CROSS-DEPARTMENT PROGRAMS

WIKIMEDIA
F O U N D A T I O N

Platform Evolution

Key Staff:
Corey Floyd, Cindy Cicalese,
Kate Chapman, Adam Baso,
Daniel Kinzler, Subramanya
Sastry, Olga Vasileva, Sam Smith

Teams:
Core Platform Team
Audiences
Technology
WMDE

Platform Evolution

Evolve our technology platform and development processes to *empower* the Wikimedia Movement

WIKIMEDIA
FOUNDATION

First Outcome

Develop a 3-5 year technical roadmap for our platforms
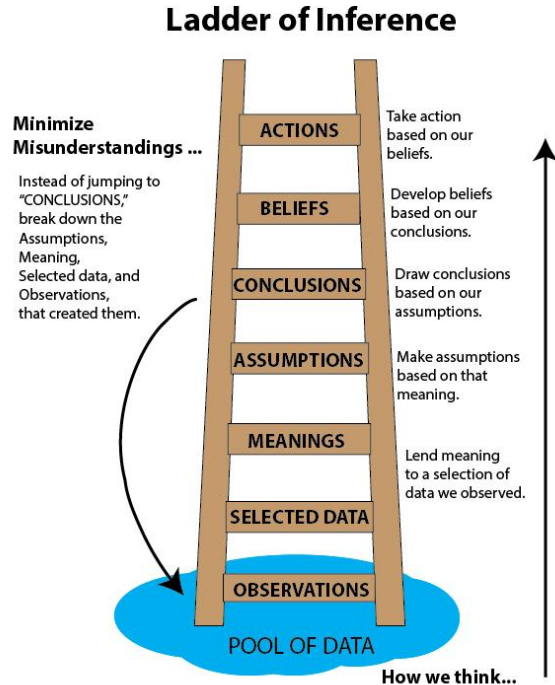
WIKIMEDIA
FOUNDATION

First Outcome

Develop a 3-5 year technical roadmap for our platforms

...with buy-in from within our organization and movement

WIKIMEDIA
FOUNDATION

Applying our tools

Ladder of Inference

Minimize Misunderstandings ...

Instead of jumping to "CONCLUSIONS," break down the Assumptions, Meaning, Selected data, and Observations, that created them.

ACTIONS — Take action based on our beliefs.

BELIEFS — Develop beliefs based on our conclusions.

CONCLUSIONS — Draw conclusions based on our assumptions.

ASSUMPTIONS — Make assumptions based on that meaning.

MEANINGS — Lend meaning to a selection of data we observed.

SELECTED DATA

OBSERVATIONS

POOL OF DATA

How we think...

Model developed by Chris Argyris

- Gather data on stakeholder needs
- Categorize and filter data to identify focus areas
- Key stakeholders meet at Tech Conf to:
  - Seek answers to open questions
  - Achieve shared understanding
  - Propose solutions
  - Document conclusions
- Post Tech Conf: agreements lead to action

WIKIMEDIA FOUNDATION

Getting data from those who know

Stakeholders

Stakeholders

- WMF Audiences and Technology Teams
- WMF Community Relations
- WMDE Engineering (Product, Engineering, and Community)
- TechCom
- MediaWiki Stakeholders' Group
- Various community members and 3rd party users

Our pool of data

Platform stakeholders
- More than 2 dozen interviews
- Written and structured feedback

Product input from both WMF Audiences and WMDE Engineering

Audience Technology Working Group Report

Tech Conf attendee pre-surveys

Filtering and Categorizing

Improving our platform to support modern clients
(APIs, Storage, Web frontend, Content/Presentation)

how do we (who does what — needs clarification) build new products and maintaining existing ones?

What is our platform(s), what are our products? Long term strategy discussion (3-5 years).

Sharing and interactions, people, organization and team responsibilities, code structure.

Surrounding infrastructure of the wikimedia platform (continuous integration, logging, pre-production environments, code review, tooling, etc)

Features that might not belong anywhere specific

Identifying our focus areas

- Define our products, users, and use cases
- Work together, make decisions, and assign responsibilities
- Architect our code for change and sustainability
- Increase our technical capabilities to achieve our strategy
- Enable our engineers, staff, and volunteers to achieve our goals easier and faster

WIKIMEDIA
FOUNDATION

Wikimedia Technical Conference
Portland, October 22-25, 2018

October 2018 **Core Platform** PE

OUTCOME: Stakeholders have a clear understanding of our technology stack and the plan to better scale, maintain and test it

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
Develop a technical roadmap for the platform	**LAST QUARTER**	
	Core Platform, Technology Department, Audiences, WMDE, Community	- Gathered data through stakeholder **interviews and data collection** - Identified **themes, scope, and key questions** for PE - Develop session materials for **TechConf**
	NEXT QUARTER	
	Core Platform, Technology Department, Audiences, WMDE, Community	**TechConf** - Develop an **Technical Roadmap** - Design and implement a **Session Service**

STATUS: OBJECTIVE IN PROGRESS

Audiences Programs

Scale, maintain and test our technology stack
- Speed up unit test execution and increase code coverage
- Reduce deployment complexity

Parser Unification
- End duplicative development efforts by combining Parsoid and the PHP parser

Enables faster feature development and iteration

OUTCOME: Stakeholders have a clear understanding of our technology stack and the plan to better scale, maintain and test it

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
Refactor presentation layer away from business logic code in MediaWiki	**LAST QUARTER**	
	Audiences	- Automate asset bundling in MobileFrontend/Minerva - Speed up unit test execution and increase code coverage
	NEXT QUARTER	
	Audiences	- Automate asset bundling in MobileFrontend/Minerva - Speed up unit test execution and increase code coverage

STATUS: OBJECTIVE IN PROGRESS

16

OUTCOME: Stakeholders have a clear understanding of our technology stack and the plan to better scale, maintain and test it

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
Parser Unification Plan	**LAST QUARTER**	
	Audiences	High-level Goal: Prototype and evaluate a Parsoid port to PHP ● Identify work to frontload before a port ● Implement testing & performance features in Parsoid ● For media, PHP Parser generates HTML similar to Parsoid's output
	NEXT QUARTER	
	Audiences	● Port additional token and DOM transformations to PHP and evaluate performance ● Identify a PEG.js replacement for the PHP Parsoid port ● Production use of Parsoid Language Converter for read views of most languages (not yet Chinese)

STATUS: OBJECTIVE IN PROGRESS

17

Structured Data on Commons

Program lead:

Amanda Bittaker

Key staff:

Sandra Fauconnier, Cormac Parle, Ramsey Isler, Cindy Cicalese, Matthias Mullie, Pam Drouin, Alex Stinson, Mark Holmquist, Brad Jorsch, Ben Vershbow, Stanislav Malyshev, Erik Bernhardson, Deb Tankersley, Erika Bjune, Lydia Pintscher, Daniel Kinzler, Adam Shorland, James Forrester

Teams:

Program Management
Core Platform
Wikidata (WMDE)
Search Platform
Multimedia
Community Programs
Research
Technical Collaboration

WIKIMEDIA
FOUNDATION

Make Commons more used by
making
adding,
finding,
reusing
media and metadata easier
by making it machine readable

WIKIMEDIA
FOUNDATION

5 outcomes

(and one corollary outcome)

1. Make editing metadata easier

   (and enable multilingual statements)

2. Make search more effective

3. Satisfy GLAM use case

4. Allow license compliant reuse

5. Support tool development

WIKIMEDIA
FOUNDATION

3 steps

1. Build the infrastructure
2. Build the products
3. Get people to use them

MCR live on Commons!

A non-free Adventure Time gif went here.
On giphy: https://giphy.com/gifs/excited-adventure-time-4CFjzDix8jacE

CHECK IN
Oct 2018

TEAM/DEPT
Technology: Core Platform

PROGRAM
Structured Data on Commons

GOAL: Build the infrastructure

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
1. Commons contributors can edit metadata more easily, in any MediaWiki-supported language	**LAST QUARTER**	
	Core Platform team	T194750 Deploy Structured Data on Commons baseline ⏱
	THIS QUARTER	
	Core Platform team	Wrapping up last MCR work for SDC file captions and statements on Commons

STATUS: OBJECTIVE AT RISK

CHECK IN
Oct 2018

TEAM/DEPT
Technology: Search Platform

PROGRAM
Structured Data on Commons

GOAL: Build the infrastructure

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
2. People searching for media on Commons can find more relevant results faster	**LAST QUARTER**	
	Search Platform team with support from Multimedia and Wikidata teams	Specific media search filters for Wikidata/Wikibase and the relationships to the topics they represent will be better supported using structured data and other techniques.
	THIS QUARTER	
	Search Platform team with support from Multimedia team	Allow search by type of license

