International license
Any restrictions to use by non-academics: No

List of abbreviations

DDE - Data Discovery Engine
FAIR/FAIRness - Findability, Accessibility, Interoperability, Reusability

N3C - National COVID Cohort Collaborative
NIAID - National Institute of Allergy and Infectious Diseases

Declarations

Ethics approval and consent to participate

Not Applicable

Consent for publication

Not Applicable

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study; however, the Outbreak.info schema generated is available at https://discovery.biothings.io/registry and from gihub at https://github.com/outbreak-info/outbreak.info-resources/tree/master/yaml.

Competing interests

The authors declare that they have no competing interests

Funding

The development of the Data Discovery Engine was supported by the National Center for Advancing Translational Sciences, as part of the National Center for Data to Health (5 U24 TR00230) award to CW. Work on Outbreak.info was supported by National Institute for Allergy and Infectious Diseases (5 U19 AI135995-02), the National Center for Data to Health (5 U24 TR00230) and Centers for Disease Control and Prevention (75D30120C09795).

Authors' contributions

MC developed the front-end of the DDE Schema Playground with feedback from GT, LDH, and JM. XZ and JX developed the backend of the DDE Schema Playground with feedback from MC. CW and AIS guided the overall development of the DDE Schema Playground. GT developed the Outbreak.info schema with feedback from LDH and JM. GT wrote the manuscript with feedback from LDH, AIS, and CW.

Acknowledgments

We thank Ben Rush for his suggestions early on in the development of the Outbreak.info schema.

References

Boeckhout M, Zielhuis GA, Bredenoord AL. The FAIR guiding principles for data stewardship: fair enough? Eur J Hum Genet. 2018 Jul;26(7):931-936. doi: 10.1038/s41431-018-0160-0. Epub 2018 May 17. PMID: 29777206; PMCID: PMC6018669.

Bruskiewich, R., Deepak, Moxon, S. and Mungall, C., 2021. Biolink Model. [online] Biolink Model. DOI:10.5281/zenodo.1242670. Available at: <https://biolink.github.io/biolink-model/> [Accessed 1 September 2021].

Canham S, Ohmann C. A metadata schema for data objects in clinical research. Trials. 2016 Nov 24;17(1):557. doi: 10.1186/s13063-016-1686-5. PMID: 27881150; PMCID: PMC5122021.

CD2H. 2021. Center for Data to Health. [online] Available at: <https://cd2h.org/> [Accessed 10 August 2021].

Franck Michel & The Bioschemas Community. Bioschemas & Schema.org: a Lightweight Semantic Layer for Life Sciences Websites. In proceedings of the Biodiversity Information Standards (TDWG) 2018 Annual Conference, Dunedn, New Zealand, August 2018. (Poster). https://doi.org/10.3897/biss.2.25836

Garcia L, Giraldo O, Garcia A, Rebholz-Schuhmann D. Biotea-2-Bioschemas, facilitating structured markup for semantically annotated scholarly publications. Genomics Inform. 2019 Jun;17(2):e14. doi: 10.5808/GI.2019.17.2.e14. Epub 2019 Jun 19. PMID: 31307129; PMCID: PMC6808630.

Gray A, Gobel C, Jimenez RC, and Bioschemas Community. Bioschemas: From Potato Salad to Protein Annotation. ISWC 2017, CEUR-WS.org, online http://ceur-ws.org/Vol-1963/paper579.pdf

Haendel M, Chute C, Bennett T, Eichmann D, Guinney J, Kibbe W, Payne P, Pfaff E, Robinson P, Saltz J, Spratt H, Suver C, Wilbanks J, Wilcox A, Williams A, Wu C, Blacketer C, Bradford R, Cimino J, Clark M, Colmenares E, Francis P, Gabriel D, Graves A, Hemadri R, Hong S, Hripscak G, Jiao D, Klann J, Kostka K, Lee A, Lehmann H, Lingrey L, Miller R, Morris M, Murphy S, Natarajan K, Palchuk M, Sheikh U, Solbrig H, Visweswaran S, Walden A, Walters K, Weber G, Zhang X, Zhu R, Amor B, Girvin A, Manna A, Qureshi N, Kurilla M, Michael S, Portilla L, Rutter J, Austin C and Gersing K. 2020. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. Journal of the American Medical Informatics Association, 28(3), pp.427-443.

