

Perspective: Towards Automated Tracking of Content and Evidence Appraisal of Nutrition Research

Chen Yang,¹ Dana Hawwash,¹ Bernard De Baets,² Jildau Bouwman,³ and Carl Lachat¹

¹Department of Food Technology, Safety and Health, Ghent University, Ghent, Belgium; ²KERMIT (Research Unit of Knowledge-based Systems), Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium; and ³Netherlands Organization for Applied Scientific Research, Zeist, The Netherlands

ABSTRACT

Robust recommendations for healthy diets and nutrition require careful synthesis of available evidence. Given the increasing volume of research articles generated, the retrieval and synthesis of evidence are increasingly becoming laborious and time-consuming. Information technology could help to reduce workload for humans. To guide supervised learning however, human identification of key study characteristics is necessary. Reporting guidelines recommend that authors include essential content in articles and could generate manually labeled training data for automated evidence retrieval and synthesis. Here, we present a semiautomated approach to annotate, link, and track the content of nutrition research manuscripts. We used the STROBE extension for nutritional epidemiology (STROBE-nut) reporting guidelines to manually annotate a sample of 15 articles and converted the semantic information into linked data in a Neo4j graph database through an automated process. Six summary statistics were computed to estimate the reporting completeness of the articles. The content structure, presence of essential study characteristics as well as the reporting completeness of the articles are visualized automatically from the graph database. The archived linked data are interoperable through their annotations and relations. A graph database with linked data on essential study characteristics can enable Natural Language Processing in nutrition. *Adv Nutr* 2020;11:1079–1088.

Keywords: STROBE-nut, reporting guidelines, graph database, research semantics; ontology; standardization

Introduction

Unhealthy diets and poor nutrition are the leading risk factors for poor health worldwide (1). Recommendations to improve diets and nutrition require the rigorous and timely assessment of evidence (2). A systematic review of published research articles is an essential process to summarize the evidence, but involves time-consuming processes such as literature retrieval, review, and data extraction (3). Machine learning and ontologies are increasingly used to classify, store, and retrieve research output (4, 5). Various scholars are piloting (semi-) automated methods of evidence synthesis (6). Natural Language Processing (NLP) enables computers to process human language and could help to extract relevant content from research articles (6–8). NLP typically operates in a supervised manner, using machine learning models that are trained using manually labeled data. The required labeled data for NLP could be provided by authors that annotate articles according to reporting guidelines.

To include essential information in research output, authoritative reporting guidelines (9) are widely used

in biomedical sciences. An extension of the STROBE-Nutritional Epidemiology (STROBE-nut) statement (10) was developed to enhance the reporting completeness of nutrition research (11). The use of reporting guidelines is recommended by many journals, including those in nutrition (12). Authors are requested to provide information on the presence of essential content in the text during manuscript submission. This information, typically submitted as supplementary material, becomes redundant after peer reviewing.

There is considerable interest in developing a virtual research infrastructure to advance food and nutrition research (5), and such an infrastructure could enable large-scale data analysis. So far, most efforts have been directed at the annotation, storage, and reuse of individual-level numeric data (e.g., anthropometric and food intake data of individuals). However, research article text represents a wealth of accumulated knowledge and evidence. Additional efforts are needed to annotate and store the text of research articles in a virtual research infrastructure. To enhance

(re)use of research output, the FAIR principles are proposed to ensure data are Findable, Accessible, Interoperable, and Reusable. The FAIR principles are also appropriate to manage the text of research articles (13).

Here, we developed and tested a semiautomated approach to retrieve, harmonize, and analyze the reporting completeness of the articles reported according to STROBE-nut. By annotating the articles according to the STROBE-nut reporting guidelines, the essential characteristics of articles are converted into identifiers that can be processed by computers. When applied at scale, the approach could facilitate retrieval, archiving, accession, and use of nutrition research knowledge in a global research environment. **Supplemental Table 1** clarifies the terms used from computer science.

Methods

Annotation of articles

We used a sample of scientific articles that were reported according to the STROBE-nut reporting guidelines (10) (**Supplemental Table 2**). On 9 April 2019, the 82 citations of 2 STROBE-nut publications (10, 14) were retrieved by DH using Google Scholar (15). Articles that were duplicates ($n = 2$), inaccessible ($n = 5$), gray literature ($n = 6$), published in languages other than English ($n = 2$), not citing STROBE-nut ($n = 2$), and citing STROBE-nut as background information ($n = 47$) were excluded. The remaining articles ($n = 18$) citing STROBE-nut as a guideline for reporting were selected for this study. Finally, 15 of these articles (16–30) were included as they were made available under the Creative Commons Attribution 4.0 License (31), which permits the unrestricted use of article content (**Supplemental Table 3**). All articles were published in Extensible Markup Language (XML) for automated data extraction.

DH and CL reviewed the content and indicated manually where the STROBE-nut items were reported in the text.

Supported by the European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI) which is part of the Joint Programming Initiative (JPI) “A healthy diet for a healthy life” and is funded by national funding agencies in 9 European countries. The work in this manuscript is funded by The Research Foundation - Flanders, grant number GOD4815N. CY is funded by a scholarship from the Chinese Scholarship Council. CL received funding from Bioversity International for the work on standards for dietary assessment. Funders had no role in the conduct of this study or in the content of this article.

Author disclosures: The authors report no conflicts of interest. CL and DH are part of the STROBE extension for nutritional epidemiology (STROBE-nut) coordination group and were authors of the STROBE-nut guidance articles.

Perspective articles allow authors to take a position on a topic of current major importance or controversy in the field of nutrition. As such, these articles could include statements based on author opinions or point of view. Opinions expressed in Perspective articles are those of the author and are not attributable to the funder(s) or the sponsor(s) or the publisher, Editor, or Editorial Board of *Advances in Nutrition*. Individuals with different positions on the topic of a Perspective are invited to submit their comments in the form of a Perspectives article or in a Letter to the Editor.

