Uncovering Alternative Metrics:

Data Mining Wikipedia for Evidence of Public Engagement and Impact

Lars Alvik, Peter Neish, Sally Tape

Acknowledgment of country

What we will cover

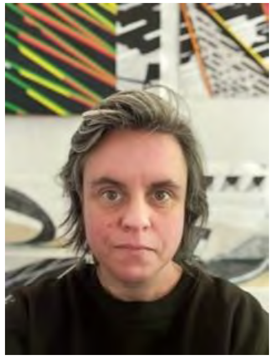
- Background on Wikipedia ecosystem
- References in Wikipedia and disinformation
- Metric and Alternative Metrics
- Data mining Wikipedia
- Where to from here

Introduction, who are we?

Peter Neish

Program Manager
Stewardship and Open Research
University of Melbourne
Wikimedia Australia

Sally Tape

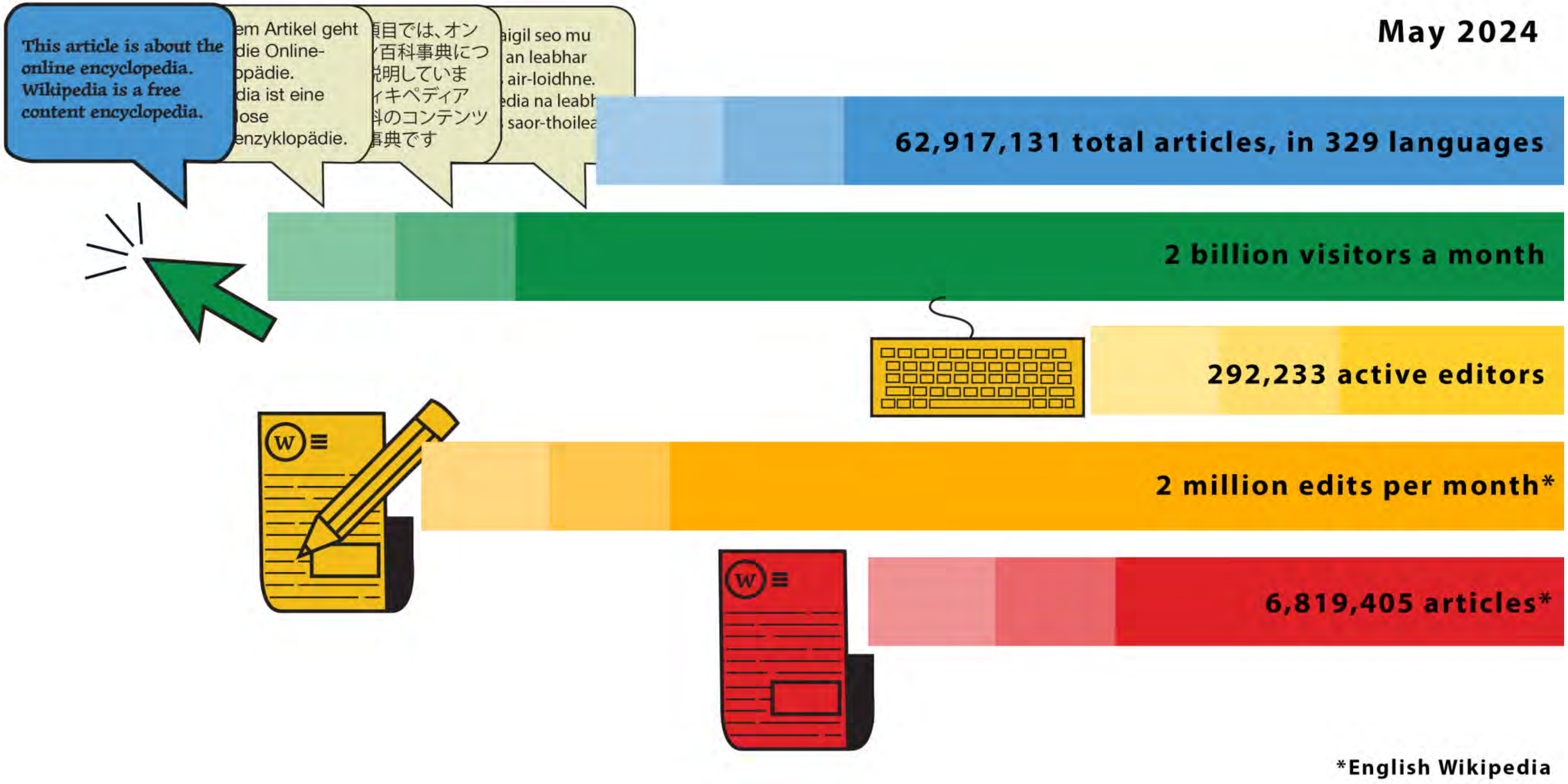


Open Research Support Specialist
University of Melbourne

Lars Alvik

Digital Curation Technical Specialist
University of Melbourne

Background on Wikipedia



https://meta.wikimedia.org/wiki/List_of_Wikipedias

https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

Background on Wikipedia

Highly discoverable

FAIR

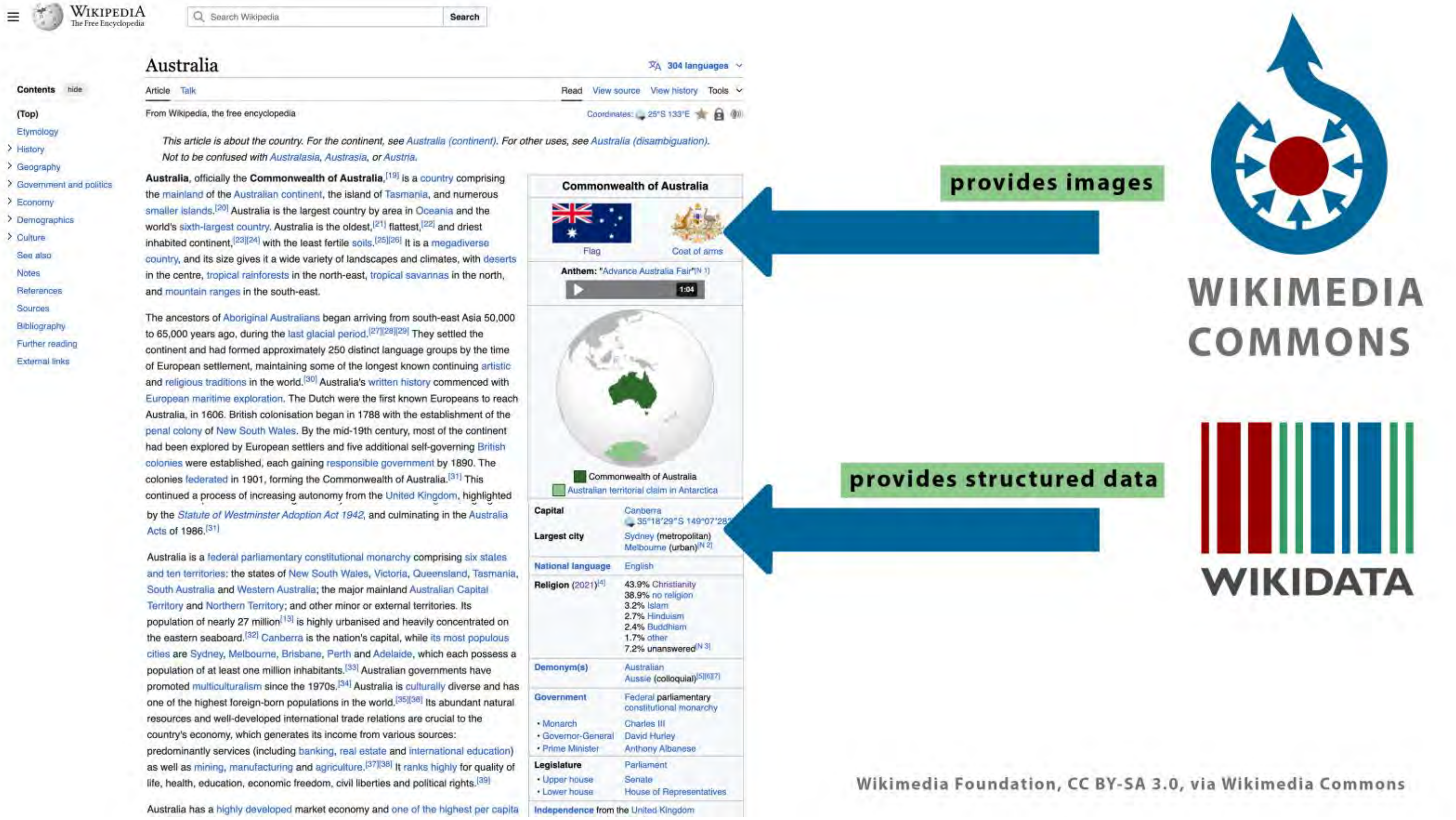
Easy to read
& understand

Impact &
integration

IMAGES: EnWIKI20 Logo Wugapodes, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Background on Wikipedia

- Structured knowledge banks
- Verifiable information
- Integration



Australia

From Wikipedia, the free encyclopedia

Coordinates: 25°S 133°E

This article is about the country. For the continent, see Australia (continent). For other uses, see Australia (disambiguation).
Not to be confused with Australasia, Austrasia, or Austria.

Australia, officially the **Commonwealth of Australia**,[19] is a country comprising the mainland of the Australian continent, the island of Tasmania, and numerous smaller islands.[20] Australia is the largest country by area in Oceania and the world's sixth-largest country. Australia is the oldest,[21] flattest,[22] and driest inhabited continent,[23][24] with the least fertile soils.