STATUS: OBJECTIVE ON TRACK

CHECK IN
Oct 2018

TEAM/DEPT
Audiences: Multimedia

PROGRAM
Structured Data on Commons

GOAL: Build the products

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
3.1 1: Commons contributors can edit metadata more easily, in any MediaWiki-supported language	**LAST QUARTER**	
	Multimedia team with support from WMDE and Search and Core Platform teams	○ Improve and update the code of user interface elements for multilingual file captions (this includes integrating OOUI with Wikibase JavaScript elements) ○ Develop working prototypes for adding/editing/removing "depicts" statements on UploadWizard and file page; expand on depicts search prototype as necessary. ○ Complete documentation for the new search features and specifications that have been added for SDC
	THIS QUARTER	
	Multimedia team with support from WMDE and Search and Core Platform teams	○ Deploy multilingual file captions on Commons in October ○ Refine API to manage adding/editing depicts values for 3rd party tools ○ Deploy depicts and other statements to Beta Commons

STATUS: OBJECTIVE AT RISK

Alpha of file captions on file page

File:Rolled up, lying sleeping cat by Jean Bernard.jpg

Jump to: navigation, search

| File | File history | File usage |

Size of this preview: 800 × 486 pixels. Other resolutions: 320 × 194 pixels | 3,072 × 1,867 pixels.
Original file (3,072 × 1,867 pixels, file size: 1.54 MB, MIME type: image/jpeg)

Summary

Description	**English:** Rolled up lying, sleeping cat by Jean Bernard (1775-1883). Original from the Rijks Museum. Digitally enhanced by rawpixel.
Date	1812-1-1
Source	https://www.rawpixel.com/board/481252/jean-bernard
Author	Jean Bernard

Template:Works created more than 200 years ago

Licensing

Template:PD-Art

File history

Click on a date/time to view the file as it appeared at that time.

	Date/Time	Thumbnail	Dimensions	User	Comment
current	21:41, 3 October 2018		3,072 × 1,867 (1.54 MB)	TheRealNightcrafter (talk | contribs)	User created page with UploadWizard

- You cannot overwrite this file.

File usage

There are no pages that use this file.

Structured data

Captions ✎

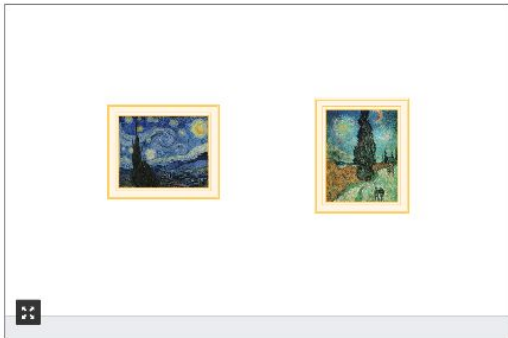| English | Sketch of a rolled up cat by artist Jean Bernard |
| German | Skizze einer eingerollten Katze des Künstlers Jean Bernard |
| Dutch | Tekening van een opgerolde kat door kunstenaar Jean Bernard |

∧ See fewer languages
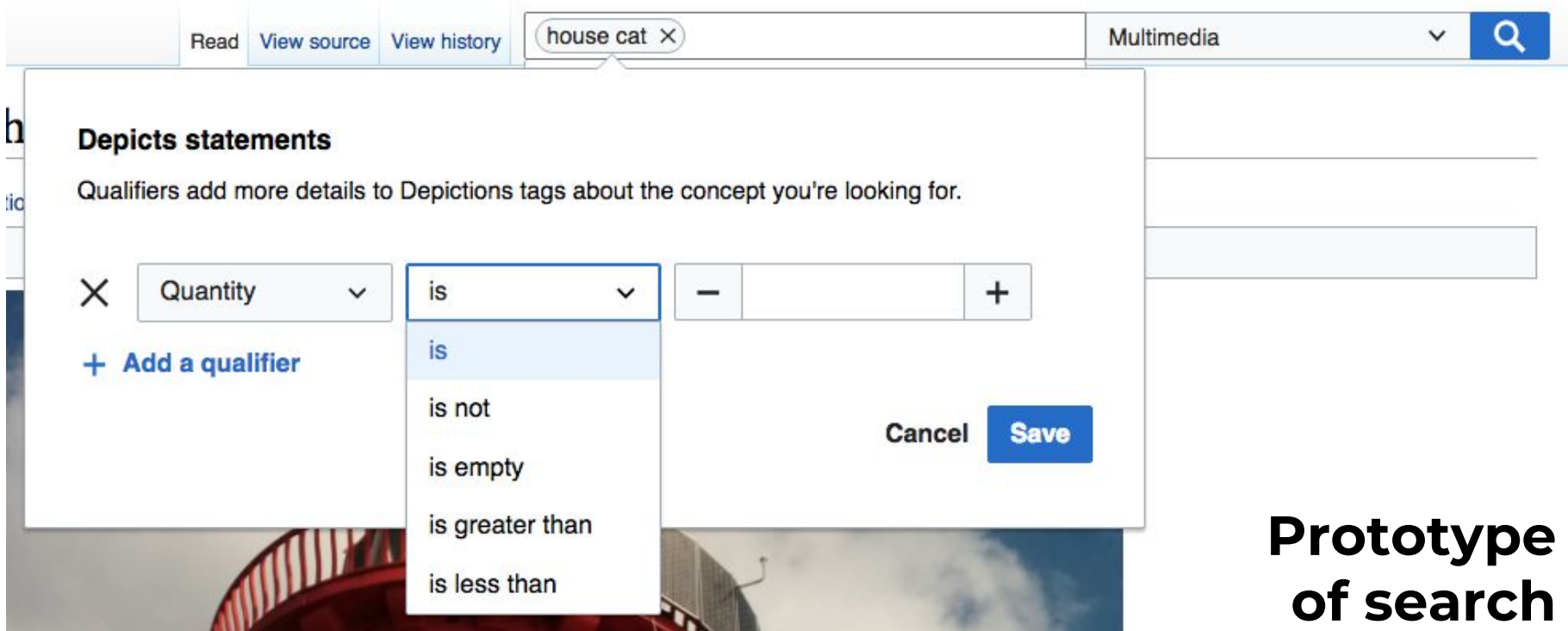
Prototype of search using depicts

Clickable mocks of file page with depicts statements

CHECK IN
Oct 2018

TEAM/DEPT
Community Engagement: TC

PROGRAM
Structured Data on Commons

GOAL: Build the products

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
5.1 The Wikimedia communities, GLAM partners, and developers are fully on board with the Structured Data project. They participate in the different stages of planning and development, and adopt the new features.	**LAST QUARTER**	
	Technical Collaboration team with support of Multimedia team	○ Send 1 or 2 newsletters and host 1 or 2 IRC office hours ○ Facilitate on-wiki conversations about design and prototypes ○ Facilitate on-wiki conversations about structured licenses
	THIS QUARTER	
	Technical Collaboration team	○ Communicate multilingual file captions release, depicts and other statements designs, etc. to communities ○ Monitor community feedback on multilingual file captions and surface issues and wins to Multimedia team ○ Facilitate community conversation about property creation process for Commons properties on Wikidata, arrive at defined property creation process for Commons property requests ○ Present SDC update at WikiConNA and talk to key program organizers about their Commons programming and SDC.

STATUS: OBJECTIVE ON TRACK

CHECK IN
Oct 2018

TEAM/DEPT
Technology: Research

PROGRAM
Structured Data on Commons

GOAL: Build the products

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
4.2 We will develop better understanding of existing needs for Structured Commons… [read more]	**LAST QUARTER**	
	Research team working with Multimedia	Interview individual and organizational re-users of Commons media to current workflows and impact of structured data (T190228) (Report on Meta.)
	THIS QUARTER	

STATUS: OBJECTIVE ON TRACK

CHECK IN
Oct 2018

TEAM/DEPT
Community Engagement: GLAM

PROGRAM
Structured Data on Commons

GOAL: Get people to use them

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
3. GLAM institutions have more reliable and less challenging workflows for sharing media files and metadata on Wikimedia Commons as part of content partnerships, allowing for increased GLAM participation.	**LAST QUARTER**	
	GLAM team working with Multimedia and Technical Collaboration teams	○ Facilitate GLAM-SDC conversations at Wikimania, including facilitating 2 sessions. ○ Define scope and timeline for a shortlist of GLAM pilot projects, including conversations at Wikimania on this topic. ○ Support GLAM-related discussions as part of the community discussions around Structured Licenses.
	THIS QUARTER	
		○ Announce relevant ongoing releases, design proposals and conversations and solicit feedback from GLAM community ○ Present SD update at the GLAM-Wiki conference & talk to key program organizers about SDC opportunities. ○ GLAM sector SDC outreach at CIDOC conference (Heraklion), Shaping Access conference (Berlin), and Baltic Audiovisual Archival Council conference (Tallinn)

STATUS: OBJECTIVE ON TRACK

Privacy, Security & Data Management

Program leads:
Tony Sebro
John Bennett

Teams:
Advancement
Analytics
Audiences
Community Engagement
Finance & Administration
Legal
Technology
Trust & Safety

WIKIMEDIA
FOUNDATION

Security achievements for Q4

We're Hiring
- **Completed Job Descriptions to round out hiring**

Security Incident Response Table Tops
- **Held our 1st TTX (Intrusion Simulation)**

Security reviews as requested

Continue to **enhance security** capabilities and processes to align with changing compliance regimes and best practice.

