Fostering Scientific Meta-Analyses with Knowledge Graphs : a Case-Study

Ilaria Tiddi ✉[0000−0001−7116−9338], Daniel Balliet, and Annette ten Teije

Vrije Universiteit Amsterdam
The Netherlands
{i.tiddi, d.p.balliet, annette.ten.teije}@vu.nl

Abstract. A meta-analysis is a Science of Science method widely used in the medical and social sciences to review, aggregate and quantitatively synthesise a body of studies that address the same research question. With the volume of research growing exponentially every year, conducting meta-analyses can be costly and inefficient, as a significant amount of time and human efforts needs to be spent in finding studies meeting research criteria, annotating them, and properly performing the statistical analyses to summarise the findings. In this work, we show these issues can be tackled with semantic representations and technologies, using a social science scenario as case-study. We show how the domain-specific content of research outputs can be represented and used to facilitate their search, analysis and synthesis. We present the very first representation of the domain of human cooperation, and the application we built on top of this to help experts in performing meta-analyses semi-automatically. Using few application scenarios, we show how our approach supports the various phases meta-analyses, and more in general contributes towards research replication and automated hypotheses generation.

Keywords: Knowledge Graphs · Meta-Analysis · Science of Science · e-Science

1 Introduction

Systematic literature reviews and meta-analyses in particular are a scientific method used in a number disciplines ranging from (bio-)medical to social sciences to summarise vast amounts of research outputs on a specific topic [27]. In a nutshell, a meta-analysis exploits statistical models to quantify, aggregate and compare evidence from a set of (dozens, sometimes hundreds) experimental studies addressing the same research question, in order to derive generalisable conclusions [11]. In this way, meta-analyses offer a snapshot of a research topic, supporting research transparency, reproducibility and re-usability – a more and more urgent topic in various research disciplines [19].

Performing meta-analyses is a knowledge-intensive process that can take months, sometimes years, due to methodological and technical barriers. And with the volume of research outputs growing exponentially every year, this problem is becoming more and more difficult [3]. A new meta-analysis can require authors to spend a significant amount of time and effort to find studies that

meet their criteria, identify the evidence in them, annotate their contents and statistically aggregate their results, before reaching any significant conclusion. Moreover, meta-analyses can be large in scope, and planning for time and human resource allocation can be a hard task. This calls for new methods to help researchers in summarising scientific evidence in a more automated way, and more in general to facilitate publication and sharing of large bodies of scientific findings.

Our motivation stems from the COoperation DAtabank[1] (CODA or DataBank henceforth), a large data repository aiming at analysing the entire history of laboratory and field research on human cooperation using social dilemmas. The goal of the DataBank is to encourage and facilitate sharing experiments as well as null findings (that tend to be hardly ever published), and consequently reduce the publication bias that currently affects the area [15]. Over the last 5 years, a small pool of domain experts manually annotated approx. 3,000 studies collecting 60 years of research publications with experimental settings, measured/manipulated variables of observation, and quantitative results, with the goal of establishing an open access database that researchers worldwide could consult to identify studies to include in their systematic literature reviews, as well as to directly conduct their own statistical (meta-)analyses.

In this work, we show how semantic technologies, which provide support for scaling, reuse, and interoperability, can be exploited to tackle the scalability, methodological and technical issues of conducting meta-analyses. Using a social science scenario, we show how the content of research outputs can be represented using semantic descriptions, and how to leverage this structured, domain-specific knowledge to facilitate search, analysis and synthesis of research outputs. Our main contributions are (1) the first structured representation of the field of human cooperation, that researchers from the field can easily reuse and extend; and (2) a Science of Science application to help experts in performing meta-analyses semi-automatically, supporting the correct evaluation and interpretation of research conclusions. We discuss on the multiple benefits of our approach using few use-cases that demonstrate how the various phases of the meta-analytic process can be facilitated and, more in general, how this can significantly contribute to research replication and automated hypotheses generation.

2 Background and Related Work

We introduce here the basic notions of scientific meta-analyses, the best practices and current applications, and overview the semantic approaches supporting scientific research.

Principles of Meta-Analysis. A meta-analysis is a process used to synthesise knowledge from studies addressing the same research question by using comparable methods. Each study may observe a relation between two (one independent, one dependent) variables, which can be quantified as an effect size. Meta-analytic

[1] https://amsterdamcooperationlab.com/databank/

techniques are then used to estimate an overall effect size average through aggregating the effect sizes observed in single studies [22]. Effect sizes can represent the differences observed between the experimental variations (treatments) of the independent variable and the dependent variable (such as a standardised difference between means d), or could also be the relation between two measured variables (such as a correlation coefficient ρ). In order to derive the overall estimate, a researcher first frames a problem statement, defining the research question, inclusion criteria, independent and dependent variables of observation etc., and then collects relevant studies (both published and non-published material) across scientific sources. Conducting a meta-analysis then consists in: (1) *Coding*, i.e. annotating the studies with the relevant characteristics, including independent and dependent variables and effect sizes; (2) *Analysis*, i.e. estimating the overall effects using fixed and random effects models, determining heterogeneity in the studies, assessing publication bias, conducting moderator analyses through meta regression, performing statistical power analysis; (3) *Interpretation*, i.e. the presentation of the obtained results along with conclusions and graphical support, often including graphs such as forests, funnel, violin/scatter-box plots. These steps make a meta-analysis significantly different from a literature review conducted, for instance, in computer science, as the researcher numerically "pools" results from the studies (i.e. the effects sizes) and arrives at a statistical summary that can be integrated into the narrative of a publication.

