Trends in article creation

Aaron Halfaker

in collaboration with

Jodi Schneider, Bluma Gelley, Subham Soni, Matthew Flaschen, and Steven Walling
/who halfak

Aaron Halfaker
Research Scientist, Wikimedia Foundation

Think big. Measure what you can.

About me
Hi. I'm Aaron Halfaker. I'm a computer scientist. I've been a Wikipedian since 2008. See my publications. In 2011, I started working with the Wikimedia Foundation as a researcher. See my volunteer account, User:EpochFail, for my non-staff work.

My work
I use the scientific method to explore the structure and function of Wikipedia. My goal is to identify problems/opportunities and to use this knowledge to improve/extend the software. See The Rise and Decline for an example of my research and Snuggle for an example of a tool for Wikipedians that I'm actively developing.

Contact me
- E-mail: ahalfaker@wikimedia.org
- Website: http://halfaker.info
- Twitter: http://twitter.com/halfak
- On IRC, I go by "halfak".
You see, one thing is, I can live with doubt and uncertainty and not knowing. I think it's much more interesting to live not knowing than to have answers which might be wrong.

I have approximate answers and possible beliefs and different degrees of certainty about different things, but I'm not absolutely sure of anything and there are many things I don't know anything about, such as whether it means anything to ask why we're here and what the question might mean.

I might think about it a little bit and if I can't figure it out, then I go on to something else, but I don't have to know an answer, I don't feel frightened by not knowing things, by being lost in a mysterious universe without having any purpose, which is the way it is really in so far as I can tell.
Outline

1. Background: Article creation workflow
Outline

1. Background: Article creation workflow
2. Results: Newcomer article creation
Outline

1. Background: Article creation workflow
2. Results: Newcomer article creation
3. Results: Articles for Creation (enwiki)
Outline

1. Background: Article creation workflow
2. Results: Newcomer article creation
3. Results: Articles for Creation (enwiki)
4. Science → New technologies
The Beginning as the End
The Beginning as the End

1. Anonymous editors should be able to create articles.
The Beginning as the End

1. Anonymous editors should be able to create articles.
2. Blocking pre-publication review is expensive.
1. Anonymous editors should be able to create articles.
2. Blocking pre-publication review is expensive.
3. Article drafting space lacks critical technological support.
Part 1. Background
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”

Article

NPP

CSD

AFD

Survived

Deleted
Workflows

“Direct to Main”

Article

Survived

Deleted
Workflows

“Direct to Main”

draft? → Article → Survived

Article → 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**

   My user page

   My article title
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

Main space
- Article
  - Survived
  - Deleted

Draft

safe space
Workflows

Draft space

- Direct to Main
- User space drafts

Draft

Main space

- Scrutiny!
- Article
  - Survived
  - Deleted

safe space
Welcome to Articles for Creation!

Welcome to Articles for Creation! If you don't have a Wikipedia user account, consider registering an account now so that you can create articles yourself. If you choose not to register, or you have a conflict of interest, but have an idea for a new article and some references, you can create one here and it will be reviewed and considered for publication. If you have an idea for the title of an article, but no content for the article itself, please make a request at Wikipedia:Requested articles. If you already have a Wikipedia user account, you can also use the Article Wizard to help you create your article. To nominate an existing draft or user sandbox for review at Articles for Creation, add the code `{{subst:submit}}` to the top of the draft or sandbox page.

Click here to create an article now!

Recently created articles

- Mike Thomas (author) by OliviaOwen (talk · contribs)
- Manor House (Sutton Courtenay) by CoatesinChocolate (talk · contribs)
- Exit Festival by Barrett at C3 (talk · contribs)
Articles for Creation (enwiki)

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 hojilion 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
Articles for Creation (enwiki)

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.

Brion -- December, 2005
Wikipedia:Articles for creation

From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by Uncle G (talk | contribs) at 09:15, 2 January 2006. It may differ significantly from the current revision.

The easiest way to create a new article is to create an account then log in - no personal information is necessary.

If you prefer not to create an account, you can post requests for article creation here. Registered users who see these requests can create articles for you. Remember your requested topic must be suitable for Wikipedia.

Please read Wikipedia:Your first article before making a proposal - it will help you avoid the classic "newbie" mistakes, and save your hard work being deleted.

Template:AFC instructions

You can get more information on using Wikipedia at the Welcome page.

Recently created articles:

• Created on December 31: Bergwind, Léon-Bernard Glot
• Created on December 30: Hsu Shu-Cheng, Huvaten Fushi
• Created on December 28: Richard Bradford (actor)
• Created on December 24: Richard E. Fleming
• December 15 creations: Miller effect
The easiest way to create a new article is to create an account then log in - no personal information is necessary.

If you prefer not to create an account, you can post requests for article creation here.

January, 2006
Amending our Terms of Use:
Please comment on a proposed amendment regarding undisclosed paid editing.

Wikipedia:Articles for creation

From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by Masly (talk | contribs) at 00:38, 1 February 2007. It may differ significantly from the current revision.