Supplemental Tables 1–4 are available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at <https://academic.oup.com/advances/>.

Address correspondence to CL (e-mail: Carl.Lachat@UGent.be).

Abbreviations used: API, Application Programming Interface; DIKW, Data, Information, Knowledge, Wisdom; DOI, Digital Object Identifier; EQUATOR, Enhancing the QUALity and Transparency Of health Research; FAIR, Findable, Accessible, Interoperable, Reusable; NLP, Natural Language Processing; STROBE-nut, STROBE extension for nutritional epidemiology; XML, Extensible Markup Language; UI, user interface.

Five of the included articles were randomly selected by DH. CL and DH reviewed the reporting completeness of the 5 articles independently according to the STROBE-nut reporting guidelines, and highlighted the placement of the article content described according to the STROBE-nut items. Disagreements were resolved through discussion until a consensus was reached. The remaining articles were assessed by DH and uncertainties were resolved based on the consensus among DH and CL.

Identification of components to enable automated tracking

We reviewed components required to enable automated tracking of nutritional research in a graph database. The components include existing concepts for information management, research infrastructures, programming languages, software, publisher interfaces, visualization tools, etc.

Automated processing of annotations

A Python (32) module was developed to process the structure and STROBE-nut annotations of the articles in XML format (<https://github.com/cyang0128/Nutritional-epidemiologic-ontologies/tree/master/strobenut>). Three types of functions are included in the module: 1) “Input.py”: provides 2 functions to extract and store article metadata through an Application Programming Interface (API; i.e., a virtual interface used to retrieve the requested data from web servers) or in local XML files respectively; 2) “Annotate.py”: annotates the reporting completeness of articles stored in a Neo4j graph database and reported according to STROBE-nut reporting guidelines; and 3) “Figure.py”: shows the statistics of reporting completeness of article(s) and the reporting frequency of different STROBE-nut items. All code was tested in Python (3.7.4) through Jupyter Notebook (6.0.1) (33).

Development of a graph database for semantic information

We conducted 3 subsequent actions to illustrate the development and application of a graph database for the management of reporting completeness of articles according to the STROBE-nut reporting guidelines.

Step 1: automated information extraction and storage in the graph database.

The metadata of articles such as “title,” “abstract,” “Digital Object Identifier” (DOI), “keywords,” and content structure (e.g., Background, Methods, Results, Discussion, etc.) were retrieved and extracted by using Python. For the articles ($n = 10$) published by “Springer Nature,” the metadata was obtained from the “Springer Nature API Portal” (34) (collection: “openaccess,” result format: “jats”). For the remaining articles ($n = 5$), the corresponding XML files were downloaded and accessed through a local path. The following functions defined in 3 existing Python modules were used to extract the metadata: 1) “lxml” (35) and “re” (36) to locate article metadata in XML and

2) “py2neo” (37) to store the extracted data in a Neo4j knowledge graph database. Several classes and relations were used to explain the extracted metadata as well as their relations:

Classes (i.e., categories of terms):

- 1) “Article” describes the type of scientific article. The 15 selected articles were arranged under this class, and their titles and DOIs were used as “property” to identify the articles uniquely.
- 2) “Abstract” describes the summary of a scientific article.
- 3) “Keyword/Keywords set” describes a word/a set of words about the scientific area of a scientific article.
- 4) “Section/Subsection” describes different parts of a scientific article such as “Methods,” “Results,” etc. as well as their subsections (e.g., statistical method).

Relations:

- 1) “section/subsection” describes the relation between a text and the resource of the text physically. Therefore, “section” was used to describe the relation between an article and the article’s sections (e.g., Methods), and “subsection” was used to describe the relation between sections of the article (e.g., Methods) and its subsections (e.g., participants recruitment).
- 2) “hasAbstract” describes the relation between an article and its abstract.
- 3) “hasKeyword” describes the relation between an article and its set of keywords.

Step 2: annotation of reporting guidelines in the graph database.

To archive information regarding the reporting completeness of the selected articles, Python functions defined in “Py2neo” (37) and “Cypher” (38) were used to construct the new function (i.e., “Annotate.py”) in Python. The function converts the annotation on the article to the virtual annotation in the graph database. Several classes and relations were selected from existing ontologies. An ontology represents a set of categories and terms that are interlinked by relations and annotated by properties. In the present study, the ontology is used to describe the annotation done according to STROBE-nut reporting guidelines:

Classes (categories):

- 1) “STROBE-nut article” indicates that an article is reported according to the STROBE-nut reporting guidelines. Therefore, the 15 selected articles were arranged under this class as well, and their DOIs were used as a “property” to identify the articles uniquely.
- 2) “STROBE-nut section/subsection” describes different parts of a scientific article such as “Methods,” “Results,” etc. as well as the subsections proposed in the STROBE-nut reporting guidelines.
- 3) “STROBE-nut item” describes the 24 items proposed in the STROBE-nut reporting guidelines to guide transparent reporting (39).

Relations:

- 1) “STROBE-nut” is chosen to link the class “Article” and “STROBE-nut article” defined in the graph database. Semantically, it clarifies that the article is reported according to the set of the STROBE-nut items in the STROBE-nut reporting guidelines.
- 2) “section/subsection” is used to describe the relations between “STROBE-nut article,” “STROBE-nut section,” and “STROBE-nut subsection.”

Step 3: automated calculation of reporting characteristics.

Statistics of reporting completeness for articles were generated from the graph database. The following functions were included: 1) visualization of the reporting completeness of a single article; 2) representation of the overall reporting completeness for the meta-analysis of the articles, and 3) representation of the reporting statistics of the STROBE-nut items. Using the STROBE-nut reporting guidelines, the graph database of the 15 articles was developed using the same classes (categories) and relations, which enables automatic calculation of these statistics.

