

# Product Discovery

April 2017 quarterly check-in



**WIKIMEDIA**  
FOUNDATION

All content is © Wikimedia Foundation & available under [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/), unless noted otherwise.

**The Wikimedia projects  
have a wealth of knowledge and  
content that is *not* easily discoverable  
by our readers.**

**The Discovery Department's mission  
is to make it discoverable.**

# How do we make the knowledge discoverable?

- **Search relevance improvements:** help users find what they're looking for
- **Cross-project searching:** expose content from sister projects to Wikipedia users
- **Cross-language searching:** let our users know we have content in their language
- **Search interface improvements:** rich interface to showcase our search results
- **Wikidata Query Service:** explore the structured data in Wikidata by querying it
- **Wikipedia Portal:** the jumping-off point for open knowledge in Wikimedia
- **Maps:** interact with the content in Wikimedia projects spatially and temporally

# Team information and KPIs

Team size

14 full-time employees, 1 contractor

Time spent

Strengthen: 40% Focus: 30% Experiment: 30%

User engagement

Start of Q3: 36% End of Q3: 36% YoY: **+2.86%**

Zero results rate

Start of Q3: 22% End of Q3: 22% YoY: **-24.1%**

**WIKIMEDIA**  
FOUNDATION

# Making search results more relevant: machine learning



WIKIMEDIA  
FOUNDATION

[Learning to rank](#)

wieżyczka	wieżycz ..	15	7	[1 wieżowy][1 wieżowa][1 wieżyczce][1 wieżyczek][5 wieżyczka][3 wieżyczkami][2 wieżyczki][2 wieżyczką][1 wieżyczkę]
wieżyc	wież ..	176	4	[6 Wieży][2 Wieża][20 wież][90 wieży][20 wieżą][38 wieżę]
wig	wig ..	16	3	[9 WIG][1 Wig][1 Wigami][3 Wigów][2 wigów]
wigierski	wigierski ..	2	2	[1 Wigierskiej][1 wigierski]
wigilia	wigili ..	10	3	[4 Wigilia][1 wigilia][2 wigilii][3 wigilię]
wigilijny	wigilijn ..	3	2	[2 wigilijna][1 wigilijną]
wik	wik ..	16	3	[2 Wik][3 Wika][9 Wiki][2 wiki]
wikariat	wikaria ..	22	5	[1 Wikariacie][2 Wikariat][6 wikariat][3 wikariatem][7 wikariatu][3 wikariatów]
wikariusz	wikariusz ..	84	6	[2 Wikariusz][1 Wikariusze][32 wikariusz][21 wikariusza][2 wikariusze][20 wikariuszem][2 wikariuszowi][4 wikariuszy]
wiking	wiking ..	6	2	[3 Wiking][3 wikingów]
wikipedia	wikipedi ..	10	2	[8 Wikipedii][1 wikipedia][1 wikipedii]
wiktor	wiktor ..	126	5	[80 Wiktor][36 Wiktora][7 Wiktorem][1 Wiktorze][2 Wiktorów]
wiktoria	wiktori ..	33	4	[16 Wiktoria][15 Wiktorii][1 Wiktoria][1 Wiktorię]
wiktoriański	wiktoriański ..	10	4	[1 Wiktoriański][1 wiktoriański][1 wiktoriańskich][5 wiktoriańskiej][2 wiktoriańskim]
wiktorowicz	wiktorowicz ..	5	2	[4 Wiktorowicz][1 Wiktorowicza]
wil	wil ..	75	5	[1 WiL][3 Wil][3 Wilka][1 Wilkami][8 Wilki][42 Will][8 wilka][2 wilkami][2 wilki][5 will]
wilan	wilan ..	14	2	[11 Wilanowie][3 Wilanów]
wilanowski	wilanowsk ..	11	4	[1 Wilanowska][6 Wilanowskiej][1 Wilanowskim][1 wilanowski][2 wilanowskiej]
wilanowy	wilanow ..	6	2	[4 Wilanowa][2 Wilanowo]
wilczkowy	wilczkow ..	4	2	[1 Wilczkowa][3 Wilczkowo]
wilczur	wilczur ..	2	2	[1 Wilczur][1 Wilczura]
wilczy	wilcz ..	10	4	[2 Wilcze][2 Wilczego][4 Wilczym][1 wilcze][1 wilczych]