[25][26] It is a megadiverse country, and its size gives it a wide variety of landscapes and climates, with deserts in the centre, tropical rainforests in the north-east, tropical savannas in the north, and mountain ranges in the south-east.

The ancestors of Aboriginal Australians began arriving from south-east Asia 50,000 to 65,000 years ago, during the last glacial period.[27][28][29] They settled the continent and had formed approximately 250 distinct language groups by the time of European settlement, maintaining some of the longest known continuing artistic and religious traditions in the world.[30] Australia's written history commenced with European maritime exploration. The Dutch were the first known Europeans to reach Australia, in 1606. British colonisation began in 1788 with the establishment of the penal colony of New South Wales. By the mid-19th century, most of the continent had been explored by European settlers and five additional self-governing British colonies were established, each gaining responsible government by 1890. The colonies federated in 1901, forming the Commonwealth of Australia.[31] This continued a process of increasing autonomy from the United Kingdom, highlighted by the *Statute of Westminster Adoption Act 1942*, and culminating in the Australia Acts of 1986.[31]

Australia is a federal parliamentary constitutional monarchy comprising six states and ten territories: the states of New South Wales, Victoria, Queensland, Tasmania, South Australia and Western Australia; the major mainland Australian Capital Territory and Northern Territory; and other minor or external territories. Its population of nearly 27 million[13] is highly urbanised and heavily concentrated on the eastern seaboard.