Python functions (i.e., functions defined in the “Figure.py”) were executed for the 15 articles to generate the figures. We used an example for each function to illustrate the approach. First, 4 bar charts are used to show the articles with the highest/lowest reporting completeness, and the frequently/rarely reported STROBE-nut items, respectively. Second, a radar chart shows the reporting completeness of a single article. The reporting completeness of information in 5 sections proposed in the STROBE-nut reporting guidelines (i.e., Title/Abstract, Methods, Results, Discussion, Other information) is presented in percentages, and the overall reporting completeness among the 5 sections is also shown on the top of the radar chart. Third, a pie chart is used to show an example of reporting completeness classification.

Results

A graph database for nutrition research knowledge

A graph database is a database that uses relations to link terms to archive and represent semantic information (e.g., phrases or sentences) (40, 41). The Semantic Web, as an extension of the World Wide Web, is being developed as the most comprehensive graph database to handle global data. Unique identifiers are given to identify terms as well as their relations. By tracking the identifiers, semantic information in a graph database can be accessed, harmonized, integrated, and visualized (42). For instance, the DOI was introduced in 2000 as a persistent and unique identifier of a scientific publication (43). The DOI of a research article is an indirect link to the article and also provides information regarding authors, journal, etc. through an API (40).

Neo4j has been the most commonly used graph database for several years (44). Moreover, Neo4j is compatible with the Python programming environment through the module

“neo2py,” which can be integrated in other Python modules (e.g., plotting library, web framework, rule engine, etc.) for the visualization, statistics calculation, etc. of a graph database.

Figure 1 shows the components that can be used to develop a graph database for nutrition articles, i.e., an underpinning theory, software, research infrastructures, and publisher interfaces. A more detailed description of the technology used is included as **Supplemental Table 4**.

To convert the semantic information of nutrition articles into code, the semantic information needs to be reorganized in a machine-readable format. The DIKW (Data, Information, Knowledge, Wisdom) pyramid is a classical model for information classification in computer science (45). For nutritional epidemiology, the DIKW pyramid is a suitable approach to manage information (41), and provides the theoretical basis to classify the semantic information processed and stored in a graph database. In this study, the semantic information of the articles is classified as “Information” in the DIKW pyramid in **Figure 1**.

To manage the semantic information of nutrition research in a graph database, reporting guidelines provide a template to define relations between the terms. EQUATOR (Enhancing the QUALity and Transparency Of health Research) (46) and FAIRsharing (47) are recommended as repositories to retrieve relevant reporting guidelines. Being the most comprehensive ontology repository of medical science, BioPortal (48) is recommended for retrieving the corresponding ontologies. For this study, Ontology for Nutritional Epidemiology (ONE) (39) was used.

To retrieve the semantic information of scientific articles, various publishers [e.g., Multidisciplinary Digital Publishing Institute (MDPI) (49), Frontiers (50), and Public Library of Science (PLOS) (51)] provide user interfaces (UI). The UIs enable interactions between a human user and a computer system to retrieve scientific articles in machine-readable formats. In addition, Springer Nature provides an API to facilitate data extraction (34). Python (32) was selected to manage the semantic information in the graph database.

Graph database visualization

A set of linked data is visualized in the Neo4j browser (37). We describe the visualization for only 1 article (17) as the graphs are similar for all scientific articles in nutritional epidemiology. **Figure 2** shows the content structure of the scientific article stored in the graph database. The metadata of this article is visualized using different colors and connected by human- and machine-readable relations. The article consists of 5 sections, 11 subsections (6 subsections under “Methods” and 5 subsections under “Discussion”), 1 abstract, and a set of 3 keywords. Different relations were used to connect all the components extracted from its XML file. **Figure 3** shows the annotation of the article’s reporting completeness according to the STROBE-nut reporting guidelines. Out of a total of 24 STROBE-nut

items, 18 items are reported in the article, as indicated with brown circles. The sections and subsections proposed in the STROBE-nut reporting guidelines are visualized using red and blue circles, respectively. A class labeled as “Harmonized article” was used to gather all the STROBE-nut annotation components. In addition, the relation “STROBE-nut” is used to connect the STROBE-nut annotations of the article text and the content structure of the article. Through the annotation, the proposed sections of the manuscript in which the STROBE-nut items were reported (i.e., title/abstract, methods, result, discussion in the present article) are defined.

Automated calculation of reporting completeness statistics

In **Figure 4**, 6 different python functions are presented to summarize the reporting completeness of the articles ($n = 15$) stored in the Neo4j database. **Figure 4.1** and **4.2** illustrate the articles with the highest and lowest reporting completeness according to the STROBE-nut reporting guidelines. Among the 15 articles, the maximum number of STROBE-nut items reported is 18 and the minimum is 5. The DOIs of the articles are shown and link each article with the presence of essential study characteristics.

Figure 4.3 shows the reporting completeness of 1 article (24). The STROBE-nut items in the section “Title/Abstract” and “Other sections” are fully reported. About 45% STROBE-nut items are reported in “Results,” and <40% STROBE-nut items are reported in “Methods.” There are no STROBE-nut items reported in the section “Discussion.” Overall, ~40% STROBE-nut items are reported in the article.

Figure 4.4 and **4.5** present the frequency of reporting for each STROBE-nut item in the Neo4j database. Among the 24 STROBE-nut items, nut-8.1 (15/15), nut-22.1 (15/15), and nut-8.5 (14/15) are frequently reported, whereas nut-8.4 (1/15), nut-17 (3/15), and nut-12.1 (3/15) are rarely reported. Moreover, nut-12.3 (0/15) is not reported in any of the selected articles.

Figure 4.6 illustrates the proportion of articles (percent) organized in 4 quantiles of reporting completeness. More than half (60%) of the articles reported 50–75% STROBE-nut items, whereas ~7% of the articles in the database reported <25% STROBE-nut items.