April 2017

**Discovery / Search Back End****Product****GOAL: Maintain key search functions and enhance capacity**

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Improve search by researching and load testing machine learning to automatically re-rank search results	<b>LAST QUARTER</b>	
	Community Engagement, Team Practices Group, and Analytics	Added real-time index updates to completion suggester; upgraded CirrusSearch and Logstash to ElasticSearch 5.
	<b>NEXT QUARTER</b>	
	Community Engagement, Team Practices Group, and Analytics	Improve search by researching and load testing machine learning to automatically rank search results by building out data munging and data pipeline to develop new models.

**STATUS: OBJECTIVE IN PROGRESS**

April 2017

**Discovery / Search Back End****Product****GOAL: Maintain key search functions and enhance capacity**

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Improve support for different languages by researching and deploying new language analyzers	<b>LAST QUARTER</b>	
	Community Engagement, Analytics, and selected Wikipedias	Research and test new language analyzers to find improvements and deploy as needed; deploy recent updates to TextCat.
	<b>NEXT QUARTER</b>	
	Community Engagement, Analytics, and selected Wikipedias	Perform research spikes to find new analyzers for different languages and test them to see if improvements are noted. Then, deploy the new analyzers to the selected Wikipedias.

**STATUS: OBJECTIVE IN PROGRESS**

**What you learned working on this objective or workflow:****Did anything unexpected (good or bad) happen?**

There were unexpected changes to the behavior of Elasticsearch associated with the upgrade to Elasticsearch 5. This created a number of bugs that took extra time to fix, delaying other goals (e.g., real-time completion suggester upgrade and language analyzer deployment) while taking development time away from some goals (e.g., real-time completion suggester upgrade).

The changes to Elasticsearch 5 that enabled the real-time completion suggester upgrade were necessary but not sufficient to enable it. Once it was possible on the Elasticsearch side, we discovered significant additional complexity on the CirrusSearch side.

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

Be more careful about having goals that depend on other large, complex goals with many unknowns (like an Elasticsearch upgrade) in the same quarter.

**Did you struggle with anything?**

The added complexity of the real-time completion suggester re-opened the question of prioritizing the upgrade. Before it looked like an easy change to enable a moderately desirable increase in functionality. Now the development cost has increased significantly, and the delays caused by the Elasticsearch upgrade have pushed it into the next quarter.



# Search results page: sister project snippets



## Birds of a **Feather**

Birds of a **Feather** (1989 - 1998) is a British sitcom, that aired on BBC One, about two sisters who are left to fend for themselves after their husbands

 [More results from Wikiquote](#)

## The Four **Feathers**

The Four **Feathers** (1902) by A. E. W. Mason  
79484The Four **Feathers**A. E. W. Mason1902 The Four **Feathers** To MISS ELSPETH ANGELA CAMPBELL June 19,

 [More results from Wikisource](#)

[See more results from Multimedia](#)

1:39 PM 79%  
sistersearch.wmflabs.org/w/index

### President of Pakistan

The **President** of Pakistan (Urdu: صدر مملکت پاکستان — Šadr-e Mumlikat-e Pākistān, Urdu pronunciation: [səḍˤr-eː ˈmʊm.lɪkət̪-e paːkɪs.t̪ɑːn]), is the  
33 KB (3,271 words) - 16:10, 19 February 2017

