

WHAT'S IN IT FOR ME? INCENTIVIZING THE USE OF INSTITUTIONAL REPOSITORIES

Varsha Khodiyar

Data Publishing Manager, *F1000Research*
@vkf1000

f1000research.com
@f1000research

Open Repositories 2014, Helsinki

ABOUT FACULTY OF 1000



The Seer of Science Publishing

Vitek Tracz was ahead of the pack on open access. Now he wants to rewrite the rules of peer review

LONDON—"Nobody reads journals," says science publisher Vitek Tracz, who has made a fortune from journals. "People read papers." Tracz sees a grim future for what has been the mainstay of scientific communication, the peer-reviewed print journal. Within the next 10 years, he says, it will cease to exist.

This prophecy ought to carry weight. Over the past 3 decades, Tracz, chairman of a conglomerate called the Science Navigation Group, has helped transform the world of science publishing. His most notable creation to date may be BioMed Central, the first for-

Tracz "always has many irons on the fire; he likes to experiment. That's unlike the rest of science publishers who are quite conservative and work on standardizing, consolidating, and reducing costs," says Matthew Cockerill, managing director of BioMed Central, which Tracz sold in 2008. By contrast, he says, "Vitek doesn't believe in business plans, but in ideas."

Now, the revolutionary, who calls himself "shy" and "un-neat," is stirring up what could become one of the biggest controversies yet in scientific publishing. Tracz is setting out to shake the very foundations of contemporary science by abolishing anonymous peer review.

Michelin Guide of science

Tracz was born in 1940 in a Polish village then occupied by the Soviet Union, and soon afterward his family joined relatives in Siberia, where his father worked in a mine. After the war they made it back to Poland, where Tracz, as an undergraduate at the University of Warsaw, tried his hand at architecture for a year and then switched to mathematics. Before he completed his degree, Tracz's family emigrated to Israel, where he continued his math studies. A year later, he moved to London and studied cinematography at the Slade School of Art. He put down roots and launched Medi-Cine, a company that made educational films for medical doctors. His enthusiasm for filmmaking soon waned, however. Tracz sold Medi-Cine and started up Gower Medical Publishing, which printed full-color medical atlases (at a time when most textbooks were in black and white) and assembled slide collections for lecturers.

Tracz grew bored of textbooks, too. In the early

1980s novel- tions i medic "The like w You n cially later s vicr, a of Ox Tr



profit open-access publisher. The pioneering site, founded in 2000 in London, has grown into an empire with more than 250 biology and medicine journals in its stable.

BioMed Central earned Tracz a reputation as a visionary. "He's one of the most important publishers of the last decade," says Michael Eisen, a biologist at the University of California, Berkeley, and co-founder of the Public Library of Science (PLOS), a nonprofit open-access publisher that launched its first journal in 2003.

net in the 1990s coul he launched BioMe that included a libra HMS Beagle, name to South America. biologists, and doc Two years later, at th the site for an undis closed the site in 20

downloaded from www.sciencemag.org on October 4, 2013

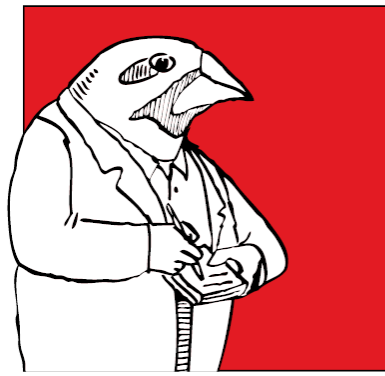
The Seer of Science Publishing

Science 4 October 2013:

Vol. 342 no. 6154 pp. 66-67

DOI: 10.1126/science.342.6154.66

<http://www.sciencemag.org/content/342/6154/66.full.pdf>



WHAT IS *F1000RESEARCH*?

F1000Research is an open access journal for life scientists that accepts all scientifically sound articles, ranging from single findings, case reports, protocols, replications, and null or negative results to more traditional articles.

