4. MAPPING SCHEMA.ORG VOCABULARY TO SUPPORT JSON-LD

In this pilot project, we aim to use the popular Schema.org vocabulary as the dereferencing tool to support Research Graph schema's definition. In the context mapping, every term is mapped to IRI (Internationalized Resource Identifiers [RFC3987])) in the context so that it is unambiguously identified by an IRI and all values representing IRIs are explicitly marked by the keywords. IRIs are fundamental to Linked Data as that is how most nodes and properties are identified. Figure 6 presents a dataset example in the JSON-LD using the mapping in the context (see the beginning part of the file).

```
{"@context": {
          "@vocab":"http://schema.org/",
          "key": {
               "@id":"http://schema.org/mainEntityOfPage",
                  "@type": "@id"
                     },
          "source":"http://schema.org/sourceOrganization",
          "local_id": {
               "@id":"http://schema.org/disambiguating
                    Description",
                  "@type": "@id"
                       },
          "last_updated":
               "http://schema.org/dateModified",
          "title": "http://schema.org/headline",
          "authors_list": "http://schema.org/author",
             "doi": {
               "@id":"http://schema.org/sameAs",
               "@type": "@id"
               },
          "publication_year":
               "http://schema.org/datePublished",
          "megabyte": "http://ls
              .org/contentSize"
          },
 "@type": "Dataset",
 "key":
      "http://researchgraph.org/nci/f3525_9322_8600_7716/",
 "source": "National Computational Infrastructure",
 "local_id": "f3525_9322_8600_7716",
 "last_updated": "2014-12-31",
 "url":
      "http://pid.nci.org.au/dataset/f3525_9322_8600_7716",
 "title": "Coupled Model Intercomparison Project
      (CMIP5)",
 "authors_list": "Evans, Ben",
 "doi": "http://dx.doi.org/10.5072/29/5874605e6b57f",
 "datePublished": "2014-12-31",
 "license":
      "http://dapds00.nci.org.au/thredds/fileServer/licenses/
  license_ua6.txt",
 "megabyte": "1,500,000"
}
```

Figure 6: Dataset from NCI in JSON-LD format

The complete mapping between Research Graph schema and Schema.org for Research Graph registry objects (Dataset, Publication, Researcher and Grant) is available at GitHub Repository[12]. The summary is available in Table 1. The mapping serves as a term definition file and referenced within JSON-LD files. The context file for the Research Graph can be retrieved from GitHub Repository.

The corresponding one-on-one relationship is defined in the context section of the JSON-LD file. However, adding the mapping in front of each metadata file is tedious and may affect our existing workflow. JSON-LD provides an alternative way to save the mapping as a separate file so that it can be referenced at the beginning of JSON-LD. The content of the mapping can be modified without affecting the actual JSON-LD file itself. In the following modified JSON-LD file (see Figure 7), the context is referenced by adding a single line at the beginning of the JSON file to convert an existing JSON file to a JSON-LD document with minimum interruption to our existing data processing pipeline.

```
{"@context":
     "https://raw.githubusercontent.com/researchgraph/
  schema/master/json-ld/context.jsonld",
 "@type": "Dataset",
 "key":
      "http://researchgraph.org/nci/f3525_9322_8600_7716/",
 "source": "National Computational Infrastructure",
 "local_id": "f3525_9322_8600_7716",
 "last_updated": "2014-12-31",
 "url":
      "http://pid.nci.org.au/dataset/f3525_9322_8600_7716",
 "title": "Coupled Model Intercomparison Project
      (CMIP5)",
 "authors_list": "Evans, Ben",
 "doi": "http://dx.doi.org/10.5072/29/5874605e6b57f",
 "datePublished": "2014-12-31",
 "license":
      "http://dapds00.nci.org.au/thredds/fileServer/
  licenses/license_ua6.txt",
 "megabyte": "1,500,000"
}
```

Figure 7: Concise version of dataset from NCI in JSON-LD format.

5. EXTENSIONS TO SCHEMA.ORG

As part of the mapping exercise we have observed challenges in the following areas: We have managed to successfully map

- Publication ⇒ schema.org/ScholarlyArticle
- Researcher ⇒ schema.org/Person
- Dataset ⇒ schema.org/Dataset

However, the closest that we find to *grant* is schema.org/Action. This is not an exact match, and it is ambiguous. **Therefore we suggest adding a new type to Schema.org for the Research Grants or Research Projects.**

Furthermore, there is no precise method for including common identifiers – ORCID, DOI, Scopus ID(s) and PURL. These identifiers are the key enablers in linking scholarly communications, and it is essential to be able to capture and link them in different registry objects. In the current mapping we are using schema.org/sameAs for all of these identifiers. The implication is that in a large scale graphs with millions of nodes, searching a particular identifier can lead to a technical challenge. **Therefore we believe adding explicit properties for ORCID, DOI and other common**

