Role of Wikimedia in the era of Open Science, Barcelona, 16th April 2018
What is the problem you're trying to solve? [edit]

"Improve biomedical content within Wikimedia, by building an algorithmic version of the medical references guideline." Wikimedia's medical content is hugely popular, with medical pages on English Wikipedia being visited 2.2 billion times in 2017. This makes it imperative that the information is kept accurate and current. However, creation of well-referenced biomedical content within Wikimedia is challenging due to the volume of appropriate secondary literature, totaling tens of thousands of papers per year. This greatly complicates finding appropriate references, and verifying that the statements made are actually supported by those references.

Two subproblems we aim to solve are that (a) specialist editors cannot currently easily access and screen potentially relevant papers due to volume. Once papers are identified, (b) currently it can't easily be connected to multiple articles in Wikipedia and entries in Wikidata in ways that would help verify the relevance of the content and claims.

We believe these problems can mostly be resolved through the combination of machine automation and input by editors that we propose here in the ScienceSource platform.

summary

ScienceSource makes a formal description and algorithm out of the MEDRS guideline on reliable sources, and applies it to the referencing of biomedical facts on Wikidata, and elsewhere.

target

Wikidata, Wikispecies, and infoboxes in the medical area that invoke Wikidata statements.

type of grant

Software
Peter Murray-Rust

From Wikipedia, the free encyclopedia

Peter Murray-Rust (born 1941) is a chemist currently working at the University of Cambridge. As well as his work in chemistry, Murray-Rust is also known for his support of open access and open data.

Contents
1 Education
2 Research
3 Achievements
4 References
5 External links

Education [edit]

He was educated at Botham School and Balliol College, Oxford. After obtaining a Doctor of Philosophy with a thesis entitled A structural investigation of some compounds showing charge-transfer properties, he became lecturer in chemistry at the (new) University of Stirling and was first warden of Andrew Stewart Hall of Residence. In 1982, he moved to Glaxo Group Research at Greenford to head Molecular Graphics. Computational Chemistry and later protein structure determination. He was Professor of Pharmacy in the University of Nottingham from 1996–2000, setting up the Virtual School of Molecular Sciences. He is now Reader Emeritus in Molecular Informatics at the University of Cambridge and Senior Research Fellow Emeritus of Churchill College, Cambridge.

Research [edit]

His research interests have involved the automated analysis of data in scientific publications, creation of virtual communities, e.g. The Virtual School of Natural Sciences in the Globewide Network Academy, and the Semantic Web. With Henry Rzepa, he has extended this to chemistry through the development of markup languages, especially Chemical Markup Language. He campaigns for open data, particularly in science, and is on the advisory board of the Open Knowledge International and a co-author of the Panton Principles for Open scientific data. Together with a few other chemists, he was a founder member of the Blue Obelisk movement in 2005.

In 2002, Peter Murray-Rust and his colleagues proposed an electronic repository for unpublished chemical data called the World Wide Molecular Matrix (WWMM). In January 2011, a symposium around his career and visions was organized, called Visions of a Semantic Molecular Future. In 2011, he and Henry Rzepa were joint recipients of the Herman Skolnik Award of the American Chemical Society. In 2014, he was awarded a Fellowship by the Shuttleworth Foundation to develop the automated mining of science from the literature.
**ScienceSource**

**Problem:** “Improve biomedical content within Wikimedia, by building an algorithmic version of the medical references guideline.“

**Solution:** “We propose the combination of a new platform, ScienceSource, a community working on it, and software development.”

**ScienceSource** makes a formal description and algorithm out of the MEDRS guideline on reliable sources, and applies it to the referencing of biomedical facts on Wikidata, and elsewhere.

ScienceSource: Approach

“The ScienceSource platform will be a collaborative MediaWiki site. It will collect and convert up to 30,000 of the most useful Open Access medical and bioscience articles and convert them.

We will work with two Wikimedia communities (Wiki Med and WikiJournal) to develop machine-assisted human-reviewing. The wiki platform will facilitate the decision-making process, driven by the human reviewers.”

ScienceSource: Approach (2)

“Articles will be annotated with terms in WikiFactMine dictionaries. In this project, those dictionaries will include, for example, diseases, drugs, genes. This not only means that the useful terms are highlighted, but they are also linked to entries in Wikidata and therefore to any relationship that is described in Wikidata. Thus “aspirin” links to d:Q18216 with synonyms, disease targets, chemistry, etc.”