WIKIMEDIA
FOUNDATION

Security achievements for Q1

Ensure the high-quality protection and security of our infrastructure and data.

1. Updated policy to end of the **1 character password**
2. Define **Security Awareness** content
   a. Securing your digital identity
   b. Application Security principles
3. Expanded **Content Security Policy** use
4. Scoped **penetration testing** for English Wikipedia
5. Identified elements for **Security release**
6. Identified and Scoped Analytics **assessment**
   a. Threat Model + Vulnerability Assessment + Risk assessment + Framework assessment + Policy = method to **balance risk**
7. Performed 2 **Security Incident Response** exercises
8. More **security reviews**

Ninja Portrait
Image:Zach Dischner

WIKIMEDIA
FOUNDATION

Community Health

Program leads:
Trevor Bolliger
Patrick Earley
Jacob Rogers

Teams:
Anti-Harassment Tools
Trust & Safety
Legal

WIKIMEDIA
FOUNDATION

Anti-Harassment Tools

Building software for moderators to reduce and fairly resolve incidents of harassment.

WIKIMEDIA
F O U N D A T I O N

Anti-Harassment Tools team focus areas

🔍 **Detection** — Tools to identify and flag harassing behavior.

🗣️ **Reporting** — Improved ways to report incidents of harassment.

⚖️ **Evaluation** — Tools to help volunteers evaluate harassment cases.

✋ **Remedies** — Methods to remove or respond to troublesome actors.

97.4%

English Wikipedia administrators confident with their skills to mitigate <u>vandalism</u>

39.3%

English Wikipedia administrators confident with their skills to mitigate <u>harassment</u>

WIKIMEDIA
FOUNDATION

Anti-Harassment Tools

- **Partial Blocks**
  - Sitewide blocks are not always the appropriate response to some situations.
  - Smaller, more tactical blocks may defuse situations while retaining constructive contributors.
  - The goal of this project is to give wiki administrators a more robust set of tools to be able to better respond to different user conflict situations.
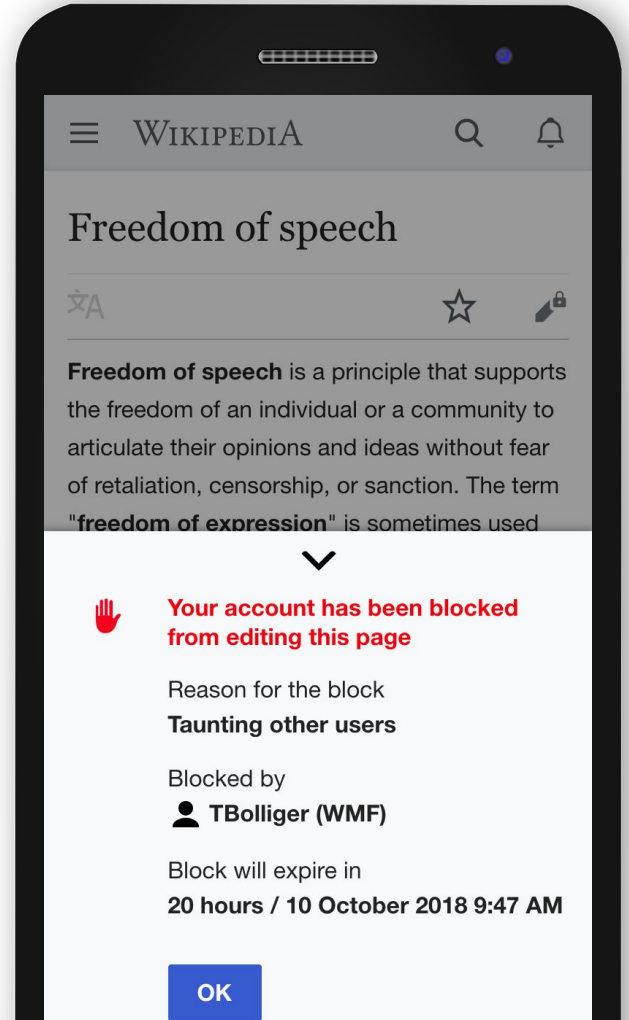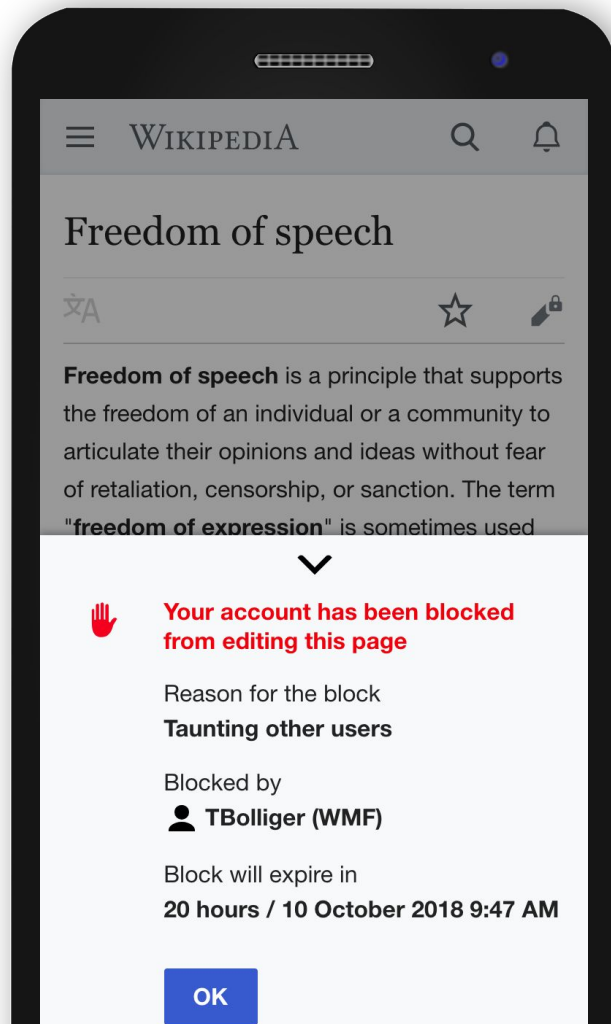
Anti-Harassment Tools

- **Block frequency across all wikis**
  - 370 sitewide blocks are set every hour on average
  - 3,482,751 sitewide blocks are active right now
  - We expect block usage to increase with the introduction of Partial Blocks
  - In Q2, our team will implement recurring reports of this data.

- **Effectiveness of blocks**
  - Do temporary sitewide and partial blocks retain constructive contributors?
  - Are partial blocks effective at stopping abusive behavior?
  - In Q2, our team will measure baselines and implement recurring reports of this data.

Data: May 8 2018 via T190328

CHECK IN
Oct 2018

TEAM/DEPT
Anti-Harassment Tools

PROGRAM
CDP:CH1

WIKIMEDIA
FOUNDATION
ANNUAL PLAN

GOAL: Develop better tools to reduce harassment in our communities

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
Support prioritization and development of improved tools, processes and policies to mitigate harassment of contributors	**LAST QUARTER**	
	Trust & Safety **Program lead:** Trevor Bolliger	✓ Built alpha version of Partial Blocks ✓ Held successful Wikimania workshop about Reporting System ✓ Hired Claudia Lo, our Design Researcher. ✖ Did not run 2018 admin confidence survey. ✖ Did not train users to use the Interaction Timeline.
	NEXT QUARTER	
	Trust & Safety **Program lead:** Trevor Bolliger	● Release partial blocks to 1+ production communities & measure their effectiveness. ● Agree to process for Reporting System & begin consultation. ● Design filters for the Interaction Timeline ● Gather more data about our work: implement weekly reports & run the admin confidence survey.

