The Cross-Publisher Originality Screening Initiative

Rachael Lammey
Product Manager, CrossRef
October/November 2014

cross check
Powered by iThenticate

crossref

crossref

DOIs FOR RESEARCH CONTENT

crossref

I'm going to show you our CrossCheck service, but before I do that I'd like to give you a little bit of background on CrossRef - what CrossRef is, and why we are running an originality screening service. CrossRef was founded fourteen years ago to solve the problem of broken links. The web is all about links, but links break. This is annoying if you're browsing the web and want to follow an interesting link, but in the context of scholarly publishing it becomes more than annoying - if you can't follow a citation from one paper to another you're being hampered in your research. CItation linking is one of the greatest benefits of online publishing, but it really does need to be reliable. ...and publishers were finding that web sites changed, content moved, and links that they had put into their articles stopped working.
So they started a multi-publisher initiative to solve this problem of broken links.

10.1098/rstl.1665.0001

This is done using the DOI - the Digital Object Identifier, which I'm sure many of you are familiar with. A CrossRef DOI is simply a unique identifier for a piece of content. Once assigned, it doesn't change. It is to all intents and purposes a meaningless number, but it allows that piece of content to be located on the web.

User clicks on CrossRef DOI reference link in Journal A

User accesses cited article in Journal B

Tani, N., N. Tomaru, M. Araki, AND K. Ohba. 1996. Genetic diversity and differentiation in populations of Japanese stone pine (*Pinus pumila*) in Japan. *Canadian Journal of Forest Research* 26: 1454–1462.[CrossRef]
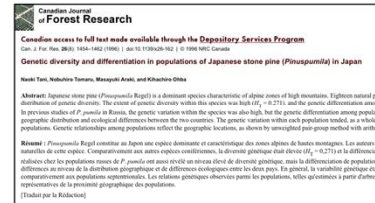
cross ref DOI directory returns URL

And it works like this: publishers use CrossRef DOIs to link to content, usually from the references at the end of articles. Users click on those DOI-based links and are referred via the CrossRef database to the cited article at it's correct location on the web. If content moves the publisher only has to update the CrossRef database once, and all of the publishers that are linking to their content using CrossRef DOIs will be redirected to the content in its new location.

- 4,208 publishers and societies
- 68,212,303 content items with DOIs
  - 28,582 journals
  - 343,924 books
  - 26,787 conference proceedings

A few numbers for you to give some idea of how CrossRef has grown in the fourteen years since its launch...
Books are the fastest growing at the moment - most publishers have assigned DOIs to their journals and journal archives, but more and more are now starting to assign them to their books, and to register their book metadata with CrossRef. Publishers are also registering components - 274,000 so far.

90,000,000

cross ref

Every month there are around 90 million clicks on CrossRef DOI links, so 90 million citations resolved to content.

crossref

- 24 staff (7 UK, 17 USA)
- 16 person Board of Directors, cross-section of publishers
- Working groups and committees

crossref

And a little about us as an organisation.
*UK and Boston, MA
*Publishers of all sizes, subjects, and nationalities - you can see the list on our website if you're interested.
*One of the reasons we achieve so much with such a small staff is that we're lucky to be supported by a network of working groups and committees.

So let's look at how much of a problem plagiarism is in scholarly publishing. It's certainly not a new problem, but is it one that's getting worse? It has never been easier to search across vast amounts of content in online publications and databases. And with more content being produced than ever before it's much harder for reviewers to have thoroughly read everything in their field.

Deja vu
A study of scientific publication ethics

Home | Browse | Report | Help | Statistics | Contact Us | Team | References

Powered by eTBLAST
Innovation Labs
Virginia Bioinformatics Institute

Deja Vu: a Database of Highly Similar Citations*

Click this link to begin browsing entries , or click the "Browse" button above and follow the instructions. To access the entries discovered by the SIP method, click SIP entries

We value your feedback. Please take one minute to take a brief survey (Click here). We appreciate your support.

Join the discussion of scientific publication ethics on COPE.

Deja vu is a database of extremely similar Medline citations. Many, but not all, of which contain instances of duplicate publication and potential plagiarism. Deja vu is a dynamic resource for the community, with manual curation ongoing continuously, and we welcome input and comments.

In the scientific research community plagiarism and multiple publications of the same data are considered unacceptable practices and can result in tremendous misunderstanding and waste of time and energy. Our peers and the public have high expectations for the performance and behavior of scientists during the execution and reporting of research. With little chance for discovery and decreasing budgets, yet sustained pressure to publish, or without a clear understanding of acceptable publication practices, the unethical practices of duplicate publication and plagiarism can be enticing to some. Until now, discovery has been through serendipity alone, so these practices have largely gone unchecked.

Latest News

2010-01-27 - Deja vu in Clinical Chemistry
An article about Deja vu has been published in Clinical Chemistry in January 2010. Read it.

2009-11-09 - Deja vu update
Deja vu database has recently been updated. A full text similarity ratio determined from manual examination has been assigned to each verified entry in the database. Users can filter the entries by specifying a range of full text similarity ratio. Classification of entries were also changed. Duplicate classification was removed. New classifications "Examined" and "Medline Issue" were added.

2009-05-27 - Nature Medicine News
Deja vu is in Nature Medicine News. Read it.

