

Beyond writing enthusiast contributors

- **Knowledge Stewards** (Active volunteers)
 - Some can be connected, some can be disconnected
 - [Text editors](#)
 - [program volunteers/leaders](#)
 - [Bureaucrats](#)
 - [Tech volunteers](#)
 - [Photo contributors](#)
 - [Translators](#)
- **Wikimedia Newcomers**
 - [Brand new editors](#)
 - Attending 1st wikimedia event
 - New donors

Institutional Knowledge Stewards

- [GLAM](#)
- [Educational Organizations](#)
- [Partner Research Institutions](#)
- Research Funders (NIH)

Distributors

- WPO partners (telecoms)
- Technology partners
- [Wikipedia Library Partners](#)

Infrastructure Stewards (Plumbers)

- [WMF Staff](#)
- [Affiliates](#)
- [WMF Board](#)
- [Committees \(FOC, AFFCOM\)](#)

Connected Global Influencers

- [Journalists](#)
- Policymakers
- Bloggers

Rule Makers (Influencers)

- Internet regulators
- Governments

Curious Information Seekers (Proactive Learners)

- Rabbit Hollers
- Deep Learners

Casual Learners

- Fact finders
- Context finders

Unknowing Learners

- [Syndication readers \(knowledge graph, siri\)](#)
- [Googlers](#)

Passive Distributors (Content Distributors)

- Publishers
- Content Syndicators

Money Contributors

- [Individual Donors](#)
- [Major Donors](#)
- [Endowment contributors](#)
- [Donors who donate to affiliates](#)

People We're Not Reaching

- [People not yet online](#) (4B)
- [Potential users \(readers / editors\)](#)
- [Readers in censored environment](#)
- People with disabilities
- Readers with special needs

Others:

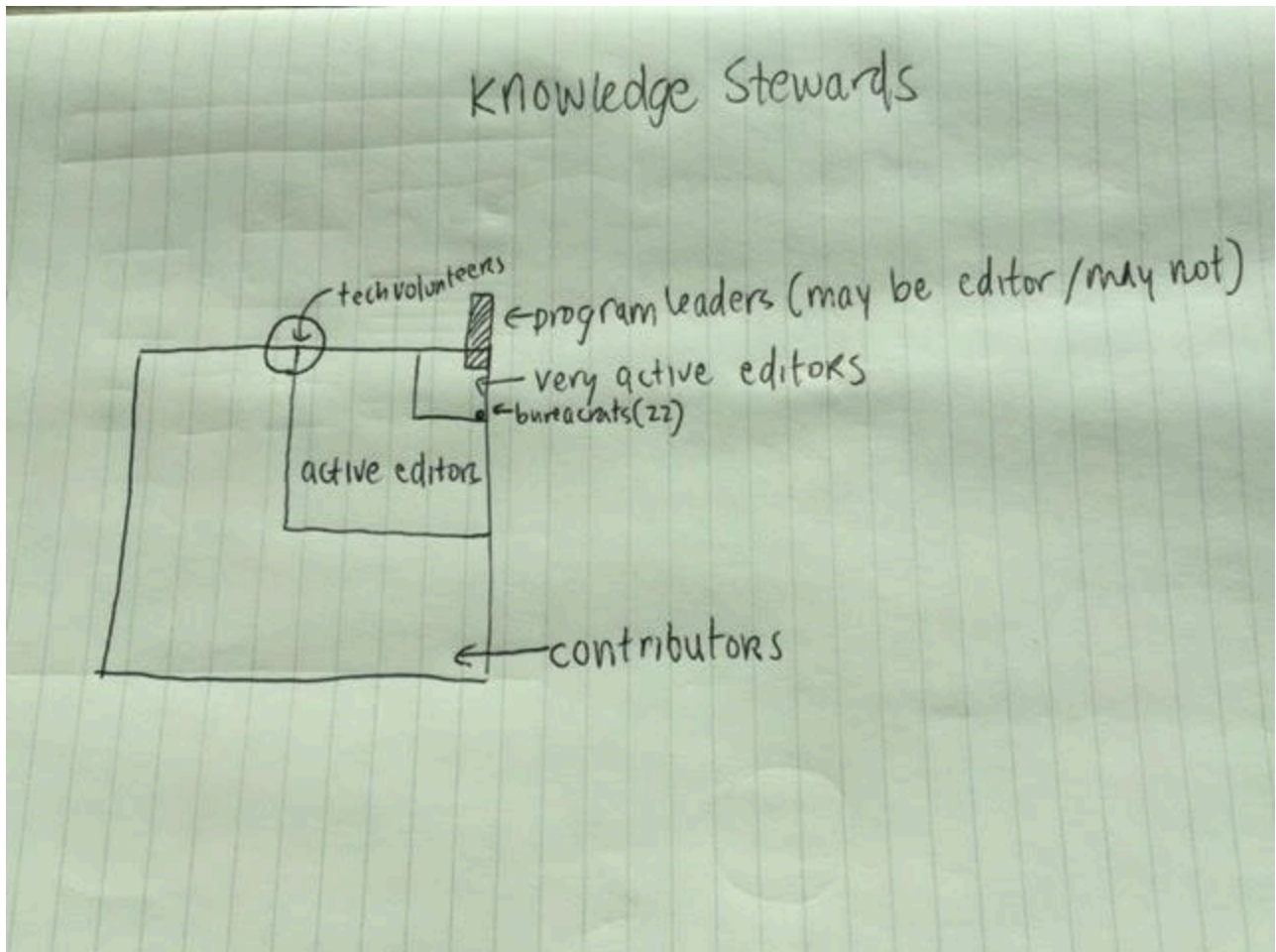
- Vandals
- Trolls
- [API Consumers](#)
- [Grantmakers - wikimedia](#)
- [Orgs using Media Wiki](#)

