RESPONSIBLE DATAFORUM

Responsible Data Reflection Stories 9

A collection of real-life examples of the risks that are faced when using data in advocacy work, along with mitigation strategies to overcome these challenges.

Opening the wrong data

The case of an open data hackathon and the police force.

Context

In early 2015, a group of open data advocates organised a hackathon, in partnership with a number of local government bodies, and a state police force. Hackathon organisers encouraged their state partners to “open” their data–that is, to publish it online in a machine-readable format, under an open license.

Hackathons partnering with government entities is not uncommon–for the government bodies, they provide good opportunities to see their data being used by their ‘target audience’, and for developers. For those attending the hackathon, the opportunity to access previously unavailable data is an added incentive to attend, as it potentially opens up the opportunity to gain new insights about their societies.

The challenge

A police officer from the police department made the decision to open up two of their datasets for those at the hackathon to use; one on mugging data, and another on mobile phone calls records. This second dataset contained a lot of personally-identifiable information; metadata of when calls took place, to whom and from whom, and their location.

Though the police force had been aiming at helping the hackathon organisers to get useful data for their event, the person responsible for this had inadvertently published online a dataset that should not have been opened, putting it available for download on their site. A number of attendees at the hackathon downloaded the data, not all of them realising the implications of the data that they then held.

A group of privacy advocates, who had heard that the hackathon was involving the government bodies, intervened after seeing the data that had been published. By going to the venue of the hackathon, they convinced the police force to take down the dataset, though it still took a few hours to do so.

In addition to the risky content of the data, it turned out that the data had been gathered as part of a terrorism investigation; so, privacy implications for individuals aside, potentially also held security risks for publishing.

Lessons learned

The hackathon organisers resolved to make sure that only data that should be opened should be published as a result of their open data advocacy; but things aren't quite that simple. Given that the police force heard their calls to open data, then published the data on their own site, there was effectively little that they could do after realising that inappropriate data had been made public, except for asking them to take it down.

Clearly, in this case, the person/persons charged with choosing data to open from the police force were unaware of the consequences of their actions, and there is little that a group of hackathon organisers can do about this except for offering more advice where needed.

A better lesson, perhaps, would be to be very clear in initial contact with government entities as to what kinds of datasets they might be looking for, and asking for a first opportunity to 'check' the data prior to publishing. The risk then lies in creating too many barriers for them to overcome before publishing, but on balance, especially with institutions dealing with sensitive data, perhaps this is a preferable outcome.

The police force in question never mentioned the mishap publicly, instead choosing to take the data down from their site without comment. The fact remains, though, that a number of people had downloaded it during the few hours it was online, and if it hadn't been for the watchful eye of privacy advocates in the region, the data may well have stayed online for much longer, with unforeseen consequences.

This publication is part series found at https://responsibledata.io, produced by the engine room's Responsible Data Program, 2016.