Methods and Applications. While meta-analyses are now established in the field, they are still seen as a controversial tool, as even small methodological violations can lead to misleading conclusions [20]. Researchers have argued that significant conclusions can only be derived from meta-analyses with large number of studies, while smaller meta-analyses can only support framing new research hypotheses [14]. In response to this, a number of methodologies across disciplines have been published to assist experts in deriving reliable conclusions, e.g. the Cochrane Handbook by the Campbell organisation [23], the York Centre for Reviews and Dissemination guidelines for health care [29], the Evidence for Policy and Practice Information and Co-ordinating Centre[2]. A considerable amount of statistical expertise is also needed to avoid deriving incorrect conclusions. A number of statistical tools are now available to overcome the technical barriers, resulting in tools and libraries such as RevMan [8], Comprehensive Meta-Analysis [4], Stata, and R packages such as *meta*, *rmeta* and *metafor*. Finally, with the volume of research outputs growing exponentially, identifying relevant studies and annotate the evidence can require significant efforts. As a result, a number of research projects emerged in the latest years with the goal of making large bodies of research findings openly available, offering summaries of scientific evidence and, more in general, automating meta-analyses [1, 5].

Supporting Science with Semantic Technologies. Semantic web technologies have been used to provide an interoperable and machine-interpretable infrastructure for scientific enquiry in the context of Semantic e-Science [7].

[2] https://eppi.ioe.ac.uk/cms/

Ontology engineering techniques have widely been employed to formally organise domain-specific knowledge and provide interactive, semantically-enhanced data querying, exploration, and visualisation in a number of cross-disciplinary settings [12, 24, 25]. A large number of vocabularies to semantically describe research outputs have been proposed, e.g. document publication [9, 16]; provenance, versioning, attribution and credits through research objects[3] or nanopublications [21]; description of research hypotheses [18] and scientific datasets [6]. The benefit of using controlled vocabularies to describe research outputs guarantees them to be findable, exchangeable and interpretable across different applications (interoperability).

Another strand of research has focused on capturing knowledge from scientific processes, in order to support design and management of workflows [26], identify common patterns and motifs in them [13, 17], or recommend activities to tackle the cold start problem of experimental design [10]. These works demonstrated that research reproducibility and support to frame research hypotheses can be supported by semantically describing and mining workflow components.

A semantic approach could be used support conducting meta-analyses. Domain vocabularies and descriptions could express the scientific knowledge contained in research outputs to facilitate search, analysis and synthesis. At the same time, replication of published results and support to derive the correct conclusions could be offered by the relying on semantic technologies, that enable scalability, interoperability and reuse. In the following, we show how our hypothesis was tested to support research replication and automated hypotheses generation in a social science scenario.

3 Motivating Scenario and Contribution

The COoperation DAtabank (2015-2020) is a large-scale effort involving a trained team of international researchers working with the goal of representing and publishing an open-access repository of over 60 years of research on human cooperation using social dilemmas. Social dilemmas are social situations that involve a conflict of interests and people must choose between doing what is best for themselves or what is best for the collective, either a dyad or group [30]. In these situations, there is always one choice that results in the best outcome for each individual, regardless of what others choose to do. However, if everyone decides to behave this way, then the entire group receives a worse outcome, relative to when everyone decides to do what is best for the group. Common social dilemma paradigms used to study cooperation include the prisoner's dilemma, public goods dilemma, and the resource dilemma. Cooperation in these situations is operationalised as deciding to do what it best for the collective.

In the DataBank, around 3,000 studies from the social and behavioural sciences published in English, Chinese, and Japanese were annotated with more than 60 cooperation-related features. These features can be grouped in three categories, i.e. (a) characteristics of the sample participating in the study (e.g.

[3] http://www.researchobject.org/

sample size, average age of sample, percentage of males, country of participants), (b) characteristics of the experimental paradigm (structure of the social dilemma, incentives, repeated trial data, etc.), and (c) quantitative results (e.g., mean levels of cooperation, variance in cooperation, and effect sizes with cooperation). In this scenario, the CODA experts are required to annotate and include new data (i.e. findings gathered from researchers worldwide) to the dataset in a continuous way. This can be inconvenient, costly and time-consuming, especially as data are not always directly accessible [2]. In the long-term, we aim at supporting CODA's maintainers in capturing and sharing knowledge more efficiently. Starting with the assumption that scholars that consult the repository online act as domain experts, the solution we target is to crowdsource the meta-analyses that users conduct online to automatically enrich, fix and update the dataset. The procedural nature of the meta-analyses allows in fact to model them as scientific workflows of sequential activities, that we wish to capture and use to update the dataset, in a way that data maintainers do not have to input new data themselves. Besides relieving the workload of the dataset maintainers, collecting workflows could benefit data consumers, as the expertise of previous scholars could support new users when performing their own analyses.