[32] Canberra is the nation's capital, while its most populous cities are Sydney, Melbourne, Brisbane, Perth and Adelaide, which each possess a population of at least one million inhabitants.[33] Australian governments have promoted multiculturalism since the 1970s.[34] Australia is culturally diverse and has one of the highest foreign-born populations in the world.[35][36] Its abundant natural resources and well-developed international trade relations are crucial to the country's economy, which generates its income from various sources: predominantly services (including banking, real estate and international education) as well as mining, manufacturing and agriculture.[37][38] It ranks highly for quality of life, health, education, economic freedom, civil liberties and political rights.[39]

Australia has a highly developed market economy and one of the highest per capita

Commonwealth of Australia

Anthem: "Advance Australia Fair"[N 1]

🟩 Commonwealth of Australia
🟩 Australian territorial claim in Antarctica

Capital	Canberra 35°18′29″S 149°07′28
Largest city	Sydney (metropolitan) Melbourne (urban)[N 2]
National language	English
Religion (2021)[4]	43.9% Christianity 38.9% no religion 3.2% Islam 2.7% Hinduism 2.4% Buddhism 1.7% other 7.2% unanswered[N 3]
Demonym(s)	Australian Aussie (colloquial)[5][6][7]
Government	Federal parliamentary constitutional monarchy
• Monarch	Charles III
• Governor-General	David Hurley
• Prime Minister	Anthony Albanese
Legislature	Parliament
• Upper house	Senate
• Lower house	House of Representatives
Independence from the United Kingdom	