- [Researchers of Wikipedia](#)
- Kiwix readers (offline)

Beyond writing enthusiast contributors

Knowledge Stewards (Active volunteers)

Some can be connected, some can be disconnected



Knowledge	Work with /	May be	Are not	May become
-----------	-------------	--------	---------	------------

Stewards	support / Amplify			
	Institutional Knowledge Stewards	Wikimedia Newcomers	People we're not yet reaching	N/A
		Money Contributors		

Text editors

Contributors (>10 times since starting to edit):

<https://stats.wikimedia.org/EN/TablesWikipediansContributors.htm>

Active Wikipedians (>5 edits/month): <https://stats.wikimedia.org/EN/TablesWikipediansEditsGt5.htm>

Very Active Wikipedians (>100 edits/month):

<https://stats.wikimedia.org/EN/TablesWikipediansEditsGt100.htm>

Program volunteers

From Maria: Program leaders - Wikipedians who are running some kind of Wikimedia programs.

(Rough estimate of 150~) 80 people on mailing list a while ago, and more every year. We are serving Learning Day - serving about 50-60 people. 100-150.

Bureaucrats

For English Wikipedia, there are currently 22 bureaucrats.

Source: <https://en.wikipedia.org/wiki/Wikipedia:Bureaucrats>

Tech volunteers

From Andre:

The number of active people who commit changes to the Gerrit code review tool and merge into our Git code repos is about 200-250:

https://www.mediawiki.org/wiki/Community_metrics#Active_Gerrit_code_review_users_per_month

But there is also a lot of code that is hosted on the wiki pages themselves, such as Templates, Gadgets/Scripts, and Modules, mostly written in Lua and JavaScript, so there are many more people.

And for that we do not have any reliable numbers unfortunately. For example, I asked Brad Jorsch last week how to get user names of most active Lua template editors across WMF sites, and the answer was "The best you might do would be to look at the top editors in the Module namespace on different wikis".

Photo contributors

~30,000 active users

Source: <https://commons.wikimedia.org/wiki/Commons:Statistics>

Translators

1000-10,000

Sources:

https://translatewiki.net/wiki/Project:MediaWiki_localisation_in_the_50_most_spoken_languages

<https://en.wikipedia.org/wiki/Wikipedia:Translation>

<https://translatewiki.net/wiki/Translating:MediaWiki>

https://www.mediawiki.org/wiki/Localisation_statistics

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

Wikimedia Newcomers

Wikimedia Newcomers	Work with / support / Amplify	May be	Are not	May become
		Proactive Learners	People we're not yet reaching	Knowledge Stewards
		Money Contributors		Money Contributors

