

Data Hunters and Data Publishing on a Shoestring

Submitted on 14 Jun 2011 by

This is my first blog post here so allow me to introduce myself. I'm the project manager of the Danish open data initiative which we call 'Open Data Innovation Strategy' (ODIS). I work at the Danish National IT & Telecom Agency in a division which deals with emerging technologies and trends and their application in public (and to some extent private) sector digitisation. This includes, among other things, cloud computing, open source software, open standards, Gov 2.0 and open data.

We launched ODIS in early 2009 and a very brief summary of our activities goes something like this: We've built a data catalogue, we've run a data contest, we've held a data conference, we've published easy-to-read guidelines and we've hosted a data camp.

In 2010 we established a small unit, or corps, of 'Data Hunters' to strengthen the ODIS initiative. The Data Hunters have three main tasks which all serve to open up government data for re-use: They respond to data re-users' requests for help to locate and get access to data, they go 'hunting for data' on their own initiative, and they help public bodies wanting to open up their data. You can think of them as the open data boy scouts.

The Data Hunters have their own community forum on our platform for collaboration, Digitaliser.dk, where data re-users can write to the Data Hunters for assistance in getting access to data. The Data Hunters call this service the 'Data Wish List' and, although they make no guarantees that they'll be able to persuade public bodies to open up the requested data, in several cases they have in fact managed to help re-users get data. Even when they don't succeed in getting public bodies to open up the requested data, the data re-users appreciate and respect their efforts.

In 2010 the Data Hunters were also responsible for making a data inventory for the Ministry of Science and its underlying agencies (of which the National IT & Telecom Agency is one of four). Our minister, the Minister for Science, decided to practise what she preaches and called for the Ministry's data to be opened up. Every organisational division were asked to provide an annotated list of their datasets which the Data Hunters then divided into three categories: 'Non-publishable' due to privacy, confidentiality or national security concerns or third-party rights, 'readily publishable' which required little or no preparation, and 'deferred publishable' which required more work.

The ODIS initiative doesn't deal at all with personal data, confidential data, classified data or data covered by third-party rights and so datasets which were deemed non-publishable for these reasons were completely disregarded. For the Danish Ministry of Science, non-publishable datasets constituted only approximately 5% of the total number of datasets; this may be different in other public bodies. However, one thing we learned which is more likely to apply in most public agencies was this: When asked about it, public agencies knew exactly which datasets were non-publishable, i.e. those who had personal data, knew it was personal data which couldn't be opened up.

In order to speed matters up and quickly make some data available, the Data Hunters first focused on the readily publishable aka 'easy' datasets. It turned out that quite a few of these datasets were in fact already available on public body websites – although often in hard-to-find locations or in impractical formats – and simply needed a little bit of polishing or some spotlight. Once the easy datasets were all properly listed in the Data Catalogue and groomed for re-use, the Data Hunters turned their attention to the 'deferred publishable' datasets which required more effort. The majority of datasets within this category have since then been published and we're now down to relatively

few datasets which will be costly, time-consuming or simply tricky to publish: These we have listed in the Data Catalogue to make them visible to re-users, but we probably won't publish them, unless there's considerable demand for it.

In the Data Hunters' own words the whole process of opening up our data was "a lot easier, a lot quicker and much cheaper than anticipated", but we're all quite convinced that we could easily have made it difficult, slow and costly if we'd approached it differently. The ODIS initiative, like other open data projects, is navigating uncharted waters and for a job like that we believe we need a very agile approach and a lot of collaboration (bordering on co-creation) with all stakeholders. We simply don't have the time to work it all out before we act - or the money to create near-perfect solutions the way most public bodies prefer.

Instead we preach and practise 'smarter, better, faster, cheaper' – the credo of Peter Corbett, CEO in iStrategyLabs and the man behind the US data contest Apps for Democracy. In data publishing terms this translates into something like 'publish data as is – publish data now – make improvements if demand warrants it – watch what happens and adapt'. So we tell public bodies, who are new to the idea of open government data, to start out by identifying some 'easy' datasets, listing them in the Data Catalogue, and publishing the data as is or with as little rework as possible. And we tell them to do it now. Then we encourage them to take it from there, step-by-step. Having no political attention, no formal incentives and – consequently – virtually no budget, our reality is that open data must gain traction and prove its worth on a shoestring. We have to make it as simple as possible.

I've spoken to civil servants from quite a few European countries who, like I do, work to open up public sector data. But I'd like to get to know many more of you and I'd like to learn from you. What exactly do *you* do to find data and to make it available for re-use? Do you find that data publishing is expensive, or technically complicated? Time-consuming? Worth the effort?