Key features:

- Publication within a week
- Transparent, post-publication peer review
- All data included
- Accepts non-traditional article types



PEER REVIEW

FACULTY*of*1000

F1000Prime

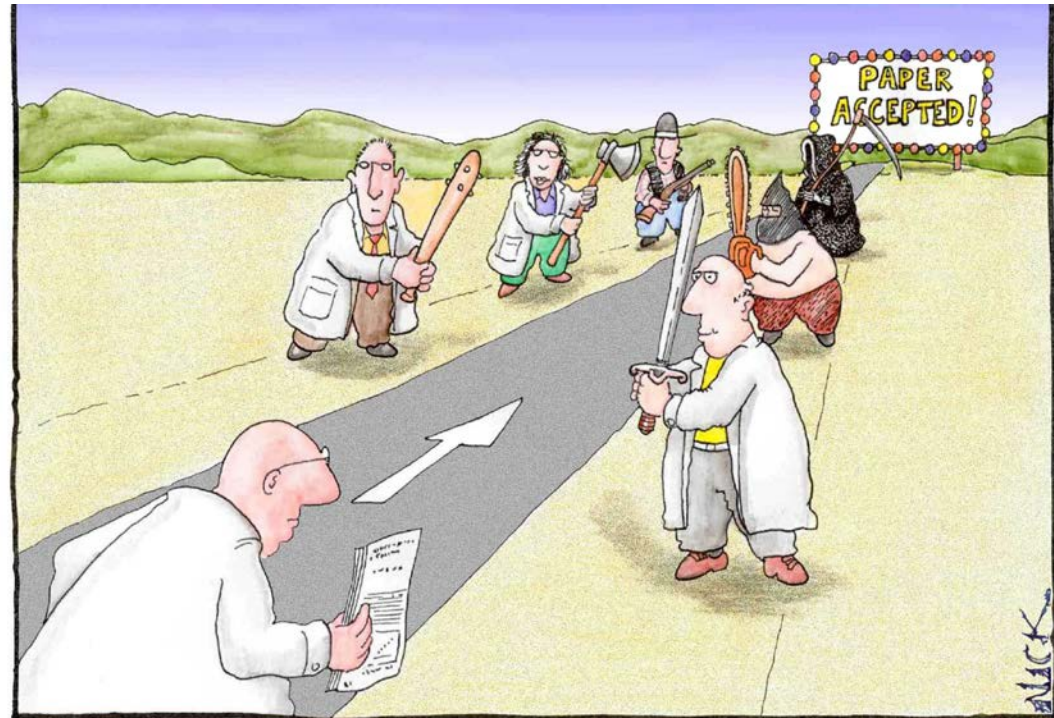
F1000**Research**

F1000Posters



ISSUES WITH TRADITIONAL PEER REVIEW SYSTEM

- Lack of transparency
 - Who are the reviewers?
 - What happened with this paper before it was accepted?
- Lack of accountability
 - Anonymous reviews
 - Editorial decisions may not reflect reviews
- Inefficiency
 - Re-reviewing the same work at different journals
- Delays
 - incidental (reviewing takes time)
 - deliberate (reviewers delaying competitor papers)



Most scientists regarded the new streamlined peer-review process as "quite an improvement."