2009-05-22 - On Science NEWS FOCUS
A large article "Plagiarism Sleuths" published in Science News Focus

cross

And finally a study that has been reported and commented on in a number of places. A team led by Harold Garner and now based at the Virginia Bioinformatics Institute analysed the contents of Medline using a piece of text comparison software called eTBLAST.   By comparing texts against each other computationally they identified pairs of articles with high levels of similarity.

Where the software threw up matches they went on to manually compare the documents - as you can see here the areas of this article that appear in an earlier article have been highlighted. And they have written up their findings in each case in the deja vu database.

http://spore.vbi.vt.edu/dejavu/

cross**ref**

   You can access the database with their findings at the URL I've listed here. At the time that this article appeared in Science in May of 2009 year there had been at least 48 retractions of suspicious papers that were flagged to editors as a result of this project.

Deja Vu

- 238 pairs of articles with high similarity and no shared authors

- 1602 pairs with high similarity and one or more shared authors

- 588 pairs that eTBLAST flagged that are in fact legitimate

August 2009

The team went on to contact the authors and journals involved and documented their responses - some of them were written up in Science in March 2009.

"I fully endorse your proposal that publishers everywhere use text recognition software. This will be a useful at-source deterrent, as you point out... biomedical journals should at least have the standards widely used in today's high schools!"

Author of plagiarised article found by Garner et al. using eTBLAST

cross**ref**

This is a quote from one of the authors whose work was plagiarised. READ SLOWLY.
He is referring to plagiarism detection software used in schools, which I'll come back to a little later.

- 2006: CrossRef board raises plagiarism as area of concern

- Late 2007/ early 2008: pilot with seven publishers and technology partner iParadigms

- June 2008: CrossCheck launched

So we started to develop CrossCheck. We ran a pilot towards the end of 2007 and the start of 2008 with 7 major publishers and a technology partner iParadigms, and on successful conclusion of the pilot we launched the live CrossCheck service in June last year. So it's been running now for around 7 years.

cross
check
Powered by iThenticate

Plagiarism Detection

cross ref

The first thing that I always say when I talk about CrossCheck is that although we call it a plagiarism detection service, it doesn't actually detect plagiarism.
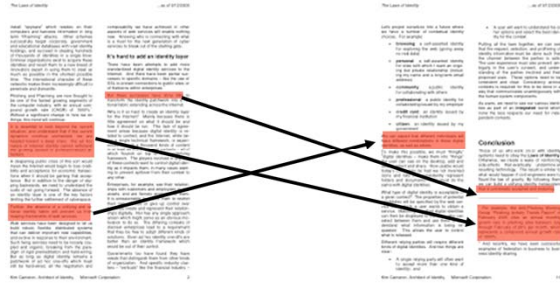
A machine cannot detect plagiarism. A machine can look at written text and tell you where it matches other written text, but it cannot tell you why that text matches, and this is critical. It takes a human being with a certain amount of domain expertise to analyse the results of any computer programme in order to determine the intent of the author or authors.

Legitimate Duplication

REFERENCES

1. Ferguson NM, Cummings DA, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 2005;437:209–14.
2. Longini IM, Nizam A, Xu S, et al. Containing pandemic influenza at the source. Science 2005;309:1083–7.
3. Ferguson NM, Cummings DA, Fraser C, et al. Strategies for mitigating an influenza pandemic. Nature 2006;442:448–52.
4. Mills CE, Robins JM, Bergstrom CT, et al. Pandemic influenza: risk of multiple introductions and the need to prepare for them. PLoS Med 2006;3:e135. (Electronic article).
5. Colizza V, Barrat A, Barthélemy M, et al. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med 2007;4:e13. (Electronic article).
6. Lloyd-Smith JO, Galvani AP, Getz WM. Curtailing transmission of severe acute respiratory syndrome within a community and its hospital. Proc Biol Sci 2003;270:1979–89.
7. Lipsitch M, Cohen T, Cooper B, et al. Transmission dynamics and control of severe acute respiratory syndrome. Science 2003;300:1966–70.
8. Choi BC, Pak AW. A simple approximate mathematical model to predict the number of severe acute respiratory syndrome cases and deaths. J Epidemiol Community Health 2003;57:831–5.
9. Dye C, Gay N. Epidemiology. Modeling the SARS epidemic. Science 2003;300:1884–5.
10. Donnelly CA, Ghani AC, Leung GM, et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet 2003;361:1761–6.
11. Wang W, Ruan S. Simulating the SARS outbreak in Beijing with limited data. J Theor Biol 2004;227:369–79.
12. Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world. Proc Natl Acad Sci U S A 2004;101:15124–9.
13. Meyers LA, Pourbohloul B, Newman ME, et al. Network theory and SARS: predicting outbreak diversity. J Theor Biol 2005;232:71–81.
14. Colizza V, Barrat A, Barthélemy M, et al. The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Natl Acad Sci U S A 2006;103:2015–20.
15. Viboud C, Bjornstad ON, Smith DL, et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 2006;312:447–51.
22. Bravata DM, McDonald KM, Smith WM, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Ann Intern Med 2004;140:910–22.
23. Centers for Disease Control and Prevention. Flu activity: reports and surveillance methods in the United States. Atlanta, GA: Centers for Disease Control and Prevention, 2007. (http://www.cdc.gov/flu/weekly/fluactivity.htm). (Accessed January 3, 2007).
24. Bonabeau E, Toubiana L, Flahault A. The geographical spread of influenza. Proc Biol Sci 1998;265:2421–5.
25. Viboud C, Boelle PY, Pakdaman K, et al. Influenza epidemics in the United States, France, and Australia, 1972–1997. Emerg Infect Dis 2004;10:32–9.
26. Baroyan OV, Genchikov LA, Rvachev LA, et al. An attempt at large-scale influenza epidemic modeling by means of a computer. Bull Int Epidemiol Assoc 1969;18:22–31.
27. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. New York, NY: Oxford University Press, 1991.
28. Greene SK, Ionides EL, Wilson ML. Patterns of influenza-associated mortality among US elderly by geographic region and virus subtype, 1968–1998. Am J Epidemiol 2006;163:316–26.
29. Sakai T, Suzuki H, Sasaki A, et al. Geographic and temporal trends in influenzalike illness, Japan, 1992–1999. Emerg Infect Dis 2004;10:1822–6.
30. Vergu E, Grais RF, Sarter H, et al. Medication sales and syndromic surveillance, France. Emerg Infect Dis 2006;12:416–21.
31. Bureau of the Census, US Department of Commerce. Census 2000 summary file 3. Washington, DC: Bureau of the Census, 2007. (http://www.census.gov/population/www/cen2000/). (Accessed January 8, 2007).
32. National Weather Service, National Oceanic and Atmospheric Administration, US Department of Commerce. State data. Silver Spring, MD: National Weather Service, 2007. (http://www.weather.gov/view/states.php). (Accessed January 3, 2007).
33. Langer WL. The black death. Sci Am 1964;210:114–21.
34. Noble JV. Geographic and temporal development of plagues. Nature 1974;250:726–9.
35. Riley S, Ferguson NM. Smallpox transmission and control: spatial dynamics in Great Britain. Proc Natl Acad Sci U S A 2006;103:12637–42.