Wikimedia Foundation, CC BY-SA 3.0, via Wikimedia Commons

Wikimedia Foundation, CC BY-SA 3.0, via Wikimedia Commons

The role of references in Wikipedia and disinformation

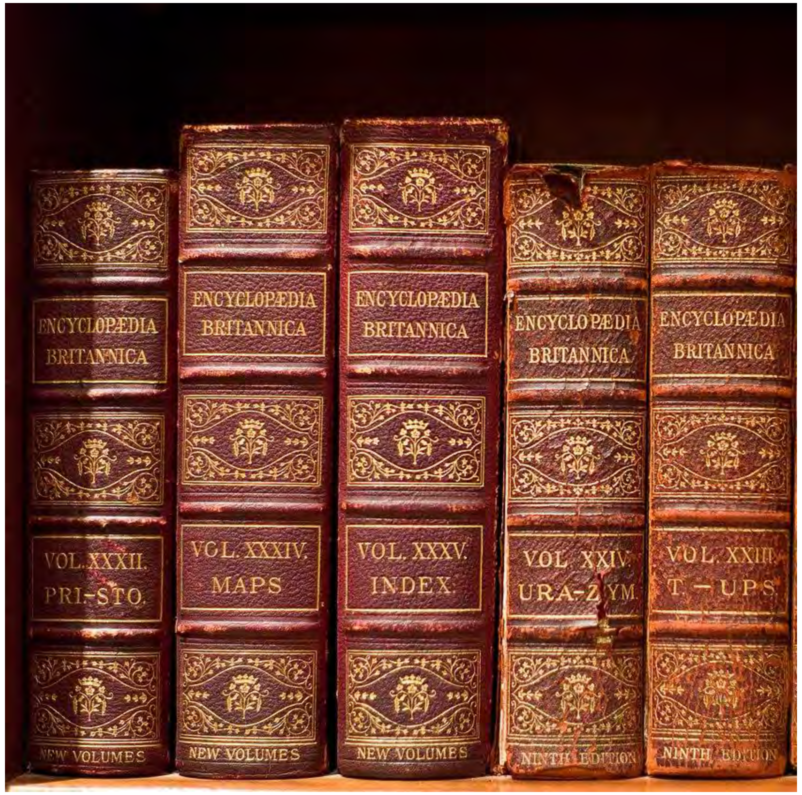
- **Ideally, everything in Wikipedia is verifiable by a reliable published* secondary source eg:**
  - **Scholarly works: monographs, articles, textbooks**
  - **Well-established News organisations**

***made available to the public in some form (could be behind a paywall)**

The fin de siècle newspaper proprietor 1894: Frederick Burr Opper, Public domain, via Wikimedia Commons