STATUS: OBJECTIVE IN PROGRESS

Trust & Safety:
Community Health Initiative

Growing Policy and Better Community Healths Metrics

WIKIMEDIA
F O U N D A T I O N

Support & Safety focus areas

🔬 **Measuring**
Understanding community health better through metrics

🐛 → 🦋 **Evolving policy**
Working with contributors to build better policy

Community Health Metrics

- Creating prototypes - exploring visualizations

- Working with Inspire grantees to get community-driven metrics

- Working with Analytics and Design to determine platform and design

- Preparing for development

edia

ns

edits

e users

onfirmed editors

day

admin

Active users per admin

100,000

75,000

50,000

25,000

0

Nov (2017) Dec (2017) Jan (2018)

CHECK IN

Oct 2018

TEAM/DEPT

Support & Safety

PROGRAM

CDP:CH 2 & 3

WIKIMEDIA
FOUNDATION
ANNUAL PLAN
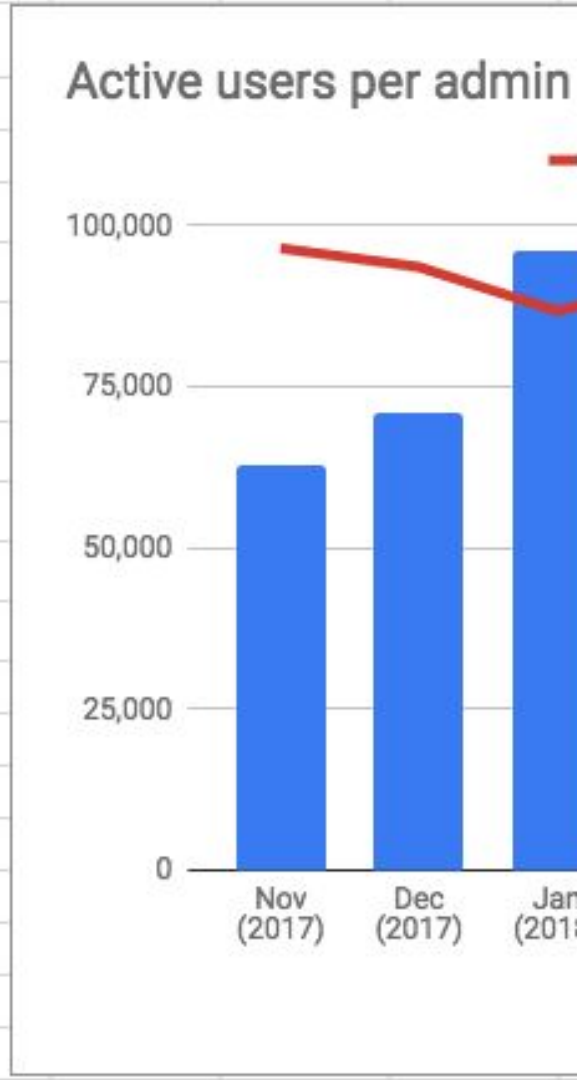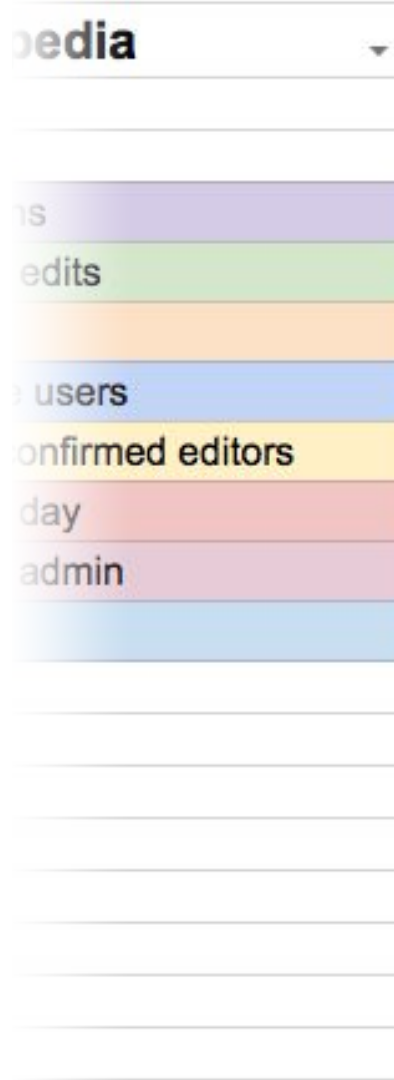
GOAL: Support better policy & enforcement growth in our communities

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Measure community health in deeper ways. Work with a large Wikimedia project on policy evolution.	**LAST QUARTER**	
	SuSa, Community Tech (AHT), Design Research **Program lead:** Patrick Earley, Trevor Bolliger, Danny Horn	Produced a map of community health measurement points, scoped for priority and development needs to create worklist. Brought Wikidata stakeholders together for consultation. Prepared policy options brief for Wikidata.
	NEXT QUARTER	
	SuSa, Community Tech (AHT) **Program lead:** Patrick Earley, Trevor Bolliger, Danny Horn	CHMK: Consult and brief community on measurement points, design options; create wireframes and prototypes for review. WD Policy: Bring Wikidata stakeholders together for consultation. Determine direction community supports, provide best practices from movement

STATUS: OBJECTIVE IN PROGRESS

Legal:
Community Health Initiative

WIKIMEDIA
F O U N D A T I O N

CHECK IN

Oct 2018

TEAM/DEPT

Support & Safety, Legal

PROGRAM

CDP:CH 2 & 3

WIKIMEDIA
FOUNDATION
ANNUAL PLAN

GOAL: Provide legal support to community members who face significant harassment

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Output 1: Community health cases receive legal support, either via direct Foundation action or funding for a volunteer action.

Output 2: Take direct legal action against users broadly disrupting the project, as appropriate | **LAST QUARTER** | |
| | Legal, SuSa
Program lead: Jacob Rogers, Patrick Earley | Continued long-term investigations. Advised on multiple Trust & Safety cases. |
| | **NEXT QUARTER** | |
| | Legal, SuSa
Program lead: Jacob Rogers, Patrick Earley, | Provide support on additional cases as necessary to protect community members. Continue to investigate legal options for particularly severe or technically adept harassment. |

STATUS: OBJECTIVE IN PROGRESS

New Readers

Program lead: Anne Gomez

Key Staff:
Abbey Ripstra, Jorge Vargas, Jack Rabah, Dan Foy, Zack McCune, Satdeep Gill, Olga Vasileva, Charlotte Gauthier, Alex Wang

Teams:
Communications
Community Resources
Design Research
Partnerships & Global Reach
Programs (Audiences)
Readers

WIKIMEDIA
F O U N D A T I O N

Evolution for FY1819

Primary audience:
People who use the internet
but do not yet read Wikipedia
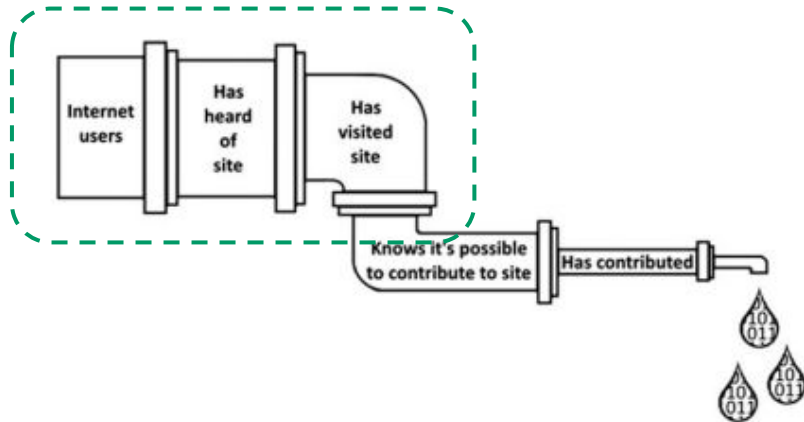
Secondary audience:
People in situations without
internet access

Primary audience:
Online
Outcomes 2-5

WIKIMEDIA
FOUNDATION

Outcome 2:
Awareness

{
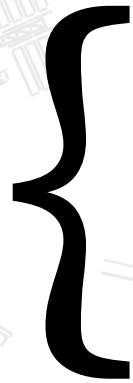
Recognition
I have heard of Wikipedia
Use Value
I use Wikipedia for ...
Attribution
I got it from Wikipedia

WIKIMEDIA
FOUNDATION

Oct 2018   Communications, Global Reach, Audiences   New Readers

GOAL: More people in India know what Wikipedia is.

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
General marketing campaigns Outputs: - Aud (PM) 2 & 3 - Comms 2 - GR&P 2, 3, & 4	**LAST QUARTER**	
	Communications, Global Reach, Wikimedia India volunteers, Finance, Happy mcgarrybowen (advertising agency)	- Evaluation of campaign impact - Optimization and extension of campaign

STATUS: NEARLY COMPLETE

Video impact

33% → **40%**

have heard of Wikipedia

The video earned **3.98 million total views**, with 2.9 million within Madhya Pradesh.

9% of internet users in Madhya Pradesh said they had seen the video.

66% of those said it was the first time they had heard of Wikipedia.

Video links:
- Facebook
- YouTube

CHECK IN

Oct 2018

TEAM/DEPT

Communications, Global Reach, Audiences

PROGRAM

New Readers

GOAL: More people in Mexico know what Wikipedia is.

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
General marketing campaigns Outputs: - Aud (PM) 2 & 3 - Comms 2 - GR&P 2, 3, & 4	**LAST QUARTER**	
	Communications, Global Reach, Wikimedia Mexico volunteers, Finance, ISLA	Marketing materials finalized (script and song written and produced, and main video finalized) by EOQ. Launching is listed for October. Baseline phone survey completed.
	NEXT QUARTER	
	Communications, Global Reach, Wikimedia Mexico volunteers, Finance, ISLA (advertising agency)	Complete production and launch video marketing campaign in Mexico. Complete endline phone survey in Mexico.