(diff) - Previous revision | Latest revision (diff) | Newer revision → (diff)

This page has a backlog that requires the attention of willing editors. Please remove this notice if and when the backlog is cleared.

This page allows unregistered and new users to create new articles with the assistance of experienced Wikipedians. The easiest way to create a new article, however, is to create a free Wikipedia user account and then log in. No personal information is necessary to register, and registered users can create new articles without the hassle or delay of the Article for Creation process.

If you have an idea for a good article subject, but do not want to provide text for the article, you should make an article request instead. Alternatively, if you want to discuss and plan an article go to the drawing board.

Most articles are declined... don't let this happen to you!

The majority of submitted articles are declined, because they are not suitable for Wikipedia. Read Wikipedia:Your first article before writing your article, to learn how to write it well and read below to avoid some common mistakes other people make.

Unregistered users: Submitting an article

Follow these directions to submit an article for creation.

Registered users: Reviewing submissions

Registered Wikipedians should follow these directions to review submitted text and create new 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
Workflows

Draft space

- Direct to main
- User space drafts
- Articles for creation
  - Draft
  - Pending
  - Mentorship

Main space

- Article
  - Survived
  - Deleted
  - Draft
  - Accepted
  - Declined
Workflows

Draft space
- Direct to main
- User space drafts

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

Main space
- Scrutiny!
  - Article
    - Survived
    - Deleted
Workflows

- Direct to Main
- User space draft -- Safe space
- Article for Creation -- Formalized mentorship
Part 2. Article creation research
Questions

1. Who’s creating all the articles?
2. Whose articles survive?
3. What effect does the workflow have?
Methods!
“surviving articles”
“surviving articles”
“surviving articles”

“editor classes”

time

registration

-day

day-week
“surviving articles”

“editor classes”

Time

Registration

day

day-week

week-month
“surviving articles”

“editor classes”

time

registration

-day

day-week

week-month

month-

week-

day-

hour

minute

Lifetime (edit diff, log scaled)
“surviving articles”

“editor classes”

- day
- week
- month
“surviving articles”

“editor classes”
Autocreated?
Autocreated?

English Wikipedia

“new user”
Autocreated?

English Wikipedia → German Wikipedia

“new user” → “autocreated”
Autocreated?

English Wikipedia → German Wikipedia → Commons

“new user” → “autocreated” → “autocreated”
Workflows examined
Workflows examined

<table>
<thead>
<tr>
<th>Direct to Main</th>
<th>User space</th>
<th>AfC</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

English only
Q1. Who is creating all the articles?
Q1. Who is creating all the articles?
Q1. Who is creating all the articles?

1. 55-85% Experienced Wikipedians

* non-enwiki
Q1. Who is creating all the articles?

1. 55-85% Experienced Wikipedians
2. 10-20% Anons*

* non-enwiki
Q1. Who is creating all the articles?

1. 55-85% Experienced Wikipedians
2. 10-20% Anons*
3. 5-15% Newcomers

* non-enwiki
Q2. Whose articles survive?
Wat
RFC Comments:

Sticking your finger into an electric shredder is not a sensible experiment, we know exactly what the result will be and confirming the prediction will be painful. [...] And you think the deletion backlogs are bad now? Just wait until IP creation is open. It's like opening big can of worms.
Q2. Whose articles survive?
Q2. Whose articles survive?

- More experience → More survival
Q2. Whose articles survive?

- More experience → More survival
- Anons > Newest newcomers
Q2. Whose articles survive?

- More experience → More survival
- Anons > Newest newcomers
- All Japanese editors (?)
Q3. What effect does the workflow have?
Q3. What effect does the workflow have?

<table>
<thead>
<tr>
<th>Direct to Main</th>
<th>User space</th>
<th>AfC</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

English only
German Wikipedia
Hypothesis: Drafts reserved for articles that might not make it. (???)
English Wikipedia
<table>
<thead>
<tr>
<th>Original namespace</th>
<th>English Wikipedia</th>
<th>User draft</th>
<th>AfC draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Direct to Main</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Direct to Main < User space < AFC

???
Q3. What effect does the workflow have?
Q3. What effect does the workflow have?

- Drafts → Survival*
Q3. What effect does the workflow have?

- Drafts → Survival*
  - Especially for newcomers
Q3. What effect does the workflow have?

- Drafts $\rightarrow$ Survival*
  - Especially for newcomers
  - German is weird (?)
Q3. What effect does the workflow have?

- Drafts → Survival*
  - Especially for newcomers
  - German is weird (???)
- AfC trades productivity for quality
Q3. What effect does the workflow have?

- Drafts → Survival*
  - Especially for newcomers
  - German is weird (???)
- AfC trades productivity for quality
  - Higher survival rate (~95%)
Q3. What effect does the workflow have?

- **Drafts → Survival**
  - Especially for newcomers
  - German is weird (???)

- **AfC trades productivity for quality**
  - Higher survival rate (~95%)
  - Lower production (~50% loss for newbies)
But why?