Cartoon by Nick D Kim, strange-matter.net



A POST-PUBLICATION APPROACH TO PEER REVIEW

F1000Research

Submission to publication: **DAYS**



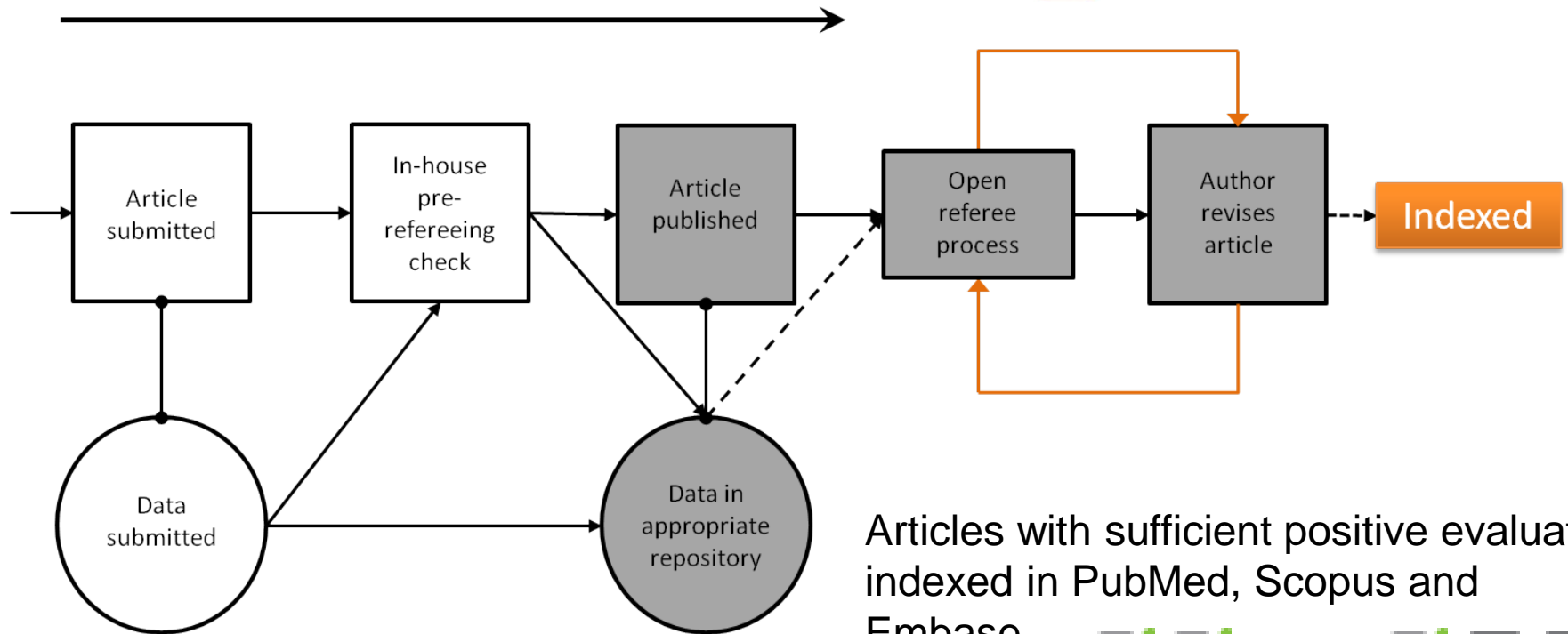
Approved



Approved with reservations



Not approved



Articles with sufficient positive evaluations indexed in PubMed, Scopus and Embase.



or



DATA PUBLISHING AT F1000RESEARCH

FACULTY*of*1000

F1000Prime

F1000**Research**

F1000Posters



DATA POLICY

- Inclusion of all data underlying the conclusions in all articles.
- ‘Openly accessible’ – apply the principles of the Budapest Open Access Initiative* (originally created for scholarly articles) to scholarly data
 - Free to view/access
 - Free to download
 - Free to re-analyse
 - Free to modify
- Community norms to be applied regarding acknowledgement and citation of data.
- Allow publication of data as a ‘Data Note’ article

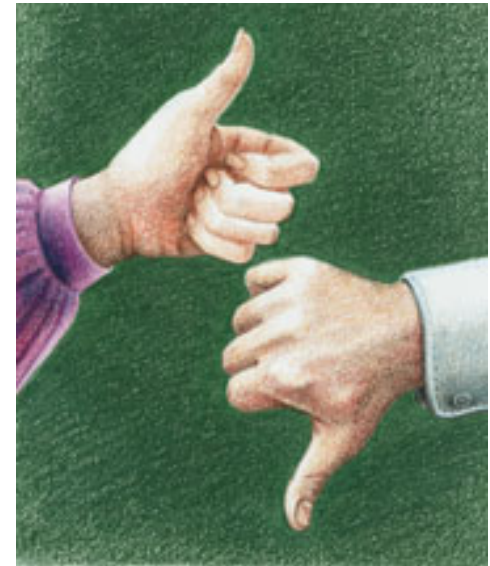
DATA PEER REVIEW

Internal pre-publication checks:

- Storage (discipline-specific repository where possible)
- Format
- Layout and labelling
- Adequate data?
- Adequate protocol information?

Referees are asked to check:

- Methods were appropriate?
- Format/structure usable?
- Data limitations and sources of error included?
- Adequate information to enable potential replication?
- Does the data 'look' OK?




DATA VIEWER







RESEARCH ARTICLE

Low budget analysis of Direct-To-Consumer genomic testing familial data [v1; ref status: indexed, <http://f1000r.es/NmzOoW>]

Gustavo Glusman¹, Mike Carias², Rafael Jimenez³, Daniel Swan⁴, Bastian Greshake⁵, Jong Bhak⁶, Darren W Logan⁷, Manuel Corpas⁸

