Haplo Research Manager

- Implementing maDMPs in Haplo

Tom Renner
Lead Developer (Repository)

Research Manager

Introducing Haplo

- Based in London, UK
- Develop research information management systems for HEIs
  - PhD Manager, Ethics Monitor, Funding, Repository
- All modules can coexist within one application

maDMPs - (our) motivation

- Haplo a natural home for DMPs
- System includes full workflow support, versioned document stores, document handling
- To take advantage of these, DMPs need to be machine-actionable
- Standards based approach critical for modular flexibility of the system

DMPs in Haplo

- Open Repositories 2019: Feedback from this WG, made aware of the group's outputs
- Use the common standard to ensure DMP functionality is:

1) Compatible with sector
2) Modular within Haplo
3) Actionable within and outside Haplo

DMPs in Haplo

Prof Westbrook Reynolds

Edit

JSON | Permissions: User roles

Home

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Researcher

Name	Prof Westbrook Reynolds
Contact category	Researcher
Research institute	Music
ORCID	https://orcid.org/4040-4618-9943-1220
Unit of assessment	12. Engineering
Email address	westbrook.reynolds@example.org

WEB PROFILE

Researcher profile

View public profile

Download CV

Doctoral Researchers

Research outputs

Projects

Ethics

Finding a global network of biomedical relationships

Edit | Attach file | JSON

Home

University of Example
Humanities
Media
Science and Technology

Committees
Upcoming meetings

REF Units of Assessment

Calendar
Past events

Using this demo

Guides

Project

Title	Finding a global network of biomedical relationships
Researchers	
Principal Investigator	Prof Westbrook Reynolds
Objectives	To find labeled, weighted networks of chemical-gene, gene-gene, gene-disease, and chemical-disease relationships based on single sentences in PubMed abstracts. All raw dependency paths are provided in addition to the labeled relationships. PART I: Connects dependency paths to labels, or "themes". Each record contains a dependency path followed by its score for each theme, and indicators of whether or not the path is part of the flagship path set for each theme (meaning that it was manually reviewed and determined to reflect that theme). PART II: Connects sentences to dependency paths. It consists of sentences and associated metadata, entity pairs found in the sentences, and dependency paths connecting those entity pairs
Funder	EPSRC
Unit of assessment	12. Engineering

DATA MANAGEMENT PLAN

Edit

View

Download

Milestones and actions

Project meetings

Contract approval

PROPOSALS

Finding a Finding a global network of biomedical relationships (proposal)

Add Proposal

Haplo

University of Example

Quick search | Search | Browse | Recent | Add | SUPPORT | Help

Data management plan forms

Home

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Data management plan | DMP Datasets | Costs

Plan title *

Data Management Plan: Finding a global network of biomedical relationships

Description

This project aims to find labeled, weighted networks of chemical-gene, gene-gene, gene-disease, and chemical-disease relationships based on single sentences in PubMed abstracts. This will generate the following, for each record:
PubMed ID

Language *

EN

Save and continue | Save for later

Data management plan | DMP Datasets | Costs

Haplo

University of Example

Quick search | Search | Browse | Recent | Add | SUPPORT | Help

Data management plan forms

Home

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Data management plan | **DMP Datasets** | Costs

Title

Either search for a pre-existing dataset

unl

Unlabelled network of disease-chemical relationships

Or enter the name manually here

Description of the dataset

Type of dataset *

-- select --

Keywords to describe the dataset

Enter one keyword per box

Add another

Does this dataset contain personal data? *

Yes

Haplo

University of Example

Quick search | Search | Browse | Recent | Add | SUPPORT | Help

Data management plan forms

Home

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Data management plan | **DMP Datasets** | Costs

Title

Either search for a pre-existing dataset

Or enter the name manually here

Labeled, weighted networks of chemical-disease relationships

Description of the dataset

from April 30, 2016. If you're interested in that network, it can be found in Version 1 of this repository. We will be releasing updated networks periodically, as the PubTator community continues to release new versions of named entity annotations for Medline each month or so.

Type of dataset *

Software

Keywords to describe the dataset

Enter one keyword per box

Add another

Does this dataset contain personal data? *

Yes

View data management plan

Home

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Show changes

Data management plan

Plan title

Data Management Plan: Finding a global network of biomedical relationships

Description

This project aims to find labeled, weighted networks of chemical-gene, gene-gene, gene-disease, and chemical-disease relationships based on single sentences in PubMed abstracts. This will generate the following, for each record:

PubMed ID

Sentence number (0 = title)

First entity name, formatted

First entity name, location (characters from start of abstract)

Second entity name, formatted

Second entity name, location

First entity name, raw string

Second entity name, raw string

First entity name, database ID(s)

SECTIONS

Data management plan

DMP Datasets

Costs

VERSIONS

22 Oct 2019 (SUPPORT)

09 Oct 2019 (Westbrook Reynolds)

View data management plan

Home

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Changes from: 22 Oct 2019 | Filter questions

Data management plan

Plan title

\+ Finding a global network of biomedical relationships

\− Data Management Plan: Finding a global network of biomedical relationships

Description

This project aims to find labeled, weighted networks of chemical-gene, gene-gene, gene-disease, and chemical-disease relationships based on single sentences in PubMed abstracts. This will generate the following, for each record:

PubMed ID

Sentence number (0 = title)

First entity name, formatted

First entity name, location (characters from start of abstract)

Second entity name, formatted

Second entity name, location

First entity name, raw string

Second entity name, raw string

First entity name, database ID(s)