In order to achieve this, our first goal is to make the DataBank available to the field to allow exploring data and conducting meta-analyses with it. Following similar approaches, we use an ontology engineering approach to represent the DataBank as a structured dataset, describing both the bibliographic information and the domain-specific knowledge from the collected studies, and then to build semantically-enhanced application to perform meta-analyses. Our work has two contributions, namely (i) we provide a detailed semantic description of the domain of human cooperation, so far never published in a structured way, that researchers from the field can easily reuse and extend; and (ii) we build a tool to conduct meta-analyses semi-automatically, reducing the time researchers need to test their new hypotheses. More in general, our work shows how semantic technologies can be used to tackle the limitations of meta-analyses at different levels (search, analysis and synthesis), fostering research reproducibility while facilitating the framing and testing of research hypotheses.

4 Performing Meta-Analyses over Knowledge Graphs

In order to allow conducting meta-analyses over a knowledge graph, we follow two simple steps: first, we deal with the generation of the DataBank, by describing the research studies and their content and generating a knowledge graph from this; second, we focus on building the application conduct meta-analyses on it. In the following, we will imagine the example of a meta-analysis performed to study the impact (effect) of using communication (independent variable) over cooperation (dependent variable) in a control group.

4.1 DataBank Generation

The first step is gathering the raw data annotated by the CODA team and build a structured knowledge graph from it. The dataset consists a series of CSV ta-

bles, roughly divided by topic, where published studies are annotated according to the features described in Section 3, including both generic information (study characteristics such as country or year of publication) and specific characteristics (information relevant to cooperation games, e.g. types of priming or incentives given to the study participants). We therefore divide this task in three steps, i.e. establishing a general schema for papers, DOIs, authors, experiments (domain-independent knowledge), providing a more fine-grained model to describe the cooperation situations (domain-specific knowledge), and populating the knowledge graph accordingly.

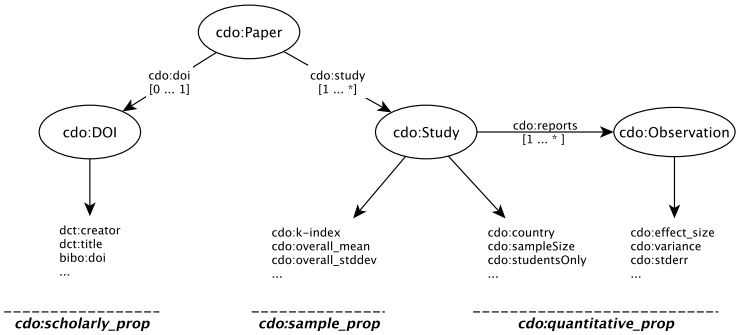


Fig. 1. Domain-independent schema for data annotation.

Modelling Domain-Independent Knowledge. Figure 1 presents the domain-independent schema we used[4], where a publication consists of a `cdo:Paper` that includes an arbitrary set of `cdo:Study`, i.e. experiments performed in different settings and with different goals. In the example of Listing 1.1, for instance, resource `:ENG00073` represents a paper written by H. Varian in 1999 reporting his experimental study `:ENG00073_1` where 48 students from the US played a prisoner's dilemma game. Additional metadata about the paper such as publication date, authors etc. are collected directly by dereferencing the paper's DOIs and by including the collected triples as properties of a `cdo:DOI` class. We then define these properties as `cdo:scholarly_prop`. In our example, the DOI allows to gather the paper's scholar information, such as his author H. Varian and its year of publication (1999).

Each `cdo:Study` has also specific properties, which we divided in `cdo:sample_prop` and `cdo:quantitative_prop` depending if they represent information about the study sample settings or the experimental quantitative/statistical information. For example, `cdo:country`, `cdo:sample_size`, `cdo:country`, `cdo:studentsOnly` and `cdo:game` are subproperties of `cdo:sample_props`, while `cdo:overall_coop` (that measures the overall participants' cooperation rate across different tests in an experiment) is a quantitative property defined as subproperty of `cdo:quant_prop`. As said, a `cdo:Study` reports one or more tests, modelled as `cdo:Observation`.

[4] All CODA's namespaces are used here for illustrative purposes. We recommend to follow the links at https://amsterdamcooperationlab.com/databank/ for the most recent versions.

The significance of the tests is estimated in terms of statistical `cdo:quantitative_prop` of an observation, e.g. effect size values, standard errors, variance, etc. In our example, study `:ENG00073_1` reports one observation called `:ENG00073_1.1.1.2.7`, reporting a measured effect size value of ∼0.60 and a standard error of ∼0.24.

```
1  @prefix cdo: <http://data.coda.org/coda/vocab/> .
2  @prefix    : <http://data.coda.org/coda/resource/> .
3  @prefix dct: <http://purl.org/dc/terms/>
4
5  :ENG00073 a cdo:Paper ;
6          cdo:includes  :ENG00073_1  ;
7          cdo:doi <http://dx.doi.org/10.1073/pnas.96.19.10933>  .
8   <http://dx.doi.org/10.1073/pnas.96.19.10933> a cdo:DOI ;
9          dct:title "Preplay contracting in the Prisoners' Dilemma";
10         dct:creator [ a foaf:Person;
11                       foaf:name "H. Varian" ].
12         dct:date "1999"^^xsd:gYear .
13 :ENG00073_1 a cdo:Study;
14         cdo:country :USA  ;
15         cdo:game :prisoner-s-dilemma  ;
16         cdo:sampleSize "48"^^xsd:integer ;
17         cdo:studentOnly "true"^^xsd:boolean ;
18         cdo:overall_coop "0.413"^^xsd:double ;
19         cdo:reports :ENG00073_1.1.1.2.7 ;
20 :ENG00073_1.1.1.2.7 a cdo:Observation;
21         cdo:effect_size "0.6057"^^xsd:double ;
22         cdo:stder "0.2431"^^xsd:double .
23         cdo:compares (:ENG00073_1.1.1, ENG00073_1.1.2)
24 :ENG00073_1.1.1 a cdo:Treatment.
25         cdo:communicationBaseline "True"^^xsd:boolean .
26 :ENG00073_1.1.2 a cdo:Treatment ;
27         cdo:communicationBaseline "False"^^xsd:boolean .
28         cdo:realCommunication "True"^^xsd:boolean .
29         cdo:communicationContent :promise .
30 :promise a cdo:CommunicationContent .
31 :prisoner-s-dilemma a cdo:Game .
```

Listing 1.1. Example of CODA resources.