ScienceSource: Approach (3)

“Advantages of this route to semi-automation (putting humans in the loop) include full documentation of the rationale for decisions, discussions in annotation form rather than prose, and the combination of human and machine inputs on the same footing. Features will include

(i) metadata based on [d:Wikidata:Source Metadata](https://meta.wikimedia.org/wiki/Grants:Project/ScienceSource), i.e. the rapid development of WikiCite,

(ii) potential incorporation of topic, experimental method, and key components relating to medical trials, e.g. [w:Consolidated Standards of Reporting Trials](https://meta.wikimedia.org/wiki/Grants:Project/ScienceSource) (CONSORT) metadata.”
Project goals

The overarching goal of the project (outcome) is to create a high-quality corpus of Wikidata-annotated biomedical articles that will be used by Wikimedians in the Wiki Med and WikiJournal communities and the wider world. In other words, the aim is good coverage on the ScienceSource wiki and Wikidata of recent biomedical review literature. Such a corpus will assist Wikimedians in writing and referencing medical content, to a high standard, and closely linked with Wikidata's science and metadata content. It will also add to the prominence of Wikidata and the WikiCite initiative. Most importantly, infobox content in medical areas drawn in from Wikidata will be more reliable and of improved quality.

<table>
<thead>
<tr>
<th>Specific Goal Description</th>
<th>Wikimedia Project benefit</th>
<th>Wikimedia community benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Import new referenced facts into Wikidata, conforming with the medical references guideline, and improve the quality of referencing of existing biomedical statements in Wikidata, by adding references, or replacing existing references.</td>
<td>WikiMed, WikiJournal</td>
<td>WikiMed, WikiJournal</td>
</tr>
<tr>
<td>Metadata improvement for Wikidata's items on biomedical papers.</td>
<td>WikiMed, WikiJournal</td>
<td>WikiMed, WikiCite</td>
</tr>
<tr>
<td>Build a working ScienceSource community, as a participatory technical platform.</td>
<td>WikiMed, WikiJournal</td>
<td>WikiMed, WikiJournal</td>
</tr>
</tbody>
</table>

Project impact

How will you know if you have met your goals?

<table>
<thead>
<tr>
<th>Specific Goal Description</th>
<th>Measurement criteria</th>
<th>Actions taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Import new referenced facts into Wikidata, conforming with the medical references guideline, and improve the quality of referencing of existing biomedical statements in Wikidata, by adding references, or replacing existing references.</td>
<td>Process 30,000 downloaded papers, including all WikiJournal papers within the biomedical scope.</td>
<td>Acquire open access papers by downloading to the platform. Export from the annotations structure to RDF, and upload to Wikidata. Track by wiki histories.</td>
</tr>
<tr>
<td>Metadata improvement for Wikidata's items on biomedical papers.</td>
<td>Metadata import to Wikidata, 15,000 statements.</td>
<td>Acquire open information on e.g. publication type, retraction, topics. Import to Wikidata via bot, track by edit summary.</td>
</tr>
<tr>
<td>Build a working ScienceSource community, as a participatory technical platform.</td>
<td>Annotations on platform: 3,000 contributions.</td>
<td>Store annotations, in a format compliant with the W3C standard. Use wiki norms to develop community control, wiki tools to track community work.</td>
</tr>
</tbody>
</table>

Continuing impact
Community engagement

Past engagement

ContentMine has extensively engaged the Wikimedia community in the past four years. Peter Murray-Rust delivered a keynote talk at Wikimania 2014 and at the Wikipedia Science Conference 2016, where we also ran a hands-on workshop. We have introduced Wikidata to audiences in over 50 presentations at UK and international meetings, highlighting it as a key resource for those searching for scientific knowledge and as the future of large-scale scientific data curation.

We contributed to linking Wikimedia with mining the scientific literature and online, for example at the July 2017 Cambridge text and data mining conference. For example, we produced a regular weekly blogpost series aimed at librarians and the Cambridge research community and a monthly newsletter delivered by English Wikipedia Mass message. We also produced and maintained the central WikiFactMine "hub" on Wikidata and totally revamped the Wikidata Sources page to help support WikiCite. ContentMine has contributed resources to the regular Cambridge Wikimedia meetups.

Participation in the ScienceSource project would be promoted by the same mixture of regular communications, training and workshops, and participation in conferences and meetups.

Wikimania 2018 Workshop

We aim to have a beta system by Wikimania 2018 that can be tried out by Wikimedians. We will run at least one introductory workshop in Capetown.

Wikidata 6th birthday event

We will have support from WMUK in organising a local event for Wikidata's 6th birthday in October, working title "Data modelling for Wikidata".

