

# Workshop on Open Citations 2018

University of Bologna  
Bologna, Italy  
3 September 2018

## Introduction

David Shotton

[david.shotton@opencitations.net](mailto:david.shotton@opencitations.net)



Oxford e-Research Centre  
University of Oxford, UK



# Welcome !

---

- Two days of formal presentations and one hack day for experimentation
- Those participating: 60 people from 11 countries,  
of whom 24 are speakers and 7 are presenting posters
- Organizers and host



Europe PMC



DIPARTIMENTO DI FILOLOGIA  
CLASSICA E ITALIANISTICA

- Sponsors



- Local organizing committee: Silvio Peroni and Marilena Daquino

# The importance of citations

798 M. A. Jobling *Review. Genetic impact of recent events*

people get upset. When our project was announced through a press release, comments appeared on the website of the British National Party, a far-right British political organization that opposes immigration and favours ‘voluntary repatriation’. One contributor described our project as: ‘Another government-funded (sic) justification for immigration which will say there is no such thing as indigenous British’, and our research group as ‘a bunch of Marxist parasites who have never done a proper day’s work’. Well, proper or not, there is much work to do, and we Marxist parasites have high hopes that it will produce new insights.

My work is supported by a Wellcome Trust Senior Fellowship in Basic Biomedical Science (grant no. 087576).

## REFERENCES

- 1 1000 Genomes Project Consortium, Durbin, R. M. *et al.* 2010 A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. (doi:10.1038/nature09534)
- 2 Barbujani, G. & Colonna, V. 2010 Human genome diversity: frequently asked questions. *Trends Genet.* 26, 285–295. (doi:10.1016/j.tig.2010.04.002)
- 3 Mardis, E. R. 2008 Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9, 387–402. (doi:10.1146/annurev.genom.9.081307.164359)
- era. *Annu. Rev. Genomics Hum. Genet.* 12, 245–274. (doi:10.1146/annurev-genom-090810-183123)
- 14 Underhill, P. A. & Kivisild, T. 2007 Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet.* 41, 539–564. (doi:10.1146/annurev.genet.41.110306.130407)
- 15 Ragoussis, J. 2009 Genotyping technologies for genetic research. *Annu. Rev. Genomics Hum. Genet.* 10, 117–133. (doi:10.1146/annurev-genom-082908-150116)
- 16 Pritchard, J. K., Stephens, M. & Donnelly, P. 2000 Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- 17 Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A. & Feldman, M. W. 2002 Genetic structure of human populations. *Science* 298, 2381–2385. (doi:10.1126/science.1078311)
- 18 Li, J. Z. *et al.* 2008 Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100–1104. (doi:10.1126/science.1153717)
- 19 Novembre, J. *et al.* 2008 Genes mirror geography within Europe. *Nature* 456, 98–101. (doi:10.1038/nature07331)
- 20 Lao, O. *et al.* 2008 Correlation between genetic and geographic structure in Europe. *Curr. Biol.* 18, 1241–1248. (doi:10.1016/j.cub.2008.07.049)
- 21 Pool, J. E. & Nielsen, R. 2009 Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* 181, 711–719. (doi:10.1534/genetics.108.098095)
- 22 Moorjani, P. *et al.* 2011 The history of African gene flow into Southern Europeans, Levantines, and Jews.

# Creating citations is an essential scholarly activity

---

- A citation is created by an author's **performative act of citing** a published work that is relevant to the current work, typically made by including a **bibliographic reference** in the reference list of the current work
- While the act of creation may for the author be the work of a moment, once created, a citation becomes a **permanent directional conceptual link** from the citing bibliographic work to a cited work

# Creating citations is an essential scholarly activity

---

- A citation is created by an author's performative act of citing a published work that is relevant to the current work, typically made by including a bibliographic reference in the reference list of the current work
- While the act of creation may for the author be the work of a moment, once created, a citation becomes a permanent directional conceptual link from the citing bibliographic work to a cited work
- It permits an author to **give credit** to another person's endeavours that have played a part in the development of the author's own ideas or results
- Direct citation is **a key indicator** of a publication's significance

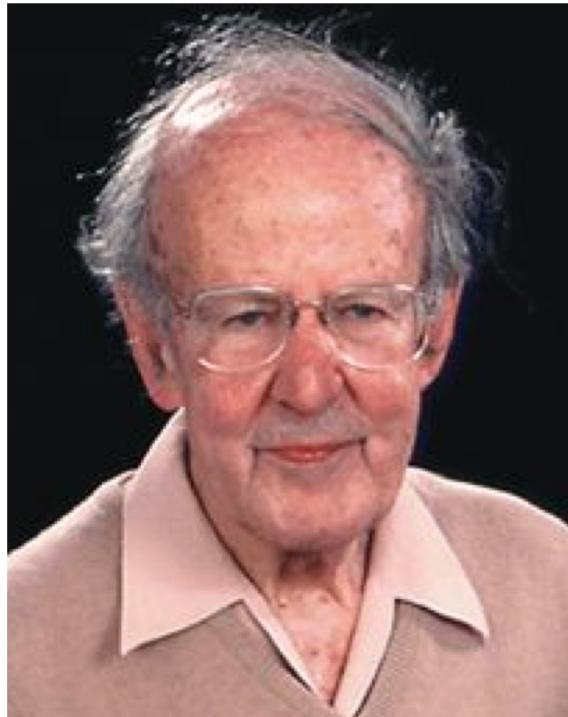
# Creating citations is an essential scholarly activity