cross ref

There are legitimate reasons why text might be the same in two documents. Here's a bibliography section which will almost-certainly be repeated in numerous places. A mathematical proof might be repeated in order to be extended, and so on... a human can spot some of these examples very easily whereas a piece of software cannot.

So let's take a look at how CrossCheck actually works by starting with a simple overview. You have a manuscript or document that you want to check for originality. You submit it to the CrossCheck system, where it is broken down and analysed, then compared against a database of other scholarly publications. Wherever overlapping or similar passages of text are found, they are highlighted in a report for an editor to take a look at.

cross check

Powered by iThenticate

- Software that analyses and compares text

- Database of content to check text against

cross ref

To effectively screen research material you need to compare it with other research material, and most of that is in publications that are on many different publisher platforms and often behind access control. So even if you find a match using Google Scholar you will still need to go to the publisher's website to see the abstract, which may or may not contain the matching text. If it doesn't, you need to get access to the full text, which may or may not involve paying, and so on and so forth.

PRESS ● E-MED LTD. ● ECOLOGICAL SOCIETY OF AMERICA ● ECOMED PUBLISHERS ● EDICIONES DOYMA ● EDP SCIENCES ● THE ELECTROMAGNETICS ACADEMY ● ELECTRONIC JOURNAL OF BIOTECHNOLOGY ● ELSEVIER ● EMERALD (formerly MCB UP) ● EMW PUBLISHING ● THE ENDOCRINE SOCIETY ● ENGLISH LITERATURE IN TRANSITION ● ENVIRONMENTAL HEALTH PERSPECTIVES ● EPI SCP (EL PROFESIONAL DE LA INFORMACION) ● EPP PUBLICATIONS LTD. ● EQUINE VETERINARY JOURNAL LTD. ● EQUINOX PUBLISHING LTD. ● ERDKUNDE ● ESTONIAN ACADEMY PUBLISHERS ● EUROPEAN ASSOCIATION OF AQUATIC MAMMALS ● EUROPEAN ASSOCIATION OF CARDIOTHORACIC SURGERY ● EUROPEAN OPTI_____ROPEAN RESPIRATORY SOCIETY ● FACULTY OF FAMILY PLANNING AND REPRODUCTIVE _____ENTAL PRESS LTD. ● FEDERATION OF AMERICAN SOCIETIES FOR EXPERI_____INSTEIN INSTITUTE FOR MEDICAL RESEARCH ● FIBRE DIFFRACTI_____FORUM FOR GLOBAL HEALTH PROTECTION ● FRAUNHOFER-I_____FRONTIERS IN BIOSCIENCE ● FUNDACION INFANCIA Y API_____E LTD ● GAZI UNIVERSITY - TECHNICAL EDUCATION FACUL_____OGICAL SOCIETY OF AMERICA (GSA) ● GEOPHYSICAL CENT_____CIENCEWORLD ● GUILFORD PUBLICATIONS ● GUNTHER EY_____GmbH/VIEWEG PUBLISHING ● THE ALAN GUTTMACHER IN_____TH PRESS, INC. ● HEALTH AFFAIRS (PROJECT HOPE) ● _____ECTRONIC & AUDIO-VIDEO PRESS ● HINDAWI PUBLISHIN_____HISTORIC BRASS SOCIETY ● HOGREFE & HUBER PUBLISHIN_____ETY ● HUMANA PRESS ● IFIS PUBLISHING ● ILLUMINA_____) ● IM PUBLICATIONS ● IMPROBABLE RESEARCH IN_____Y CENTER FOR GENOMICS AND BIOINFORMATICS ● INDIA_____NIVERSITY PRESS ● INFORMA UK LIMITED ● INSTITUTE F_____T SCIENCES (INFORMS) ● INSTITUTION OF CHEMICAL ENGIN_____HNOLOGY (IET) ● THE INSTITUTE OF ELECTRICAL AND ELECTRONIC_____CHEMISTRY AND BIOCHEMISTRY, ACADEMY OF SCIENCES OF THE CZ_____L OPTICS ● INSTITUTE OF PHYSICS (IOP) ● INSTITUTE OF PURE AND AP_____OF SOUTHEAST ASIAN STUDIES ● INTEGRATED OCEAN DRILLING PROGRAM-_____AL ● INTEL CORPORATION ● INTELLECT ● INTERNATIONAL ASSOCIATION FOR BRIDGE_____ENGINEERING (IABSE) ● INTERNATIONAL ASSOCIATION OF HYDROLOGICAL SCIENCES ● INTERNATIONAL BONE AND MINERAL SOCIETY ● INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL ● INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE ● INTERNATIONAL GLACIOLOGICAL SOCIETY ● INTERNATIONAL HORMESIS SOCIETY ● INTERNATIONAL JOURNAL OF PSYCHOANALYSIS, LONDON ● INTERNATIONAL LIFE SCIENCES INSTITUTE ● INTERNATIONAL PHYCOLOGICAL SOCIETY ● INTERNATIONAL READING ASSOCIATION ● INTERNATIONAL SOCIETY OF ENDOVASCULAR SPECIALISTS ● INTERNATIONAL UNION OF CRYSTALLOGRAPHY (IUCr) ● INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY (IUPAC) ● IWA PUBLISHING ● JAPAN LASER PROCESSING SOCIETY ● JAPAN SCIENCE AND