Brand new editors

At least 10 edits since arriving: <https://stats.wikimedia.org/EN/TablesWikipediansNew.htm>

Attending 1st wikimedia event

New donors

Institutional Knowledge Stewards

Institutional Knowledge Stewards	Work with / support / Amplify	May be	Are not	May become
	Wikimedia Newcomers	Knowledge Stewards	Passive Distributors	Money Contributors
		Connected Global Influencers		Distributors

GLAM

For GLAM its a much broader criteria (and includes some of the library collaborators), with fuzzier numbers. Current GLAM organization partners across the movement, who are actively working on content collaboration in some way are probably in the 100-300 range, depending on the threshold of partnership you are looking at. Organizations that would do more with the movement, if directly asked and given a simple activity or capacity to do it (like a Wiki Loves event or Art and Feminism -- or who had a volunteer collaborator on hand) is somewhere in the ballpark of 400-1000 distinct organizations, with the vast majority being in Europe and North America -- and directly connected to organizations like DPLA and Europeana. Sometime next year, I want to spend some time trying to figure this out, but the measure would be vast. Professionals are aware of GLAM-Wiki, its probably upward of 50-60% everyone in digital/web GLAM work, though how much they understand is another question.

Educational Organizations

6,000 students - worked with 300 universities. Worked in 272 classes currently. About a 150 universities. We track faculty members rather than universities. Independent, not a chapter, don't fit into model because we're a spin-off. Content that students are adding 1% of all content, 10% of the content to academic subject areas that are stub/start range.

Source: <https://dashboard.wikiedu.org/>

Programs:80

Start Year:2005

Collective Years:318

Funded Programs:27

Students:46,926

Educators:4,602

Volunteers302

Admins / Sysops / Stewards:25

Program Managers (Paid Staff):8

Countries:91

Cities:539

Universities:528

Academic Departments:44

Secondary Schools:224

Primary Schools:35

Adult Schools:11

Senior Programs:3

Total Schools (All Levels):938

WikiCamps:8WikiClubs:17Government Partnerships:13Wikimedia Projects:10Languages:50Articles

Created:148,835Articles Improved:36,055Articles Translated:7,200Files Uploaded:22,502Bytes

Added:543,634,584Bytes Deleted:1,393,217

https://outreach.wikimedia.org/wiki/Education/Historic_data/Table

Partner Research Institutions

From Dario:

1. **research orgs/institutions we have formal collaborations with** (as in: covered by our collaboration policy and with an MOU on file): the list is [here](#).
2. **research orgs/institutions who contribute to Wikimedia research and we collaborate informally with:** the list is very large but we don't have it written anywhere, probably in the order of 50+.
3. **research conference committees/societies we substantially contribute to** (as part of professional service, by participating in steering/program committees, review panels etc): we don't have a list but can produce it easily, it's probably in the order of 10-15 orgs.
4. **non-research orgs/institutions** (libraries, scholarly tech, industry partners) we collaborate with informally for data, services or outreach: a long list much broader in scope, probably in the 50+ range

- Distributors**
- Research Funders (NIH)
 - WPO partners (telecoms)
 - Technology partners

Wikipedia Library Partners

So TWL works closely with >70 partners either through our paid staff or one of the volunteer coordinators (see the list at https://meta.wikimedia.org/wiki/The_Wikipedia_Library/Databases).

From here it's all my guesstimating:

Through library outreach like #1lib1ref and our partnership with IFLA (<https://blog.wikimedia.org/2016/09/19/engaging-libraries-wikipedia/>) , we probably provide direction communication and encouragement to somewhere between several dozen and 100 library/Open Access/knowledge community orgs, and have the ear/positive opinion of >1000 (maybe even >1500) individual librarians between chapter and WMF outreach, and the programs like the Catalan Public Library Network activities
--https://outreach.wikimedia.org/wiki/GLAM/Case_studies/Catalonia%27s_Network_of_Public_Libraries)

Infrastructure Stewards (Plumbers)

Infrastructure Stewards	Work with / support / Amplify	May be	Are not	May become
	Wikimedia Newcomers	Wikimedia Newcomers	People We're not Yet Reaching	
	Institutional Knowledge Stewards			
	Distributors			
	Connected Global Influencers			
	Rule Makers			
	Money Contributors			
	People We're not Yet Reaching			

