



Toolforge webservices are in the final stages of [migrating to the toolforge.org domain](#) .  
Please help us clean up older documentation referring to tools.wmflabs.org!

# Incident documentation/20150406-Flow

[< Incident documentation](#)

## Contents [\[hide\]](#)

- [1 Summary](#)
- [2 Timeline](#)
- [3 Conclusions](#)
- [4 Actionables](#)

## Summary

We added a new field to track content length so we could report accurate byte count changes in diffs. After the field was added and being populated, we had to run a maintenance script, `FlowUpdateRevisionContentLength.php`, to populate old rows.

Flow's `RevisionStorage` class had a bug that caused `update()` to store a NULL blob to External Storage, then update `rev_content` (the content field), to point to that NULL. It first calculated a column diff. After that, there was logic to check whether an External Storage URL was present. Since it was not, it wrongly inserted a NULL entry (since `rev_content` had been stripped in the column diff phase because it was unchanged). It then updated `rev_content` to the URL (that pointed to NULL), along with a related field (`rev_flags`).

There was also validation to prevent `rev_content` from being changed for existing revisions. However, this only ran in the column diff phase, so it did not validate the changes made by Flow's External Store methods.

As a result, all content data before February was set to point to a NULL entry in External Store (data where the length was already correct was not affected, because the data code exited earlier if there were no changes). However, the correct data still existed orphaned in External Store.

## Timeline

- 2015-04-06 - Erik Bernhardson started running `FlowUpdateRevisionContentLength.php`. This continued until sometime on 2015-04-07 - <https://phabricator.wikimedia.org/T90443> [↗](#).
- 2015-04-09 - Matt Flaschen reported a bug where content was intermittently not displaying properly on `office.wikimedia.org`.
- 2015-04-10 - While investigating the bug, Erik determined that the data had been changed point to the NULL External Store rows. The team discussed the issue, began fixing it, and notified the communities.
- 2015-04-10 - 2015-04-14 - Erik wrote a script that reverse-engineered the Flow External Store rows by comparing the length (after decoding to wikitext) and accounting for which were already claimed by standard wikitext revisions. After finding the orphaned rows, the script updated Flow's database to point to them again, as well as fixing the `rev_flags` column. This was reviewed, tested, and merged.
- 2015-04-14 - We ran the script and recovered the vast majority of data. Most exceptions are test posts.

## Conclusions

This bug was not caught earlier because existing revision content is not updated in the normal user flow (new revisions are created). This scenario was not covered by our unit tests, and it was missed when testing the `FlowUpdateRevisionContentLength.php` script, because it only triggered when External Store, an optional feature, was in use.

It also did not affect past maintenance scripts because those used direct database access (`dbw`), whereas this was the only one that uses the Flow data layer (`put()`).

This incident revealed that our testing, both automated and manual, were insufficient, and that we lacked adequate backup coverage.

[Main page](#)  
[Recent changes](#)  
[Server admin log \(Prod\)](#)  
[Server admin log \(RelEng\)](#)  
[Deployments](#)  
[SRE/Operations Help](#)  
[Incident status](#)

[Cloud VPS & Toolforge](#)

[Cloud VPS documentation](#)

[Toolforge documentation](#)

[Request Cloud VPS project](#)

[Server admin log \(Cloud VPS\)](#)

[Tools](#)

[What links here](#)

[Related changes](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Cite this page](#)

[Print/export](#)

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

## Actionables

- Flow data missing on Wikimedia production wikis - <https://phabricator.wikimedia.org/T95580> - Done
- Recover Flow data on officewiki - <https://phabricator.wikimedia.org/T96951> - Done
- Improvements to Flow backups - <https://phabricator.wikimedia.org/T95835>
- Fix the External Store code to never insert NULL blobs - <https://gerrit.wikimedia.org/r/#/c/203479/> - Done
- Fix and test the update() method - <https://phabricator.wikimedia.org/T95869> - Done
- Set up appropriate testing environments for External Store
  - Beta Cluster - <https://phabricator.wikimedia.org/T95871>
  - MediaWiki-Vagrant - <https://phabricator.wikimedia.org/T95870>
- Require running all maintenance scripts (read or write) on a public wiki (such the Beta Cluster) with as close to production's configuration as possible (including External Store), and checking the results, before running in production.
- Publicly released dumps. Though not a backup, it is related and we are already planning on doing this for other reasons - <https://phabricator.wikimedia.org/T89398>

Category: [Incident documentation](#)

This page was last edited on 24 April 2015, at 02:11.

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. See [Terms of Use](#) for details.

[Privacy policy](#) [About](#)  
[Wikitech](#)

[Disclaimers](#) [Code of Conduct](#) [Developers](#) [Statistics](#) [Cookie statement](#) [Mobile view](#)