#### Results from sister projects

**président**  
also: **president** and **President** præsidēt (obsolete) Latin praesidens. See président, -ent.  
IPA(key): /pʁɛ.zi.dɑ̃/ **président** m (plural **présidents**, feminine  
[\[w\]](#) [More results from Wiktionary](#)

**Pinyin/President**  
Search **president** zongtoong  
[More results from Wikibooks](#)

**President of China** lunches with Brazilian **President**  
Writing an article Hu Jintao, the **President** of the People's Republic of China had lunch today with

[new search results design](#) *and* [in real life](#)

# Explore similar and Wiktionary widget

## Rainbow

For other uses, see **Rainbow** (disambiguation). A **rainbow** is a meteorological phenomenon that is caused by reflection, refraction and dispersion

68 KB (8,429 words) - 19:29, 26 March 2017 • [explore similar](#)▼

### RELATED PAGES



Circumhorizontal arc  
optical phenomeno...



Circumzenithal arc



Atmospheric optics

### CATEGORIES

Atmospheric optical phenomena

Atmospheric sciences  
Heraldic charges

Lucky symbols  
Rainbow

### IN OTHER LANGUAGES:

af:Reënboog

am:ቀበሌ ገጽ

an:Arco de Sant Chuan

ar:قوس قزح

arc:Ⲁⲓⲛⲁⲓⲛⲁⲓ

as:বামধনু

ast:Arcu la vieya

ay:Kurmi

az:Göy qurşağı

ba:Йәйғор

## rainbow

(General American [↗](#)) IPA<sup>(key)</sup>: /ˈreɪnbou/, /ˈreɪmbou/; enPR: rānˈbō, rāmˈbō

(Received Pronunciation [↗](#)) IPA<sup>(key)</sup>: /ˈreɪnbəʊ/  
Audio (US)(file) [🔊](#)

### Noun (English)

A multicoloured arch in the sky, produced by prismatic refraction of light within droplets of rain in the air.

Any prismatic refraction of light showing a spectrum of colours.

A wide assortment; a varied multitude.

An illusion, mirage.

A curveball, particularly a slow one.

In Texas hold 'em or Omaha hold 'em, a flop that contains three different suits.

Rainbow trout.

### Adjective (English)

Multicoloured.

Made up of several races or ethnicities, or (more broadly) of several cultural or ideological factions. LGBT.

Composed entirely of different suits.

### Verb (English)

To pattern with many colours, like a rainbow.

more

Definition from wiktionary.org

[Explore Similar Wiktionary widget](#)



**GOAL: Maintain key search functions and enhance capacity**

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Refine and test new design of search results page with occasional deployments	<b>LAST QUARTER</b>	
	Community Engagement, Analytics and select community Wikipedias: Persian, Polish, Italian, Catalan, Russian, German, Arabic and French	Have a working prototype in <a href="#">wmflabs</a> and run at least 2 A/B tests on selected Wikipedias on the search results page <ul style="list-style-type: none"> <li>Sister project snippets (to be deployed in April)</li> </ul>
	<b>NEXT QUARTER</b>	
	Community Engagement, Analytics and selected community Wikipedias: Persian, Polish, Italian, Catalan, Russian, German, Arabic and French	Test out new functionality with at least 2 A/B tests, analyze results, gather feedback and deploy to all Wikipedias on the search results page <ul style="list-style-type: none"> <li>Explore similar</li> <li>Wiktionary widget</li> </ul>

**STATUS: OBJECTIVE IN PROGRESS**

April 2017

**Discovery / Search Front End**Product**What you learned working on this objective or workflow:****Did anything unexpected (good or bad) happen?**

- We went through various iterations of the sister project snippets in the search results and ended up with a version that more closely aligned with how search results are currently shown.
- When we asked the community for volunteer wikis to test on, we were surprised that Persian was first to ask to be included in the tests!
- We also found that German and Italian wikis ended up with overall less clickthrough when the sister project snippets were shown in testing and we're not sure why.

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

Going through the various iterations might have seemed a bit on the slow side, but we're happy with where we ended up.