Son exome files	1052 views	4 shares	25 downloads
Showing 1/7: Son's Aligned Bam File.bam			
1	FCB021RACXX:4:1208:7911:79502#CAGATCAT 147 1 11941 0 90M - 11883 -147 CTTCCCGTGTCCCTTTCCAA		
2	FCD044UACXX:4:2205:3896:171755#CAGATCAT 99 1 12059 0 90M - 12212 242 ACTGGAGTGGAGTTTTCC'		
3	FCB021RACXX:4:1205:8439:53145#CAGATCAT 99 1 12154 0 90M - 12167 102 ACCACAACCAGGCATAGGGG		
4	FCD044UACXX:4:1205:1748:199749#CAGATCAT 163 1 12165 0 90M - 12203 127 GCATAGGGGAAAGATTGG		
5	FCD044UACXX:4:2103:5744:184901#CAGATCAT 99 1 12167 0 90M - 12274 196 GTAGGGGAAAGATTGGAGG		
6	FCB021RACXX:4:1205:8439:53145#CAGATCAT 147 1 12167 0 90M - 12154 -102 ATAGGGGAAAGATTGGAGG		
7	FCD044UACXX:4:1205:1748:199749#CAGATCAT 83 1 12203 0 90M - 12165 -127 TCAACTTCTCTACAACC		
8	FCD044UACXX:4:2205:3896:171755#CAGATCAT 147 1 12212 0 90M - 12059 -242 CTCACAACCTAGGCCAG		
9	FCD044UACXX:4:2103:5744:184901#CAGATCAT 147 1 12274 0 90M - 12167 -196 CCCTCGCTCCAGCAGC'		
10	FCB021RACXX:4:1107:6633:165696#CAGATCAT 99 1 12275 0 90M - 12313 127 CCTCGCTCCAGCAGCTGG		

To see the rest of the document click on the  icon

figshare 1 / 7     Share  Cite  Download

The Fastq files represent the raw exome data for the son. The BAM files are derived from the fastq files by aligning the reads using a Burrows-Wheeler Aligner (BWA). The BAM file (.bam) is the binary version of a tab-delimited text file that contains sequence alignment data. The BAM file index (.bai) provides fast random access to the BAM file. The compressed VCF file (.vcf.gz) describes variant calls of the data in text format.

- Preview large datasets prior to downloading
- View data without leaving the article
- Usage statistics provided
- Legends and DOIs for data

DATA CITATION AND DISCOVERABILITY

Strasser C, Kunze J, Abrams S, Cruse P (2014) **DataUp: A tool to help researchers describe and share tabular data [v1; ref status: approved with reservations 1, <http://f1000r.es/2n7>]**
F1000Research 2014, 3:6

Data and software availability

Data

Figshare: DataUp manuscript data, doi: [10.6084/m9.figshare.884625](https://doi.org/10.6084/m9.figshare.884625)²⁴.

Software

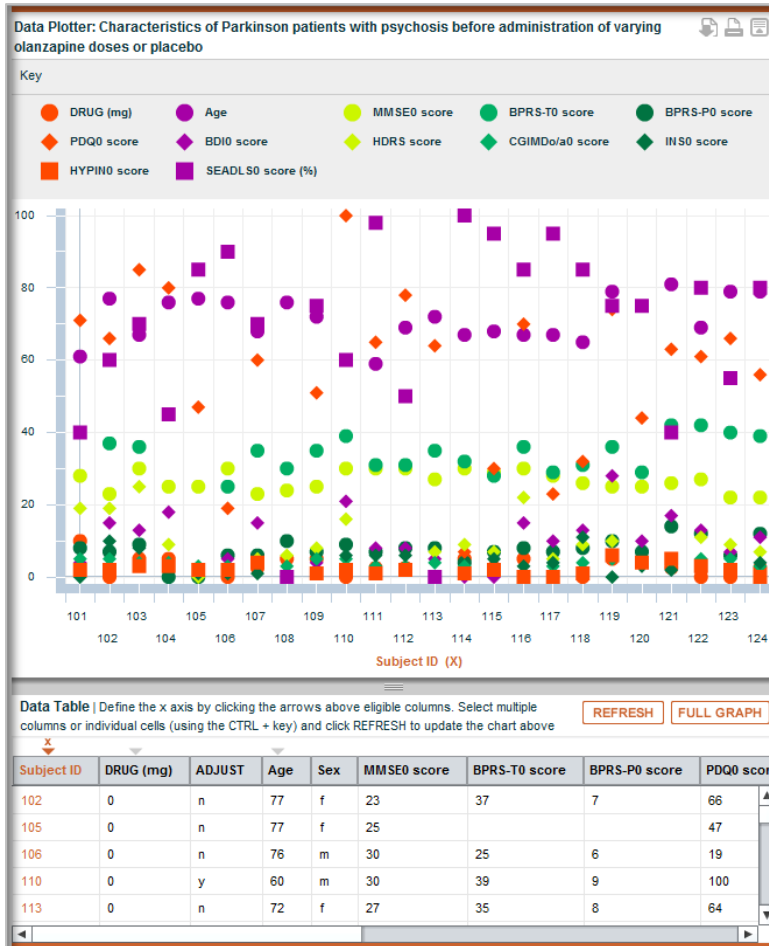
Zenodo: The DataUp source code package, doi: [10.5281/zenodo.7639](https://doi.org/10.5281/zenodo.7639)²⁵.