STATUS: ON TRACK - IN PROGRESS (NEW PROJECT)

Wikipedia awareness among internet users

USA
87%

Mexico
45%

Brazil
39%

France
84%

Nigeria
27%

Iraq
19%

India
33%

Japan
64%
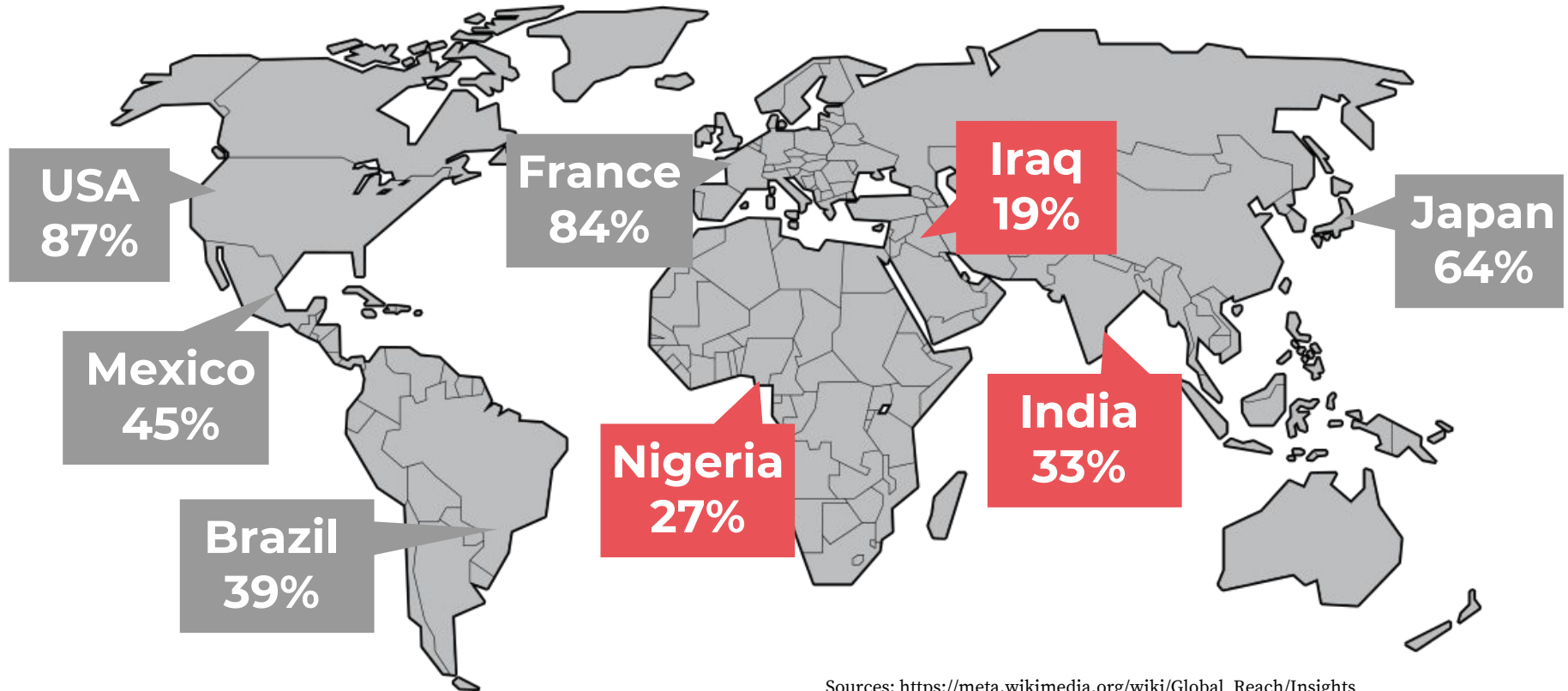
Sources: https://meta.wikimedia.org/wiki/Global_Reach/Insights

Wikipedia awareness among internet users

USA
87%

Mexico
45%

Brazil
39%

France
84%

Nigeria
48%

Iraq
25%

India
40%

Japan
64%
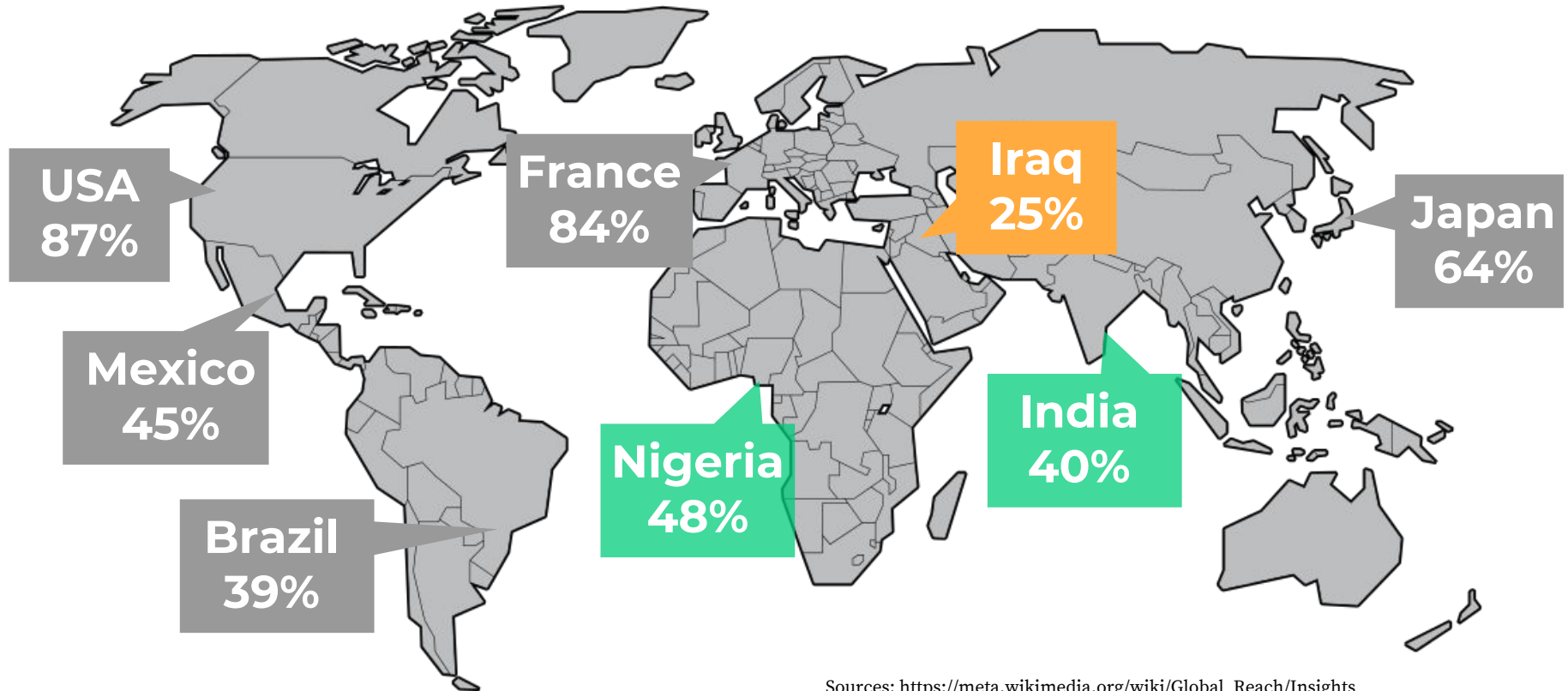
Sources: https://meta.wikimedia.org/wiki/Global_Reach/Insights

Marketing at a glance

	Iraq	Nigeria	India
Audience reached	5 million	17 million	4 million
Awareness result	32% increase *From 19% → 25%*	71% increase *From 28% → 48%*	21% increase *From 33% → 40%*
Promotion costs	*Asiacell donated media*	*$12,270 media spend*	*$18,700 media spend*
Campaign efficiency	N/A	$0.0026 per view	$0.015 per view

WIKIMEDIA
FOUNDATION

What you learned working on this objective or workflow:

Did anything unexpected (good or bad) happen?

- Our cost effectiveness is shockingly high. On average, we are reaching people for less than $0.01 worldwide. This is in part due to the charitable rates that creative agencies have offered us.
- As a region, India has proven very difficult for project organization and impact. Our cost effectiveness was lowest in India, with digital media channels very expensive to buy reach.
- Community feedback seems best suited to detail. Things like copy/script review they excel at, but creative strategy and emotional content are challenging for feedback.

What would you do differently if you had to do it again?