WMF Staff and contractors

280

Source: https://wikimediafoundation.org/wiki/Staff_and_contractors

Affiliates

112 affiliates overall, https://meta.wikimedia.org/wiki/Wikimedia_movement_affiliates;

70 user groups, https://meta.wikimedia.org/wiki/Wikimedia_user_groups;

41 chapters, https://meta.wikimedia.org/wiki/Wikimedia_chapters;

1 thematic organization, https://meta.wikimedia.org/wiki/Wikimedia_thematic_organizations;

Size of each: ranges from 3-20,000 people



I checked with Comm Resources on affiliate member counts: "No - chapters don't have their members lists publicly available in most cases so we can't - would be a good reporting item though"

Comm Resources handles affiliate report tracking and such

we don't really aggregate data from the affiliates and many report it on places other than Meta. There are also safety concerns for a few groups, and the user groups are not really required to report that level of detail. Some data is in the reports shared here: <https://meta.wikimedia.org/wiki/Reports>

AffCom has talked about posting tables to help aggregate data - but reporting the data would be voluntary

WMF Board

Number: 10 seats, 8 current members

Source: https://wikimediafoundation.org/wiki/Board_of_Trustees

Committees (FDC, AFFCOM)

FDC: 9 voting members, 2 Board liasions, 3 staff, 1 ombudsperson (non-voting)

https://meta.wikimedia.org/wiki/Grants:APG/Funds_Dissemination_Committee

AFFCOM: 12 voting members, 2 Board liasions, 6 staff, 2 community advisors:

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

Connected Global Influencers

Connected Global Influencers	Work with / support / Amplify	May be	Are not	May become
	Infrastructure Stewards	People we're not reaching		

Journalists

Sam: Sure happy to help! I'd say roughly speaking, we converse with about 100 - 300 different journalists in a given year. Juliet -- Does that sound right to you?

I'm mostly basing this off of the number of press requests we get in a given year which I keep track of [here for FY 16 - 17](#) and an archived spreadsheet [here for FY 15 - 16](#).

Policymakers

Bloggers

Rule Makers (Influencers)

Rule Makers	Work with / support / Amplify	May be	Are not	May become
-------------	-------------------------------	--------	---------	------------

