

Tab Groups Feature Onboarding Rollout

Author(s): Stefan Smagula

Target launch week: As of Feb 18:

expose a small percentage of Firefox users to an experiment in Release 137. Then rollout to N% afterwards.

Rollout Link:

<https://experimenter.services.mozilla.com/nimbus/tab-groups-feature-onboarding-rollout/summary>

QA for the Rollout Request: **TK**

QA Jira Ticket: <https://mozilla-hub.atlassian.net/browse/QA-3236> **Eng Jira Ticket:** **TK**

Data Science Jira Ticket: [DS-3881](#) for the feature metric design

Open Questions

- ☒ ~~Shell / Stefan find the ticket Vanessa was assigned~~
- ☐
- ☐

Getting Started

- ☒ ~~**Summarize your observations**~~ and motivations for the rollout.
 - ☐ What do you need to watch with this rollout? **Success Metric(s)**.
 - ☐ Who should be in this experiment? AKA **Targeting Considerations**
 - ☐ Describe the **Rollout change**
 - ☐ Lay out the Schedule you are considering. Come to data science **office hours** if you have questions: [Desktop](#) is on Wednesday or [Mobile](#) is on Thursday.
 - ☐ Define your **feature for experimentation using the Nimbus API**. After reading docs - for help go to #ask-experimenter
 - ☐ Enter experiment details into **experimenter** (include a link to this brief)
 - ☐ Put the Rollout into **“Preview”**.
 - ☐ File **QA Jira ticket** “Issue Type” = rollout (include link to experimenter)
 - ☐ Look through your Risks section and determine if any Sign-offs are needed
 - ☐ After QA, click **Launch** for your rollout. Request a reviewer in [#ask-experimenter](#)
 - ☐ After Launch, check **Live Monitoring** to verify your rollout is enrolling well.
 - ☐ Experiment owners get email reminders to end enrollment and end the experiment. Request changes in #ask-experimenter
 - ☐ Check your **Rollouts** OpMon (operational monitoring) dashboard link
-

Rollout Details

Summary of Observations and the Problem

Introduce the first iteration of tab organization to Firefox

See Figma prototypes here:

- 📄 Tab Grouping Spec : lo-fi explorations
- 📄 Tab Grouping Spec : visual reference

See Epic: TK

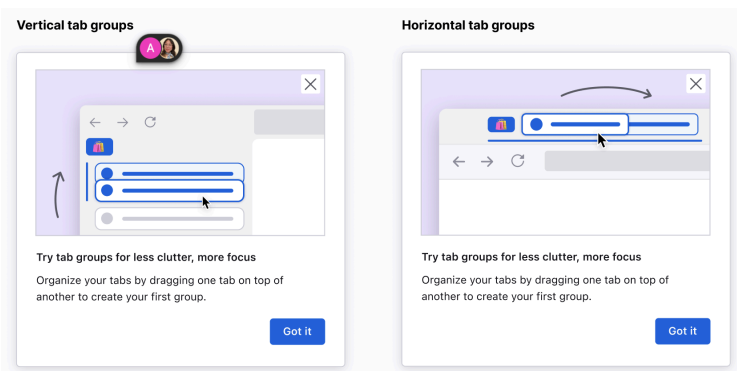
Reference Material

Provide links to prior documents related to your experiment, including dashboards, metrics, studies, and bugs, as well as prior experiments that inform this work.

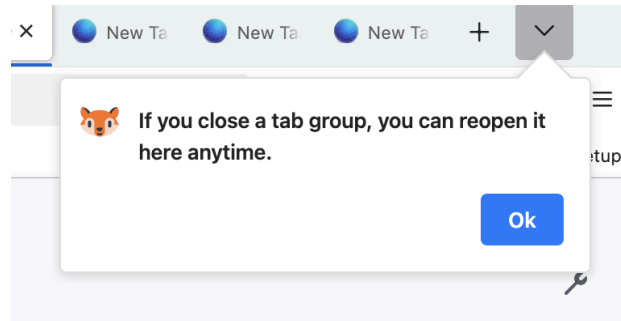
- 📄 **Tab Groups Feature Onboarding**
- 📄 **QA TICKETS tk**