Discussion

Here, we described an approach to apply the STROBE-nut reporting guidelines as digital annotations for research articles in nutritional epidemiology. We demonstrated a feasible and semiautomated method to annotate, retrieve, archive, link, and process essential study characteristics and the reporting completeness of nutrition research. To the best of our knowledge, this is the first attempt to apply the FAIR principles to reporting guidelines in biomedical sciences.

The present work illustrates the potential to use information on article content as a metric to assess articles. Previous attention has been directed to article-level instead of

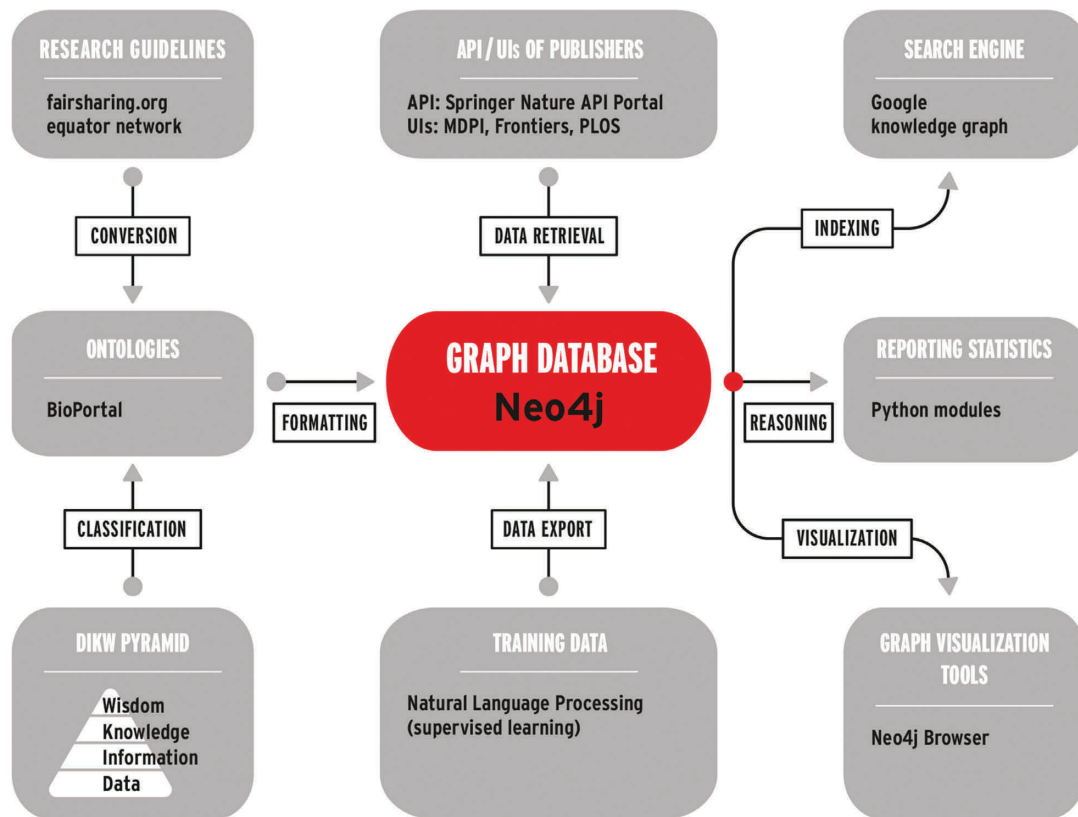


FIGURE 1 The development, enrichment, and implementation of a graph database to track nutritional research knowledge. API, Application Program Interface; DIKW, Data, Information, Knowledge, Wisdom; Neo4j, the Community Edition of the Neo4j graph database management system; MDPI, Multidisciplinary Digital Publishing Institute; PLOS, Public Library of Science; UIs, user interfaces.

journal-level indicators to assess the added value of research (52, 53). These metrics, known as altmetrics, however, currently mainly deal with scholarly citations and references in social media (54, 55). The present approach paves the way for an appreciation of article content regarding its reporting completeness. For instance, a radar chart (Figure 4.3) could visualize the metric for a single article, and indicate the reporting completeness of different sections (e.g., title, introduction, methods, results, etc.) of an article. However, for literature reviews, it might be more appropriate to use summary statistics (Figure 4) to describe the overall reporting completeness of selected articles.

The present study used annotations done by experts involved in the development of STROBE-nut (DH and CL). For application at scale however, authors of articles are best placed to provide the annotations. Increasingly, nutrition journals encourage authors to use article-based reporting guidelines such as STROBE-nut to provide information on the reporting completeness of their scientific articles (12). To support authors, publishers could provide a user-friendly UI (e.g., a website) to collect data on reporting completeness during the submission process. The utility of reporting guidelines to date is mainly geared at improving the completeness of the manuscript at the time of submission and facilitation of peer review. Converting information on

reporting completeness to linked content of articles enables new applications that add value for authors, journals, and search engines to classify and retrieve information. The approach presented is the first use case in this regard and could unlock new applications to increase the use of and adherence to reporting guidelines.

For application at scale, machine learning approaches can automate the annotation approach and generate annotations directly from the submitted manuscript text. A fully automated approach however, may affect the accuracy of the annotations and metrics. In a first instance, machine learning approaches should be used to suggest relevant terms to facilitate the annotation process during submission. Feedback from authors on the suggested terms will be collected to verify the quality of the annotation process. Gradually, an iterative process between the machine learning suggestions and author feedback will improve the accuracy of the automatic annotation process.

