Open Data Comes to Market

Report of a discussion seminar at the Open Data Institute, 14th Feb, 2013

Kieron O'Hara, University of Southampton

Introduction: context

This report summarises the discussion at the seminar *Open Data Comes to Market: the Mysterious Case of the Disappearing Apps*, which brought together developers, small businesses, government and other public sector officials to try to develop an understanding of what the ideal open data market would look like in the crime and criminal justice sector, in terms of the balances between public and private, small developers and large systems, and different ways of presenting data as accessible information. The seminar consisted of a panel (chaired by Kieron O'Hara, with Hadley Beeman [TSB], Nici Hosfield [Home Office], Dan Lewis [UKCrimestats] & Chris Yiu [Policy Exchange]), followed by an open plenary discussion.

There were several drivers for the meeting, including the interest of the Open Data Institute in understanding and facilitating the use of open data, the drive of the Cabinet Office to improve takeup of open data, issues for third parties which have been brought to the attention of various organisations by Dan Lewis of UKCrimestats and the academic study of the use of open data by groups of citizens conducted by the EPSRC project SOCIAM.

The questions flagged as important at the beginning of the seminar were:

- How should the crime data market be structured if it is to deliver innovative services?
- How should police.uk best coexist with alternative service providers to support innovation while ensuring the reliable delivery of trusted information?
- Is there a roadmap to reach the 'ideal' market?

The discussion was extremely constructive, with several points of consensus and some areas of disagreement. This brief report will summarise these points, and conclude with suggested ways forward.

Note on terminology. I will discriminate between 'data' and 'information'. Data is the set of variable values (e.g. expressed in Excel or CSV and downloadable from police.uk or from the API for mostly additional qualitative as well as numerical data), and information is an interpretation of the data, such as a map on police.uk. The implication is that a process that outputs information (such as displaying data on a map) adds value to the data, which is the premise of the open data economy.

Consensus

The government has been far-sighted in publishing open data. People are now better-informed about crime at the level of their street than they have ever been. Police.uk has engaged people to look at crime data, creating familiarity and data literacy on the part of concerned citizens, which private sector information providers probably would not have been able to do in the early stages (had the government only produced data and left information provision to the market from day 1). This is not to say that the current system is ideal or perfect, but the fact that we are even having this debate reflects well on the government and well on the team which has produced police.uk and the associated data.[1]

The government's job is not only to foster a development market. It needs to get information out to the citizen, and sites like police.uk and data.gov.uk have been successful at doing that. However, the government is in a dominant position – it is the major data provider and its information providers (like police.uk) have advantages of resources and effective brands. It should not abuse the power that it has, but equally it must innovate in its communications. Government has to try to be better and/or cheaper with digital technology.

Publishers, app builders and users all need reasons to work in the system. There are different types of open data apps – some providing information from government, some delivering services to and on behalf of government, some delivering services independent from government, and others creating efficiencies within the public sector via better conclusions from the data. All these are part of the picture of information flow between government and citizen.

Different sectors have different requirements. For instance, the kinds of monetisation acceptable to citizens will vary across sectors, and similarly the role of choice between information providers will also vary. Choice is not always appropriate. There will be more hard cases than we would ideally like.

There are various services which need to be performed in advance of data release – for instance, anonymisation and geographical 'vaguing'.

---

[1] As noted in the introduction (and as emphasised by Dan Lewis in comments on an earlier draft of this report), there have been many drivers to this debate, and some argue that the police.uk team have not been very responsive to concerns of third parties. However, there was certainly consensus that the production of data *per se* has been a major advance on the previous transparency situation in this area.

Disagreement

There were two major areas of disagreement. The first concerned what counted as discriminatory interference with the market for information, and the second concerned data quality.[2] I do not intend to judge these issues, but to lay them out so the disagreements are clear.

Discrimination. Although there seems to be an argument concerning whether police.uk illegitimately interferes with the market (i.e. whether it and its support system discriminates against developers), and whether the Public Sector Information Regulations are being properly applied, actually it is a deeper point. There is a conceptual division about what counts as non-discriminatory publication of information or data.

At present, the data arrive at a central point from various police forces. The data are then smoothed to make a single dataset. When the API is updated, police.uk and every app built using the API is updated at the same time. The CSV datafiles are released at this time too, and so those developers which do not use the API then need to download the data and update their apps manually.