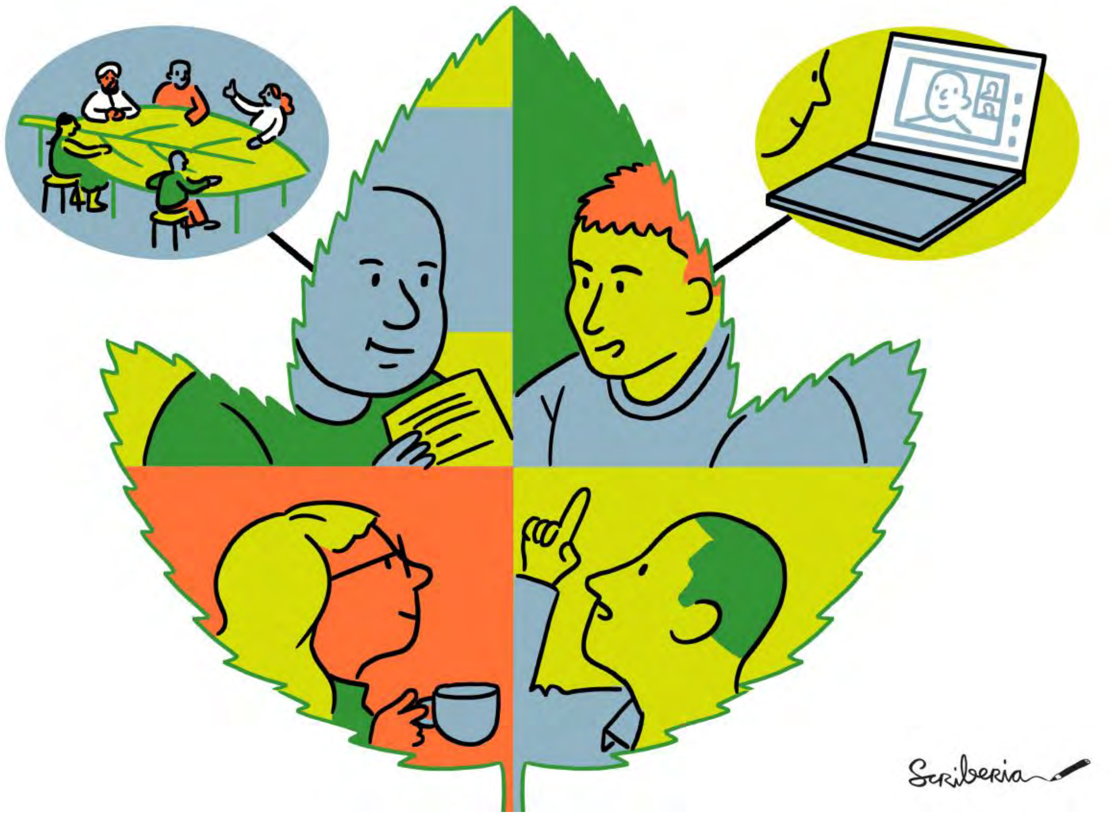
The role of references in Wikipedia and disinformation

- Wikipedia is not a for original research
- Wikipedia is versioned and any vandalism or misinformation can be readily rolled back
- Bots and people can monitor and review content
- Studies continually find that Wikipedia is equal or more reliable than scholarly textbooks and published Encyclopaedias



Joi Ito., CC BY 2.0 via Wikimedia Commons

Why is Wikipedia interesting from a University perspective?