This is where CrossRef is able to bring something new to a service such as this. We already have working relationships with over 4000 publishers and societies, and can circumvent the need for so many bilateral agreements to create a comprehensive database of content.

PRESS ● E-MED LTD. ● ECOLOGICAL SOCIETY OF AMERICA ● ECOMED PUBLISHERS ● EDICIONES DOYMA ● EDP SCIENCES   ● THE ELECTROMAGNETICS ACADEMY ● ELECTRONIC JOURNAL OF BIOTECHNOLOGY ● ELSEVIER ● EMERALD (formerly MCB UP) ● EMW PUBLISHING ● THE ENDOCRINE SOCIETY ● ENGLISH LITERATURE IN TRANSITION ● ENVIRONMENTAL HEALTH PERSPECTIVES ● EPI SCP (EL PROFESIONAL DE LA INFORMACION) ● EPP PUBLICATIONS LTD. ● EQUINE VETERINARY JOURNAL LTD. ● EQUINOX PUBLISHING LTD. ● ERDKUNDE ● ESTONIAN ACADEMY PUBLISHERS ● EUROPEAN ASSOCIATION OF AQUATIC MAMMALS ● EUROPEAN ASSOCIATION OF CARDIOTHORACIC SURGERY ● EUROPEAN OPTICAL SOCIETY ● EUROPEAN RESPIRATORY SOCIETY ● FACULTY OF FAMILY PLANNING AND REPRODUCTIVE HEALTH CARE ● FDI WORLD DENTAL PRESS LTD. ● FEDERATION OF AMERICAN SOCIETIES FOR EXPERIMENTAL BIOLOGY (FASEB) ● THE FEINSTEIN INSTITUTE FOR MEDICAL RESEARCH ● FIBRE DIFFRACTION REVIEW ● FISHERIES SCIENCES.COM ● FORUM FOR GLOBAL HEALTH PROTECTION ● FRAUNHOFER-INSTITUT FUR MATERIALFLUSS UND LOGISTIK ● FRONTIERS IN BIOSCIENCE ● FUNDACION INFANCIA Y APRENDIZAJE ● FUTURE DRUGS ● FUTURE MEDICINE LTD ● GAZI UNIVERSITY - TECHNICAL EDUCATION FACULTY ● THE GENETICS SOCIETY OF AMERICA ● GEOLOGICAL SOCIETY OF AMERICA (GSA) ● GEOPHYSICAL CENTER OF THE RUSSIAN ACADEMY OF SCIENCES ● GEOSCIENCEWORLD ● GUILFORD PUBLICATIONS ● GUNTHER EYSENBACH, JMIR PUBLICATIONS ● GWV FACHVERLAGE GmbH/VIEWEG PUBLISHING ● THE ALA_____HARVEY WHITNEY BOOKS CO. ● THE HAWORTH PRESS, INC. ● HEALTH AFFAIRS_____UBLICATIONS ● THE HIGHER EDUCATION ELECTRONIC & AUDIO-VIDEO PRESS ● H_____TION ● THE HISTOCHEMICAL SOCIETY ● THE HISTOCH BR____SOCIETY ● HOGREFE &_____● HUMAN FACTORS AND ERGONOMICS SOCIETY ● HUMAN PRES_● IFIS PUBL_____EERING SOCIETY OF____RT____C____NA)___M PUBLICATIONS ● IMPROBA_____CIENCE ENTERPRISE ● IN_ANA_NILEPSI_CE___ER FOR GENOMICS AND BIOINFORMATICS ● INDIANA UNIVERSITY MATHEMATICS JOURNAL ● INDIANA UNIVERSITY PRESS ● INFORMA UK LIMITED ● INSTITUTE FOR OPERATIONS RESEARCH AND THE MANAGEMENT SCIENCES (INFORMS) ● INSTITUTION OF CHEMICAL ENGINEERS ● INSTITUTION OF ENGINEERING AND TECHNOLOGY (IET) ● THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS (IEEE) ● INSTITUTE OF ORGANIC CHEMISTRY AND BIOCHEMISTRY, ACADEMY OF SCIENCES OF THE CZECH REPUBLIC ● INSTITUTE OF PHYSICAL OPTICS ● INSTITUTE OF PHYSICS (IOP) ● INSTITUTE OF PURE AND APPLIED PHYSICS (IPAP) ● INSTITUTE OF SOUTHEAST ASIAN STUDIES ● INTEGRATED OCEAN DRILLING PROGRAM- MANAGEMENT INTERNATIONAL ● INTEL CORPORATION ● INTELLECT ● INTERNATIONAL ASSOCIATION FOR BRIDGE AND STRUCTURAL ENGINEERING (IABSE) ● INTERNATIONAL ASSOCIATION OF HYDROLOGICAL SCIENCES ● INTERNATIONAL BONE AND MINERAL SOCIETY ● INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL ● INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE ● INTERNATIONAL GLACIOLOGICAL SOCIETY ● INTERNATIONAL HORMESIS SOCIETY ● INTERNATIONAL JOURNAL OF PSYCHOANALYSIS, LONDON ● INTERNATIONAL LIFE SCIENCES INSTITUTE ● INTERNATIONAL PHYCOLOGICAL SOCIETY ● INTERNATIONAL READING ASSOCIATION ● INTERNATIONAL SOCIETY OF ENDOVASCULAR SPECIALISTS ● INTERNATIONAL UNION OF CRYSTALLOGRAPHY (IUCr) ● INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY (IUPAC) ● IWA PUBLISHING ● JAPAN LASER PROCESSING SOCIETY ● JAPAN SCIENCE AND

