Accept, decline, postpone

An exploration of Article for Creation efficiency

Aaron Halfaker
WMF

Jodi Schneider
INRIA

Bluma Gelley
NYU
Whoami

Name: Aaron Halfaker

Occupation: Research Scientist @ WMF

Contact: halfak in #wikimedia-research
Outline

1. Recap: Wikipedia article creation
2. Problem: AfC → Decreased productivity
3. Results:
   - Pre-publication review is too slow
   - Linus’ Law breaks down (no collaborators)
4. Implications: Draft namespace & Process literacy
How to create an article
How to create an article

1. Go to Wikipedia
How to create an article

1. Go to Wikipedia
2. *Find* an article that doesn’t exist
How to create an article

1. Go to Wikipedia
2. *Find* an article that doesn’t exist
How to create an article

1. Go to Wikipedia
2. *Find* an article that doesn’t exist
How to create an article

1. Go to Wikipedia
2. *Find* an article that doesn’t exist
3. Add some content
How to create an article

1. Go to Wikipedia
2. Find an article that doesn’t exist
3. Add some content
4. Hit “save”
How to create an article

Done!
Workflows

“Direct to Main”

Article
Workflows

“Direct to Main”

Article → NPP → CSD → AFD → Survived
Workflows

“Direct to Main”
Workflows

“Direct to Main”

- Article
  - Survived
  - Deleted
Workflows

“Direct to Main”

draft? → Article

Survived → Deleted
Workflows

Draft space

Direct to Main

User space drafts

Main space

Article

Survived

Deleted

Draft
User space draft

1. **User:EpochFail**/Article that doesn’t exist
User space draft

1. **User:EpochFail**/Article that doesn’t exist

2. *iterate, make mistakes and fix them*
User space draft

1. **User:**EpochFail/Article that doesn’t exist

2. *iterate, make mistakes and fix them*

3. Move to Main when you’re ready
User space draft

1. **User:EpochFail/Article** that doesn’t exist

2. *iterate, make mistakes and fix them*

3. Move to Main when you’re ready

Cool! Right?
**Workflows**

- **Draft space**
  - Direct to Main
  - User space drafts
  - Draft (safe space)

- **Main space**
  - Article
    - Survived
    - Deleted
Workflows

Draft space

Direct to Main

User space drafts

Draft

safe space

Main space

Scrutiny!

Article

Survived

Deleted
Articles for Creation

Never mind. I just did the whole thing manually. --Khoikhoi 05:08, 10 December 2005 (UTC)

Anon page creation restriction [edit]

Jimmy’s had me disable creation of new pages by anonymous users on en.wikipedia as an experiment. Any logged-in user can still create articles, and any anon can still create talk pages. Anons can still edit existing pages, but to create a new article page will require first logging in.

This is one in a series of experiments on cutting vandalism without cutting too much into the ability of people to get things done; with a hojillion articles already, creating new ones is less of a priority than it was two or three years ago, while tuning up existing articles is quite important.

The message shown can be edited at MediaWiki:Nocreatetext and MediaWiki:Nocreatetitle. - - Brion 19:07, 5 December 2005 (UTC)

Interesting. It should probably say “by users who are not logged in” on that page, though - if someone wants to create a page, they can log in and do it.

I don't entirely see the motivation. Someone else said -- and I agree -- that new pages really aren't as much of a problem as vandalism to existing pages. People just get more upset at new pages because they have such a high profile on AID. If page creation could be “reverted” without five days of discussion, like vandalism can be, this kind of option wouldn't be necessary. rspeer 19:49, 5 December 2005 (UTC)

One minor detail: creation of talk pages for referenced-but-as-yet-uncreated articles
This page allows *unregistered* and new users to create new articles with the assistance of experienced Wikipedians.

February, 2007
This page allows unregistered and new users to create new articles with the assistance of experienced Wikipedians.

February, 2007
Workflows

Draft space

Direct to main

User space drafts

Articles for creation

Main space

Article

Survived

Deleted

Draft

Pending

Accepted

Declined

Draft
Workflows

Draft space

- Direct to main
- User space drafts

Articles for creation

- Mentorship

Draft

Pending

Accepted

Declined

Main space

Article

Survived

Deleted
Workflows

Draft space
- Direct to main
- User space drafts

Articles for creation
- Mentorship
  - Draft → Pending
  - Declined
  - Accepted

Main space
- Scrutiny!
  - Survived
  - Deleted
The problem

AfC leads to decreased newcomer productivity
What's productivity?

Assumptions:

- Wikipedia needs new articles
- Articles that survive review are desirable
What’s productivity?

Assumptions:

- Wikipedia needs new articles
- Articles that survive review are desirable
What’s productivity?