Hollmann S, Kremer A, Baebler Š, Trefois C, Gruden K, Rudnicki WR, Tong W, Gruca A, Bongcam-Rudloff E, Evelo CT, Nechyporenko A, Frohme M, Šafránek D, Regierer B, D'Elia D. The need for standardisation in life science research - an approach to excellence and trust. F1000Res. 2020 Dec 4;9:1398. doi: 10.12688/f1000research.27500.2. PMID: 33604028; PMCID: PMC7863991.

Holub P, Kohlmayer F, Prasser F, Mayrhofer MT, Schlünder I, Martin GM, Casati S, Koumakis L, Wutte A, Kozera Ł, Strapagiel D, Anton G, Zanetti G, Sezerman OU, Mendy M, Valík D, Lavitrano M, Dagher G, Zatloukal K, van Ommen GB, Litton JE. Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health. Biopreserv Biobank. 2018 Apr;16(2):97-105. doi: 10.1089/bio.2017.0110. Epub 2018 Jan 23. PMID: 29359962; PMCID: PMC5906729.

Hruby GW, Hoxha J, Ravichandran PC, Mendonça EA, Hanauer DA, Weng C. A data-driven concept schema for defining clinical research data needs. Int J Med Inform. 2016 Jul;91:1-9. doi: 10.1016/j.ijmedinf.2016.03.008. Epub 2016 Apr 2. PMID: 27185504; PMCID: PMC4873775.

Jauer ML, Deserno TM. Data Provenance Standards and Recommendations for FAIR Data. Stud Health Technol Inform. 2020 Jun 16;270:1237-1238. doi: 10.3233/SHTI200380. PMID: 32570597.

Koesten L, Vougiouklis P, Simperl E, Groth P. Dataset Reuse: Toward Translating Principles to Practice. Patterns (N Y). 2020 Nov 4;1(8):100136. doi: 10.1016/j.patter.2020.100136. PMID: 33294873; PMCID: PMC7691392.

Matthew B. Jones, Stephen Richard, Dave Vieglais, Adam Shepherd, Ruth Duerr, Doug Fils, Lewis McGibbney. (2021). Science-on-Schema.org v1.2.0 (Version 1.2.0). Zenodo. https://doi.org/10.5281/zenodo.4477164

Outbreak.info. 2020. Research Library. [online] Available at: <https://outbreak.info/resources> [Accessed 10 August 2021].

Papadiamantis AG, Klaessig FC, Exner TE, Hofer S, Hofstaetter N, Himly M, Williams MA, Doganis P, Hoover MD, Afantitis A, Melagraki G, Nolan TS, Rumble J, Maier D, Lynch I. Metadata Stewardship in Nanosafety Research: Community-Driven Organisation of Metadata Schemas to Support FAIR Nanoscience Data. Nanomaterials (Basel). 2020 Oct 15;10(10):2033. doi: 10.3390/nano10102033. PMID: 33076428; PMCID: PMC7602672.

Profiti G, Jimenez RC, Zambelli F et al. Using community events to increase quality and adoption of standards: the case of Bioschemas [version 1; not peer reviewed]. F1000Research 2018, 7(ELIXIR):1696 (document) (https://doi.org/10.7490/f1000research.1116233.1)

Sansone SA, Gonzalez-Beltran A, Rocca-Serra P, Alter G, Grethe JS, Xu H, Fore IM, Lyle J, Gururaj AE, Chen X, Kim HE, Zong N, Li Y, Liu R, Ozyurt IB, Ohno-Machado L. DATS, the data

tag suite to enable discoverability of datasets. Sci Data. 2017 Jun 6;4:170059. doi: 10.1038/sdata.2017.59. PMID: 28585923; PMCID: PMC5460592.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18. Erratum in: Sci Data. 2019 Mar 19;6(1):6. PMID: 26978244; PMCID: PMC4792175.