The current work contributes to other initiatives. For example, the “Springer Nature SciGraph” (34) is a Linked Open Data (i.e., linked and freely available online data) platform that aggregates data from publications of Springer Nature. “Springer Nature SciGraph” links data from reliable sources and presents how information is interconnected in the articles. Moreover, Elsevier provides an API, “the

\$ MATCH (n:Article{DOI:'10.1186/s12966-017-0567-y'}) RETURN n

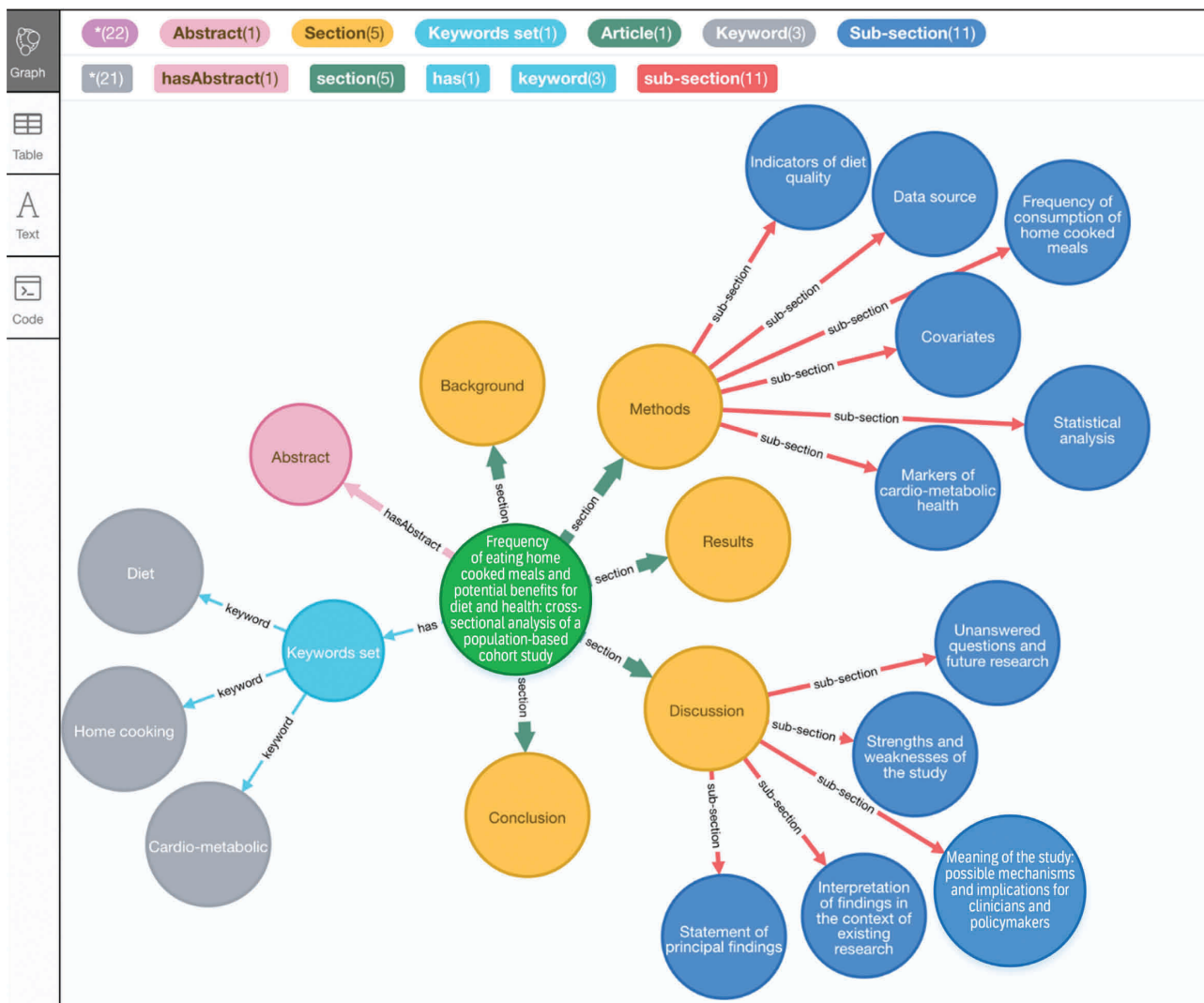


FIGURE 2 Visualization of 1 article's metadata (17) in the graph database. STROBE-nut, STROBE extension for nutritional epidemiology.

Elsevier Developer,” for developers to gain access to linked data of journals, books, data, abstracts, etc. published by Elsevier (56). Here, the use of linked data for study reporting completeness could be instrumental to manage the content of scientific publications thereby improving the findability and accessibility of research findings.

To promote the use of a graph database, the encoding of XML-based scientific articles according to reporting guidelines is an interesting prospect. XML is a markup language with setting rules and vocabulary identifiers. It produces documents that are both human-readable and machine-readable. Using the STROBE-nut ontology to annotate scientific articles in XML, the annotated nutritional study content enables knowledge queries, integration, and inferences. In our study, essential study characteristics of the articles were converted to the machine-readable STROBE-nut identifiers. Users can use these identifiers as well as their

relations to find, filter, and integrate the content of the articles in a virtual research infrastructure. Moreover, rules can be developed to reason on the annotated study characteristics. For instance, when using the reported STROBE-nut items as inclusion and exclusion criteria for a systematic review, customized reasoning rules can be set to automate the selection of articles.

According to the quality rating of linked data introduced by Berners-Lee (57), the graph database presented here obtains a “4 star” rating: the database consists of human-readable information of the used vocabulary; is machine-readable and enables semantic reasoning; has been linked to existing linked data sources (e.g., DOIs, etc.); and finally, the metadata about the vocabulary (i.e., definitions, uses, authors, references of reporting guidelines, etc.) are available. To achieve the full “5 star” status for linked data, the developed graph database should be connected to linked data

\$ MATCH (n:`STROBE-nut article`{DOI:'10.1186/s12966-017-0567-y'}) RETURN n



FIGURE 3 Visualization of the STROBE-nut annotations of an article (17) in the graph database. STROBE-nut, STROBE extension for nutritional epidemiology.

networks that are in use. This would require combined efforts from different stakeholders involved in the development of research guidelines, and the production and publication of research findings. Research groups such as the EQUATOR network (46) and FAIRsharing (58) should promote the submission of reporting guidelines as linked data, and make the linked data available online for people to see, understand, adopt, and use. DOI Registration Agencies could register the reporting completeness information as linked data of DOI

and enable the API for information query and linked data reuse.