- Increase media spend to "sustain" campaigns and extend the reach of the creative messages. We are in a position to reach millions more people for just a few thousand dollars of investment.
- Hire paid media specialist/marketing analyst into New Readers team. We would be MUCH more informed and confident of investment impact if we were managing all paid distribution internal to the Foundation.
- We need data analysis of site traffic for proper understanding of long term effects and attribution of changes.

Did you struggle with anything?

Coordination! It's hard working across time zones, cultures, and expertise. This is where much of the "genius" of Wikipedia marketing has come from, but it is very demanding in time, effort, and patience.

Community marketing pilots

Building on what we've learned through the Inspire campaign grants and general marketing efforts, we are planning to develop capacity for social media marketing within communities.

Satdeep Gill will work with 2-3 communities in India to:

- Develop an understanding of online marketing within the communities
- Create a training toolkit to be delivered along with awareness rapid grants
- Understand cost of small-scale marketing to inform future awareness grants design

![Wikimedia Foundation logo]
WIKIMEDIA
F O U N D A T I O N

CHECK IN

Oct 2018

TEAM/DEPT

Communications

PROGRAM

New Readers

GOAL: More people in Nigeria visit Wikipedia.

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Product marketing Output: Comms 3	**LAST QUARTER**	
	Communications, Audiences Android, Audiences Programs	Execute product marketing campaign for offline functionality in the Android app, particularly Reading Lists, with intent to understand if offline features drive adoption/retention in data-constrained markets.

STATUS: COMPLETE

What you learned working on this objective or workflow:

Did anything unexpected (good or bad) happen?

- We were not able to track retention of these users due to vendor configuration of account for promotion.
- Creative work was not exceptional, and likely could have been completed in house at lower cost
- Very high cost of staff time due to challenges with vendor follow through (up to 3 hours/week for 6 weeks)

What would you do differently if you had to do it again?

- Hire contractor to execute campaign internally with more oversight rather than working through agency
- Product marketing should be a capacity that's developed within the Foundation. It doesn't require same local expertise as general marketing and working with external teams has high management cost.
- Should have dedicated payment account for marketing spend to facilitate WMF ownership.

Did you struggle with anything?

- Coordination with agency was difficult and costly in terms of staff time

Outcome 3:
Discovery

{ ???

WIKIMEDIA
FOUNDATION

Outcome 4:
Retention

GOAL: More people in Mexico find Wikipedia useful and return to it.

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Create custom landing pages for marketing campaigns to support gradual introduction to Wikipedia. Output: Audiences 2	NEXT QUARTER Audiences Design, Communications, Global Reach	Deliver custom landing pages with tracking. Landing pages will be linked from awareness campaign video in Mexico.

STATUS: ON TRACK - IN PROGRESS (NEW PROJECT)

Outcome 5:
Syndication

CHECK IN

Oct 2018

TEAM/DEPT

Global Reach & Partnerships

PROGRAM

New Readers

GOAL: More people in target countries/languages read Wikimedia content outside of Wikimedia sites

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Pursue partnerships for inclusion of Wikipedia content on other sites, in the form of context cards. Outputs: - GR&P 5 - Audiences 4	**LAST QUARTER**	
	Global Reach, Audiences Programs	Confirm viability of hovercard integration on 3rd party sites. Run small pilot to determine whether technical integration is feasible.

STATUS: OBJECTIVE PAUSED

Secondary audience:
Offline
Outcome 1

WIKIMEDIA
FOUNDATION

CHECK IN

Oct 2018

TEAM/DEPT

Global Reach & Partnerships

PROGRAM

New Readers

GOAL: People without internet access can read Wikipedia. (WOMP)

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Complete offline medical pilot in Nigeria that started in the previous fiscal year, including report about possibility to scale. Output: P&GR 1	**LAST QUARTER**	
	Global Reach, Programs, Internet in a Box, The Africa Center	All logistics and planning for project deployment will be finalized (including kick-off and metrics). TAC will hire a project manager, and produce a quarterly financial report that includes an expenditure to-date, balance remaining, and project expenses for the next quarter.
	NEXT QUARTER	
	Global Reach, Programs, Internet in a Box, The Africa Center	Concrete project plan including overall timeline, milestones, data collection plan, metrics/targets for project. (Not publicly reported Q2 goal)

STATUS: OBJECTIVE BEHIND SCHEDULE

CHECK IN

Oct 2018

TEAM/DEPT

Communications, Global Reach, Audiences

PROGRAM

New Readers

What you learned working on this objective or workflow:

Did anything unexpected (good or bad) happen?

- WOMP:Hiring project manager has been slower than anticipated, which will have impact on remainder of WOMP timeline.

What would you do differently if you had to do it again?

- Ideally, include a tighter scoped project plan as part of the grant agreement. Due to rush with political cycle in Nigeria, this wasn't possible, and it's now causing some challenges.
- Ensure the timelines and plan of action is set beforehand. The term of reference and MOU guidelines were not enough.

Did you struggle with anything?

- Lack of fully dedicated point of contact made for staggered and slow communication.

CHECK IN

Oct 2018

TEAM/DEPT

Audiences (program management)

PROGRAM

New Readers

GOAL: People without internet access can read Wikipedia. (Kiwix)

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Direct support of Kiwix, which is essential central infrastructure for the majority of offline educational resources. Support through funding and advising (strategy, user experience, product, etc.). Output: Aud (PM) 1	**LAST QUARTER**	
	Kiwix, Programs	Hire contract NodeJS and Python developers, as well as GSOC intern to complete foundational work for Q2 goals (release of topical ZIM files, Android updates, etc.). Execute hackathon to advance foundational work.
	NEXT QUARTER	
	Kiwix, Programs	- Deliver strategic plan for Kiwix, focused on serving reusers. - Release auto-updating, small topical ZIM files suitable for mobile (from WMF research)

STATUS: OBJECTIVE ON TRACK

What you learned working on this objective or workflow:

Did anything unexpected (good or bad) happen?

- WOMP:Hiring project manager has been slower than anticipated, which will have impact on remainder of WOMP timeline.

What would you do differently if you had to do it again?

- Ideally, include a tighter scoped project plan as part of the grant agreement. Due to rush with political cycle in Nigeria, this wasn't possible, and it's now causing some challenges.
- Ensure the timelines and plan of action is set beforehand. The term of reference and MOU guidelines were not enough.

Did you struggle with anything?

- Lack of fully dedicated point of contact made for staggered and slow communication.

Staffing product team

Build out product team to develop iteratively, running lots of experiments quickly and cheaply. Team must have experience working with multilingual, data conscious, consumers using low cost devices.

Initial plan:
- 1 Product Manager
- 1 Designer
- 3 Engineers
- Coverage for data analysis from existing team

Current plan:
- 1 Product Manager with highly relevant experience (Q2)
- Agency in target market to build, test, and measure prototypes

Risk mitigation. This is a new team working on new customers - hiring an agency allows flexibility to learn some lessons at a lower cost rather than the expense of hiring 5 FTEs, where decisions are costly to change.

WIKIMEDIA
FOUNDATION

CHECK IN
May 2018

TEAM/DEPT
New Readers

SCORECARD
Mobile traffic

		Pageviews (mobile web, non-Opera*)			
Country	**Language**	**Q1**	**Q4**	**QoQ**	**YoY**
Nigeria	English	62,098,193	61,350,470	**1%**	**11%**
India	English	1,914,925,665	1,758,454,251	**9%**	**21%**
	Hindi	146,454,034	107,740,023	**36%**	**74%**
Mexico	Spanish	611,860,445	645,436,727	**-5%**	**-1%**
	English	96,757,334	93,760,819	**3%**	**2%**

* a bug in Opera browser's IP assignment from February 6 - April 17, 2018 means that QOQ and YOY data for Q3/4 must exclude Opera traffic.
In 2017, Opera mini traffic was 55% of Nigeria's mobile traffic and 7% of India's.