[12]github.com/researchgraph/Schema/tree/master/json-ld

Table 1: Mapping between Research Graph and Schema.org

A. Mapping for Research Graph mandatory properties

Research Graph Schema	Schema.org Type	Property
key	Thing/CreativeWork/Article/SholarlyArticle Thing/Person, Thing/Action Thing/CreativeWork/Dataset Thing/Action	Schema.org/mainEntityOfPage
source	same as above	Schema.org/publisher for Publication and Dataset Schema.org/affiliation for Researcher and Grant
local_id	same as above	Schema.org/disambiguatingDescription
last_updated	same as above	Schema.org/dateModified

B. Mapping for Research Graph optional properties

Research Graph Schema	Schema.org Type	Property
Publication	Thing/CreativeWork/Article/SholarlyArticle	
title	same as above	Schema.org/headline
doi	same as above	Schema.org/sameAs
publication_year	same as above	Schema.org/datePublished
url	same as above	Schema.org/url
authors_list	Thing/Person	Schema.org/author
Researcher	same as above	
full_name	same as above	Schema.org/name
first_name	same as above	Schema.org/givenName
last_name	same as above	Schema.org/familyName
url	same as above	Schema.org/url
Dataset	Thing/CreativeWork/Dataset	
title	same as above	Schema.org/headline
doi	same as above	Schema.org/sameAs
publication_year	same as above	Schema.org/datePublished
url	same as above	Schema.org/url
license	Thing/CreativeWork	Schema.org/sameAs
megabyte	Thing/CreativeWork/MediaObject	Schema.org/contentSize
Grant	Thing/Action	
title	same as above	Schema.org/headline
Participant_list	same as above	Schema.org/agent
start_year	same as above	Schema.org/startTime
end_year	same as above	Schema.org/endTime
url	same as above	Schema.org/url
funder	Thing/Organization	Schema.org/funder

identifiers to Schema.org/{ScholarlyArticle, Dataset, Person} can extend the functionality of research infrastructures that leverage JSON-LD. A similar extension has already been proposed by BioSchemas community[13]. Their code is available on GitHub[14].

6. CONCLUSION AND FUTURE WORK

In this paper, we have presented a pilot project for adding JSON-LD support for Research Graph data. This project can enable an improved interoperability of connected Research Graph nodes including but not limited to publications from CERN, Dryad, datasets from figshare, da|ra, NCI, Research Data Australia and grants from Australian funders to third party services. We hope the new capability improves the discoverability and reusability of the Research Graph database.

Schema.org is a key enabler in transforming various XML files to JSON-LD. However, our preliminary work identifies a need for extending Schema.org to support widely used identifiers such as DOI, ORCID and PURL. As we are at the early stages of this project, we need feedback and direction from the community and collaboration in this domain, particularly from the service providers who have an interest in research metadata and enabling interoperability between research data infrastructures using JSON-LD. If you are interested in this project, please contact us.

It is promising to demonstrate the possibility and values of converting current Research Graph records into the JSON-LD format. We are getting one step closer to making Research Graph semantically accessible, searchable, and actionable across the web. We will develop an API to con-

[13]http://bioschemas.org/community/index.html
[14]https://github.com/BioSchemas

vert our existing database into the JSON-LD format if it is endorsed by the community to be a useful practice.

7. ACKNOWLEDGMENTS

We would also like to show our gratitude to the Martin Fenner (DataCite, Hannover, Germany - orcid.org/0000-0003-1419-2405) for sharing his insightful comments with us during this research work, and we thank the reviewers for their constructive comments to improve the quality of this work.

8. REFERENCES

[1] A. Aryani. Data description registry interoperability wg: Interlinking method and specification of cross-platform discovery. Technical report, Research Data Alliance, December 2016.

[2] K. Börner, M. Conlon, J. Corson-Rikert, and Y. Ding. Vivo: A semantic approach to scholarly networking and discovery. *Morgan-Claypool*, p.1-175, 2012.

[3] M. Fenner, M. Crosas, J. Grethe, D. Kennedy, H. Hermjakob, P. Rocca-Serra, R. Berjon, S. Karcher, M. Martone, and T. Clark. A data citation roadmap for scholarly data repositories. *Cold spring harbor laboratory*, 2016.

[4] K. Hanson, S. Morrissey, A. Birkland, T. Dilauro, and M. Donoghue. Using rmap to describe distributed works as linked data graphs: Outcomes and preservation implications. *13th International Conference on Digital Preservation, Bern, October 3-6, 2016*, 2016.

[5] A. Sinha, Z. Shen, Y. Song, H. Ma, D. Eide, B. Hsu, and K. Wang. An overview of microsoft academic service (mas) and applications. *In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, 243-246*, 2015.

[6] J. Wang, A. Aryani, B. Evans, M. Barlow, and L. Wyborn. Graph connections made by rd-switchboard using nci's metadata. *D-Lib Magazine*, Volume 23(1/2), January/February 2017.