Bitbucket: Source code for the DataUp Excel add-in and web application, <https://bitbucket.org/dataup/>.

References

22. Abrams S, Cruse P, Kunze J, *et al.*: Curation micro-services: A pipeline metaphor for repositories. *J Digit Imaging*. 2011; **12**(2). [Publisher Full Text](#)
23. DataUp: Further Development and Community Building. *eScholarship*. 2013. [Reference Source](#)
24. Strasser C, Cruse P, Kunze J, *et al.*: DataUp manuscript data. *Figshare*. 2014. [Data Source](#)
25. Strasser C, Cruse P, Kunze J, *et al.*: The DataUp source code package. ZENODO 2014. [Data Source](#)

IN-ARTICLE DATA MANIPULATION



RESEARCH ARTICLE

A fixed-dose randomized controlled trial of olanzapine for psychosis in Parkinson disease [v1; ref status: indexed, <http://f1000r.es/1au>]

Michelle J Nichols^{1,2}, Johanna M Hartlein^{1,3}, Meredith GA Eicken^{4,5}, Brad A Racette³, Kevin J Black^{1,3,6,7}

DATA AS A FIRST CLASS RESEARCH OBJECT

FACULTY *of* 1000

F1000Prime

F1000**Research**

F1000Posters

FORMAL CREDIT FOR DATA SHARING

Feeding into currently recognized scholarly outputs

4.5 Data Notes

Data Notes in *F1000Research* are brief descriptions of scientific datasets that include details of why and how the data were created; they do not include any new analyses or results.

Benefits:

- Appropriate credit for data producers with a citable publication
- Data accessible from repository
- Data independently discoverable via bi-directional linking
- Data available in usable form
- Potential increase in ‘value’ of data, as increasing numbers of studies are carried out

INSTITUTIONAL DATA REPOSITORIES

Data must be hosted by a stable and recognised open repository. Using such a repository ensures that your dataset continues to be available in a useable form in the future.

4.1 Non-exhaustive list of *F1000Research*-approved repositories

Below is a list of repositories that have already been approved for hosting data alongside an *F1000Research* article. If you are an author who wishes to use a repository not already on this list, or you manage a repository that you would like included on the list, please [contact us](#).

Data Type	Approved Repository
Nucleotide or Protein sequence data; Microarray, GWAS or sequencing data; Biologically relevant chemical substances; Sequence-based reagents	Any appropriate NCBI - or EBI -based repository. Note: please submit your data directly and use the unique accession ID for your data in your manuscript.

<http://f1000research.com/data-preparation>

Repository Accreditation based on guidance from PREPARDE
<http://proj.badc.rl.ac.uk/preparde>

A PROPOSAL

Aim – Encourage bioscience researchers to share data and gain scholarly credit for doing so

Suggestion:

- Share metadata between institutional data repository and journal
- Encourage publication of orphan data in institutional repositories

Benefits:

- Researcher only needs to input metadata once
- Researcher gains scholarly credit via formal publication of data paper
- Increases data discoverability and drives traffic back to the repository
- Institution able to demonstrate broader range of output to funding agencies

CHALLENGES AND POTENTIAL SOLUTIONS

- Limited uptake of institutional data repositories by bioscience researchers
 - ✓ Provide option of gaining a formally recognized data publication
- Concept of sharing data as part of publication is novel
 - ✓ Increasingly mandated by funding agencies and journals, so becoming more accepted
- Data-specific metadata standards will be required
 - ✓ Researchers starting to specify standards, e.g. BioSharing.org
- Should be compatible across publishers
 - ✓ *F1000Research* involved in multiple cross-publisher initiatives
- Who could fund development?
 - ✓ Potential interest from funding agency
- Require institutional bioscience data repository to work with us
 - ✓ ?



Email: varsha.khodiya@f1000.com

Twitter: [@vkf1000](https://twitter.com/vkf1000) / [@f1000research](https://twitter.com/f1000research)

Leave a comment on our blog: <http://blog.f1000research.com/>

F1000Research

FACULTYof1000

F1000Prime

F1000Research

F1000Posters