And this is precisely what we've done - we've facilitated the indexing of full text content from CrossRef publishers who join CrossCheck, and with iParadigms have put this content into a database to screen documents against. Just to talk about our technology partner for this project for a moment. Several technology providers were considered when we were planning CrossCheck. The one that we decided to work with is an organisation called iParadigms, who are based in Northern California. Their proven technology is probably best known as powering the Turnitin plagiarism screening tools for higher education. Turnitin is used widely in the UK and USA and I believe in several other European countries. Since 2003 it has also been available in the form of a commercial product called iThenticate. And it's the iThenticate system that is made available to CrossCheck participants.
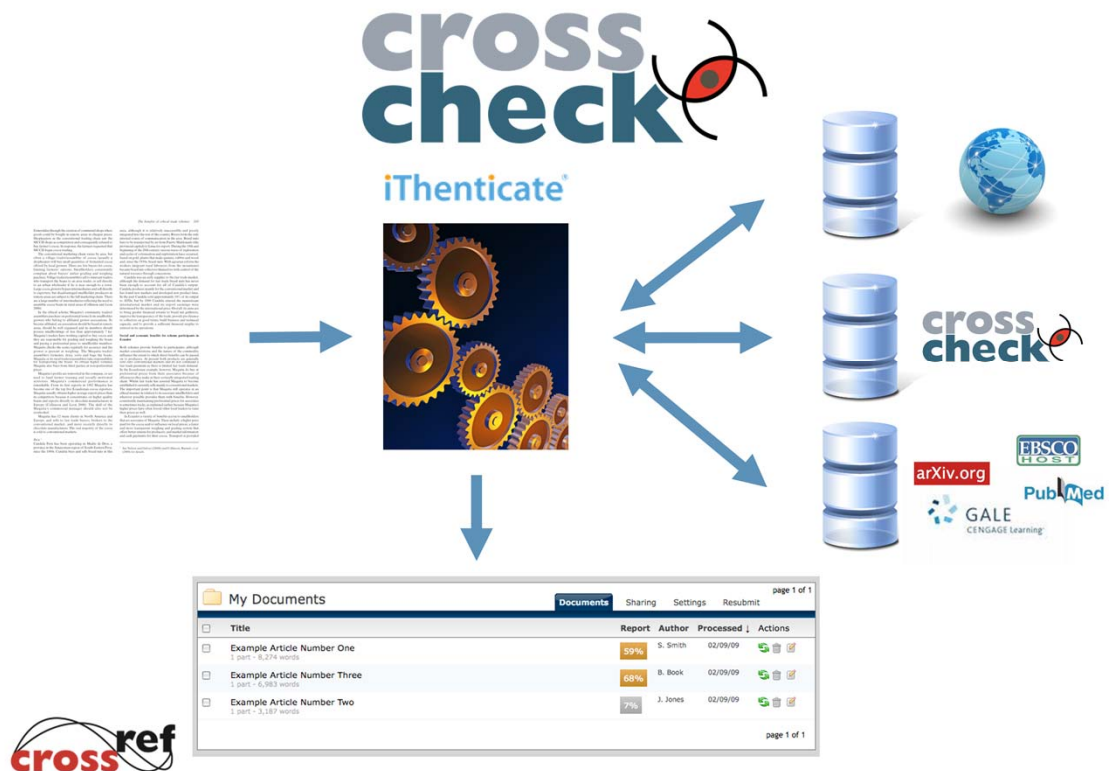
So to look at the process in a little more detail: you submit your manuscript to the iThenticate system, and it is by default checked against three databases of content. It is checked against web content - iThenticate indexes web pages in much the same way as a search engine, but with the added advantage that they keep an archive of web pages going back eight years.