As an application for nutrition research, we used a sample of articles that reported using STROBE-nut. The approach presented can relatively easily be expanded to other reporting guidelines such as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (59), STROBE (60), CONSORT (CONsolidated Standards of Reporting Trials) (61), or ARRIVE (Animal Research: Reporting of In Vivo

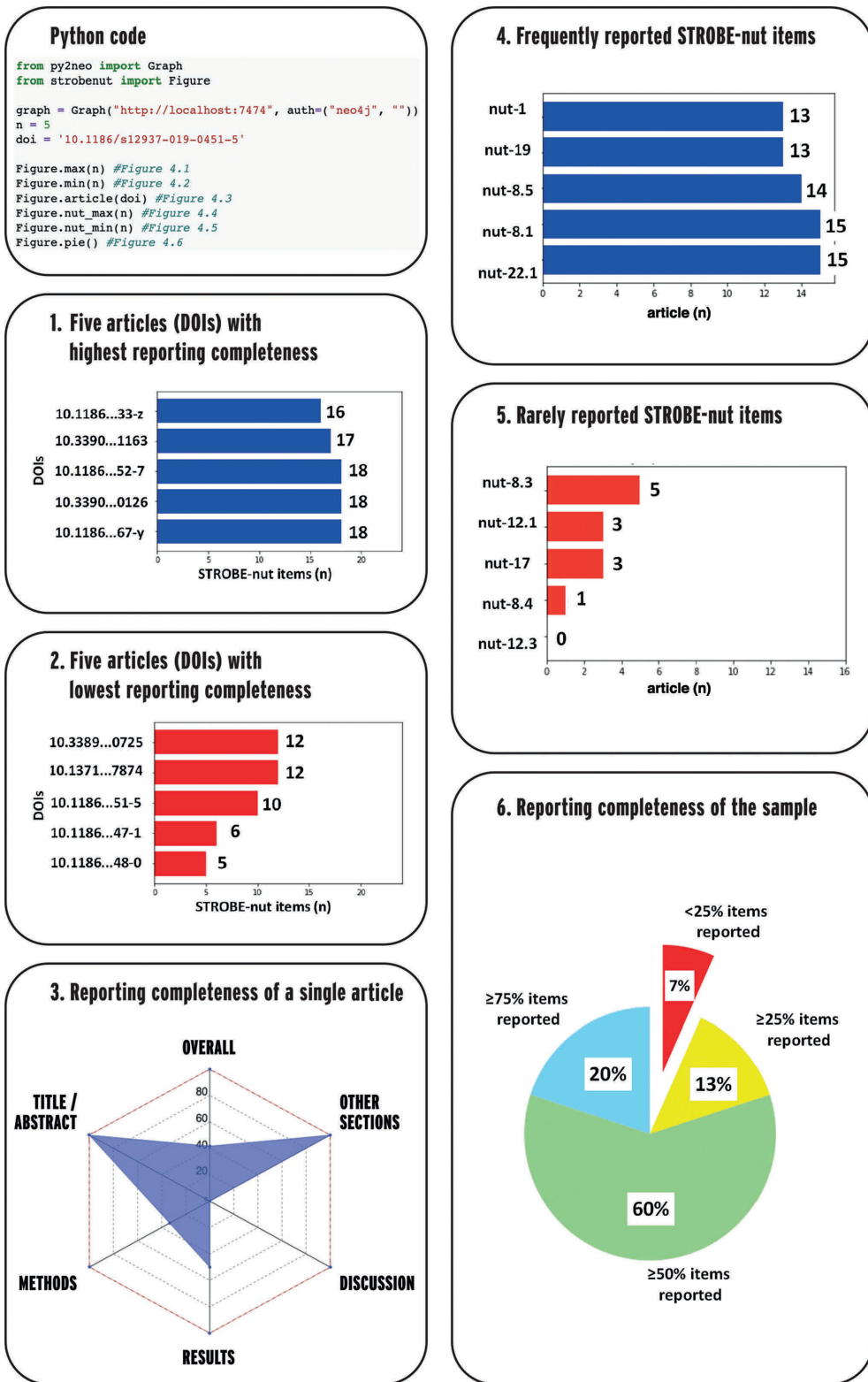


FIGURE 4 Statistics regarding article reporting completeness and essential study content from the graph database. Legend: DOI: Digital Object Identifier. Full DOIs are available from Supplementary Table 3. STROBE-nut, STROBE extension for nutritional epidemiology.

Experiments) (62). This would, however, require the development of relevant ontologies and substantial involvement of other guideline developers, which was beyond the scope of the present study.

To apply the present approach at scale, a culture of linked data in nutrition research needs to be fostered. Eventually, a new workforce of researchers will be required to apply information technology for nutrition research. Basic knowledge regarding the use of ontologies, open science, and FAIR data needs to be integrated in the curriculum of students and researchers.

Acknowledgments

We thank Patrick Kolsteren for comments on an earlier version of the manuscript.

The authors' responsibilities were as follows— CY, CL, and BDB: conceptualization; CY, DH, and CL: investigation; CY and CL: methodology; CY: software; CL: supervision; CY: writing of original draft; CY, DH, BDB, JB, and CL: review and editing; and all authors: read and approved the final manuscript.

