

BitTorrents for Science

Egon Willighagen 

Published April 18, 2010

Citation

Willighagen, E. (n.d.). In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/0z9nh-6ed21>

Keywords

Opendata

Copyright

Copyright © Egon Willighagen 2010. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The idea has been lingering in the air for a long time now: sharing large science data sets using [bittorrent](#). Over the past couple of years I have seen a lot of science related software being distributed over torrents, and the use in open source in general is abundant. Given a good network of so-called seeders, download times go down dramatically, and the overall energy consumption goes down too, as data has to follow a much shorter path.

It could very well be that the uptake of this technology for sharing data is only now coming about because only recently we started caring about Open Data licenses, which formally take care of rights of redistribution, which is obviously crucial to setting up a torrent network. Initiatives like the [Panton Principles](#) are changing this, even though we had a good deal of Open Source-licensed data for many years already.

So, with an increasing amount of Open Data, the time was now right, according to the authors [Morgan](#) and [Jonathan](#), to set up BioTorrents, and publish a paper in PLoS ONE: *BioTorrents: A File Sharing Service for Scientific Data* (doi:[10.1371/journal.pone.0010071](https://doi.org/10.1371/journal.pone.0010071)). I have to admit, that I do not particularly like the design of the website, and I think it could do with more social web integration, but importantly, they provide a *tracker*. Trackers are key parts (well, they are being made obsolete, though I am not up-to-date with the state of that evolution), and work as a service discovery hub. Additionally, the website gives means to find data, and allow categorising torrents.

It is worth nothing that the uptake has been minimal so far, since the [idea was posted last October](#). But it is slowly being picked up, or at least [blogged about](#).

How to make BioTorrent work?

The success of BioTorrent will very much depend on the user base. This is common to social web applications, and a recent accidental loss of torrents is unforgivable; well, personally, I was happy to upload [my torrent](#) once more, but would not have done that if I had many torrents uploaded already. Torrent content is distributed, but the tracker information is not. Backup, backup, backup. Oh, and backup :) It happens to the best of us. Additionally, it is worth realising that the service needs to give something back to the user. Traditionally, I always thought this had the be of actual use, but a recent post by Rich actually suggested that even a [game mechanic](#) may be enough. Indeed, websites like [ChemPedia.com](#) and [Blue Obelisk eXchange](#) implement this by means of personal karma, allowing people to compete in high score lists. Also, APIs to integrate with other tools are crucial, such as personal RSS feeds to allow posting my new torrents to, for example, [FriendFeed](#) and [Identicia](#).

But, the by far most important feature for BioTorrents will be to set up a reliable network of seeders. I already [mentioned this on FriendFeed](#) where I suggested university libraries to get involved. Ideally, every library will act as seeder in torrent networks, so that typically, you will download the data directly from your local library, instead of the other end of the world. For data sets of GB size or larger, this is going to have an important environmental impact, on top of the much higher download speed.