The manuscript is checked against the CrossCheck database, which contains the content from all of the participating CrossCheck publishers.

And it's also checked against a growing repository of online and offline content that iThenticate is gathering and indexing, including datbases from Gale and Ebsco, and sites such as PubMed and Arxiv.org.

And as before, matches retrieved by comparison with these databases are pulled into a report for an editor to examine in more detail.

What's in the

iThenticate Database Content

STM Publishers:	Content Partners:	Internet:
38 Million	**92 Million**	**45 Billion**

STM Publishers:

38 Million

Journal articles, conference proceedings and books via CrossCheck powered by iThenticate from 530+ leading scientific, technical and medical (STM) publishers, including:

- American Chemical Society
- American Institute of Physics
- American Physical Society
- Elsevier
- IEEE
- Institute of Physics
- Lippincott Williams & Wilkins
- Nature Publishing
- Ovid
- Oxford University Press
- Sage Publications
- Springer
- Taylor & Francis
- Wiley Blackwell

Content Partners:

92 Million

Online and offline subscription content and research titles from 30 leading aggregators, databases and content providers, including:

- ABC CLIO
- Cengage Learning
- EBSCOHost: 2.7m periodicals, biographies, brochures, encyclopedias, magazines, journals, books, and abstracts
- Emerald Journals
- Gale: 86m articles
- Pearson, McGraw-Hill and Wiley: 2,000+ academic textbooks
- ProQuest: 300,000+ theses and dissertations
- PubMed/MedLine: 1.4m abstracts and citations; medical resources
- SAGE Reference: 160+ encyclopedia titles

Internet:

45 Billion

iThenticate's proprietary Internet crawler is comparable to major search engines. Archived back nearly a decade, iThenticate currently crawls 10 million web pages per day.

crossref   http://www.ithenticate.com/content

Why would something not be in the database?

- Not all publishers are CrossRef members

- Not all CrossRef members participate in CrossCheck: http://www.crossref.org/crosscheck_members.html

- Indexing needs to be enabled

- Timing – usually a few weeks between publication and indexing

- Needs a DOI!

crossref

This is the screen that you see when you've uploaded one or more manuscripts to iThenticate. You can see the article titles on the left, author and date processed on the right. The Report column with the square buttons beneath tells you what percentage of text within the manuscript has been found to match text in other documents.

Explain highlighting and thresholds.

And you get to this, which is the first of four different report manipulations available - this one is called the Similarity Report: Manuscript on left, matches on right from highest to lowest. Scroll up and down to compare.
URLs (plus date) or citation depending on database.   Links.
Ability to exclude a match if you know it's not relevant.   Click on the left to see side by side report

Point out print icon and filter and exclude items.
Show link to Document Viewer and touch on report view

Here you can see the two matching pieces of content side by side:
Glance feature, full source view
Importantly - entire article or piece of content on the right.
Scroll up and down and have opportunity to establish the context.
This is another benefit of using CrossCheck. We negotiated with iThenticate to allow users to see the complete article where there is matching text. If you use the commercial ithenticate product directly you only see snippets of text either side of the match. But we feel that it's important for publishers and editors to be able to see more than that.

one possible reason for these **results is that the ratings reflected** people's **attitude towards local Government rather than being the result of their interaction with the particular web site.**

One possible reason for the **results is that the** satisfaction **ratings reflected** [1] subjects' **attitude towards** their country's **local Government rather than being the result of their interaction with the particular web site.**

crossref

You might have spotted in the previous examples that the technology isn't just looking for word for word matches. The way that it breaks the text down allows it to spot passages of text with word substitutions, so it is looking for similar as well as identical text. In this example you can see that some of the words have been very subtly substituted or moved but the technology still picks them up.

Additional Features

Exclude Quotes

Exclude Bibliography

Exclude Small Matches

Section Exclusion & Phrase Exclusion

Filters & Settings

FILTERS

Exclude Quotes

Exclude Bibliography

Exclude sources that are less than:

○ [ ] words
○ [ ] %
● Don't exclude by size

Exclude matches that are less than:

○ [ ] words
● Don't exclude

Exclude Sections:

Abstract

Methods and Materials

Includes variations: Methods, Method, Materials, Materials and Methods

▼ Apply Changes

A Document Viewer overview is available at:

cross**ref**

Some additional features that you should be aware of: it's possible to exclude certain things to help reduce background noise. You can opt to exclude anything that's included in quotation marks. You can exclude the reference section, and you can choose not to be shown any matches below a certain number of words - so perhaps strings of fewer than 25 words would not be shown.