Knowledge Integrity

Trust, Verifiability, Citations

WIKIMEDIA
FOUNDATION

Knowledge Integrity

Program lead:
Jake Orlowitz

Teams:
1. Community Programs (CE)
2. Research (Tech)
3. Wikidata (WMDE)
4. *Citoid (Audiences)*
5. *Internet Archive*

Key Staff:
1. Ben Vershbow, Jake Orlowitz, Sam Walton
2. Miriam Redi, Bahodir Mansurov, Jonathan Morgan, Dario Taraborelli, Leila Zia, Sarah Rodlund
3. Lydia Pintscher
4. Marcella Wagner, Marielle Volz
5. Mark Graham

WIKIMEDIA
FOUNDATION

References are the foundation of **Wikimedia's quality and trustworthiness.**

WIKIMEDIA
FOUNDATION

But trust is a **complex** **evolving** **contested** target

WIKIMEDIA
FOUNDATION

Trust is a **networked** phenomenon

Wikimedia is at the heart of that network

WIKIMEDIA
FOUNDATION

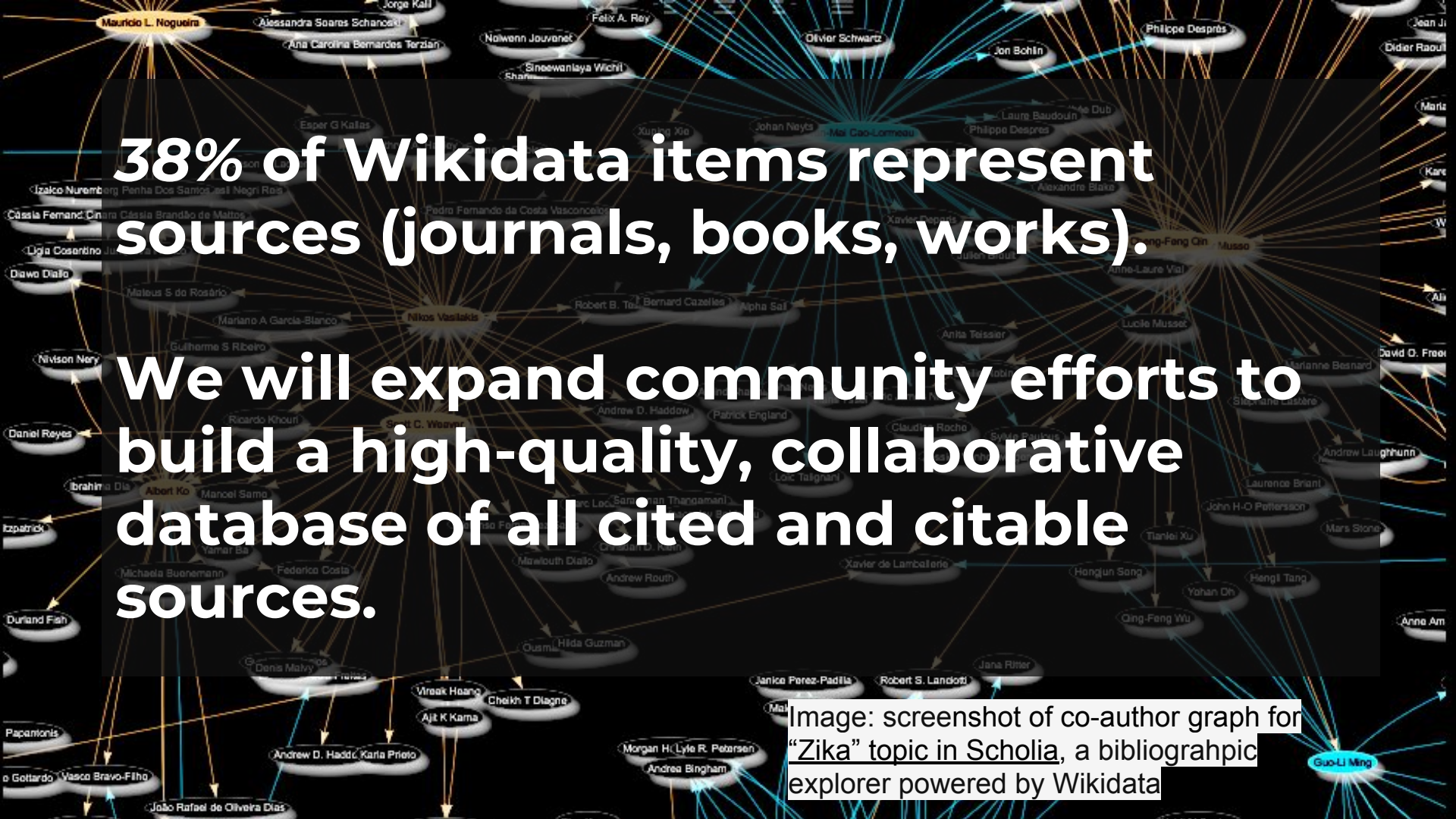# How can we improve our citation systems and practices to become **essential infrastructure of the trustable web**?

Wikipedia sends *tens of millions* of people to external sources each year.

We will conduct research to understand why and how readers leave our site.

38% of Wikidata items represent sources (journals, books, works).

We will expand community efforts to build a high-quality, collaborative database of all cited and citable sources.

Image: screenshot of co-author graph for "Zika" topic in Scholia, a bibliograhpic explorer powered by Wikidata

The Internet Archive has fixed over *9 million* dead links on Wikipedia.

We will enable instantaneous archiving of *every* link on all Wikipedias.

#1Lib1Ref reaches *6 million* people on social media.

We will bring #1Lib1Ref to Wikidata and more languages, spreading the message that references improve quality.

Youtube & Facebook are leveraging Wikipedia to help users assess sources and subjects.

We will help the public understand Wikipedia's mechanisms for reliability and why & how they can trust it wisely.

5 levers

Awareness and public literacy

Knowledge Integrity

Research

Outreach and partnerships

Infrastructure and tooling

Access and preservation
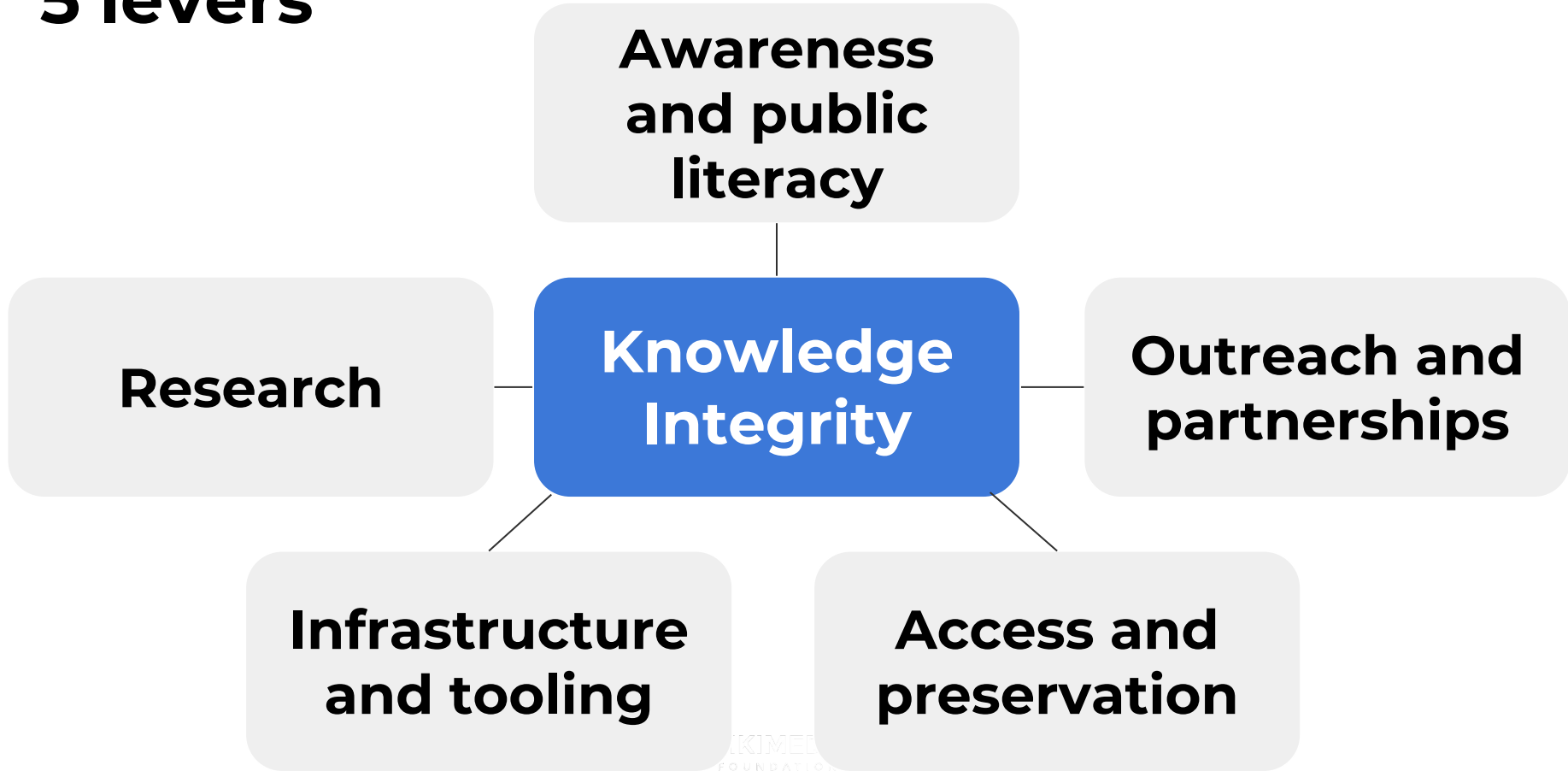
1. Conduct **research** to understand how readers access sources and to help contributors improve citation quality.

