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

    - Majority of our work is open-source

- All modules can co-exist within one application, and work from the same data

Research Manager

maDMPs – (our) motivation

- Haplo a natural home for DMPs, as it can manage the full life-cycle of research

- 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

Research Manager

DMPs in Haplo

- Use the RDA DMP Common Standard for machine-actionable Data Management Plans
- This has helped ensure DMP functionality is:
    1) Compatible with sector
    2) Modular within Haplo
    3) Actionable within and outside Haplo

Research Manager

DMPs in Haplo

Haplo

Quick search

Search | Browse | Recent | Add

Susi Davies ⚙ | Help

Dr Susi Davies

Researcher

Name	Dr Susi Davies
Contact category	Researcher
Research institute	Computer Science
ORCID	ⓘ https://orcid.org/7047-5452-4474-6221 (unauthenticated)
Unit of assessment	11. Computer Science and Informatics
Email address	susi.davies@example.org

Biography

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Teaching

Curabitur pretium tincidunt lacus. Nulla gravida orci a odio. Nullam varius, turpis et commodo pharetra, est eros bibendum elit, nec luctus magna felis sollicitudin mauris. Integer in mauris eu nibh euismod gravida. Duis ac tellus et risus vulputate vehicula.

WEB PROFILE

Researcher profile

View public profile

Download CV

Doctoral Researchers

Research outputs

Projects

Ethics

Finding a global network of biomedical relationships derived from text

Edit ⌄   📥

Attach file

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 derived from text
Research institutes	
Lead	Computer Science
Collaborator	Life Sciences
Researchers	
Principal Investigator	Dr Susi Davies
Objectives	To find a 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). The themes themselves are listed below and are in our paper (reference below).

PART II: Connects sentences to dependency paths. It consists

Milestones and actions

Project meetings

Contract approval

Costing

Ethics

DATA MANAGEMENT PLAN

Add

PROPOSALS

The power number government the (proposal)

Haplo

🔍 Quick search   **Search** | **Browse** | **Recent**   **Add**

Susi Davies ⚙ | **Help**

Home

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Accessibility

Add data management plan

Data management plan   Datasets   Costs   Ethical issues

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.
Each record contains the following information:

Language *

English ⇕

Save and continue   Save for later

Data management plan   Datasets   Costs   Ethical issues

Quick search | Search | Browse | Recent | Add

Susi Davies ⚙ | Help

Add 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

Accessibility

Data management plan | **Datasets** | Costs | Ethical issues

Title *

Labelled, weighted networks of chemical-disease relationships

Description of the dataset

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.

Type of dataset *

Archive ⇕

Expected size of dataset

1TB + ⇕

Keywords to describe the dataset

Enter one keyword per box

[   ]

[Add another]

Does this dataset contain personal data? *

🔘 Yes

Haplo

Quick search

Search   **Browse**   **Recent**   **Add**

Susi Davies ⚙ | Help

Add 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

Accessibility

This form has some data prefilled from proposals related to your project. If anything is incorrect you can edit it here before submitting.

Data management plan   Datasets   **Costs**   Ethical issues

Title *

Finding biomedical relationships derived from text (proposal)

Description

To find a labeled, weighted networks of chemical-gene, gene-gene, gene-disease, and chemical-disease relationships based on single sentences in PubMed abstracts.
PART I: Will connect dependency paths to labels, or "themes". Each record

Value and currency

10,000   GBP

Remove

Add another

Save and continue   Save for later

Data management plan   Datasets   **Costs**   Ethical issues

Haplo

Quick search   **Search**   **Browse**   **Recent**   **Add**   **Susi Davies** ⚙ | **Help**

Add 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

Accessibility

> This form has some data prefilled from ethics applications related to your project. If anything is incorrect you can edit it here before submitting.

Data management plan   Datasets   Costs   **Ethical issues**

Are there ethics issues related to the data this DMP describes? *

- ● Yes
- ○ No
- ○ Unknown

Location of the report on ethical issues

`https://dev6a2.infomanaged.co.uk/85897/ethics-application-eth1920-0005`

Description

The project involves access to, or use of, material which could be classified as personally sensitive, as it reports on medical information from other stidues.

[Remove]

[Add another]

[Save and finish]   [Save for later]

View data management plan

Changes from: 17 Feb 2020 ⇅   **Filter questions**

Data management plan

Plan title

+ Deriving a global network of biomedical relationships from text
− Data Management Plan: Finding a global network of biomedical relationships derived from text

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.