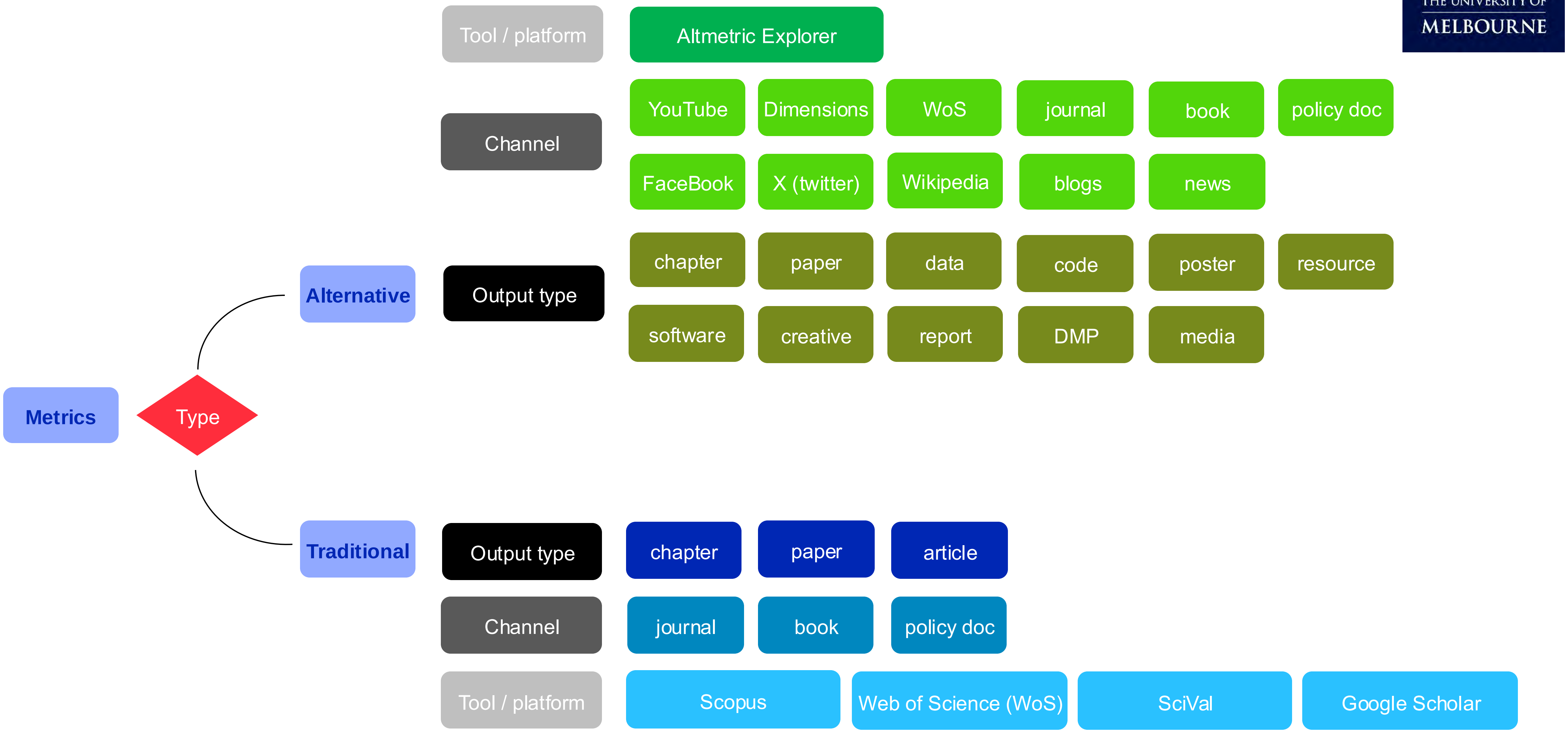


- Impact through wide readership outside of scholarly community
- Contribute to Open Knowledge
- Content related to institution alumni or areas of expertise

The Turing Way Community, & Scriberia. (2024). Illustrations from The Turing Way: Shared under CC-BY 4.0 for reuse. Zenodo. https://doi.org/10.5281/zenodo.10654984

Background on Metrics

THE UNIVERSITY OF
MELBOURNE



Metrics

Type

Alternative

Tool / platform: Altmetric Explorer

Channel: YouTube, Dimensions, WoS, journal, book, policy doc, FaceBook, X (twitter), Wikipedia, blogs, news

Output type: chapter, paper, data, code, poster, resource, software, creative, report, DMP, media

Traditional

Output type: chapter, paper, article

Channel: journal, book, policy doc

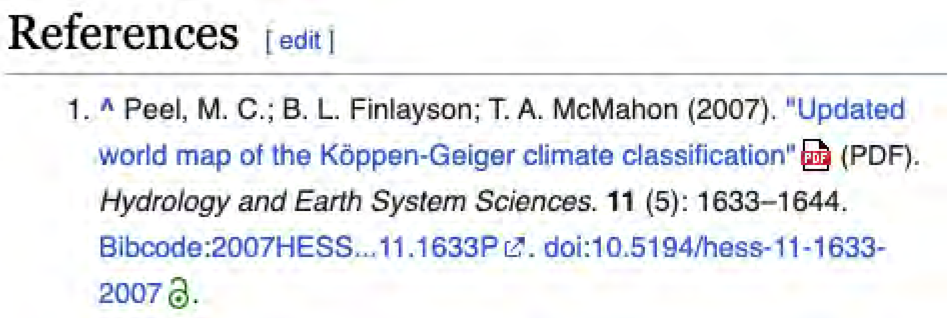
Tool / platform: Scopus, Web of Science (WoS), SciVal, Google Scholar