Shouldn’t working with mentors produce more surviving articles?
Part 3. Articles for Creation
Workflows

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

![Histogram showing page creations over time](image)
Workflows

Newcomers → AfC → Lower productivity

![Graph showing trend of newcomer drafts to AfC over calendar dates from 2009 to 2014]

![Screenshot of the creation process for an article on a wiki, including options for introduction, subject, notability, sources, content, and end. The process involves creating a draft, which will be reviewed by volunteers and moved to the article namespace if successful. It also includes a note about using Articles for Creation instead of directly creating articles in the article namespace, with guidelines and best practices.]
Workflows

Newcomers → AfC → Lower productivity
Workflows

Newcomers → AfC → Lower productivity

We strongly recommend that you use Articles for Creation.
Workflows

Newcomers → AfC → Lower productivity

![Newcomer drafts → AfC](chart1)

![AfC → Lower productivity](chart2)
Why?

Where’s the breakdown?
Hypotheses

1. AfC’s review process is too slow
Hypotheses

1. AfC’s review process is too slow

2. AfC drafts are hidden from potential collaborators (see Linus’ Law)
Linus’s Law?
Linus’s Law?

"given enough eyeballs, all bugs are shallow"

Given that enough people see an incomplete article, all potential contributions to that article will be easy for someone.
Linus’s Law?

visibility is critical to open collaboration
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/sad promotion articles which have been multiply declined and are relentlessly re-submitted. See the numerous example 6. 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 plowing through 2500 uncategorized submissions whose titles are often no indication of the 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, organisation, or product)
- Category: ARC submissions (Biographical)
- Category: ARC submissions (Biographical)
- Category: ARC submissions (Science and technology)
- Category: ARC submissions (Places and transportation)
- Category: ARC submissions (Fiction and the arts)
- Category: ARC submissions (Indigensible or unsaleable 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

- Draft
- Pending
- Declined
- Accepted

![Graph showing time between creation and submission](image)
1. AfC’s process is to slow

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

- 93% of reviewed drafts are submitted for review within one week of creation
- 29% of drafts are never even submitted
1. AfC’s process is to slow
1. AfC’s process is to slow

Nope
1. AfC’s process is to slow

Nope

Regular backlog drives keep review time short.
1. AfC’s process is to slow

Nope

Regular backlog drives keep review time short.

Process breakdown -- drafts not submitted for review
2. No collaborators (Linus’ Law)
2. No collaborators (Linus’ Law)

Pre-AfC
~3%

Post-AfC
~30%
2. No collaborators (Linus’ Law)

Activity by non-creating editor

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

Pre-AfC
~3%

Post-AfC
~30%

Activity 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)
2. No collaborators (Linus’ Law)

**Supported.** AfC Drafts see less collaboration.

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

1. AfC’s review process is too slow **Unsupported**, but process breakdown (Templates too hard?)

2. Less collaboration (Linus’ Law) **Supported**
Part 4.
Science → New technologies
Anons are better than we thought
Anons are better than we thought

- Revisit anon article creation?
Anons are better than we thought

- Revisit anon article creation?
- Draft namespace (enwiki)
  - Anons can create articles, but can’t publish.
Anons are better than we thought

- Revisit anon article creation?
- Draft namespace (enwiki)
  - Anons can create articles, but can’t publish.
- Need to understand anons better
  - Our intuitions are failing us
  - Anons != vandals
  - Anons > newest newbies
AfC trades productivity for quality
Wikipedia:BOLD, revert, discuss cycle

The Wikimedia Foundation's 2014 steward election has started. Please vote.

While this essay is not a policy or guideline itself, it is intended to supplement the Wikipedia:Consensus and Wikipedia:Be bold pages, which should be deferred to in case of inconsistency between that page and this one.

This page in a nutshell: Making bold edits is encouraged, as it will result in either improving an article, or stimulating discussion. Therefore, if your edit gets reverted, do not revert again. Instead, use the opportunity to begin a discussion with the interested parties to establish consensus.

The BOLD, revert, discuss cycle (BRD) is a proactive method of reaching consensus. It can sometimes be useful for identifying objections, keeping discussion moving forward and help towards breaking deadlocks. Care and diplomacy should be exercised. Some editors will see any reversion as a challenge, so be considerate and patient.
1. **Boldly** do what you think is right
2. If someone disagrees, they will **revert**
3. Have a **discussion** about what to do
If you let people do what they want,
If you let people do what they want, they will tend to do the right thing.
Trade productivity for quality

- Insight of WP:BRD
Trade productivity for quality

- Insight of WP:BRD
  - Most people will do the right thing if you let them
Trade productivity for quality

- Insight of WP:BRD
  - Most people will do the right thing if you let them

- Workflow designers/developers:
  - Drafts need **visibility** for collaboration
  - Don’t block publication
Thanks

Wikipedia article creation


Schneider, J., Gelley, B., Halfaker, A. Accept, decline, postpone: How newcomer productivity is English Wikipedia is reduced by pre-publication review. (2014) OpenSym.
You went too far.
GO BACK!
NO!