



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/20131104-Parsoid

[< Incident documentation](#)

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

- [1 Summary](#)
- [2 What happened](#)
- [3 Timeline](#)
- [4 Recommendations](#)
- [5 See also](#)

## Summary

The Parsoid cluster corrupted the charset in wikitext converted from HTML for about 55 minutes after a broken Parsoid deployment.

## What happened

After several delays due to npmjs.org server issues and broken file permissions on tin, we deployed Parsoid version d7b556f25353 at 19:45 UTC. This version included a host of changes including the new [busboy](#) form data parser library. We switched to this library as the bodyParser library we used so far is deprecated in the connect framework and suffers from [a potential DOS vector with temporary file creation on file upload requests](#).

At 20:04 UTC, there was a first IRC reports about VisualEditor edit corruption on the French Wikipedia. At 20:14, we became aware of this. The issue looked a lot like charset corruption. Due to some issues with git-deploy revert this took a while to revert. At 20:40 UTC, Parsoid was reverted to the pre-deploy version and the edit corruption stopped. At 20:47, both the code and the dependencies were fully reverted and Parsoid was back up.

The issue was caused by the busboy library defaulting to ISO-8859-1 encoding for incoming POST data rather than UTF8, which corrupted incoming HTML to Wikitext serialization requests from VisualEditor. This affected the full posted HTML if it included non-ASCII chars, so corruption was not limited to edited parts of the page.

## Timeline

All times in UTC

- 19:45 Parsoid d7b556f25353 deployed, no issues in logs
- 20:04 First error reports in visualeditor channel
- 20:14 Parsoid developers informed
- 20:30 First revert attempts fail
- 20:40 Parsoid cluster down after successful config repository revert
- 20:47 Parsoid cluster back up with old version after the code is reverted too

## Recommendations

- create HTTP-based Parsoid API tests in addition to parser tests and [round-trip tests](#) so that the full stack is tested. Tracked in [bug 56590](#). **✓ Done**
- Manually test VisualEditor on non-English wikis before and after deployment. Added to [the deployment instructions](#). **✓ Done**
- allow Parsoid deployers to quickly stop Parsoid machines to avoid further corruption in case of emergencies.
- consider sanity checks on wikitext diff vs. HTML diff size in VE to flag corruptions at any point along the way including the VisualEditor PHP extension
- maybe only list non-active commits in the `git deploy revert` dialog and default to the previous deploy to reduce the potential for mistakes

[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](#)

## See also

- Corresponding bug: <https://bugzilla.wikimedia.org/56583>   **Done**

Category: [Incident documentation](#)

This page was last edited on 7 February 2014, at 19:28.

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

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

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

[Wikitech](#)