Background on Metrics

- Metrics from Wikipedia are tracked through 'mentions' on-line

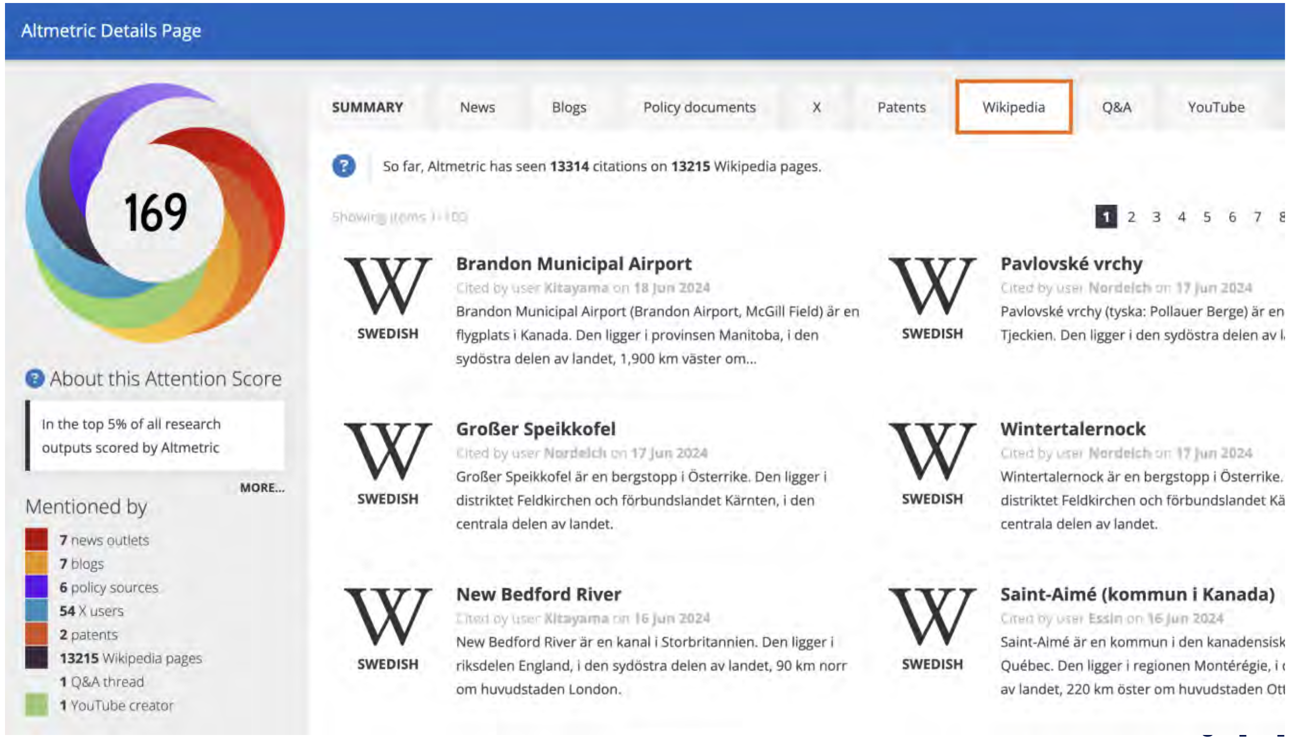
- Wikipedia articles are published in 300+ different languages

- Metrics are currently recorded on the Altmetric Explorer
for 34 Wikipedia languages



References [edit]

1. ^ Peel, M. C.; B. L. Finlayson; T. A. McMahon (2007). "Updated world map of the Köppen-Geiger climate classification" (PDF). *Hydrology and Earth System Sciences*. **11** (5): 1633–1644. Bibcode:2007HESS...11.1633P. doi:10.5194/hess-11-1633-2007.



So far, Altmetric has seen **13314** citations on **13215** Wikipedia pages.

Showing items 1-100

About this Attention Score

In the top 5% of all research outputs scored by Altmetric

MORE...

Mentioned by

7 news outlets
7 blogs
6 policy sources
54 X users
2 patents
13215 Wikipedia pages
1 Q&A thread
1 YouTube creator

SWEDISH
Brandon Municipal Airport
Cited by user **Kitayama** on **18 Jun 2024**
Brandon Municipal Airport (Brandon Airport, McGill Field) är en flygplats i Kanada. Den ligger i provinsen Manitoba, i den sydöstra delen av landet, 1,900 km väster om...

SWEDISH
Pavlovské vrchy
Cited by user **Nordelch** on **17 Jun 2024**
Pavlovské vrchy (tyska: Pollauer Berge) är en Tjeckien. Den ligger i den sydöstra delen av l

SWEDISH
Großer Speikkofel
Cited by user **Nordelch** on **17 Jun 2024**
Großer Speikkofel är en bergstopp i Österrike. Den ligger i distriktet Feldkirchen och förbundslandet Kärnten, i den centrala delen av landet.