**Did you struggle with anything?**

- What is relevant results - it can change with the person that is looking at the search results. Is it better to call them 'interesting' results?
- We also struggled with getting real world users to look at the tests, most of our early testing was done exclusively with A/B tests and guerilla canvassing, until we recently received access to [usertesting.com](https://www.userTesting.com).

# Wikipedia Portal

- Translations updated
- Statistics updated
- Wikipedia apps
- RTL bugs fixed
- Minor improvements to colors/placement (small pixel changes)

Wikipedia Portal in Arabic. The page features a search bar at the top right with 'EN' and a dropdown menu. Below the search bar is a navigation bar with a dropdown menu. The main content area is a grid of 12 icons representing various Wikipedia projects and services. At the bottom, there is a footer with the Creative Commons Attribution-ShareAlike license information.



**Wikipedia apps are now available:**

Download for iOS on the App Store

Download for Android on Google Play

View full list of available Wikipedia apps

April 2017

**Discovery / Portal**Product**GOAL: Improve discovery of Wikimedia projects**

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Maintenance and bug fixes (as needed)	<b>LAST QUARTER</b>	
	Community Engagement	Updated article statistics and newly translated text on a regular basis, added WMF app links, fixed bugs.
	<b>NEXT QUARTER</b>	
	Community Engagement	Update article statistics and newly translated text on a regular basis and fix bugs as needed

**STATUS: OBJECTIVE IN PROGRESS**

April 2017

**Discovery / Portal**Product**What you learned working on this objective or workflow:****Did anything unexpected (good or bad) happen?**

- There were no set goals for the Portal this quarter, but we did complete many tasks, routine updates and bug fixes. A few of the bug fixes corrected the RTL display of various elements of the page. The team learned a lot about multi-lingual display and RTL considerations, all of which are very important to enhancing the readability of the Portal.
- Statistics and text translations were updated on a somewhat regular basis.
- Based off of data analysis, the Wikimedia mobile application links were found to be useful and thus, were kept on the Portal page after a 2 month test.

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

More precautions would have been taken to prevent the page from becoming unreadable due to javascript errors (safeguards were put in place after a bug that exposed this weakness had surfaced).

**Did you struggle with anything?**

Portal deployments were stalled significantly this quarter (almost a month without updates) because of a bug in the deployment script (scap3) as well as some caching issues that caused another nasty bug to surface for several days.



# Analysis

## A Test Of Cross-wiki Search: Helping Users Discover Content On Wikipedia's Sister Projects

Erik Bernhardson *Senior Software Engineer, Wikimedia Foundation*  
Jan Drexwielak *User Experience Engineer, Wikimedia Foundation*  
Dan Garry *Product Manager (Search Related), Wikimedia Foundation*  
Mikhail Popov *Data Analyst, Wikimedia Foundation*  
Deb Tankersley *Product Manager (Analytics, Search Friends), Wikimedia Foundation*

Wikimedia Engineering's Discovery's Search team ran an A/B test from 9 Feb 2017 to 22 Feb 2017 to assess the effectiveness of performing cross-wiki searches and showing Catalan, Italian, Persian, and Polish Wikipedia's users results from sister projects such as Wikisource and Wikiquote. We found that while the overall engagement with the search results was higher for the two test groups compared to the control group, there was not sufficient evidence to definitively say that the additional search results increased user engagement. We suspect that a critical UX design issue – links shown in black, rather than standard blue – and the particular languages this test was deployed on (resulting in a higher zero results rate than seen across other languages) had a negative effect on the results. Our path forward is to fix the existing issues and perform a follow-up test that includes additional languages.

