Publish or Perish:
Realising Google Scholar’s potential to democratise citation analysis

Professor Anne-Wil Harzing
Middlesex University

Dr. David Adams
Tarma Software Research
Presentation outline

1. Brief historical overview of “citizen bibliometrics”, i.e. use of bibliometrics by non-experts
2. How Publish or Perish and Google Scholar have democratised citation analysis
3. Publish or Perish users: who are they and how do they use PoP?
4. Publish or Perish version 5: key new features
5. What’s next for citizen bibliometrics?
1. A potted history of citizen bibliometrics

[with no pretence at completeness]
In the dark old days [before 2004], we had...

- As a data source
  - The Thomson Reuters Web of Science
  - The Thomson Reuters Web of Science
  - The Thomson Reuters Web of Science

- As metrics
  - The Thomson Reuters Journal Impact Factor
  - The Thomson Reuters Journal Impact Factor
  - The Thomson Reuters Journal Impact Factor

- Academia in the grip of a commercial monopoly
So why was this so bad?

- Data source: Thomson Reuters grew from a Western (Life) Sciences based paradigm and did not make much effort to adapt to the changing academic landscape

- Metrics: Journal Impact Factor is designed for journals not individual articles or academics
  - The average # cites/paper in a journal says very little about an individual article in that journal [promise over proof]
  - JIF 2-year timeframe is much too short for most disciplines outside the Life Sciences
  - JIF is easy to manipulate by unscrupulous editors, manipulation became increasingly common
2004: The start of the age of enlightenment?

Enter... Scopus by Elsevier
- Better coverage in Engineering, Social Science & Humanities than WoS
- Much more modern technology base (e.g. allows search for non-Western scripts, better disambiguation)
- More appropriate journal metric indicators (SNIP, SJR)
- BUT: (like WoS) huge subscription fees [and provided by “big bad” Elsevier]

Enter... Google Scholar
- Much broader coverage in all disciplines
- Free: accessible to anyone with internet access
- BUT: interface rather cumbersome to use
- AND: doesn’t calculate any metrics beyond citations
2005-2011: I can see more clearly now?

- 2005: Jorge Hirsch introduces the h-index, which takes the academic world by storm
- 2006: Publish or Perish 1.0 is released
  - Initially used mainly as a tool to calculate an academic's h-index based on Google Scholar data
- 2007: Publish or Perish 2.0 is released
  - Many additional features expanding its use cases
- 2007: Anne-Wil starts her systematic research on Google Scholar (Harzing, 2007a/b; Harzing & van der Wal, 2008, 2009)
  - 2009 paper compares Web of Science JIF with GS h-5 nearly 5 years before the metric was even introduced by Google Scholar 😊
- 2008: EC3 group in Granada starts studying Google Scholar
- 2010: Publish or Perish 3.0 with multi-query center is released
- 2010: Publish or Perish book is published
  - 2011: Publish or Perish book re-published in 3 parts
2012-2016: Free bibliometrics is becoming mainstream

- 2012: Google Scholar introduces GS Profiles with h-index and h5 index
- 2013: Publish or Perish v4 is released with streamlined interface and Microsoft Academic Search (MAS version 1) support
- 2013-current Anne-Wil continues research on:
  - Web of Science (Harzing, 2013a/2015): Problematic categorization & disambiguation
  - Google Scholar (Harzing 2013b/2014): Coverage is improving for all disciplines
  - Alternative metrics (Harzing, Alakangas & Adams, 2014): hla: individual, annual h-index
  - Google Scholar vs WoS/Scopus (Harzing & Mijnhardt, 2015; Harzing & Alakangas, 2016): GS performs better than WoS/Scopus in all disciplines
  - Microsoft Academic (Harzing, 2016, 2017): Serious alternative to GS, WoS and Scopus
- 2014: Google Scholar digest blog started by EC3 Research Group
- 2016: Publish or Perish v5 is released (200th or so public release)
  - New: GS Profile Search and Microsoft Academic (version 2) Search
  - New: Exporting full reference formats + heuristic classification of publication types
  - Publish or Perish tutorial published, helpfile completely overhauled and extended
2. How PoP & GS have democratised citation analysis

[Some examples and illustrations]
GS democratises access to research materials

- Anurag Acharya at Association of Learned and Professional Society Publishers Sept 2015

- Articles/publications stand on their own rather than be part of a journal; this benefits:
  - Older articles
  - Regional articles
  - Publications in non-elite journals
  - Non journal publications such as dissertations, book chapters

- Spread of attention outside elite group of journals
  - Good ideas can come from anywhere
  - Insight is not limited to the well-funded ...
  - ... or to the well-published
PoP & GS democratise access to citation data (1)