Assumptions:
- Wikipedia needs new articles
- Articles that survive review are desirable

Productivity:
- The rate of production of surviving articles
Workflows

Newcomers → AfC → Lower productivity
Workflows

Newcomers $\rightarrow$ AfC $\rightarrow$ Lower productivity
Workflows

Newcomers → AfC → Lower productivity

Newcomer drafts → AfC

Articles for Creation

Direct to Main
Workflows

Newcomers → AfC → Lower productivity

We strongly recommend that you use Articles for Creation.
Workflows

Newcomers → AfC → Lower productivity

Newcomer drafts → AfC

AfC → Lower productivity
Why?

Where’s the breakdown?
Hypotheses

1. AfC’s review process is too slow

2. No collaboration (Linus’ Law)
Methods!
Methods!

AfC uses templates to track status
Methods!

AfC uses templates to track status

{{AFC submission|??}}
Methods!

AfC uses templates to track status

Find AfC drafts:
MySQL Database

Parse templates:
XML Dump & API (deleted pages)

{{AFC submission|??}}
Methods!

Reading draft histories and discussions

A different kind of "Drive" [next]

Per the problems listed above concerning backlog drives and the need to solve some of the other ARC problems first, I suggest a drive whereby ARC members manually go through all the current submissions and categorize them. Perhaps have separate drives for each of the two parent categories proposed below.

Category:ARC submissions by review status. This would allow reviewers to prioritize (if they wish) articles which have never been reviewed. Currently, the queue is hopelessly clogged with self-promotion/paid promotion articles which have been multiply declined and are relentlessly re-submitted. See the [previous example] - Sub-categories:

• Category:Previously declined ARC submissions (for submissions which have been reviewed at least once and declined)
• Category:Unreviewed ARC submissions (for submissions which have never been reviewed)

Category:ARC submissions by subject. The categories below are based on the general categories used by AIB. Once this is done, reviewers could more easily find and choose to review submissions in their area(s) of expertise instead of randomly going through 2500 uncategories submissions whose titles are often not indicative of the type of subject matter. It would also enable members of subject-based WikiProjects to easily scan these categories to see if there are some drafts worth developing/accepting. Note that in some cases more than one category could be added to a draft. An example would be biographical articles which would also have the category for the area in which the person is active/known. Sub-categories:

• Category:ARC submissions (Media and music)
• Category:ARC submissions (Organisation, corporation, or product)
• Category:ARC submissions (Biographical)
• Category:ARC submissions (Biographical)
• Category:ARC submissions (Society topics)
• Category:ARC submissions (Science and technology)
• Category:ARC submissions (Places and transportation)
• Category:ARC submissions (Fiction and the arts)
• Category:ARC submissions (Indusassemble or unclassifiable topic)

An example of a submission in two categories would be Wikipedia talk:Articles for creation/L.J. Sealey (ARC submissions (Biographical) and ARC submissions (Fiction and the arts))
Results

One hypothesis at a time
1. AfC’s process is to slow
1. AfC’s process is to slow
1. AfC’s process is to slow
1. AfC’s process is to slow
1. AfC’s process is to slow

65.5% of reviews happen within 24 hours
1. AfC’s process is to slow
1. AfC’s process is to slow
1. AfC’s process is to slow

93% of reviews are submitted for review within one week of creation.
1. AfC’s process is to slow
1. AfC’s process is to slow

Nope

Qualitative work:
Regular backlog drives keep review time short.
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)

![Graph showing page creations over calendar dates. The graph indicates a significant increase in page creations from 2011 onwards, with a peak in 2013. There are two notable sections: one with approximately 3% AfC and another with approximately 30% AfC.](image-url)
2. No collaborators (Linus’ Law)

Contributions by non-creating editor

- Revisions
- Bytes added
- Unique registered
- Unique anons
2. No collaborators (Linus’ Law)

Contributions by non-creating editor

- Revisions
- Bytes added
- Unique registered
- Unique anons
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)

Supported. AfC Drafts see less collaboration.

Qualitative work:
Onus on the creator alone to make productive changes to drafts.
Summary

1. AfC’s review process is too slow  
   Unsupported

2. Less collaboration (Linus’ Law)  
   Supported
Implications for design of new features
Draft namespace

- Review can be fast (but there’s other problems)
  - Reviewers lack subject expertise
  - Incentive to be restrictive
Draft namespace

- Review can be fast (but there’s other problems)
  - Reviewers lack subject expertise
  - Incentive to be restrictive

Proposal: No publication review. Focus on “productive review”.

Not just review, but contribution.
Draft namespace

- If drafts are hidden, Linus’ Law breaks down
Draft namespace

- If drafts are hidden, Linus’ Law breaks down.

Proposal: Make drafts visible
Draft namespace

- If drafts are hidden, Linus’ Law breaks down.

Proposal: Make drafts visible
Thanks!

AfC Process

Work log: https://meta.wikimedia.org/wiki/Research_talk:AfC_processes_and_productivity