You should be aware that the first two of these features work on fairly hard and fast rules, so there need to be opening and closing quotation marks for a quote to be spotted and excluded, and the exclude bibliography feature relies on there being a recognisable section heading for

the references to identify it at the end of a document. So some documents will slip through these filters.

Limitations:
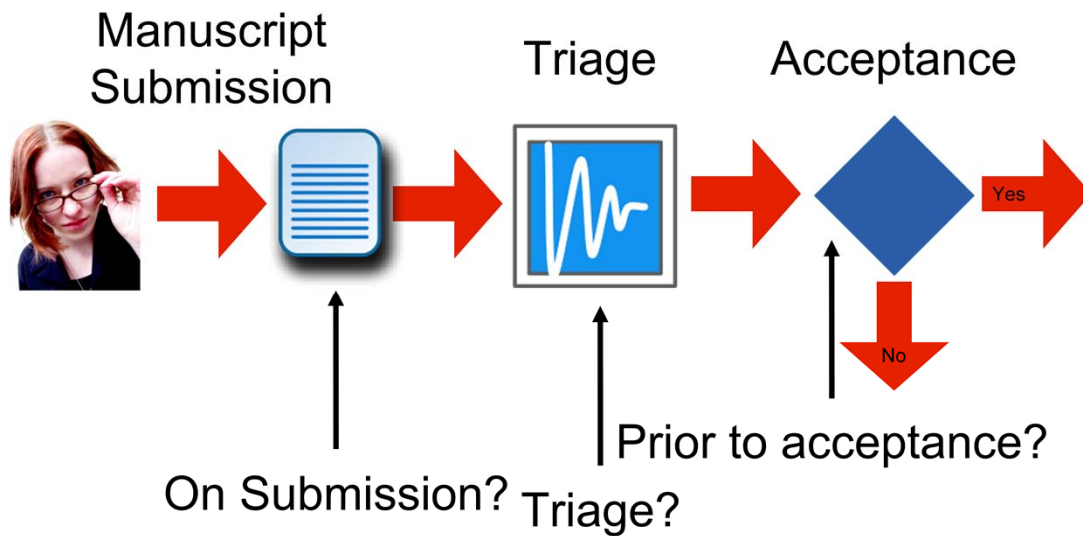photos or images
graphs and tables
formulae
Text only

Accepted Formats

We currently support file upload for the following document types:

Word

Word XML

Text

PostScript

PDF

HTML

WordPerfect WPD

RTF

crossref

Also copy and paste

Manuscript Submission — On Submission?

Triage — Triage?

Acceptance — Prior to acceptance?

Yes / No

crossref

Three obvious places where you might want to do plagiarism screening
1) On submission   2) At some defined point in the review and editorial process - - and obviously this is a massive over-simplification not in the least because this process varies widely form publisher to publisher
3) Just prior to acceptance
And we have CrossCheck members taking each of these approaches - trend moving towards on submission.

cross check

Powered by iThenticate

- 615 publishers
- Over 41 million content items indexed
  - 121,000+ titles
- 170,000+ manuscripts checked each month

cross ref

The progress of CrossCheck to date.
Very comprehensive database - can see list of titles on our website.

ACTA Press ● American Academy of Pediatrics ● American Association for the Advancement of Science (AAAS) ● American Association on Intellectual and Developmental Disabilities (AAIDD) ● American Diabetes Association ● American Geophysical Union (AGU) ● American Institute of Physics (AIP) ● American Physical Society (APS) ● American Psychological Association ● American Roentgen Ray Society ● American Society for Clinical Investigation ● American Society for Microbiology ● American Society for Nutrition ● American Society of Civil Engineers ● American Society of Neuroradiology (ASNR) ● American Society of Plant Biologists ● American Statistical Association ● American Thoracic Society ● Ammons Scientific ● Annual Reviews ● Association for Computing Machinery (ACM) ● Australian Academic Press ● BioMed Central ● BioSci[   ]hing Group ● British Institute of Non-Destructive Testing (BI[crossref.org/crosscheck   ]Clinic Journal of Medicine ● Co-Action Publishing ● Commonwealth Forestry Association ● Croatian Medical Journal ● CSIRO Publishing ● Digital Science Press (Urotoday International Journal) ● Earthquake Engineering Society of Korea ● EDP Sciences ● Electronics and Telecommunications Research Institute (ETRI) ● Elsevier ● Environmental Health Perspectives ● European Respiratory Society Journals ● Expert Reviews Ltd ● Fundacion Infancia & Aprendizaje (FIA) ● Future Medicine Ltd ● Future Science Ltd ● Geological Society of America ● Hindawi Publishing Corporation ● IM Publications ● IMAPS ● Inderscience ● INFORMS ● Institute of Control, Robotics and Systems ● Institute of Electrical & Electronics Engineers (IEEE) ● International Union of Crystallography (IUCr) ● IOP Publishing ● IWA Publishing ● Journal of Bone and Joint Surgery ● Journal of Histochemistry ● Journal of Neurosurgery Publishing Group ● Journal of Rehabilitation Research & Development ● Journal of Zhejiang University SCIENCE ● King Abdulaziz University Scientific Publishing Centre ● Korea Chemical Society ● Korea Ocean Research and Development Institute ● Korean Institute of Science and Technology Information (KISTI) ● Korean Pharmacopuncture Institute ● Korean Powder Metallurgy Institute ● Korean Society for Bioinformatics and Systems Biology ● Korean Society for Information Management ● Korean Society of Environmental Engineering ● Korean Society of Sport Biomechanics ● Kyung Hee Oriental Medicine Research Center, Kyung Hee Universit ● LWW / Wolters Kluwer Health ● Mary Ann Liebert ● Mayo Clinic Proceedings ● Mayo Clinic Scientific Publications ● Nature Publishing Group ● New England Journal of Medicine (NEJM/MMS) ● Oncology Nursing Society ● Optical Society of America ● Oxford University Press ● Palgrave Macmillan ● Poultry Science Association ● Proceedings of the National Academy of Sciences ● Professional