Feature onboarding

- Will improve discoverability among people who feel the pain of organizing / managing tabs
- Will improve retention among people who adopt TG, because they will grasp how to use it, and the purpose and value of the feature better
- Visuals + content: animations in progress
 - Figma file: 📄 Tab Grouping Spec



- Create group / Save and close group callouts:



- Motivation for tab groups, vision for MVP and future: [Vision for tab groups](#)
- [European Area Experiment spreadsheet](#) see for Approval to use Nimbus in EU
-

Success Metrics

Rollouts do not provide an A/B comparison. You will not see a “results page” like you do for experiments, so there are no “stat sig” changes being monitored for any metrics. There is an OpMon dashboard created that

How will you know your rollout is successful?

- Usability testing doesn't point towards unexpected issues with the feature onboarding
- Message engagement / annoyance signals
 - New users - WNP
 - Existing users - onboarding feature callout
- Feature onboarding
 - Users that have 8 or more tabs open
- Feature retention
 - People who create a group with 1 or more tabs - what % return to touch any tab groups
 - ADD PROBES
- How many people discover without onboarding -
 - Is there telemetry for that custom feature callout being shown? Ask engineers...
 -
- Any bugs found get fixed before wider release
- Guardrail metrics?
 - Possibly more engagement - URI's loaded, Active Hours
 - Does tab groups cause a reduction in search activity?
 - What is acceptable based on past experience? Accounts dropped non-monetized (so organic) searches. We held that feature while observing longer.

- At 5 weeks we will check for stat sig drops - if see will keep observing another 3 weeks.

What sources will be monitored? Usability testing results, User comments in Connect

- Often the risks you are hoping to mitigate are brand, load, or stability. So the monitoring plans involve team dashboards for specific metrics, giving comms a heads up when the launch is happening to alert for negative sentiment in channels, looking at the stability/crash dashboards, or working with SRE to monitor server logs/availability.
- The default OpMon dashboard provides absolute numbers for people in your rollout versus everyone else. If you use any targeting for the rollout - then any differences seen are likely between your targeted audience and “all” population.

Who is watching and making those decisions? The Tab Groups team, and Stefan Smagula

What would make you stop the rollout? If we discovered a bug or usability issue that cannot be remedied via a dot release or Beta uplift

What will make you move forward? No extremely negative feedback from early users in Connect and no concerning usability results

Will there be a holdback experiment? YES

What are the possible next steps based on a positive, negative, or neutral result? If there are issues, we'll disable the new bookmarking experience and users will have no feature onboarding until the issue is resolved

Targeting Considerations

Who should get this rollout?

Application	Firefox Desktop ▾ Only Desktop
Channel	Nightly ▾ Stefan: we need to test before full release, so should Nightly be one Rollout request, then Beta, then Release?

	Create Required fields are marked with an asterisk * Project * ☒ Quality Assurance (QA) Issue type * ☒ Request ☒ Request ☐ Embedded QA ☐ Nimbus/Remote delivery
Min Version	Is there a minimum Firefox version needed that code or telemetry for this feature landed? If not, use the current version or current version-1 unless you have a reason to request different. To come as soon as we have firmed this up
Max Version (optional)	
Locales	Is there copy that needs to be shown in a user's language? YES
Countries	ALL
Advanced Experiment Targeting See existing targeting here	Do you need to target specific groups of users beyond the regular targeting options? Examples: New users Only? NO Only users with preference ___ set to ___? Profile Age? Specific Operating Systems Existing users 8th tab close - because when someone opens a tab they are doing an action and in task mode. On closing they are in tab management mode and should be less disruptive / in-context. - 54% open 8 or more tabs each day. - 34% of the user base on a given day have 8 or more tabs open concurrently on the tab bar. - After around 8 tabs people get frustrated per 2021 research New users - just after completed onboarding in the What's new page Nightly callout: Users that create a tab group, or close and save a tab group