- **Wide user base:** PoP used by academics, librarians, governments, grant agencies, and research laboratories
  - Approximately half a million individual academic users
  - Thousands of libraries worldwide list the software as a free alternative to Scopus and the Web of Science
  - Government [departments]: e.g. US EPA, US Agency for International Development, Colciencias (Columbia), Poland, France
  - Grant giving agencies: e.g. SSHRC in Canada, CNRS in France
  - Research laboratories: e.g. Microsoft, Hewlett Packard, IBM

- **Wide geographical base:** PoP used in more than 100 countries
  - Prestigious Western universities: e.g. Harvard, Stanford, MIT, Oxford, Cambridge, INSEAD
  - Very popular in Italy, Poland, France, Germany, and Greece for its broader coverage and as a tool to expose nepotism in academic appointments
  - But also... used in under-resourced universities in countries such as Armenia, Botswana, Ethiopia, Lesotho, Mongolia, Paraguay, Tajikistan, Ukraine, and Uruguay
PoP & GS: Let’s hear the PoP users speak

- The more I work with GS the more I appreciate the gate those guys opened for us. PoP enhances GS and if I were working for Google, I would consider developing it further to become the equivalent of Google Earth.

- We have benefitted from the use of Harzing’s Publish or Perish software. We have been fans of your work for some time now and have used your tools to inform our own benchmarking here at Harvard Business School.

- A close colleague of mine told me an amusing anecdote last month about a social sciences meeting for upgrading Oxford University academics to professor, at which 17 department heads presented cases for members of their staff - all but one of whom used Harzing (PoP) statistics.

- I’ve referred people to your software more times than I can count, it’s recognized as the single most effective tool for calculating personal H-indices. I do not doubt that PoP is in large part responsible for the broadening interest in bibliometrics. Giving academics a tool by which to compare themselves to their peers has had a tremendous impact globally.

- As I am sure you realize, being able to demonstrate the influence of research by the Worldbank is enormously important to supporting and expanding that research. [...] I have found your software Publish or Perish to be the single most useful tool available for these purposes.

- It was a great pleasure to meet someone who has contributed so much, completely free of charge, to the development of the social sciences. Academic altruists are rare indeed, and your PoP programme is a huge advance.
PoP & GS democratise access to citation data (2)

- **Wide language base: PoP used to cover non-English language publications**
  - Web of Science and Scopus have VERY limited coverage of LOTE publications
  - Web of Science doesn’t allow searches in non-Western scripts

- **Wide disciplinary base: PoP used for bibliometric research in the Social Sciences and Humanities**
  - Thousands of published papers using Publish or Perish to analyze Google Scholar data
  - Bibliometrics can now be used in these disciplines with appropriate metrics and databases
PoP & GS make non-Anglo scholars visible

PoP & GS make Engineering & SSH disciplines visible

See also: http://www.harzing.com/blog/2016/09/citation-analysis-for-the-social-sciences-metrics-and-datasources
PoP & GS reduce bias in academic rankings


- New top-40 using GS/PoP individual annual h-index, rather than Thomson Reuters JIF
  - Reduction of traditional dominance of: Economics over Business, older over younger academic, male over female academics, and academics from prestigious institutions
  - 7 out of the academics originally in the top-10 drop out of the top-40 completely when using the hIa and Google Scholar

- Resulting ranking is more democratic
  - removal of disciplinary, age, gender, and institutional biases
  - verdict based on entire academic community
  - transparent and easy to replicate for anyone
  - more dynamic with better chances for younger academics
3. Publish or Perish users: who are they and how do they use PoP?

[Some survey data and blog posts]
What disciplinary area do users come from?

- 42.6% Social Sciences (e.g. Business, Economics, Education, Political Science)
- 21.4% Life Sciences (e.g. Medicine, Dentistry, Biology, Agriculture)
- 17.4% Engineering (e.g. Computing & Information Systems, Mechanical Engineering)
- 11.4% Sciences (e.g. Chemistry, Physics, Mathematics, Astronomy)
- 7.2% Arts & Humanities (e.g. History, Languages, Law, Music)
How long have you been using Publish or Perish?
How often do you use Publish or Perish?
What do you use PoP for [multiple answers possible]?
PoP helps with dozens of daily academic tasks

- Making your case for impact
- Looking for John Smith: disambiguate authors in Google Scholar
- Preparing a case for tenure or promotion
- Deciding where to submit your next paper
- Having a meeting with your academic hero?
- Impressing your academic interview panel
- Doing a literature review
- Evaluating your research group/department/school
- And many many more... [PoP: a Swiss army knife]
4. PoP version 5: key new features