Internet regulators

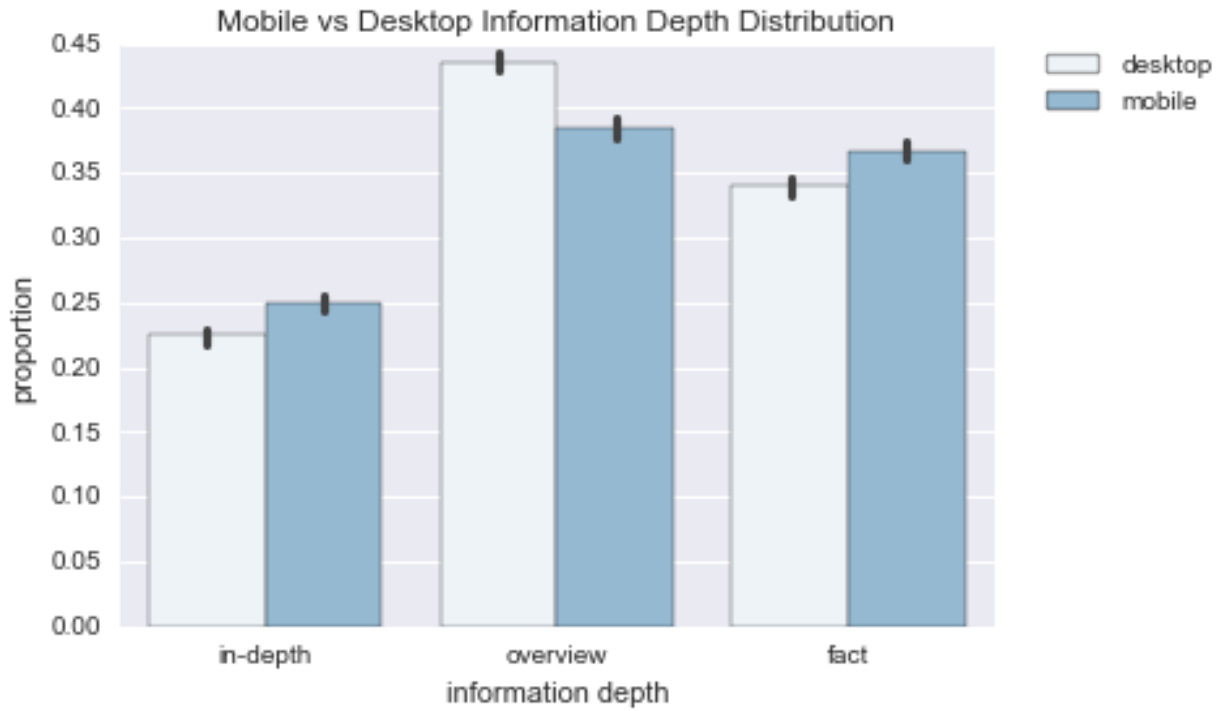
Governments

Curious Information Seekers (*Proactive Learners*)

Curious Information Seekers	Work with / support / Amplify	May be	Are not	May become

Rabbit Hole-ers

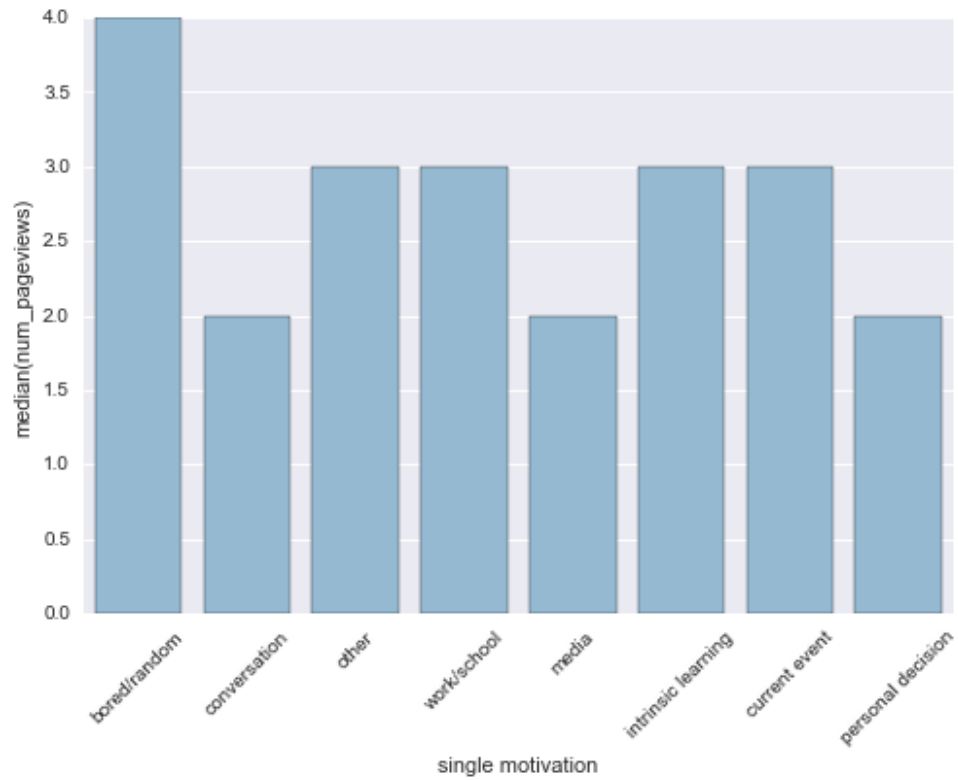
% of user by depth intention:



I would just eyeball the graph since the research team has not provided us with tables for this and there was some subsequent refinement:

We can ask research for more specific questions around breaking out the rabbit holders from the deep learners, as we have clues to this (such as whether or not the user is motivated by a school assignment or boredom).