This is a partial list of member publishers - now that we're over 500 it's hard to fit them on one page and still make them readable but you can see the complete list on the CrossRef website so don't strain your eyes trying to read this one. Just to emphasise really that they are publishers of all sizes, of multiple nationalities and representing many disciplines.

- Webinars

- CrossRef support

- Formation of guidelines, best practices etc.

  - See:
    [http://www.crossref.org/crosscheck_plagiaris m_resources.html](http://www.crossref.org/crosscheck_plagiarism_resources.html)

- User group meetings

And we are also building something of a CrossCheck community. We have a users email listserv that members are invited to use to ask questions or share experiences.
In addition to the excellent support available from the iThenticate team, there are CrossRef staff such as myself available to help with setting up and running CC. And with the supporting CC committee we're looking to create guidelines and best practices for use of CC, and template plagiarism policies for those publishers that may not have one of their own.

Advanced CrossCheck training from ME Support team

- Scheduled for November and December.
- 3 webinars:
  - Interpreting results and using the flowchart in the plagiarism policy (looking at examples).
  - Exploring different CrossCheck settings.
  - Types of CrossCheck reports.

# Dealing with Plagiarism: Cochrane's plagiarism policy

- Available in the Cochrane Editorial and Publishing Policy Resource
- Launched September 2014
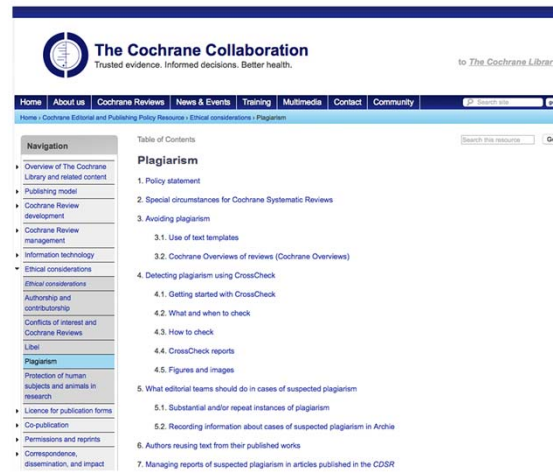
www.cochrane.org/editorial-and-publishing-policy-resource/plagiarism

If you've got a case of suspected plagiarism on your hands, it's important that it is dealt with sensitively and carefully. Cochrane offers a guide based on this. Cochrane has their own policy which is COPE-approved.

Figure 2. Flowchart: what to do if plagiarism is suspected
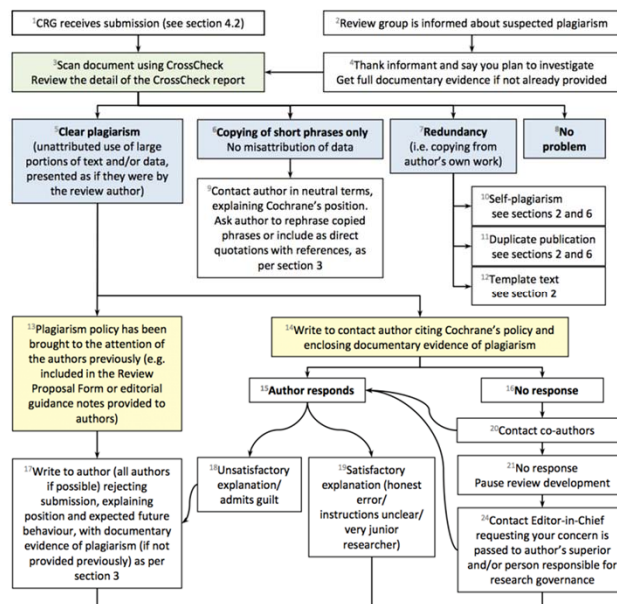
And flowcharts adapted from the COPE/Committee on Publication Ethics – a resource widely used in the publishing world. These walk you through the steps needed and the appropriate parties to contact at each point in the process to ensure the correct steps are followed.

And that's exactly why the CrossCheck service has been created - to help publishers screen for originality before it gets published. Once plagiarised content is published and out there it becomes quite a messy problem to fix. With CrossCheck we're providing tools that will hopefully mean any problems are found and sorted out much earlier on in the process. And ultimately it's about maintaining the integrity of your publications, and adding value through the publishing process.

How to access CrossCheck

- All Cochrane Review Groups can receive a login to CrossCheck.
- Contact Gavin Stewart at Wiley (gstewart@cochrane.org).

Thank You
www.crossref.org/crosscheck.html

Remind about Cochrane follow-up training.