The essence of the disagreement is this. One view is that this arrangement does *not* discriminate, because all apps are, or could be, updated at the same time. Everyone, police.uk included, gets the data at the same time (modulo use of the API), at least in respect of the monthly uploads. There are some anomalies. For example, when additional new datasets are incorporated on the site, it would be helpful to give advance notice, which is under discussion. Data from HMIC is used on but not available from police.uk, but is available from http://www.hmic.gov.uk/data/ under the OGL. The basic data about Police and Crime Commissioners on police.uk is not easily reusable, and the release of more data under the API is an option under discussion. And developers don't get access to police officers' photographs, but this is not an issue within scope here, as it is a data protection issue driven by the Home Office and police forces. So on this view, any discrimination is relatively minor and is being addressed where possible.

The opposing view is that this *does* discriminate, because the data provider has also developed one of the apps (police.uk), and so the designers of police.uk have a greater familiarity with the data and their creation. On this second view, to avoid discrimination, the data provider should not be involved in any downstream service provision – the provider should simply make the data available, and then other users (including police.uk) would have equal and independent access. In response, the police.uk team reply that the live data is ready only a short time before the API is updated and so the practical effects are nil. Furthermore, the police.uk team does not accept that using the same contractor to process the data and build and maintain the site is discriminatory under the Public Sector Information Regulations.

It was also argued that the provider's position could also be exploited because providing links to apps gives an advantage. Some stressed the importance of endorsement; government departments have strong brands, and an endorsement of some apps rather than others can be seen as skewing the market. The developers respond that although there is a link from police.uk to a version optimised for display on mobile devices; this does not correspond to an endorsement of a separate app, but rather a link between versions of the same site, and that on the http://www.police.uk/apps page several apps, not just the police.uk mobile version, are linked to.

---

[2] There were other debates and disagreements at the event, but these were the two significant ones.

In each of these two related arguments, there was a difference of opinion as to what counted as non-discrimination. Should we insist on a firewall between data provision and service provision? Or should we aim to ensure that all service providers receive the data as soon as possible and at the same time from the same source? The first condition implies the second but not vice versa, and so is the stronger. The question for policymakers is to decide whether the second condition is adequate.

Data quality. One opinion voiced was that users don't need perfect data – they basically need to be informed, and to have a sense of what is going on. Precision or high accuracy, though desirable, are not necessarily vital to that task, and should not be reasons for delaying or impeding publication.

Against that, there was a view that some issues with the data are significant. Some examples were discussed with respect to geocoding – there is no standard gazetteer of public places, and also there was a plea for the use of ONS census boundaries when reporting crime statistics, as opposed to using neighbourhood policing team boundaries. Up to 25% of certain types of crime are ambiguously placed because of failure to use standard boundaries. This is of course not an issue that can be solved solely by the police.uk team, and would require a response across the entire data-generating system.

As another type of example, there have been millions of changes to the data as a result of its relatively frequent updating, without adequate documentation. Given the nature of the data, very detailed documentation would be impracticable, but the police.uk team is working to improve the information provided about changes. Hopefully this process will be informed by dialogue with third party developers.

A related view was that, with transparency in mind, the public should be clear about what data the police have available. However, data processing is performed at the Home Office where the data from the various police forces are smoothed into a single national dataset. In that case, the public is *not* being presented with the same data as the police. Some thought it was misleading to produce a polished product when the raw data are relatively poor. The police.uk team does not see its mission as providing this kind of data, and believes that this would not be proportionate to police.uk's objectives of giving a picture of local crime at the street level.

Ways forward

In this final section I briefly suggest some implications of the debate as to how to move forward.

1. Clarity on intentions, outcomes, metrics

Providers of open data (such as police.uk) need to be clear about their intentions and what outcomes they are aiming for. Without a clear set of outcomes, and a set of significant and accurate metrics, it is hard to judge arguments about the government's role in information provision. If police.uk could be shown to be successfully addressing its requirements, then it would be easier to defend its particular role in the marketplace. If it could not, then a sensible debate about the potential for the private sector could emerge, underpinned by an understanding of where police.uk was falling down.

This is not a simple matter, in that the current objectives are relatively broad and open to interpretation (so not much help for third parties), and yet very exact objectives may fix police.uk in a particular form which may result in it failing to innovate or exploit future technological advances. Finding a middle way will be hard. However, the main driver for police.uk has generally been promoting accountability within the transparency agenda, and that driver may be in tension with other transparency drivers (such as providing economic benefit). The development of metrics to measure impact and accountability would seem to be a priority.

In general, if the objectives will inevitably be fluid, then at a minimum metrics to indicate whether the current objectives are being met should be a key requirement. This is not an easy matter to determine, and the police.uk team are already investigating how to measure impact, for example engaging in focus group analysis of stakeholders. Website analytics can of course tell something, as can website feedback, and improved communication routes (discussed below) can also provide useful feedback. Monitoring conversations on the Web (e.g. via microblogging or social media sites) may also provide insight.