Summary table for community engagement activities

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Description</th>
<th>Month</th>
<th>Effort (person months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Meetups (Cambridge)</td>
<td>Organise and deliver a three-monthly meetup, at or near ContentMine offices (10-20 people per session)</td>
<td>1 every quarter</td>
<td>0.2</td>
</tr>
<tr>
<td>C2</td>
<td>Workshops</td>
<td>Organise and deliver four workshops for users and developers, e.g. at Wikimania 2018, WikiCite, Mozilla and FORCE11, follow-up on results (20-50 people per session)</td>
<td>1 every quarter</td>
<td>0.5</td>
</tr>
<tr>
<td>C3</td>
<td>Conference presentation</td>
<td>Deliver presentation during Wikimania 2018 and WikiCite (20-100 people per session)</td>
<td>M1-M12</td>
<td>1</td>
</tr>
<tr>
<td>C4</td>
<td>Newsletter</td>
<td>Deliver a monthly newsletter, reaching wikimedians, volunteers and community members (100 people on newsletter list).</td>
<td>1 every month</td>
<td>0.2</td>
</tr>
<tr>
<td>C5</td>
<td>Project webpages and social media</td>
<td>Ensuring development work and results are communicated through ContentMine's own site, wiki page and social media to interested communities at each step in the process.</td>
<td>M1-M12</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Get involved [edit]

Participants [edit]

**Peter Murray-Rust (Founder and Director of ContentMine)**

Peter has been a Wikimedian since 2006 and delivered a keynote talk at Wikimania 2014 and Wikipedia Science Conference 2015, where CM also ran a hands-on workshop. Peter founded ContentMine as a Shuttleworth Foundation Fellow, and is the main software pipeline architect. He received his Doctor of Philosophy from the University of Oxford and has held academic positions at the University of Strting and the University in Nottingham. His research interests have focused on the automated analysis of data in scientific communities. In addition to his ContentMine role, Peter is also Reader Emeritus in Molecular Informatics at the Unilever Centre, in the Department of Chemistry at the University of Cambridge, and Senior Research Fellow Emeritus of Churchill College in the University of Cambridge. Peter is renowned as a tireless advocate of open science and the principle that the right to read is the right to mine.

**Jenny Molloy (Director of ContentMine)**

Jenny is a molecular biologist by training and manages ContentMine collaborations and business development. She spoke on synthetic biology at Wikipedia Science Conference 2015 and has been a long term supporter of open science. She is also a Director of Biomakespace, a non-profit community lab in Cambridge for engineering with biology.

**Charles Matthews (Wikimedian in Residence, ContentMine)**

Once a lecturer at DAMMS, Cambridge, Charles has been a Wikimedian since 2003, and is active also on Wikisource and Wikidata. He has been a staff member and contractor for Wikimania UK, and is co-author of "How Wikipedia Works". He was employed by ContentMine in 2017 as Wikimedian in Residence at the Moore Library, and on their WikiFactMine project, and continues to work with them as a volunteer.

Wikimedian advisors (provisional)

- Daniel Mietchen, technical and scientific
- Evolution and evolvibility, WikiJournal
- Magnus Manske, technical and scientific
- RexxR, technical and Wiki Project Med
- T Arrow, technical and WikiCite

Community notification [edit]

Early forms of this proposal have been trailed widely to Wikimedians, and significant discussions held. Advance notifications of the proposal were: Facto Post Issue 5, 17 October 2017 (see w:User:Charles Matthews/Facto Post/Issue 5 — 17 October 2017), editorial on annotations, link to d:Wikidata:WikiFactMine/Annotation for fact mining and invitation to comment. This mass message is delivered to seven WikiProject talk pages, listed at w:Wikimedia Foundation mailing list, as well as individual editors. It was supported by:

- Invitation to discuss posted to d:Wikidata Project Chat permalink
How to participate

“If you feel this is a good idea, just go the URL and leave a message. This isn't committing you to anything - it's just showing interest. Of course if you think you can either contribute or use ScienceSource that's great. If you've never edited Wikipedia before it's simple - find the "edit" button to open the wiki - switch to the new visual editor, add your message at the bottom. You can leave your name or be anonymous - whatever suits. If you get it wrong, don't worry - wikimedians or their bots will tidy it up.”

Peter Murray-Rust

Welcome to Wikidata

the free knowledge base with 46,723,457 data items that anyone can edit.

Introduction • Project Chat • Community Portal • Help

Want to help translate? Translate the missing messages.

Welcome!

Wikidata is a free and open knowledge base that can be read and edited by both humans and machines.

Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wikisource, and others.

Wikidata also provides support to many other sites and services beyond just Wikimedia projects! The content of Wikidata is available under a free license⁶, exported using standard formats, and can be interlinked to other open data sets on the linked data web.

Learn about data

New to the wonderful world of data? Develop and improve your data literacy through content designed to get you up to speed and feeling comfortable with the fundamentals in no time.