SWEDISH
Wintertalernock
Cited by user **Nordelch** on **17 Jun 2024**
Wintertalernock är en bergstopp i Österrike. distriktet Feldkirchen och förbundslandet Kä centrala delen av landet.

SWEDISH
New Bedford River
Cited by user **Kitayama** on **16 Jun 2024**
New Bedford River är en kanal i Storbritannien. Den ligger i riksdelen England, i den sydöstra delen av landet, 90 km norr om huvudstaden London.

SWEDISH
Saint-Aimé (kommun i Kanada)
Cited by user **Essin** on **16 Jun 2024**
Saint-Aimé är en kommun i den kanadensisk Québec. Den ligger i regionen Montérégie, i av landet, 220 km öster om huvudstaden Ott

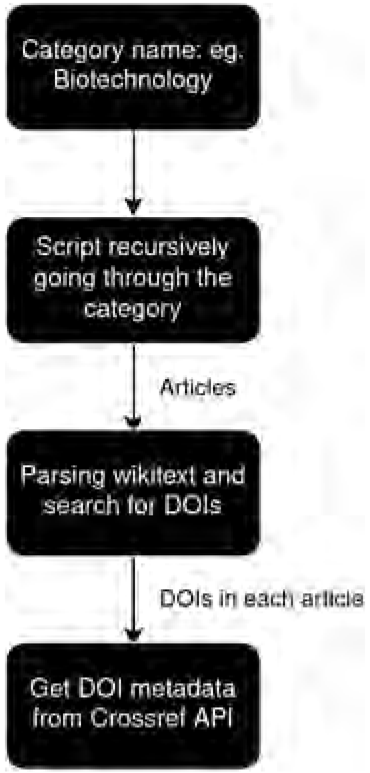
https://www.altmetric.com/explorer/outputs?identifier=10.5194%2Fhess-11-1633-2007&scope=all&show_details=3112735

What we did

- Initially we wanted to look at the whole Wikipedia and use their tools to get all links to doi.org from articles within Wikipedia. (this didn't work)
- The alternative was to download the parsed html of each article and scan through for links to doi.org. This worked, but we had to cut down on the number of articles.
- We narrowed it down to articles within the first 3 category levels of a handful of Australian science categories.

Scripts

1. **Runs through categories and saves page names to a list.**
2. **Gets the parsed text from the Wikipedia API and saves DOIs to a list.**
3. **Checks doi.org if the DOI is a Crossref or DataCite DOI.**
4. **Gets metadata about the DOI from the relevant API.**
5. **Print stats.**



Category name: eg. Biotechnology

Script recursively going through the category

Articles

Parsing wikitext and search for DOIs

DOIs in each article

Get DOI metadata from Crossref API

Findings (1 / 2)

History of Australia

- 1255 articles / 243 with DOIs (19%)
- 1526 authors listed in the DOIs
  - 446 with affiliations (29%)
  - 108 with ORCIDs (7%)

Politics of Australia

- 2768 articles / 279 with DOIs (10%)
- 1228 authors listed in the DOIs
  - 290 with affiliations (23%)
  - 82 with ORCIDs (7%)

Findings (2 / 2)

Geology of Australia

- 888 articles / 283 with DOIs (32%)
- 3649 authors listed in the DOIs
  - 980 with affiliations (27%)
  - 368 with ORCIDs (10%)

Biota of Australia

- 5961 articles / 2641 with DOIs (44%)
- 15267 authors listed in the DOIs
  - 2691 with affiliations (17%)
  - 954 with ORCIDs (6%)

Problems

- The lack of affiliations and affiliations being a free
  text field makes it hard for us to gauge the impact of our researchers.
- only a subset of Wikipedia, so we can't get a complete picture
- only DOIs, doesn't take into account references that don't use a persistent identifier
- Long process, many steps, and time consuming.

Future!

- WikiData - https://www.wikidata.org/wiki/Q21090057
- Scholia - https://scholia.toolforge.org/
- Wikicite - http://wikicite.org/