### Introduction

Within the Wikimedia Foundation's Engineering group, the Discovery department's mission is to make the wealth of knowledge and content in the Wikimedia projects (such as Wikipedia) easily discoverable. The Search team is responsible for maintaining and enhancing the search features and APIs for MediaWiki, such as language detection – i.e. if a French Wikipedia visitor searches and gets fewer than 3 results, we check if maybe their query is in another language, and if our language detection determines that the query's language is most likely German (for example), then in addition to results from French Wikipedia, they would also get results from German Wikipedia, if any.

Specifically, the Search team's current goal is to add cross-wiki searching – that is, providing search results from other (also referred to as "sister") Wikimedia projects ("wikis") within the same language. For example, if a work (e.g. a book or poem) on French Wikisource matched the user's query, that user would be shown results from French Wikisource in addition to any results from French Wikipedia.

For the users who received the experimental user experience (UX), each additional wiki's top result was shown as a box in a sidebar with a link to view more results (see Figure 1). There were two groups of users who received the experimental UX and one control group that did not:

- **Control:** This group received the baseline user experience, which only includes the search results from the wiki they are on. To make their experience comparable to the test groups with respect to latency, we performed the search across the additional indices, but did not show the results to the end user.
- **Test (Random):** This group received the experimental user experience, which includes search results from other wikis (if any were returned). The boxes holding the results (one box for each wiki) were ordered randomly.

Source code and data are available on GitHub ([wikimedia-research/Discovery-Search-Test-CrossWikiSidebar](https://github.com/wikimedia-research/Discovery-Search-Test-CrossWikiSidebar))  
Screenshot by Deb Tankersley available on Wikimedia Commons, licensed under CC BY-SA 4.0.

- Finished migration to Analytics' ReportUpdater infrastructure

← Cross-wiki search A/B test report<sup>[1]</sup>

- Wikimedia Blog post on hiring data scientists<sup>[2]</sup>



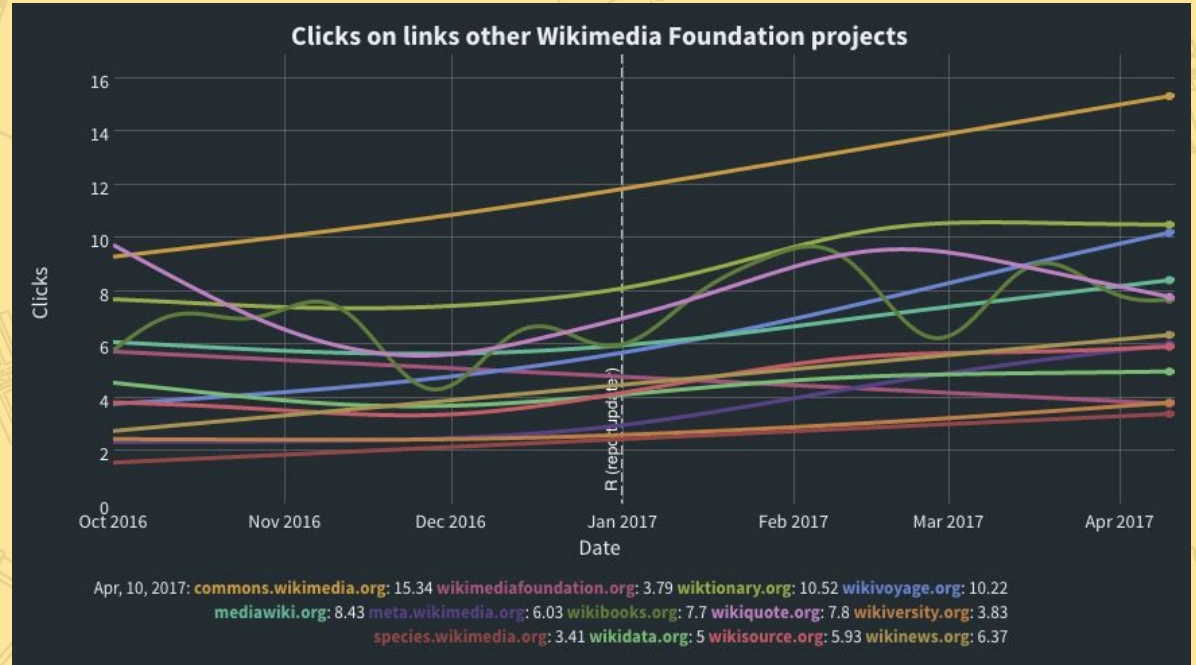
WIKIMEDIA  
FOUNDATION

[1]: [PDF on Commons](#), [web version on GitHub](#)