Item: Earth (a2i)  property: highest point (a1t)  custom value: Mount Everest (a1t)

Get involved

Learn about Wikidata

- What is Wikidata? Read the Wikidata introduction.
- Explore Wikidata by looking at a featured showcase item for author Douglas Adams⁷.
- Get started with Wikidata's SPARQL query service.
How researchers could ensure discoverability of research data through Wikipedia?

How to increase the use of Wikidata in research community?
The Open Science Fellows Program is a joint project of Wikimedia Deutschland, the Stifterverband, and the Volkswagen Foundation. It is aimed at doctoral students (m/f), post-docs (m/f), and junior professors (m/f) who want to promote their research in an open manner. As scientific partners, the Technische Informationsbibliothek (TIB), the Museum für Naturkunde Berlin, the Center for Digital Systems (CeDiS) at Free University Berlin and the Gottfried Wilhelm and University Library are participating in the program by offering a range of training opportunities. The program seeks to make aspects of the scientific process such as methodology, research data, and publications openly accessible and usable.

In this way, data and information are made available to the public – but also to other researchers – as a “knowledge commons.” A primary focus of the program is to strengthen the concept of Open Science and research within the meaning of Open Science, where transparency, reproducibility, and the comparability of scientific work are essential requirements. Above all, the program will facilitate the exchange and networking of active participants in the area of Open Science to advance the gradual dissemination of science and research. To this end, the program offers the opportunity to network with experts from different disciplines who offer insights into their open research practice. Questions, suggestions, and ideas about the program can be communicated via our discussion page.

Information about the Open Science Fellows Program in German can be found here.

Program partners [edit]

Wikimedia Deutschland works to ensure that the knowledge collected by humankind is openly accessible to everyone. Our contribution to make this vision of the future into reality is laid out in our charter. The purpose of the association is to promote the creation, collection, and distribution of open content in altruistic activities to advance education and equal access to knowledge. In addition to supporting various Wikimedian projects and their communities (e.g. Wikipedia, Wikidata, Wikiversity), we also work with educational and cultural institutions.

About 3,000 companies, business associations, foundations, and private individuals have joined forces in the Stifterverband to collectively promote science and education. With its support, the Stifterverband helps ensure the innovation infrastructure, i.e., high-performance universities, strong research institutions, and a productive exchange between industry and science. The Stifterverband supports a quarter of a million gifted young people every year. It also manages 645 foundations with total assets of over 2.7 billion euros.

The Volkswagen Foundation is an independent, non-profit foundation under private law based in Hannover. With a total funding volume of approximately 150 million euros per year, one of the country’s largest foundations overall. It allocates its funds exclusively to scientific institutions. In the more than 50 years of its existence, the Volkswagen Foundation has invested over 1 billion euros. It is consequently also one of the largest non-profit foundations in Germany under private law. The Volkswagen Foundation devotes special attention to young scientists.
Open Science Fellows Program

Goal

The primary aim of the Fellows Program is to promote the gradual opening of science and to spread the principle of collaborative knowledge production in accordance with the Wikipedia model. The program is geared towards academics from all disciplines who want to make their own research and teaching open and usable.

Program Structure

The Fellows Program supports the Fellows over a period of eight months and consists of four components: qualification, mentoring, financial support, and networking & visibility.

Qualification

In addition to theorectical basics on the topics of Open Knowledge and Open Science, the Fellows Program also shows participants how to apply them in practice. The focus is on supporting the Fellows in their work, acquainting them with methods and procedures, but also with tools they can use to promote Open Science.

Mentoring

Over a period of eight months (October 2017 to 2018), the Fellows are individually advised by mentors who already apply Open Science methods in their research. They serve as permanent contact persons for the Fellows within their research projects. The Fellows are in regular contact with their mentors. This exchange can take the form of individual mentoring or peer-to-peer mentoring in the form of self-organized working groups. For joint work, the Fellows and their mentors agree on fixed goals and courses of action at the beginning of the program.

Financial support

For the program year 2017/2018, 20 fellowships will be awarded, each endowed with 5,000 euros. These are individual and customized fellowships that provide Fellows with the freedom and resources they need to pursue projects. Travel and accommodation costs for classroom events as part of the Fellows Program will also be reimbursed as required.

Networking and visibility

The program is designed to make the subject of Open Science more visible and to promote it in scientific institutions and communities. A central component of the program is the open scientific communication among other active participants in the area of Open Science.
The future of science is Open

START YOUR RESEARCH TRAINING NOW

The FOSTER Plus bootcamp in Open Science will take place on 18-20 April at CRG, Barcelona

USE FOSTER TO:

- Access Free Courses
- Get Badges
- Earn Specialisation
- Attend live events
- Participate in the community

AVAILABLE LEARNING PATHS
Thank you! Questions?

iryna.kuchma@eifl.net
@irynakuchma
@fostersscience
Facebook: fosteropenscience