Note that the current work only focuses on using a semantic-based approach as a mean to simplify meta-analyses. In other words, vocabulary and data model are not finalised, and while alignments at schema and instance level are already under way, they remain out of this paper's scope.

Modelling Domain-Specific Knowledge. In order to allow experts to understand the factors affecting cooperation, the next step is describe the content of each study in a fine-grained way. We model observations as comparisons of one or two different `cdo:Treatment`, consisting in the experimental settings that an experimenter modifies with the goal of assessing how and if the cooperation between participants of a game varies significantly. For example, observation `:ENG00073_1.1.1.2.7` compares a treatment in which participants were not allowed to communicate (line 24–25) with a second treatment, in which participants were playing with real partners (line 28) and could only exchange promises about future game behaviours (line 29). These experimental settings, modified across different treatments of the same independent variable (IV), are fundamental to perform meta-analyses. An experimenter could be interested in observing the effects of allowing or denying communication within participants of a game,

or on the impact of specific characteristics (called moderators) such as age, gender or personality of the participants, type of communication exchanged.

We therefore take all RDF properties whose domain is the class `cdo:Treatment`, and organise them in a domain-specific taxonomy of information relative to cooperation in social dilemmas. The resulting taxonomy, shown in Figure 2, was built in a bottom-up fashion, i.e. (i) an initial list of key variables and definitions was drafted by the CODA's team given their extensive expertise in the domain; (ii) the list was used to perform an initial annotation of ∼1k studies across universities, to report potential problems and additions, and (iii) further revised by a scientific advisory board of 12 senior domain experts; (iv) existing papers were revised and new ones were annotated accordingly.

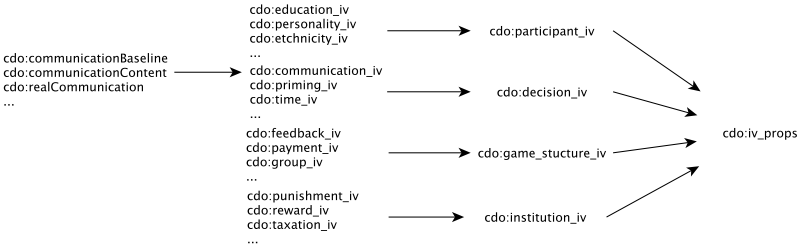


Fig. 2. Property taxonomy for annotation of domain-specific knowledge (simplified).

All properties are by definition subproperties of a generic `rdf:Property` called `cdo:iv_props`, and can be either `cdo:measured_iv` or `cdo:manipulated_iv`, depending if it consists in a discrete (e.g. type of communication) or continuous (e.g. amount of money incentive) variable. Additionally, up to four categories of properties can describe the cooperative game in a treatment, namely:

- *participant variables* (`cdo:participant_iv`), i.e. all variables related to the people taking part of a cooperation, including personal background (age, ethnicity, education), stable personality traits (e.g. HEXACO/Social Value Orientation), dynamic psychological states (e.g., emotions, moods);
- *decision variables* (`cdo:decision_iv`), i.e. all variables related to the decisions that people take during the game, e.g. intrapersonal features (priming, time constraints), or interpersonal features (communication, gossip);
- *game structure variables* (`cdo:game_structure_iv`), consisting in all variables related to the structural aspects of the game, e.g. payment of the participants, protocol for forming teams, composition of the team etc;
- *institution variables* (`cdo:institution_iv`), involving the rules and norms for participants, such as punishment, reward or taxation during the game.

Taking back the example of Listing 1.1, `cdo:communicationBaseline`, `cdo:realCommunication`, `cdo:communicationContent` are subproperties of `cdo:communication_iv`, indicating that in his study, the experimenter only manipulated communication as an independent variable. Of course, this is a rather simplified example, and treatments describe on average multiple IVs.

Knowledge Graph Population and Storage. Once defined the two parts of the schema, we create statements with the support of Python's RDFlib library, additionally dividing them across a number of named graphs (i.e. studies, papers, observations, treatments, vocabulary descriptions) for storage and querying convenience. The generated dataset is hosted as a triplyDB instance[5], allowing to easily upload and update datasets and expose them through APIs such as SPARQL, RESTful, and textual search. While the complete dataset is in the process of being iteratively published, its preliminary online version currently includes 330,655 statements, including approx. 1.1k studies and 61 specific independent variables (cfr. Table 1).

Table 1. DataBank status, as of Dec. 2019. Observations are still being computed.