[2]: <https://blog.wikimedia.org/2017/02/02/hiring-data-scientist/>



# Portal dashboard



New tracking for clicks on Wikipedia mobile app links from the Portal <sup>[1]</sup>  
and clicks from the portal to Wikipedia's sister projects <sup>[2]</sup>

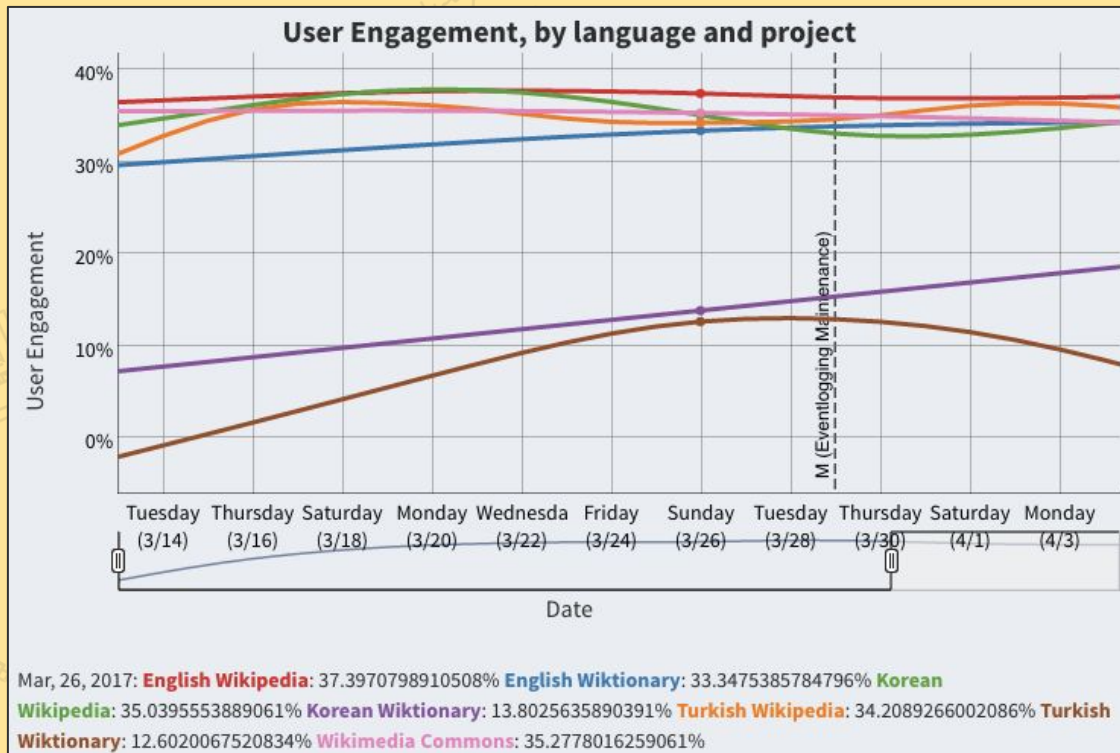


[1]: [http://discovery.wmflabs.org/portal/#app\\_links](http://discovery.wmflabs.org/portal/#app_links)

[2]: [http://discovery.wmflabs.org/portal/#sister\\_projects](http://discovery.wmflabs.org/portal/#sister_projects)