References

1. Gakidou E, Afshin A, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abdulle AM, Abera SF, Aboyans V, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet North Am Ed* 2017;390(10100): 1345–422.
2. Blake P, Durao S, Naude CE, Bero L. An analysis of methods used to synthesize evidence and grade recommendations in food-based dietary guidelines. *Nutr Rev* 2018;76(4):290–300.
3. Bragge P, Clavisi O, Turner T, Tavender E, Collie A, Gruen RL. The Global Evidence Mapping Initiative: scoping research in broad topic areas. *BMC Med Res Methodol* 2011;11:92.
4. Brown P, RELISH Consortium, Zhou YQ. Large expert-curated database for benchmarking document similarity detection in biomedical literature search. *Database* 2019;2019:baz085.
5. Snoek HM, Eijssen LMT, Geurts M, Vors C, Brown KA, Bogaardt MJ, Dhonukshe-Rutten RAM, Evelo CT, Fezeu LK, Finglas PM, et al. Advancing food, nutrition, and health research in Europe by connecting and building research infrastructures in a DISH-RI: results of the EuroDISH project. *Trends Food Sci Tech* 2018;73: 58–66.
6. Marshall IJ, Johnson BT, Wang Z, Rajasekaran S, Wallace BC. Semi-automated evidence synthesis in health psychology: current methods and future prospects. *Health Psychol Rev* 2020;14(1): 145–58.
7. Marshall C, Brereton P. Systematic review toolbox. *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering - EASE '15* 2015. p. 1–6.
8. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. *J Am Med Inform Assoc* 2016;23(1):193–201.
9. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, Howells DW, Ioannidis JPA, Oliver S. How to increase value and reduce waste when research priorities are set. *Lancet North Am Ed* 2014;383(9912):156–65.
10. Lachat C, Hawwash D, Ocke MC, Berg C, Forsum E, Hornell A, Larsson C, Sonestedt E, Wirfalt E, Akesson A, et al. Strengthening the Reporting of Observational Studies in Epidemiology-Nutritional Epidemiology (STROBE-nut): an extension of the STROBE statement. *PLoS Med* 2016;13(6):e1002036.
11. Roberts JL, Houle TT, Loder EW, Penzien DB, Turner DP, Rothrock JF. Establishing a Coherent Reporting Guidelines Policy in Health Journals. *Guidelines for Reporting Health Research: A User's Manual*. Chichester, United Kingdom: John Wiley & Sons, Ltd.; 2014.
12. Lachat C. STROBE-nut: towards better reporting in nutrition research [Internet]. [cited 1 April, 2020]. Available from: <http://www.strobe-nut.org/>.
13. Stall S, Yarmey L, Cutcher-Gershenfeld J, Hanson B, Lehnert K, Nosek B, Parsons M, Robinson E, Wýborn L. Make scientific data FAIR. *Nature* 2019;570(7759):27–9.
14. Lachat C, Hawwash D, Ocke MC, Berg C, Forsum E, Hornell A, Larsson CL, Sonestedt E, Wirfalt E, Akesson A, et al. Strengthening the reporting of observational studies in epidemiology - nutritional epidemiology (STROBE-nut): an extension of the STROBE statement. *Nutr Bull* 2016;41(3):240–51.
15. Google. Google Scholar [Internet]. [cited 9 April, 2019]. Available from: <https://scholar.google.com/>.
16. Chatelan A, Beer-Borst S, Randriamiharisoa A, Pasquier J, Blanco JM, Siegenthaler S, Paccaud F, Slimani N, Nicolas G, Camenzind-Frey E, et al. Major Differences in Diet across Three Linguistic Regions of Switzerland: Results from the First National Nutrition Survey menuCH. *Nutrients* 2017;9(11):1163.
17. Mills S, Brown H, Wrieden W, White M, Adams J. Frequency of eating home cooked meals and potential benefits for diet and health: cross-sectional analysis of a population-based cohort study. *Int J Behav Nutr Phys Act* 2017;14(1):109.
18. Lam MCL, Adams J. Association between home food preparation skills and behaviour, and consumption of ultra-processed foods: Cross-sectional analysis of the UK National Diet and nutrition survey (2008–2009). *Int J Behav Nutr Phys Act* 2017;14(1):68.
19. Chatelan A, Castetbon K, Pasquier J, Allemann C, Zuber A, Camenzind-Frey E, Zuberbuehler CA, Bochud M. Association between breakfast composition and abdominal obesity in the Swiss adult population eating breakfast regularly. *Int J Behav Nutr Phys Act* 2018;15(1):115.
20. Llanaj E, Adány R, Lachat C, D'Haese M. Examining food intake and eating out of home patterns among university students. *PLoS One* 2018;13(10):e0197874.
21. Abris GP, Kim NH, Provido SMP, Hong S, Yu SH, Lee CB, Lee JE. Dietary diversity and nutritional adequacy among married Filipino immigrant women: The Filipino Women's Diet and Health Study (FiLWHEL). *BMC Public Health* 2018;18(1):359.
22. Teasdale SB, Ward PB, Jarman R, Wade T, Rossimel E, Curtis J, Lappin J, Watkins A, Samaras K. Is Obesity in Young People With Psychosis a Foregone Conclusion? Markedly Excessive Energy Intake Is Evident Soon After Antipsychotic Initiation. *Front Psychiatry* 2018;9:725.
23. Vissers LET, Rijkssen J, Boer JMA, Verschuren WMM, van der Schouw YT, Sluijs I. Fatty acids from dairy and meat and their association with risk of coronary heart disease. *Eur J Nutr* 2019;58(7):2639–2647.
24. Pendergast FJ, Livingstone KM, Worsley A, McNaughton SA. Examining the correlates of meal skipping in Australian young adults. *Nutr J* 2019;18(1):24.
25. Hausken SES, Lie HC, Lien N, Sleddens EFC, Melbye EL, Bjelland M. The reliability of the general functioning scale in Norwegian 13–15-year-old adolescents and association with family dinner frequency. *Nutr J* 2019;18(1):20.
26. Pestoni G, Krieger JP, Sych JM, Faeh D, Rohrmann S. Cultural Differences in Diet and Determinants of Diet Quality in Switzerland: Results from the National Nutrition Survey menuCH. *Nutrients* 2019;11(1):126.
27. Krieger JP, Pestoni G, Cabaset S, Brombach C, Sych J, Schader C, Faeh D, Rohrmann S. Dietary Patterns and Their Sociodemographic and Lifestyle Determinants in Switzerland: Results from the National Nutrition Survey menuCH. *Nutrients* 2018;11(1):62.
28. Schwedhelm C, Iqbal K, Schwingshackl L, Agogo GO, Boeing H, Knuppel S. Meal analysis for understanding eating behavior:

- meal- and participant-specific predictors for the variance in energy and macronutrient intake. *Nutr J* 2019;18(1):15.
29. Clifford Astbury C, Penney TL, Adams J. Comparison of individuals with low versus high consumption of home-prepared food in a group with universally high dietary quality: a cross-sectional analysis of the UK National Diet & Nutrition Survey (2008-2016). *Int J Behav Nutr Phys Act* 2019;16(1):9.
 30. Workicho A, Belachew T, Argaw A, Ghosh S, Kershaw M, Lachat C, Kolsteren P. Adolescent pregnancy and linear growth of infants: a birth cohort study in rural Ethiopia. *Nutr J* 2019;18(1):22.
 31. Creative Commons. Attribution 4.0 International (CC BY 4.0) [Internet]. [cited 17 August, 2019]. Available from: <https://creativecommons.org/licenses/by/4.0/>.
 32. Oliphant TE. Python for scientific computing. *Comput Sci Eng* 2007;9(3):10–20.
 33. Kluyver T, Ragan-Kelley B, Pérez F, Granger BE. Jupyter Notebooks - A Publishing Format for Reproducible Computational Workflows. Paper published in Proceedings of the 20th International Conference on Electronic Publishing Positioning and Power in Academic Publishing: Players, Agents and Agendas, Göttingen, Germany, June 7, 2016 – June 9, 2016. IOS Press; 2016.
 34. Springer Nature. Springer Nature API Portal [Internet]. [cited 22 October, 2019]. Available from: <https://dev.springernature.com/>.
 35. Richter S. lxml - XML and HTML with Python [Internet]. [cited 22 October, 2019]. Available from: <https://lxml.de/>.
 36. Python. re — Regular expression operations [Internet]. [cited 22 October, 2019]. Available from: <https://docs.python.org/3/library/re.html>.
 37. Neo4j. The Py2neo v4 Handbook [Internet]. [cited 22 October, 2019]. Available from: <https://py2neo.org/v4/#library-reference>.
 38. Holzschuher F, Peinl R. Performance of graph query languages. 2013:195.
 39. Yang C, Ambayo H, Baets B, Kolsteren P, Thanintorn N, Hawwash D, Bouwman J, Bronselaer A, Pattyn F, Lachat C. An ontology to standardize research output of nutritional epidemiology: from paper-based standards to linked content. *Nutrients* 2019;11(6):1300.
 40. Summers E. DOIs as Linked Data [Internet]. [cited 9 August, 2019]. Available from: <https://inkdroid.org/2011/04/25/does-as-linked-data/>.
 41. Yang C, Baets BD, Lachat C. From DIKW pyramid to graph database: a tool for machine processing of nutritional epidemiologic research data. 2019 IEEE International Conference on Big Data (Big Data). Los Angeles: IEEE; 2019. p. 5202–5.
 42. Heath T, Bizer C. Linked data: evolving the web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology* 2011;1(1):1–136.
 43. Bates MJ, Maack MN. *Encyclopedia of Library and Information Sciences*. Third edition ed. Boca Raton: CRC Press; 2009.
 44. DB-Engines Ranking of Graph DBMS [Internet]. [cited 31 December, 2019]. Available from: <https://db-engines.com/en/ranking/graph+dbms>.
 45. Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science* 2007;33(2):163–80.
 46. The Centre for Statistics in Medicine (CSM). Equator: Enhancing the QUALity and Transparency Of health Research [Internet]. [cited 17 September, 2018]. Available from: <http://www.equator-network.org/library/translations-of-reporting-guidelines/>.
 47. FAIRsharing.org [Internet]. [cited 11 August, 2019]. Available from: <https://fairsharing.org/>.
 48. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, Jonquet C, Rubin DL, Storey MA, Chute CG, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 2009;37(Web Server issue):W170–3.
 49. MDPI-publisher of open-access journals [Internet]. [cited 31 December, 2019]. Available from: <https://www.mdpi.com/>.
 50. Frontiers - Open Access Journals [Internet]. [cited 31 December, 2019]. Available from: <https://www.frontiersin.org/>.
 51. PLOS. Public Library of Science [Internet]. [cited 31 December, 2019]. Available from: <https://www.plos.org/>.
 52. Tennant JP, Crane H, Crick T, Davila J, Enkhbayar A, Havemann J, Kramer B, Martin R, Masuzzo P, Nobes A, et al. Ten hot topics around scholarly publishing. *Publications* 2019;7(2):34.
 53. Cagan R. The San Francisco declaration on research assessment. *Dis Model Mech* 2013;6(4):869–70.
 54. Adie E, Roe W. Altmetric: enriching scholarly content with article-level discussion and metrics. *Learned Publishing* 2013;26(1): 11–17.
 55. Lin J, Fenner M. Altmetrics in evolution: defining and redefining the ontology of article-level metrics. *Information Standards Quarterly* 2013;25(2):20.
 56. Elsevier. Elsevier Developers [Internet]. [cited 14 January, 2020]. Available from: <https://dev.elsevier.com/index.html>.
 57. Janowicz K, Hitzler P, Adams B, Kolas D, Vardeman C. Five stars of linked data vocabulary use. *Semant Web* 2014;5(3): 173–6.
 58. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.
 59. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6(7):e1000097.
 60. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE-Initiative. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;4(10):e296.
 61. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
 62. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, Clark A, Cuthill IC, Dirnagl U, et al. The ARRIVE guidelines 2019: updated guidelines for reporting animal research. *bioRxiv* 2019:703181.