Knowledge	Class	Statements	Annotated Resources	Total Resources
domain independent	Papers	6,411	1,454	2,922
	DOIs	22,545	1,276	2,588
	Studies	78,523	2,095	2,102
	Observations	115,881	8,288	n/a
	Treatments	43,844	2,219	11,432
domain specific	Participant IVs	2,527	23	42
	Decision IVs	799	12	20
	GameStr. IVs	629	9	45
	Institution IVs	4,209	18	31

4.2 Conducting Meta-analyses with the DataBank

In the second phase, we build a web-based interface allowing experts to conduct their meta-analysis over the generated knowledge graph. The application, shown in Figure 3, is accessible through the main website, and allows to (i) explore the DataBank; (ii) select relevant studies and (iii) performing meta-analyses online.

Data Exploration. At first, users are presented a global overview of the data. Studies and the effect sizes they report can be explored using a number of criteria, e.g. year of publication, sample size, country of publication (as in Figure 3). At the bottom, a tabular condensed view of the data is also offered, and users are given the possibility to click on studies, papers and observations to explore their properties in-depth directly from the triplyDB instance. An additional panel at the top offers the possibility of visualising the taxonomy of independent variables described in the previous section.

Search&Selection. The left panel allows users to select the desired studies before starting their statistical computations. The selection can be performed based on independent variables that are manipulated during treatments of a study (cfr. Figure 2). In the example of Figure 3, the user selected observations from studies manipulating some `cdo:communication_iv`, and specifically studies manipulating the properties `cdo:realCommunication` and `cdo:communicationContent`, resulting in a selection of 21 observations from 6 different studies.

[5] https://triplydb.com/

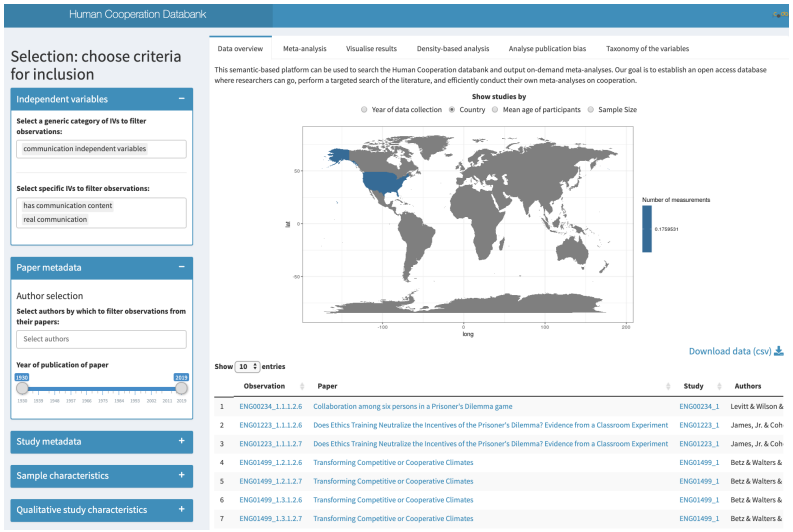
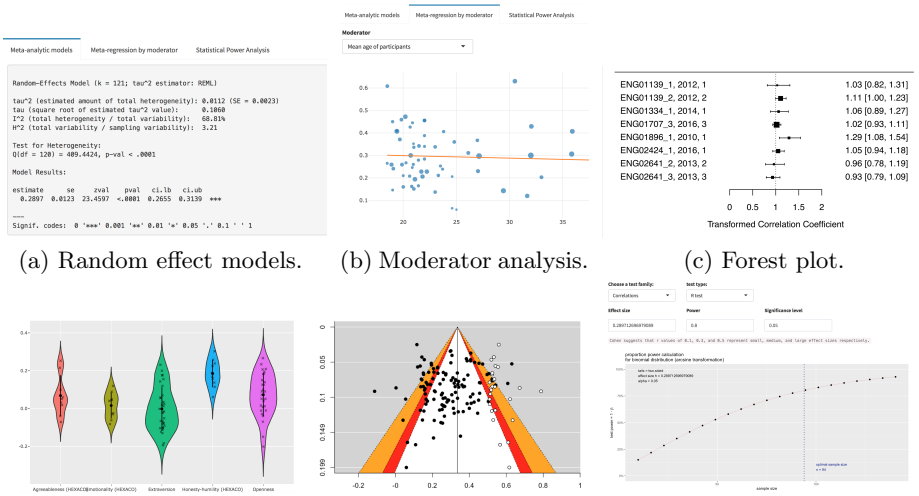


Fig. 3. Main view of the CODA web-app, allowing data visualisation (bottom), search&selection (left), meta-analytic activities (top, and Figure 4).

Multiple selection is allowed, e.g. one could additionally include `cdo:personality_iv` or `cdo:emotion_iv` variables. Studies can be additionally filtered based on sample and quantitative properties, e.g. one could choose to include in the meta-analysis only observations from studies published between 2000 and 2010, or those with at least 100 participants. In order to foster data sharing and reuse, the portion of desired data can be downloaded in tabular format. In this way, as authors are relieved from the coding step, as the provided data are already annotated, and the model can be extended according to the meta-analysis' purposes.