---

- A citation is created by an author's performative act of citing a published work that is relevant to the current work, typically made by including a bibliographic reference in the reference list of the current work
- While the act of creation may for the author be the work of a moment, once created, a citation becomes a permanent directional conceptual link from the citing bibliographic work to a cited work
- It permits an author to give credit to another person's endeavours that have played a part in the development of the author's own ideas or results
- Direct citation is a key indicator of a publication's significance
- Citations also **integrate** our independent acts of scholarship into a global knowledge network
- **Bibliometric analysis** of the flow of information and ideas through the citation network, and its changes over time, can reveal patterns of communication between scholars and the development and demise of academic disciplines

## Two of my heroes

---



Max Perutz



John Sulston

# The four phases of Open Scholarship

---

- Open source **software**, now ubiquitous – think Linux, Open Office, Firefox
- Open access **papers**
  - Many funders, including NIH and NSF, the European Commission, the Wellcome Trust, and the Gates Foundation, now make Open Access mandatory for grantees
- Open **datasets** on which scholarly articles are based
  - Much current effort to encourage deposit of research datasets
  - Historically, bioinformatics researchers have always been open, while others, notably chemists, have always been tight-fisted with their data



- **Open metadata** - machine-readable descriptions of scholarly works
  - **Metadata 2020** is a collaboration launched last autumn to promote open reusable metadata describing all research outputs, and several of its members are present at this meeting
  - At this workshop, we will concentrate on **bibliographic and citation metadata**

# Requirements for Open Scholarship

---

- FAIR principles (<https://www.force11.org/group/fairgroup/fairprinciples>)
  - To facilitate humans and machines in their discovery of, access to, integration and analysis of scholarly material, stuff should be
    - Findable
    - Accessible
    - Interoperable
    - Re-usable

# Requirements for Open Scholarship

---

- FAIR principles (<https://www.force11.org/group/fairgroup/fairprinciples>)
  - To facilitate humans and machines in their discovery of, access to, integration and analysis of scholarly material, stuff should be
    - Findable
    - Accessible
    - Interoperable
    - Re-usable
- The Initiative for Open Citations goals (<https://i4oc.org/#goals>)
  - To facilitate reuse, stuff (citations) should be
    - Structured (accessible in common, machine-readable formats)
    - Separable (metadata available without access to source publications)
    - Open (freely accessible and re-usable)
- The problem is that at present most stuff is not free to access, is subject to inconsistent, hard-to-parse licenses, and is not machine-readable

## For citations, how is the present situation imperfect?

---

- The present scholarly citation system inadequately exposes the knowledge networks that exist within the scholarly literature, linking papers, authors, funders, research projects and datasets



- Citation data are **hidden behind subscription firewalls** of commercial companies
- Academics are **not free** to use their own citation data as they please

## For citations, how is the present situation imperfect?

---

- The present scholarly citation system inadequately exposes the knowledge networks that exist within the scholarly literature, linking papers, authors, funders, research projects and datasets



- Citation data are hidden behind subscription firewalls of commercial companies
- Academics are not free to use their own citation data as they please
- In this Open Access age, it is a **scandal** that reference lists from journal articles, the core elements of the academic data cycle, are not (yet) freely available for use by the scholars who created them
- Citation data now need to be recognized as a part of the Commons – those works that are freely and legally available for sharing

## Working towards open citations

---

- We are **late** entering the fight for open citations, which has been dominated by commercial citation indexes charging very high subscriptions: WoS and Scopus
- While other indexes - Google Scholar and Microsoft Academic – are free to search, all these commercial indexes impose severe restrictions on machine access and data download for reuse

- We are **late** entering the fight for open citations, which has been dominated by commercial citation indexes charging very high subscriptions: WoS and Scopus
- While other indexes - Google Scholar and Microsoft Academic – are free to search, all the commercial indexes impose severe restrictions on machine access and data download for reuse
- **The Initiative for Open Citations** (<https://i4oc.org>), launched on 6<sup>th</sup> April 2017, has had remarkable success in persuading almost all the major scholarly publishers to **open** the reference lists they submit to **Crossref**
  - Crossref now hosts more than **half a billion** open journal article references!
- I have recently called for research funders to mandate that their grant-holders publish in journals which make their citations open
  - David Shotton (2018). Funders should mandate open citations. *Nature* **553**: 129. <http://dx.doi.org/10.1038/d41586-018-00104-7>

- We are **late** entering the fight for open citations, which has been dominated by commercial citation indexes charging very high subscriptions: WoS and Scopus
- While other indexes - Google Scholar and Microsoft Academic – are free to search, all the commercial indexes impose severe restrictions on machine access and data download for reuse
- The Initiative for Open Citations (<https://i4oc.org>), launched on April 6 2017, has persuaded almost all the major scholarly publishers to open the reference lists they submit to Crossref
  - Crossref now hosts more than half a billion open journal article references!
- I have recently called for research funders to mandate that their grant-holders publish in journals which make their citations open
  - David Shotton (2018). Funders should mandate open citations. *Nature* 553: 129. <http://dx.doi.org/10.1038/d41586-018-00104-7>
- This is the first **Workshop on Open Citations**, and we are here to work . . .
  - Enjoy!