CHECK IN

Oct 2018

TEAM/DEPT
Community Programs (CE), Research (Tech), Wikidata (WMDE)

PROGRAM
CDP:KI 1.1 & 1.2

WIKIMEDIA FOUNDATION ANNUAL PLAN

GOAL: Wikimedia contributors are better able to focus and prioritize their sourcing efforts and Product teams can build the best user experiences to support readers' learning goals and their digital literacy.

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
Research to map the "state of verifiability" of Wikimedia projects (unsourced contents, inaccessible sources). Research to understand how readers use citations (quantitatively and qualitatively).	**LAST QUARTER**	
	Research Collaborators (L3S, Stanford, EPFL)	Published datasets on citations with identifiers and their accessibility. Developed classifier to detect statements needing citations with up to 90% accuracy. Designed taxonomy of reasons why editors add citations. Launched two pilots asking community to annotate sentences with reasons why they need a citations. Collected data capturing readers' interactions with citations and footnotes. We did a first analysis of this data. We modified the schema to get better data next round.
	NEXT QUARTER	
	Research Collaborators (L3S, Stanford, EPFL)	Collect more annotations on sentences and implement a classifier to detect the reason why sentences need citations. Write a paper for The Web Conference. Collect data on citation usage for a longer period of time. Analyze collected data.

STATUS: IN PROGRESS

1. **A taxonomy of reasons why Editors add citations to sentences in Wikipedia**

Reasons for adding a citation ✎

- **Quotation**: The statement appears to be a direct quotation or close paraphrase of a source
- **Statistics**: The statement contains statistics or data
- **Controversial**: The statement contains surprising or potentially controversial claims - e.g. a conspiracy theory (see Wikipedia:List_of_controversial_issues for examples)
- **Opinion**: The statement contains claims about a person's subjective opinion or idea about something
- **Private Life**: The statement contains claims about a person's private life - e.g. date of birth, relationship status.
- **Scientific**: The statement contains technical or scientific claims
- **Historical**: The statement contains claims about general or historical facts that are not common knowledge
- **Other**: The statement requires a citation for reasons not listed above (please describe your reason in a sentence or two)

Reasons for not adding a citation ✎

- **Common Knowledge**: The statement only contains common knowledge - e.g. established historical or observable facts
- **Main Section**: The statement is in the lead section and its content is referenced elsewhere in the article
- **Plot**: The statement is about a plot or character of a book/movie that is the main subject of the article
- **Already Cited**: The statement only contains claims that have been referenced elsewhere in the paragraph or article
- **Other**: The statement does not require a citation for reasons not listed above (please describe your reason in a sentence or two)

2. A Machine learning algorithm to detect statements In need of a citation (data from Featured Articles)

Interpretable Models

72-78% accuracy

Deep Learning Models

83-89% accuracy

3. A Dataset of scholarly Publications cited in Wikipedia

Accessibility

Inferred using *Unpaywall* data linking DOIs with their Open Access version

30% of publications are OA at source 10% are paywalled but free copy exists

Topic

Inferred using *Scoring Platform's drafttopic*

Articles about Space are mostly OA (55%) Economics far less Open access (10%)

4. Characterizing Citation Usage

We collected one week of data of readers' interaction with citations and footnotes, and analyzed it

0.8% on average (1 in 125)

30% on Mathematics & Tech articles

Number of times a reader clicks on at least one external citation link when visiting a page in English Wikipedia

4. Characterizing Citation Usage

Top clicked domains in English Wikipedia

Domains (top to bottom):
- news.google.com
- www.forbes.com
- www.dailymail.co.uk
- www.independent.com
- www.imdb.com
- news.bbc.co.uk
- www.washingtonpost.com
- tools.ietf.org
- timesofindia
- www.telegraph.co.uk
- en.m.wikipedia.org
- www.ncbi.nlm.nih.gov
- www.allmusic.com
- www.youtube.com
- www.bbc.co.uk
- www.theguardian.com
- www.nytimes.com
- doi.org
- books.google.com
- web.archive.org

y-axis: *domain*

x-axis: *number of external clicks* — 0, 50000, 100000, 150000, 200000

Top clicked domain: **archive.org**, who featured us in their blog!

5. Why do readers trust Wikipedia?

We developed a survey to discover…

● How much do readers trust Wikipedia?

● How much do they trust *this* article?

● Why do they trust it?

We will run this survey across several languages to learn what features of Wikipedia articles promote trust **and why**

2. Improve **tools** for linking information to external sources, catalogs, and repositories.

WIKIMEDIA
FOUNDATION

CHECK IN

Oct 2018

TEAM/DEPT
Community Programs (CE), Research (Tech), Wikidata (WMDE)

PROGRAM
CDP:KI 2.2, 2.3, 2.5, 2.11

WIKIMEDIA FOUNDATION ANNUAL PLAN

GOAL: Contributors, tool developers and partner organizations can understand and accelerate the referencing and linking of knowledge statements to external sources, catalogs, metadata providers, and content repositories.

What is your objective?	Who are you working with?	What impact / deliverables are you expecting?
Real-time event stream tracking changes to external links and references.

Improve tools to identify and fill unsourced statements such as with Citation Hunt for #1lib1ref.

Study how to integrate, and integrate, Citoid in Wikidata | **LAST QUARTER** | |
| | Community Programs, Research, Wikidata team | Connected WMF engineering with WMDE to study Citoid integration into Wikidata. |
| | **NEXT QUARTER** | |
| | Community Programs, Research, Wikidata team | Requirements gathering and architecture analysis for the EventStream.

Integrate Worklist tool with Citation Hunt for Q3 #1lib1ref campaign. Identify participating Citoid engineers at WMDE. Present prototype Citoid gadget at Wikicite. Define work flows and begin Citoid development. |

STATUS: IN PROGRESS

185 million Wikidata statements have no external reference

138 million statements are entirely unreferenced
47 million statements are sourced only to Wikipedia

3. Ensure resources cited across Wikimedia projects are **accessible in perpetuity**.

WIKIMEDIA

FOUNDATION

CHECK IN

Oct 2018

TEAM/DEPT
Community Programs (CE), Research (Tech), Wikidata (WMDE)

PROGRAM
CDP:KI 3.7

WIKIMEDIA
FOUNDATION
ANNUAL PLAN

GOAL: Resources cited across Wikimedia projects are accessible to readers in perpetuity, thanks to technical partnerships securing their preservation and digitization.

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Internet Archive will immediately cache all resources linked from Wikimedia projects, also prioritize efforts to digitize sources cited in Wikimedia.	**LAST QUARTER**	
	Research, Community Programs, Internet Archive	Work beginning in Q3.
	NEXT QUARTER	
	Research, Community Programs, Internet Archive	Work beginning in Q3.

STATUS: IN PROGRESS

4. Grow outreach and partnerships to scale community and technical efforts to improve the structure and quality of citations.

CHECK IN

Oct 2018

TEAM/DEPT
Community Programs (CE), Research (Tech), Wikidata (WMDE)

PROGRAM
CDP:KI 4.6, 4.8, 4.9

GOAL: More knowledge professionals and other contributors are motivated to join the effort to build an open citation ecosystem, and are more able to actively improve the structure, quantity, and quality of citations on Wikimedia projects.

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Fundraise for WikiCite. Host WikiCite and support satellite events. Run 1Lib1Ref to include references for statements on Wikidata. Run #OAbot campaign.	**LAST QUARTER**	
	Research, Community Programs, Advancement	3 years funding raised for WikiCite from Sloan. WikiCite 2018 received a record number of applicants (240), and allocated 60K USD in scholarships. Wikicite Roadmap and strategic directions posted to the community. 6 satellite sessions in Q1.
	NEXT QUARTER	
	Research, Community Programs	Host WikiCite 2018. Document and report outcomes from the event. Run #OAbot campaign during October OpenAccess week. Create resource guide for Wikidata #1lib1ref participation.

STATUS: IN PROGRESS

{ } wikicite

Berkeley, 27-29 November 2018

Bay Bridge by Tehani Schroeder • flic.kr/p/oFz47p • **CC BY**

5. Increase public awareness of the processes Wikimedians follow to verify information and articulate a collective vision for a trustable web.

CHECK IN

Oct 2018

TEAM/DEPT

Community Programs (CE), Research (Tech), Wikidata (WMDE)

PROGRAM

CDP:KI 5.10

WIKIMEDIA FOUNDATION ANNUAL PLAN

GOAL: The public has increased awareness and understanding of the processes Wikimedians follow to verify and fact-check information, and on the benefits of open, auditable, linked information ecosystems.

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Create an audience map of our ecosystem and develop communications strategies for each audience.	**LAST QUARTER**	
	Research, Community Programs	Posted our annual plan, position paper, program announcement, and meta page. Wrote two popular blog posts about citations on Wikipedia which received major media coverage. Internet Archive featured the growth of dead-link fixing on their blog, which was at the top of Hacker News.
	NEXT QUARTER	
	Research, Community Programs	Write another blog post on citation research. Promote Wikicite event on social media. Use Foundation blog for KI introductory interview post.

STATUS: IN PROGRESS