Meta-analytic Activities. Once selected the desired data, the user can perform his meta-analysis using the tabs in the app. Due to space restrictions, Figure 4 shows a simplified view of the meta-analytic steps, but the reader is invited to consult the online version to replicate our example. Typical meta-analysis steps include:

1. *Fitting a meta-analytic model* (Figure 4a). This operation consists in choosing the model type (fixed, random and mixed effects), the method (maximum or restricted likelihood estimators), and the variables (single, or aggregated) to obtain the estimate of the overall population effect size. Models can also be fitted by specific moderators (e.g. mean age of the sample as in Figure 4b), corresponding to the IVs and study characteristics in the KG schema.
2. *Exploring the heterogeneity of single studies.* Using forest plots (Figure 4c), a meta-analysts can illustrate the effect sizes of the studies ordered by year of publication, using confidence intervals that reflect the precision with which effects were estimated (the narrower the confidence interval, the greater precision). Effect sizes can also be plotted to check their distribution and density



(a) Random effect models. (b) Moderator analysis. (c) Forest plot.

(d) Violin plot. (e) Funnel plot. (f) Power analysis.

Fig. 4. Example of a meta-analysis: (a) fitted models to estimate the global effect size; (b) linear regression to assess the relation between the moderator and the effect size; (c) forest plot to determine heterogeneity of effect sizes (X-axis) per study (Y-axis); (d) violins to visualise the studies distribution in details; (e) funnels to assess symmetry of the results (X-axis) based on their error degree (Y-axis); (f) power analysis to estimate the required sample size in future experiments.

using violin plots (Figure 4d), where relevant statistics such as median and its 95% confidence interval, the quartiles and outliers are also shown.

3. *Checking for publication bias.* Using funnel plots (Figure 4e), the user can plot effects sizes against the sample sizes in a symmetrical inverted funnel centred on the average effect, in a way that asymmetries in the distribution of the data should be revealed. An additional data-augmentation method called "trim&fill" can also be selected in order to estimate the number of studies that are missing from the meta-analysis.
4. *Computing power analysis.* This activity (Figure 4f) allows to derive the optimal sample size for a desired effect size (either obtained by the fitted model, or specified by the user) with a given power and p-value. Determining the optimal effect size given a desired sample size and p-value is also possible. With this operation, researchers can calculate the required sample size necessary to obtain high statistical power in future studies.

Similarly to the data selection, all meta-analytic results can also be comfortably downloaded through the interface. This is particularly beneficial to less experienced meta-analysts, as they can be relieved from the often tedious and time consuming task of writing efficient code. Additionally, all statistical computations and activities are presented sequentially in order to support methodological design – thoroughly crafted using meta-analytic experts. Finally, by allowing to compute meta-analyses online, published and unpublished meta-analyses can

also be easily reproduced, benefitting study reproducibility and transparency for the whole research field.

Implementation. The above web-app is implemented using R Shiny[6], a package for dashboard development straight from R code. The advantage of using Shiny lies mostly in the fact that we can exploit the large variety of statistical techniques (linear/nonlinear modelling, classical statistical tests etc.) and their graphical outputs (funnel, forests and violin plots) to manipulate, interact and visualise data from the DataBank knowledge graph. Data are selected through SPARQL queries stored as APIs on the triplyDB instance, allowing to further decouple the application from the dataset.

5 Usability Assessment via Use-Cases

Since neither the CODA app nor the dataset in its entirety have officially released at the time of writing, we focus here on a qualitative assessment with the domain experts using few use-cases, discussing how our approach support the various phases of the meta-analysis. A usability testing with users is under preparation, and will be held further the official release.

Offloading Data Maintainers. Here, we are interested in reducing the workload of the experts in maintaining and updating the dataset. We therefore asked the CODA team to quantify the time it takes the editorial board to include a new study in the dataset. Table 2 shows the main activities they need to perform, i.e. searching studies, skimming publications to assess if they fall under the eligibility criteria, coding studies based on the chosen characteristics, and computing the aggregate effect sizes if a work reports separate splits (e.g. a paper reporting 10 different cooperation rates, because 10 rounds of a game were experimented). We report answers of three experts E_n supervising the DataBank, as well as one annotator A of more limited expertise. These are then compared with the data provided by [28], an established reference that analysed the problem of performing meta-analyses in the medical domain.

Table 2. Time for data maintenance. Aggregation (*) is not always performed. The ranges relate to the difficulty of the studies, which can go from easy studies (small analyses from psychology/economics) to complex meta-analyses that require additional computations.

	Searching (hours)	Skimming (mins/paper)	Coding (mins/paper)	Aggregating* (mins)
E_1	80	5-10	20-90	3
E_2	160-200	3-5	20-90	3
E_3	160-240	5-10	60-120	3
A	–	10-20	45-180+	10
[28]	20-60	60	90-100	640

The table illustrates that data maintainers invest most of their efforts in searching, skimming and annotating papers, reported as the most time-expensive activities needed to be performed. A straightforward conclusion here is that the this

[6] https://shiny.rstudio.com/

workload could be significantly reduced by allowing users consulting the DataBank to upload their studies using the same annotation schema. This would allow maintainers to focus only on the light-weight refinement and validation of the uploaded data. Finally, the disproportion in time between activities in the social science and medical domain suggests that a substantial difference lies in the type of studies that experts need to analyse (e.g. lengths, regression analyses).

Improving Study Exploration. Here, we look at estimating how our approach supports meta-analysts in searching data. We asked the experts to write a set of competency questions, which can indicate how well the knowledge graph support the search phase through requirement fulfilment.

Table 3. Competency questions to assess data exploration.