SECTIONS

Data management plan

DMP Datasets

Costs

VERSIONS

22 Oct 2019 (Westbrook Reynolds)

22 Oct 2019 (SUPPORT)

09 Oct 2019 (Westbrook Reynolds)

Labeled, weighted networks of chemical-disease relationships

Home

University of Example
Humanities
Media
Science and Technology

Committees
Upcoming meetings

REF Units of Assessment

Calendar
Past events

Using this demo

Guides

Dataset

Title	Labeled, weighted networks of chemical-disease relationships
Authors	Reynolds, W.
Abstract	This release contains the annotated network for the September 15, 2019 version of PubTator. The version discussed in our paper is an older one - from April 30, 2016. If you're interested in that network, it can be found in Version 1 of this repository. We will be releasing updated networks periodically, as the PubTator community continues to release new versions of named entity annotations for Medline each month or so.
Keywords	chemical-gene chemical-disease
Year	2019
Project	Finding a global network of biomedical relationships
Data files	Labelled_Networks_2019.pdf Download Versions Preview DATA TYPE Text

The DMP for this project states that sensitive or personal data will be produced, but this dataset has non-restricted files.

Set REF submitting author

STATUS
Draft record - not yet available in public repository

Deposit directly to repository

Set embargo

REF QUALITY ASSESSMENT

Overview

```
dmp": {
 "title": "Finding a global network of biomedical relationships",
 "language": "en",
 "datasets": [
   {
     "title": "Labeled, weighted networks of chemical-disease relationships",
     "description": "This release contains the annotated network for the September 15, 2019 version of PubTator. The versio
cussed in our paper is an older one - from April 30, 2016. If you're interested in that network, it can be found in Version
s repository. We will be releasing updated networks periodically, as the PubTator community continues to release new versio
ed entity annotations for Medline each month or so.",
     "type": "software",
     "keywords": [
       {
         "keyword": "chemical-gene"
       },
       {
         "keyword": "chemical-disease"
       }
     ],
     "personal_data": "yes",
     "sensitive_data": "no",
     "data_quality_assurances": [
       "The PI will be responsible for overall quality assurance, with lead country researchers and the UK research assista
ertaking specific activities to ensure quality control.\nDetailed protocols for extracting data from secondary sourceswill
eloped, piloted, refined and agreed in Phase 2. Quality will be assured through routine monitoring by the lead country
earcher, and periodic cross-checks against the protocols by the UKbased research assistant."
     ]
   }
 ],
 "costs": [
   {
     "title": "Long-term storage",
     "description": "The cost model is twice the current cost of storage. At $1,850/usable TB, costs are estimated at
700/usable TB for the storage hardware for indefinite data retention.\nSince we anticpate producing ~5TB of data, this come
roximately $18,500",
     "currency_code": "usd",
     "value": 18500
   }
 ],
 "description": "This project aims to find labeled, weighted networks of chemical-gene, gene-gene, gene-disease, and chemic
ease relationships based on single sentences in PubMed abstracts. This will generate the following, for each record:\nPubMe
nSentence number (0 = title)\nFirst entity name, formatted\nFirst entity name, location (characters from start of
tract)\nSecond entity name, formatted\nSecond entity name, location\nFirst entity name, raw string\nSecond entity name, raw
ing\nFirst entity name, database ID(s)\nSecond entity name, database ID(s)\nFirst entity type (Chemical, Gene, Disease)\nSe
ity type (Chemical, Gene, Disease)\nDependency path\nSentence, tokenized",
 "ethical_issues_exist": "yes",
 "ethical_issues_report": "https://dev9a22.infomanaged.co.uk/86y3z/ethics-application-eth1920-0001",
 "ethical_issues_description": "I will make sure to consult with the funder to ensure our IP is correctly registered in the
ht places",
 "created": "2019-10-09T07:53:54.368Z"
```

Coming soon...

Labeled, weighted networks of chemical-disease relationships

Dataset

Reynolds, W. 2019. *Labeled, weighted networks of chemical-disease relationships.*

Title	Labeled, weighted networks of chemical-disease relationships
Authors	Reynolds, W.
Abstract	This release contains the annotated network for the September 15, 2019 version of PubTator. The version discussed in our paper is an older one - from April 30, 2016. If you're interested in that network, it can be found in Version 1 of this repository. We will be releasing updated networks periodically, as the PubTator community continues to release new versions of named entity annotations for Medline each month or

File access - Access request

EPSRC Funded Research Project

Data Management Plan

Principle Investigator

Prof Westbrook Reynolds, ORCID ID: https://orcid.org/4040-4618-9943-1220

Project Title Finding a global network of biomedical relationships

Funders

Engineering and Physical Sciences Research Council, Grant ID:

Collaborators

Describe the data that will be generated as part of this funded project

Labeled, weighted networks of chemical-disease relationships: This release contains the annotated network for the September 15, 2019 version of PubTator. The version discussed in our paper is an older one - from April 30, 2016. If you're interested in that network, it can be found in Version 1 of this repository. We will be releasing updated networks periodically, as the PubTator community continues to release new versions of named entity annotations for Medline each month or so.

Describe the process by which you will ensure the archived data is of good quality, and reusable by other researchers?
Labeled, weighted networks of chemical-disease relationships: The PI will be responsible for overall quality assurance, with lead country researchers and the UK research assistant undertaking specific activities to ensure quality control. Detailed protocols for extracting data from secondary sourceswill be developed, piloted, refined and agreed in Phase 2. Quality will be assured through routine monitoring by the lead country researcher, and periodic cross-checks against the protocols by the UKbased research assistant.

haplo.com/repository
@HaploRepo
github.com/haplo-org/haplo-repository

Tom Renner -
tom.renner@haplo.com

Research Manager