# Search dashboard

Sort projects by	Sort languages by
Volume of requests ▾	Volume of requests ▾
Project	Language
Wikipedia (97.560%)	English (45.395%)
Wikimedia Commons (1.000%)	German (12.327%)
Wiktionary (0.796%)	Spanish (6.917%)
Wikisource (0.101%)	Russian (5.270%)
Wikibooks (0.085%)	French (5.134%)
Wikiquote (0.059%)	Italian (3.970%)
Wikidata (0.047%)	Portuguese (3.580%)
MediaWiki (0.025%)	Japanese (2.929%)
Wikiversity (0.017%)	Arabic (1.902%)
Wikivoyage (0.014%)	(None) (1.353%)
Wikimedia Foundation w	Dutch (1.167%)
Wikinews (0.013%)	Polish (1.053%)
Meta wiki (0.010%)	Chinese (1.019%)
Wikispecies (0.008%)	Swedish (0.947%)
Incubator for new wikis (	Czech (0.798%)
Sources wiki (0.003%)	Persian (0.779%)
Wikimedia Chapter (0.00	Turkish (0.698%)
Wikimania 2017 (0.001%)	Korean (0.551%)
Incubator for Wikiversitie	Hebrew (0.400%)



Search KPIs and other metrics broken down by any valid language and project combination (currently in [beta testing](#))

# Updating map styles



April 2017

**Discovery / Interactive**Product**GOAL: Evolve content discovery and interactive tools on Wikimedia projects**

What is your objective / workflow?	Who are you working with?	What impact / deliverables are you expecting?
Stabilize maps and graphs code base	<b>LAST QUARTER</b>	
	Community Liaison	Continuing work on disputed borders and improving overall map style while fixing high priority bugs. The remaining team members are <a href="#">working on high priority bugs and code stabilization in an official capacity</a> only.
	<b>NEXT QUARTER</b>	
	Community Liaison	Possible release of new map styles while fixing only high priority bugs. Releasing mapframe to all Wikipedias (except those with flagged revisions) is uncertain at this point. The long-term future of the maps project is in discussions with senior management.

**STATUS: OBJECTIVE INCOMPLETE AND DISCONTINUING**

April 2017

**Discovery / Interactive**

Product

**What you learned working on this objective or workflow:****Did anything unexpected (good or bad) happen?**

The interactive team is in the process of being disbanded.

While waiting for budget considerations regarding this team's work to be decided, the remaining members rallied around the shared effort to understand and stabilize the existing map efforts. For example, technical debt was reduced by upgrading to Node 6 to standardise with other services in production. Efforts to stabilise the work for putting it into maintenance mode continue.

**What would you do differently if you had to do it again?****Did you struggle with anything?**

It was difficult to make mid- and long-term plans whilst awaiting a decision from the senior management on the future of the project. The team originally had plans to deploy maps more broadly in order to increase inter-project consistency, but we were requested to put those plans on hold. The first email announcing the Interactive team's work being paused was sent on 20th January to the Discovery mailing list. At the time of this Quarterly Check-in, a decision on the future of this project has not been communicated publicly or to the team affected.



# Discovery Roadmap FY 16-17

Legend

Search backend	Search frontend
Maps	Wikipedia Portal
Wikidata Query	

Q1			Q2			Q3			Q4		
Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Switch to BM25			Enrich data in search index			Upgrade to Elasticsearch 5			Learning to rank		
Plan for cross-project search			Enable cross-project search			Language analyzers			More language analyzers		
Map links on all wikis			Prototype new results page UI			Beta testing new results page			Deploy sister project snippets		
Mapframe on wikis (not WP)			Increase Maps Usage			Mapframe on Wikipedia			New A/B tests for search results page		
Migrate away from GeoHack			Shareable geoshapes			Improve map style			SPARQL federation support		
i18n and new page layout			Add Wikipedia app icons			Access to Wikidata API			Map styles and bug fixes (TBD)		
Unit support			Integrate graphs/maps on-wiki			Full text search functionality					