

Who Owns Crisis Data?

[Cross-posted from OpenCrisis.org]

N.B. After investigating the complaint in detail (see [this blogpost](#)), we re-opened the [resource doc](#) containing the dataset discussed below later on 09 Feb.

This is one of several posts going up as a result of OpenCrisis investigations into a complaint about one of its datasets. This post will concentrate on the ownership aspects of that complaint.

Please note that this is currently a 'holding' blogpost, so people affected can be updated on what's happening (sadly, all work on this much-used dataset has ceased until we get this all straightened out) - we'll be adding to it as we have time (we're concentrating more on making sure the dataset isn't breaking its stated objective of "first, do no harm") and as we learn more about this important issue.

Some background: starting late last year, OpenCrisis began assembling a dataset (South Sudan humanitarian data, geodatasets, Twitter, etc) for its own work, that it then shared with another group (lets call them "LovelyPeople") that was collecting information for a journalism group (let's call them 'Z') - this second collection was deemed sensitive by 'Z', so we carefully left their deployment off our list of groups working on this crisis (for the record, we do this a lot, and will continue to do it for anyone who asks us to). OpenCrisis made some of its original data (anything considered sensitive was kept private) available to people working in South Sudan via a publicly-visible but not widely-advertised spreadsheet.

Unfortunately, 5 Twitter names made their way from 'Z's sheet into the new spreadsheet, and 'Z' or one of its representatives (it's unclear which) is now threatening to sue two individuals in OpenCrisis for the reuse of 'their' data. To put this in perspective, the contribution from OpenCrisis to 'Z' was roughly 20+ Facebook addresses, 20 blog addresses, 60 multimedia records, virtually all the local media outlets cited by 'Z', virtually all the Twitter lists listed by 'Z', 50-70 Twitter names and a direct copy (credited) of the OpenCrisis crisis mapping page as it was at the time. This is all made more confusing because a third group, LovelyPeople, were also involved, and the OpenCrisis member concerned (Brendan O'Hanrahan) believed that the work with LovelyPeople was on the basis of mutual benefit, because that was stated when he joined the project.

Just to be clear on the OpenCrisis position on this dataset: 'Z's specific problem appears to be with the list of Twitter users. There are many many Twitter lists containing the data in question now - so much so that OpenCrisis stopped updating their spreadsheet list back in January - and we have no problems with removing any content that we can't prove is our own.

But we don't want to live in a world where data ownership and worrying about being sued is a concern for every mapper trying to improve the world. We might get sued, but this isn't about us. The much more important thing is resolving (or starting to resolve) the issue of data ownership when that data has been generated collectively by multiple individuals and groups.

So, who owns crisis data?

The heart of this problem is ownership of community-generated data. I have much reading and thinking to do before I can start to answer this question, but the use of agreements (even if it's agreement that all data will be shared across the community) appears to be key.

The legal position in the US appears to be clear: *"It is important to remember that even if a database or compilation is arranged with sufficient originality to qualify for copyright protection, the facts and data within that database are still in the public domain. Anyone can take those facts and reuse or republish them, as long as that person arranges them in a new way"* (Uni of Michigan's [exceptions to copyrights page](#)). That's actually a huge relief, because if verified (and IANAL), the constant work that we all do on existing crisis datasets will help us to keep them free to use.

So the issue now appears to be less of a legal one, and more of a moral and ethical one: when is it right to share data between groups, and when is it right to claim ownership?

Data Licences

Although 'ownership' of data for good is anathema to us, there is one reason why it can be good: reducing confusion about who can use what where, via licensing. We often need to say that the data we produce can be used by anyone, and say it legally and publically, and that's what open data licences do. Fortunately, there are some good "you can go use this" licences out there (e.g. [ODbL](#)), but as OSM et al know from [painful experience](#), picking the right data licence to be compatible with other people's data gathering and use can be hard.

Privacy

The privacy of individuals is extremely important in our community. When we were working through another issue raised by 'Z', we considered locking down the spreadsheet to subscribers only - only to realise that that would mean making a list of people (and emails of people) engaged in this work. Which we'd also have to protect. We're still thinking about that one.

Legal protection

We can't stress this enough: if you're running a crisis data group, then seriously consider creating a nonprofit company for it. We hate having hierarchies and official registrations too, but without the protection of being an NJ non-profit, those two individuals (and all that they and their family own)

OverCognition

Journeys through development data.

<http://overcognition.com>

would be at risk instead of this being an organisation-to-organisation thing. We've started an OpenCrisis page, [who owns crisis data](#), for links and discussion on the ownership issue. We'd love contributions of useful links and analysis for it.