[A very brief selection]
PoP Version 5: Clean out your Google Scholar Profile

- Originally bulk import of over 300 publications in my profile
- Getting an overview of the “dross” is hard in the GS Profile interface; you can only see a dozen or so publications on screen at the same time and cannot sort by journal or author
- PoP allows me to easily spot strays [which can then be merged in GSP] and incomplete references [which can then be completed in GSP], resulting in a (now) very clean profile
PoP Version 5: Spotting dirty profiles easily

- Tsinghua University for example has many dirty profiles, but the extent only becomes clear when displaying them in PoP. Here is a junior academic with publications several decades before he was born 😐
- Misappropriation of Geim’s graphene article
- First fifteen publications seem to refer to at least 10 different people in different disciplines
PoP V5: Easily compare your record across 4 data-sources

- PoP allows easy importing of Scopus and WoS data
PoP Version 5: Export full bibliographic details

- Export one, many or all results into any format!
The result? A neat list of references in a few clicks

PoP V5: Intuitive exporting of results or metrics
The result: A neat list of data for further analysis

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Name</td>
<td>Source</td>
<td>Scholars</td>
<td>Data</td>
<td>Analysis</td>
<td>Further</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Thiruvaidyaguru</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Bunsen</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Browne</td>
<td>Researcher</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Thomas</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Stephen</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Phelan</td>
<td>Co-author</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Barkham</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Sandra</td>
<td>AVAUniversity</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Ray</td>
<td>Researcher</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Roger</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Roberto</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Rachel</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Pradeep</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Phil</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Paul</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Nisam</td>
<td>Associate Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Niko</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Nick</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Neville</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Steve</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Michael</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Michael</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Michael</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Mark</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Marilyn</td>
<td>University of Melbourne</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Marie</td>
<td>Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Margaret</td>
<td>Senior Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Malcolm</td>
<td>Associate Professor</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Liz</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Ulisse</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Steve</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Seth</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Sue</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Sarah</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>John</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Jane</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>John</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Jane</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>John</td>
<td>Lecturer</td>
<td>Google Scholar</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td>A neat list</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Want more information?

- PoP tutorial (80 tips: book & online)
  - What the heck are all these metrics?
  - Present your case: Find the pearls in your record
  - Meeting an academic visitor
  - Preparing for a job interview
  - Tips for Deans and other administrators
  - Where to submit your paper?
  - Identifying key authors, journals & publications in a field
  - Bibliometric research with Google Scholar
  - Finding reviewers, examiners, keynote speakers, referees
  - Tracking a forgotten conference contact
  - And many many more...

- Publications on research evaluation

- Many white papers and presentations on PoP & Google Scholar
  - [http://www.harzing.com/publications/white-papers](http://www.harzing.com/publications/white-papers)
5. What is next for citizen bibliometrics?

[Will the landscape change further?]
Commercial providers: increasingly distracted?

  - Sold to Onex Private Equity (Canada) Barings Private Equity Asia (China), establishing independent company Clarivate Analytics
  - Mission is “enabling customers to discover, protect and commercialize new ideas, faster”
  - What does this mean for commitment to providing bibliometric data to academics?

- 1 Jan 2017: German & Taiwanese universities boycott Elsevier journals; Peruvian government stops funding Elsevier jnls
  - Will other countries follow [2016: threats in Netherlands & Finland]?
  - Will the boycott by individual academics (currently signed by nearly 17,000 individuals) be strengthened?
  - How committed is Elsevier to individual users, their emphasis appears to be on Scival institutional use
Will the free alternatives prevail?

- Microsoft Academic (MA v2): Is the Phoenix getting wings?
  - Early signs (Harzing, 2016, Harzing & Alakangas, 2017) show excellent coverage and accuracy
  - For details see: [http://www.harzing.com/blog/2016/06/microsoft-academic-search-a-phoenix-arisen-from-the-ashes](http://www.harzing.com/blog/2016/06/microsoft-academic-search-a-phoenix-arisen-from-the-ashes)
  - MS Research is very responsive to user feedback, is actively addressing problems, as well as developing new features, incl. self-managed academic profiles (I have beta-tested them)

- What is happening to Google Scholar?
  - After the very interesting 10-year anniversary articles in 2014 very little has been added to their blog and to GS functionality
  - Only 2 posts in 2015, 4 in 2016, last post half a year old
  - No significant recent news coverage on Google Scholar apart from a Nature article on the launch of Microsoft Academic “aiming to outdo Google Scholar”

- So hopefully we’ll learn more in the next presentation!
But first?

Any questions or comments?