Competency Question	Before	After
1. *Find studies on variable X*	✓	✓
2. *Find studies on N variables*	✗	✓
3. *Find studies with X and Y characteristics, comparing N variables*	✗	✓
4. *Find effects of studies with X or Y characteristics*	✗	✓
5. *Find effects comparing X and Y properties, both of variable Z*	✗	✓
6. *Find negative effects of continuous variables*	✗	✓
7. Number of Independent Variables	n/a	86

Table 3 shows that a number of requirement, particularly related to the selection of studies with multiple characteristics or using effect sizes, could not be answered with the original the dataset. This means that a user would mostly have to go through the process of extracting and re-coding the studies of interests, before performing a new meta-analysis. Additionally, the DataBank now includes over 86 independent variables (vs. the original dataset with no controlled vocabulary for independent variables), which can be used both for the study selection and the moderator analyses at finer or coarser grain (e.g. users can decide to simply consider all communication variables, or to some specific communication characteristics). This is a great opportunity for the behavioural science community, which can easily investigate different research questions in a more automated and assisted way.

Support in Performing Statistical Analyses. We are interested in understanding if our approach facilitates users performing the statistical computations needed to finalise and publish a meta-analysis. Table 4 provides an estimate of the resources necessary to compute meta-analytic models in a normal setting (i.e. when not supported by the CODA application) based on the answers of two experts that recently run a meta-analysis (ma_1, ma_2), as well as the information provided by [28]. We used lines of code as a measure for data preparation, model fitting and result plotting to show that users might be relieved from writing a significant part of code when using a semantically-enriched system, as opposed to a database of meta-analyses.

Of course, resource allocation is highly dependent on the type meta-analysis performed (i.e. number of studies analysed, complexity of the question framed,

Table 4. Resource allocation for manually running a meta-analysis.

	Manual Resources (#people)	Data prep. (#lines)	Model fitting (#lines)	Plots (#lines)	Tot (h) (time)
ma_1	1 expert 1 assistant	∼200	∼250	∼400	80-100
ma_2	1 expert	∼100	∼400	∼500	160-200
[28]	1 expert 1 statistician	–	–	–	360

number of moderator analyses...), and the same would lie when conducting meta-analyses with the support of our framework. Yet, users would be relieved from the data preparation and programming tasks, offered by the CODA app as interactive activities to be performed in a customised way. To give a baseline over the current application, a simple sequence of model fitting, heterogeneity analysis and moderator analysis takes on average 5 to 10 mins.

Fostering Reproducible Research through Recommendations. Finally our approach offers quality improvement for the research field in terms of (1) reproducibility, (2) domain description and (3) best practice workflows. First, we offer a dataset of annotated and reproduced experimental studies openly available for consultation and download (both studies and the meta-analyses computed online). Secondly, by relying on a taxonomical representation of the domain, recommendations of variables to explore and analyse can be offered to users as "if you are looking at studies where [$specific\_var_1$] variables were manipulated, you might want to explore other [$parent\_var$] as [$specific\_var$]$_{2,\ 3}$ ...", where $specific\_var_n$ are siblings of a parent variable (e.g. `cdo:communication_iv` and `cdo:priming_iv` are children of `cdo:decision_iv`). Additionally, by relying on SPARQL queries over the dataset, we can monitor the users' activities, and offer recommendations of new variables to meta-analyse based on popularity (i.e. suggesting the most popular moderators for specific variables) or anti-patterns (i.e. suggesting to choose less popular and unexplored queries for the meta-analysis). Finally, by describing meta-analysis as scientific workflows that manipulate data with specific parameters, we can collect and offer them as recommended practices to users. In this sense, the expertise of previous users could be leveraged to offer inexperienced practitioners a more automated way of performing meta-analyses – tackling to the cold start problem of designing experiments.

6 Conclusions

In this work, we have shown how to use semantic technologies to support researchers in conducting meta-analyses and summarise scientific evidence in a more automated way. Meta-analysis are a Science of Science method for research synthesis, which tend to suffer from scalability, methodological and technical issues also due to the exponentially growing volume of research. Using a social science scenario, we showed how the content of research outputs can be semantically described and used to build an application to help meta-analysts

in searching, analysing and synthesising their results in a more automated way. The use-cases we discussed have shown that the approach is beneficial at several levels of the meta-analysis, and has the potential of fostering research replication and facilitating the framing and testing of research hypotheses in the field of behavioural science.

Future work will be focused on : (i) publishing the DataBank following the FAIR principles[7], which will require alignment to existing vocabularies (e.g. Data Cubes) and linking instances to available datasets (e.g. Microsoft Academics); (ii) improving the web application with more advanced analyses, e.g. dynamics of citations, multivariate analyses, integration of cross-societal moderators from the linked open datasets (e.g GDP or GINI indices from Eurostats); (iii) implementing the collection and documentation of meta-analytic workflows using PROV[8]; (iv) evaluation through user-testing, quantifying the time they take to perform a meta-analytic tasks with and without the support of the knowledge-based recommendations, workflows and documentation.