Here's another example to give you a sense of what is available:

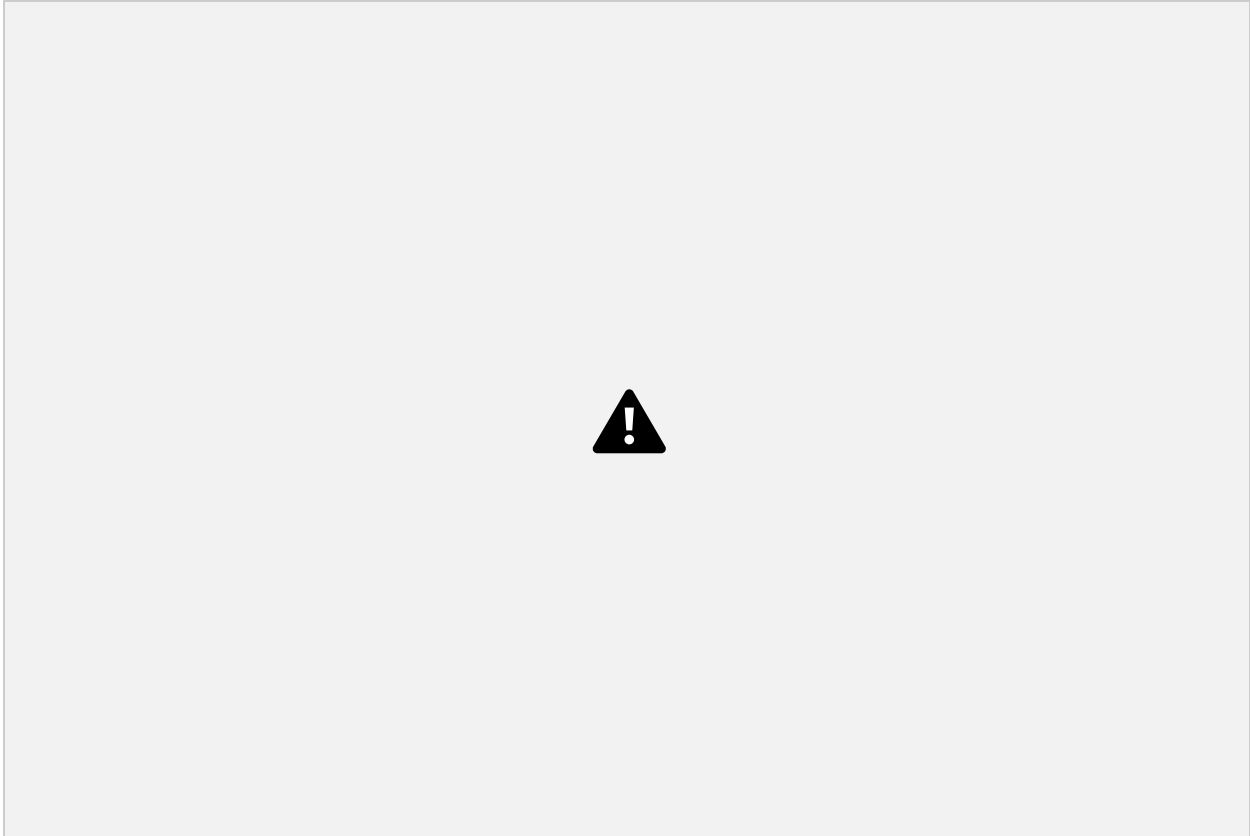


Segments:

It is worth noting that these categories came about after doing an open-text survey in 3 languages and 10,000s of users and several rounds of hand coding, so we are pretty confident about their validity.

Metrics:

The team looked at the following metrics:



Taken from the research paper attached. I think "article topics" is the only one that doesn't really work.

We definitely see differences across user segments, but not enough for them to be predictive of any individual user (since most users just look at 1 page).

Deep Learners

Casual Learners

Casual Learners	Work with / support / Amplify	May be	Are not	May become

Fact finders

- Context finders

Unknowing Learners

Syndication readers (knowledge graph, siri)

Many many billions of people

Googlers

3.5 billion searches/day.

Passive Distributors (*Content Distributors*)

Passive Distributors	Work with / support / Amplify	May be	Are not	May become

Publishers

Content Syndicators

Money Contributors

Money	Work with /	May be	Are not	May become
-------	-------------	--------	---------	------------

Contributors	support / Amplify			

Individual Donors

4.9 million donors

370,000 through email

Donation Totals by Continent

Europe: 23.4 million

US: 45.3 million

South America: 50,000

Africa: 150,000

Asia: 2.7 million

Australia: 2.7 million

Source: https://wikimediafoundation.org/wiki/2014-2015_Fundraising_Report

Major Donors

10.7 million dollars

1300 donors

Source:

https://wikimediafoundation.org/wiki/2014-2015_Fundraising_Report#/media/File:FY1415MajorGiftsKeyStats.jpg

Endowment contributors

The Wikimedia Endowment will start with less than \$1 million donated by the estate of Jim Pacha, a software engineer who died in 2014, and then build toward a goal of accumulating \$100 million by 2026.