Each record contains the following information:

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

SECTIONS

Data management plan

Datasets

Costs

Ethical issues

VERSIONS

17 Feb 2020 (Susi Davies)

17 Feb 2020 (Susi Davies)

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Accessibility

Home

< Back

New Dataset: Dr Susi Davies

Home

University of Example

Humanities

Media

Science and Technology

Committees

Upcoming meetings

REF Units of Assessment

Calendar

Past events

Using this demo

Guides

Accessibility

Please complete this form as fully as possible. Fields may be left blank for information that is not applicable to this output. The University repository will be alerted to new or edited outputs.

[Preview]   [Save]

+ Title

[⇅] [Labelled, weighted networks of chemical-gene relationship:] **−**

+ Authors   ⑦

[🔗 Dr Susi Davies]   [Davies, S.] **−**

+ Abstract   **−**

This release contains the annotated network for the September 15, 2019 version of PubTator. The version discussed in our paper, below, 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   ⑦

[   ] **−**

+ Year

[📅 yyyy]   **−**

Is this dataset in the project's DMP?

If the created dataset was listed in the project's DMP select it below, if not then select 'None of these options'. The items are listed with their expected titles and descriptions from the DMP

None of these

This dataset wasn't listed in the DMP

Labelled, 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, below, 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.

Labeled, weighted networks of chemical-gene relationships based on single sentences in PubMed abstracts

Edit ⌄   ⬇

Copy as new item   Export   Change output type

Dataset

Title	Labeled, weighted networks of chemical-gene relationships based on single sentences in PubMed abstracts
Authors	Davies, S.
Abstract	This release contains the annotated network for the September 15, 2019 version of PubTator. The version discussed in our paper, below, 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 relationships Text mining
Year	2020
Project	Finding a global network of biomedical relationships derived from text
Data files	Example.xlsx

THUMBNAIL NOT AVAILABLE   ⬇ Download   ⌥ Versions   👁 Preview

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

STATUS
Waiting for review

CURRENTLY WITH
Repository Editors

🟢 Deposit

🟡 Deposit closed access

🟡 Return to submitter

🟡 Put on hold

🔴 Reject

Set embargo

OPEN ACCESS

Open access by year   Open access by REF Unit of Assessment

Open access by Faculty   APCs overview

Open access by Department

Ingest from DOI...

USAGE STATISTICS

Overview by Faculty   Overview by Department

MANAGE LISTS

Publisher   Funder

Journal

EXPORTS

JISC APC Spreadsheet

DATA MANAGEMENT PLANS

DMP reports by project   DMP reports by PGR project

ORCID

ORCID Progress   Output ORCID Progress by Faculty

ORCID Progress by Faculty   Output ORCID Progress by Department

ORCID Progress by Department

CUSTOMISATION

Edit Repository overview   Edit Repository guidance notes

Re-order repository output types

Data management plans by project

1
Total datasets deposited

8
Total datasets expected

Export

Project	Created	Last modified	Large files expected	Sensitive data expected	Sensitive data secured	Datasets expected	Datasets linked to DMP
Finding a global network of ...	17 Feb 2020	17 Feb 2020	✓	✓		1	1
A review of laws passed in t...	17 Feb 2020	17 Feb 2020	✓		✓	1	—
Man parent that city person	17 Feb 2020	17 Feb 2020			✓	3	—
Time information place cas...	17 Feb 2020	17 Feb 2020		✓	✓	1	—
By man thing student lot pl...	17 Feb 2020	17 Feb 2020		✓	✓	2	—

Coming soon...

- Integrating with other systems that use the RDA DMP Common standard

- Depositing to the public repository

- Minting DOIs for DMPs

- Extending reporting as use cases emerge

Research Manager

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

Tom Renner
tom.renner@haplo.com

Research
Manager

Data management plans by PGR project

0
Total datasets deposited

3
Total datasets expected

Export

◄ 2019 - 2020 ►

Project	Created	Last modified	Large files expected	Sensitive data expected	Sensitive data secured	Datasets expected	Datasets linked to DMP
Government on case health...	17 Feb 2020	17 Feb 2020			✓	1	—
Community others word pre...	17 Feb 2020	17 Feb 2020			✓	2	—

Home

Univesity of Example
Humanities

Media

Science and Technology

Committees
Upcoming meetings

REF Units of Assessment

Calendar
Past events

Using this demo

Guides

Accessibility

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.

```
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 version
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 version
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 h
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 chemica
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"
```