References

1. Bergmann, C., Tsuji, S., Piccinini, P.E., Lewis, M.L., Braginsky, M., Frank, M.C., Cristia, A.: Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. Child development **89**(6), 1996–2009 (2018)
2. Berman, N.G., Parker, R.A.: Meta-analysis: neither quick nor easy. BMC medical research methodology **2**(1), 10 (2002)
3. Borah, R., Brown, A.W., Capers, P.L., Kaiser, K.A.: Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the prospero registry. BMJ open **7**(2), e012545 (2017)
4. Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R.: Introduction to meta-analysis. John Wiley & Sons (2011)
5. Bosco, F.A., Uggerslev, K.L., Steel, P.: Metabus as a vehicle for facilitating meta-analysis. Human Resource Management Review **27**(1), 237–254 (2017)
6. Brickley, D., Burgess, M., Noy, N.: Google dataset search: Building a search engine for datasets in an open web ecosystem. In: The World Wide Web Conference. pp. 1365–1375. ACM (2019)
7. Brodaric, B., Gahegan, M.: Ontology use for semantic e-science. Semantic Web **1**(1, 2), 149–153 (2010)
8. Collaboration, C., et al.: Review manager (revman)[computer program] (2014)
9. Corcho, O., Garijo Verdejo, D., Belhajjame, K., Zhao, J., Missier, P., Newman, D., Palma, R., Bechhofer, S., García Cuesta, E., Gomez-Perez, J.M., et al.: Workflow-centric research objects: First class citizens in scholarly discourse. (2012)
10. Daga, E., d'Aquin, M., Gangemi, A., Motta, E.: An incremental learning method to support the annotation of workflows with data-to-data relations. In: European Knowledge Acquisition Workshop. pp. 129–144. Springer (2016)
11. DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. Controlled clinical trials **7**(3), 177–188 (1986)
12. Dietze, H., Schroeder, M.: Goweb: a semantic search engine for the life science web. BMC bioinformatics **10**(10), S7 (2009)

[7] https://www.go-fair.org/fair-principles/
[8] https://www.w3.org/TR/2013/REC-prov-o-20130430/

13. Ferreira, D.R., Alves, S., Thom, L.H.: Ontology-based discovery of workflow activity patterns. In: International Conference on Business Process Management. pp. 314–325. Springer (2011)
14. Flather, M.D., Farkouh, M.E., Pogue, J.M., Yusuf, S.: Strengths and limitations of meta-analysis: larger studies may be more reliable. Controlled clinical trials **18**(6), 568–579 (1997)
15. Franco, A., Malhotra, N., Simonovits, G.: Publication bias in the social sciences: Unlocking the file drawer. Science **345**(6203), 1502–1505 (2014)
16. Gangemi, A., Peroni, S., Shotton, D., Vitali, F.: The publishing workflow ontology (pwo). Semantic Web **8**(5), 703–718 (2017)
17. Garijo, D., Alper, P., Belhajjame, K., Corcho, O., Gil, Y., Goble, C.: Common motifs in scientific workflows: An empirical analysis. Future Generation Computer Systems **36**, 338–351 (2014)
18. Garijo, D., Gil, Y., Ratnakar, V.: The disk hypothesis ontology: Capturing hypothesis evolution for automated discovery. In: K-CAP Workshops. pp. 40–46 (2017)
19. Gonzalez-Beltran, A., Li, P., Zhao, J., Avila-Garcia, M.S., Roos, M., Thompson, M., van der Horst, E., Kaliyaperumal, R., Luo, R., Lee, T.L., et al.: From peer-reviewed to peer-reproduced in scholarly publishing: the complementary roles of data models and workflows in bioinformatics. PLOS one **10**(7), e0127612 (2015)
20. Greco, T., Zangrillo, A., Biondi-Zoccai, G., Landoni, G.: Meta-analysis: pitfalls and hints. Heart, lung and vessels **5**(4), 219 (2013)
21. Groth, P., Gibson, A., Velterop, J.: The anatomy of a nanopublication. Information Services & Use **30**(1-2), 51–56 (2010)
22. Gurevitch, J., Koricheva, J., Nakagawa, S., Stewart, G.: Meta-analysis and the science of research synthesis. Nature **555**(7695), 175 (2018)
23. Higgins, J.P., Green, S., et al.: Cochrane handbook for systematic reviews of interventions (2008)
24. Hoekstra, R., Meroño-Peñuela, A., Rijpma, A., Zijdeman, R., Ashkpour, A., Dentler, K., Zandhuis, I., Rietveld, L.: The datalegend ecosystem for historical statistics. Journal of Web Semantics **50**, 49–61 (2018)
25. Hu, W., Qiu, H., Huang, J., Dumontier, M.: Biosearch: a semantic search engine for bio2rdf. Database **2017** (2017)
26. Khan, F.Z., Soiland-Reyes, S., Sinnott, R.O., Lonie, A., Goble, C., Crusoe, M.R.: Sharing interoperable workflow provenance: A review of best practices and their practical application in cwlprov. GigaScience **8**(11), giz095 (2019)
27. Olkin, I.: Keynote addresses. meta-analysis: Reconciling the results of independent studies. Statistics in medicine **14**(5-7), 457–472 (1995)
28. Steinberg, K., Smith, S., Stroup, D., Olkin, I., Lee, N., Williamson, G., Thacker, S.: Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. American Journal of Epidemiology **145**(10), 917–925 (1997)
29. Tacconelli, E.: Systematic reviews: Crd's guidance for undertaking reviews in health care. The Lancet Infectious Diseases **10**(4), 226 (2010)
30. Van Lange, P.A., Balliet, D.P., Parks, C.D., Van Vugt, M.: Social dilemmas: Understanding human cooperation. Oxford University Press (2014)