Exclude users enrolled in these experiments/rollouts (past or present)	
Require users to be enrolled in these experiments/rollouts (past or present)	
Sticky Enrollment?	Yes ▾ Clients remain enrolled until the rollout ends - even if they no longer meet initial targeting criteria.
Messages Only:	Target of 8+ tabs open concurrently on the tab bar and trigger when they close a tab OR when they restart/update -> callout with animation Trigger: Are there any user actions/conditions required to see the message? YES Timing: If message, when does it show to people? YES

Sizing

people who will have opened 8 or more tabs will get onboarding - that is about 53% of people per session. We are not going to base on 53% though as people may discover on their own.

We will de-risk with a smoke test - so are comfortable sizing up a bit.

We want enough people to see if we regress searches.

Population %	"Expected Number of Clients"	Users that will see the treatment?	Enrollment length	Whole Experiment Duration
Smoke test 0.5% - comfortable number of users that won't break the feature	500,000		Observe 7-14 days looking for bugs	DO NOT END until the rollout has reached 100% - put 100 days in here.

Stat-sig 5%	2.3M	1.2M	1 week	300 Manually start analysis but DO NOT END experiment until the rollout has reached 100% Note; from accounts and bookmark experiments we noted retrieving takes longer to see and often has drops initially in searching. If we see search regressions we may check in later. If there is small regression - could be mitigated by increased long term retention We could manually pull analysis again at 8 weeks if needed

Rollout Schedule

The size you start at depends on the comfort of the risk, ability to measure an impact, and how quickly you want to get it out. Is there a goal date?

Percent for	Observation goals	Deciders to move to next channel?
-------------	-------------------	-----------------------------------

Possibly Firefox Labs	There are 2 prefs - so would be how to use with Nimbus	
Nightly 100% rollout	Usability testing with user research Turn on for nightly users and look for bugs.	Stefan
Beta 100% rollout	Get feedback via Mozilla Connect users	Stefan
Optional: holdback experiment 10%	This would give your guardrail metrics at a low cost (typically holding back - just to look for anything being off. After 7 days of enrolling + 28 days observation.	
Optional: onboarding treatment experiment:	Experimenting in english is very easy - so here is a good window that you can start... If you want to iterate message/images after things are percolating... this is a good time. Could look for early message engagement signal.	
Release - 100% rollout	Release the improved and tested onboarding experience to the wider audience in Release.	Stefan

Rollout Change

Default experience of people not in rollout

What is the experience for users not in the rollout? What is the state of their message, copy, screenshots, or settings.

Rollout

Please describe the rollout experience, including images/screenshots when possible. How can QA verify that the change has happened?

Stefan: in Nightly 23? on Desktop, all users will experience the new feature onboarding. All other channels will continue to experience no onboarding

- Nimbus Feature ID: tk

Risks

Experiments and Rollouts are making remote changes to the experience of live users, often millions of people. Doing QA and answering the risk questions carefully helps reduce the chance of production incidents. These are all “soft sign-offs” - in that you can check them yourself saying you have followed the guidance and are satisfied the risk is mitigated. [Learn More](#)

If the public, users or press, were to discover this experiment and description, do you think it would negatively impact their perception of the brand? Learn more	No ▾
Does this experiment have a risk to negatively impact revenue (e.g. search, Pocket revenue)? Learn more	No ▾
Does this experiment impact or rely on a partner or outside company (e.g. Google, Amazon)? Learn more	No ▾

Keeping a running list as you encounter risks helps you recall what to mitigate.

<i>Risk Description</i>	<i>Impact</i>	<i>Probability</i>	<i>Response</i>	<i>Status Notes</i>
Usability issues	High ▾	M... ▾	Mitig... ▾	Usability testing with prospective and existing users

Results

Did it go smoothly? Did you learn anything interesting? Need to end the rollout early? Need to go slower?