Source:

<http://www.gadgetsnow.com/tech-news/Wikipedia-turns-15-year-old-gets-new-source-of-cash/articleshow/50586571.cms>

Donors who donate to affiliates

\$6.6M

Source:

https://wikimediafoundation.org/wiki/2014-2015_Fundraising_Report#/media/File:FY1415DonationSources.png

People We're Not Reaching

People We're Not Yet Reaching	Work with / support / Amplify	May be	Are not	May become

People not yet online (4B)

4 reasons people are not yet online: availability, affordability, relevance, readiness:

<https://www.weforum.org/agenda/2016/02/4-reasons-4-billion-people-are-still-offline/>

Potential users (readers / editors)

New Readers research: https://meta.wikimedia.org/wiki/New_Readers/At_a_glance

Findings from India: https://meta.wikimedia.org/wiki/New_Readers/Findings/India

Findings from Nigeria:

[https://meta.wikimedia.org/wiki/New_Readers/Findings/Nigeria#Theme: Using Wikipedia](https://meta.wikimedia.org/wiki/New_Readers/Findings/Nigeria#Theme:_Using_Wikipedia)

Readers in censored environment

From Stephen: Are you looking for an estimate of readers we could serve if Wikipedia were not censored? Or the readers we serve in regions where there is internet censorship? The first question will be tough to estimate, although we have some rough idea of the lost traffic on Chinese language Wikipedia from China for example. For the second question, we could cross reference Freedom House's annual Freedom of the Net Report (<https://www.freedomhouse.org/report/freedom-net/freedom-net-2016>) with our readership by country stats (<https://stats.wikimedia.org/wikimedia/squids/SquidReportPageViewsPerCountryOverview.htm>) to get a general idea. There are a few caveats about our readership data you have probably heard before -- we only have historical data on pageviews, not unique visitors (or devices) for non-English sites (see https://www.mediawiki.org/wiki/Wikimedia_Product#Reading). I don't believe our unique devices stats are broken out by geography, either. If we need anything specific, we'd probably need to talk to an analyst or researcher.

People with disabilities

Readers with special needs

Others:

- Vandals
- Trolls

API Consumers

From Dan Garry: However, if I had to guess, I'd say it's in the **thousands, or possibly tens of thousands**. There are a bunch of scripts and tools that users use which interact with the site through the API; those alone probably number in the hundreds. Our own apps use the API extensively. There are also tons of third-party re-users. You are welcome to quote me on my estimate. :-)

I do have some sources you can cite; although they do not answer the exact question you are asking, they do contain research in that area:

- [Search API usage](#) - Oliver Keyes, former Data Analyst, Discovery. August 2015.
- [Google referrer report](#) - Oliver Keyes, former Data Analyst, Discovery. August 2015.

From Bryan: It's not an easy question to answer in part because until about a year ago nobody was tracking anything related to Action API usage. We have a data set in hadoop now that could be used to make some reasonable estimates. We have data going back to April 2016 that buckets requests by user agent [0]. That data could be used to make some broad guesses. They are going to be pretty broad because a very large number of Action API requests either include no user agent information or have the signatures of generic sources (curl, LWP, desktop browsers, popular frameworks, ...).

