Unpaywall and the coming Age of OA Flight

@jasonpriem
NFAIS 2018
available at bit.ly/nfais-unpaywall
9 days later...
Well that was totally out of the blue!

(pun intended)
Or was it?

Those familiar with the field knew this was coming soon.

- 1853 Cayley manned glider flies
- 1890 Clement Ader makes first powered takeoff
- 1890s Lilienthal makes serial controlled flights in gliders
- 1894 Chanute publishes *Progress in Flying Machines*
- 1902 St Louis Aeronautical Exposition
- 1903 Jatho makes short powered flights
OK, let’s talk about scholarly communication

1. Open is the new default access model.
2. Value in the industry is now a level up
3. Unpaywall helps this happen now
Open is the new default access model.
Let’s look at some numbers

From a study of all 90 million scholarly articles with a DOI, using Unpaywall database.
Percent of articles that are Open Access, by publication year
%OA by publication year, plus a naive projection

Nearly half of 2015 articles are Open Access.

In 2040, all articles published that year will be Open Access. (based on current growth)

75% of articles published before the mid 1990s are behind paywalls.
%OA by publication year: a more realistic projection

Steady growth of OA from mid-90s to 2020.

OA% accelerates with 2020 mandate

OA% reaches steady state around 90% in 2030.
Market share of toll-access by year (same graph, upside down)
I’ll just put this here...
Horses and mules per capita, United States

Importantly, more-used articles are more likely to be OA
Unpaywall gets data on what articles people actually read.
Another approach: what do people cite?

“Measuring the impact of piracy and open access on the academic library services”

Muller and Iriarte looked at articles cited by Swiss researchers in the last two years.

https://archive-ouverte.unige.ch/unige:102345
Percentage of articles cited by UNIGE researchers in their 2015 & 2016 publication that are present in

- Institutional repository: 4% (4%)
- E-journals: 67% (67%)
- Print journals: 25% (24%)
- ILL / DD requests: 0% (0%)
- Open Access: 38% (41%)
- Sci-hub (CH): 0% (1%)
- None of those: 16% (11%)
Toll-access is not going away.

It’s just getting smaller.
OK then what’s the future?
Adding value on top of the OA corpus
Value is moving up the abstraction stack

Moving from articles to groups of articles.

- Information overload
- OA means less money in selling access lower in stack (articles)
- Growth of easy machine learning
- OA corpus means more raw material
The O.G.

Garfield was first to realize massive value by moving up the abstraction stack.
Today: somewhere, in an unassuming bicycle shop...
That said:
There is a little catch...
It’s open, but is it available?

OA lurks in thousands of repositories and journals. Metadata quality ranges from absent to execrable.

It’s like trying to set up a server farm before AWS. You do not want to have to be an expert in this stuff.
Unpaywall is built to scale

It’s not the Wright Flyer, it’s the beach.

- Structured data for every Crossref DOI (95 million articles)
- Built for copyright compliance from ground up
- Accurate: 98% precision, 75% recall compared to Google Scholar, independently assessed.
- Updated weekly
Who’s using Unpaywall?

OA Assessment efforts

- US National Institutes of Health
- Research Councils UK
- National projects in Chile, Switzerland, Sweden, Portugal
Who’s using Unpaywall?

- OA Button
  5k users
- Kopernio
  4k users
- Unpaywall extension
  120k users

Browser-based access tools
Who’s using this?

Link resolvers

- MIT, Harvard, British Library.
- Germany
- 1500+ libraries
- ~200k uses a day
Who's using this?

Value-add aggregation tools

- Web of Science
- Dimensions
- Verum analytics
- Catalytic
- Sparrho
- Scholarcy
- and many we can't talk about yet...
Let’s get flying!