I have an aggregation report for raw requests that runs daily which tells me that there were **16.2 billion Action API requests** in the month of October:

```
$ awk 'BEGIN {OFMT="%.0f"}; {sum+=$2}; END {print sum}'
```

```
out/monthly-2016-10.tsv
```

```
16263115791
```

It's a recommended best practice for an API user to add something to their User Agent header that would tell us how to contact them in case their client is doing something harmful on-wiki. In the same month **we saw 3.4M distinct user agents**:

```
hive (bd808)> select count(distinct userAgent)
```

```
> from action_ua_hourly
```

```
> where year = 2016
```

```
> and month = 10;
```

Total MapReduce CPU Time Spent: 20 minutes 10 seconds 250 msec

OK

_c0

3486682

We can break that down a bit more by the origin of the traffic:

```
hive (bd808)> select count(distinct userAgent) as agents, ipclass
```

```
> from action_ua_hourly
```

```
> where year = 2016
```

```
> and month = 10
```

```
> group by ipclass;
```

Total MapReduce CPU Time Spent: 23 minutes 46 seconds 480 msec

OK

```
agents ipclass
```

```
65935  wikimedia
```

```
3470109 internet
```

```
16578  wikimedia_labs
```

The 3.4M is still the rough number worth keeping based on this breakdown. (I'm actually really surprised by the number of distinct UAs that we are seeing from internal traffic. That looks like something that should be drilled down into a bit to figure out what it really means. It may be that my internal vs labs categorization is broken or it may be legit based on something that Parsoid and/or RESTBase does.)

That 3.4M number is hugely under counting the number of actors however. I know from past analysis that a shockingly low number of API requests actually send a unique User Agent header with contact information. Since we still have raw request data for October (we only keep 90 days of this), we can count unique ip addresses for October:

```
hive (wmf_raw)> select count(distinct ip)
> from ApiAction
> where year = 2016
> and month = 10;
```

Total MapReduce CPU Time Spent: 4 days 9 hours 12 minutes 21 seconds 770 msec

OK

_c0

231319583

So we have 231M unique ips that used the Action API during the month of October. Not all Action API interactions are from people who know they are interacting with the API. The mobile apps do all of their work via API requests. Searching on wiki is also typically an API intensive process with AJAX calls doing all of the type ahead suggestions. Other API hits come from various gadgets and other javascript driven sources.

So we have two numbers to start from: 3.4M and 231M. The former is almost certainly too low. The latter is definitely too high. The real number is somewhere in the middle, but with a spread of two orders of magnitude that leaves a lot of wiggle.

This data really needs an actual analyst to dig around in it. I don't have the time or legitimate credentials to do the kind of analysis that someone like Tillman could do. There was a plan to hire an analyst for this at one point

in late 2015/early 2016, but then Lila was finally convinced that we couldn't monetize the API by offering a "fast lane" for people who paid \$\$\$\$. I never heard about it again.

The Action API data also doesn't account for crawler traffic from Google, Bing, Yahoo, etc that is actually requesting content pages. **These bots are really API consumers if you are thinking about customer groups and drive traffic that is normally discarded by the Analytics team when they talk about page views and unique devices.**

[0]: <https://wikitech.wikimedia.org/wiki/Analytics/Data/ApiAction>

Grantmakers - wikimedia

From Katy:

For our Wikimania scholarship program, we see a lot of unsuccessful applicants. We typically get 600 requests for scholarships, and give about 100-150. For fiscal year 2015-2016, we had about 34 organizational grantees, 25 group grantees, and 184 individual grantees including Wikimania Scholarship 2015 recipients.

We track information in our quarterly reviews (see here:

https://docs.google.com/presentation/d/1CqnG8_Cu9zcppnJaLfyzYDz7rZEs9yHi9Hpn9cuzlQU/edit#slide=id.g129658f92d_0_0) and on Meta:

https://meta.wikimedia.org/wiki/Community_Resources/Grants_spending_analysis#Diversity_breakdown_Q1:_July-Sep_2016 for Quarter 1 for instance. That will help you see how many grantees we do per quarter

We reach roughly 30-50k individuals a quarter, indirectly & directly through grants, though that # is pulled from our QRs and those people are 'repeat customers' over the quarters.

Perhaps the most useful is [here](#) in the last fiscal year with nearly 300 grants made. How many were rejected, you ask? I don't have that at hand. Do you need that?

Orgs using Media Wiki

25,000-30,000

https://wikiapiary.com/wiki/Main_Page

Researchers of Wikipedia

From Dario:

1. **research orgs/institutions we have formal collaborations with** (as in: covered by our collaboration policy and with an MOU on file): the list is [here](#).

2. **research orgs/institutions who contribute to Wikimedia research and we collaborate informally with:** the list is very large but we don't have it written anywhere, probably in the order of 50+.

Kiwix readers (offline)