But this is a hard problem, beyond the scope of this short report to address, complicated by the fact that police.uk shouldn't be seen in isolation, but as part of a suite of tools for addressing policing problems. Feedback from other groups (within and without the Home Office) with an interest in policing ("does police.uk help us, hinder us or is it irrelevant?") would also be useful.

2. Drafting rules for interventions

Most participants agreed that there should be non-discriminatory 'rules of engagement' about publishing government information, so that developers, users and publishers know what to expect.

Rulebooks are hard to write (particularly given the variance across sectors), and the rules may exist in a messy patchwork. It is also difficult to bind the government to behaving in certain ways. And even if the government does not pursue certain potential lines of action, that will still affect the market, and might not be what consumers want.

A major part of the problem was alluded to above: how do we define a level playing field? The question of whether we need the stronger or the weaker condition against discrimination (see the italicised paragraph on p.4) is an important fault-line in the debate, and resolving this would induce much greater clarity.

It is hard to reconcile the stronger condition with the widely-held view (discussed in the Consensus section) that the government has multiple legitimate policy aims. It is not there solely to foster a market for information services off the back of open data, and so arguments that there are imperfections in such a market cannot stand on their own as an indictment of government actions.

It has other communication needs (and also of course a certain history in the provision of information). Given that, a private sector service provider needs to show not only that it has been discriminated against by some particular government action or structure, but also that the overall position is not serving to meet government's other communication requirements in the sector. It is important to take the public good into account in any arguments about the structure of information markets.

Of course, this argument would be easier to make for the private sector provider (and also easier to defend against by government) if the intentions and desired outcomes of a particular government operation were articulated more clearly, as noted above.

One final option is to explore analogies in other domains. For example, a broadcaster which owned a satellite needs to obey rules about carrying other channels, while BT provides access to the 'last mile'. It may be that there are either legal or commercial solutions from these sectors which could be applied in the world of open data and information provision.

3. Duty to be fair and mindful

An information market may well provide a wide-range of choice, but it cannot be denied that a state-backed entity makes a major difference in the marketplace. It would be absolutely wrong for a provider with state backing to invade existing markets, but some markets are more mature than others.

The role of government needs to be fair and appropriate. In many cases it has a duty to provide data and information for free. However, it also needs to recognise that it is in a dominant position, and therefore has a duty to be fair and mindful.

Hence the duty of mindfulness should be written clearly into any decision-making procedures concerning open data. Certain government open data initiatives have been very striking and secured positive media coverage – but the prospect of favourable coverage should not be the reason for an initiative. Decision-makers need to ask whether private sector providers could produce similar services – and to do this they will need to consult those suppliers. They must always ask: "could the private sector do this better?"[3] and "are we negatively affecting someone's business?"

4. Communication

Which leads to the final point – there need to be good routes of communication and points of contact between open data publishers, government service providers and private sector service providers. Dialogue needs to be clear, open and straightforward. Government needs to take the views of those demanding data and supplying information *genuinely* into account in decision-making.

---

[3] If indeed the private sector is not already doing it better.

i. There is shortly to be a forum around police.uk providing communications e.g. about changes to the data and the API, and also to give feedback on wider issues. Some have argued that this forum needs to be independently hosted (i.e. not hosted on police.uk or data.gov.uk), but this seems impractical – the official forum is surely more likely to be taken seriously if it appears on an official site. However it should (a) be openly and transparently moderated, given the lack of trust between some of the relevant actors, and (b) be clear that feedback is taken seriously, and acted upon where appropriate.

ii. There needs to be constant awareness-raising about what data there are available, what technologies have been available, and how these can be applied. Technologists and policymakers need to be in constant touch. And police.uk needs to be included as a tool to achieve a range of crime prevention objectives (and other objectives) across the Home Office.

iii. If the government is to take seriously its duty of mindfulness, there needs to be a forum in which private sector providers can feed their thoughts into decision-making.

iv. There needs to be a trusted environment for grievances to be aired and questions of discrimination judged. At present, the arbitration procedures appear arcane and ad hoc.

Sector panels and the Open Data Institute are two obvious routes for dialogue of types ii and iii to take place. The Cabinet Office and its transparency unit should consider how iv could be facilitated.

Of these routes for engagement, (i) and (iv) should be expected to involve already-committed developers. In order to get increased engagement and take-up, and greater value from the data, (iii) is perhaps the most important communication route here. There are possibilities to opening up (iii):

- One would be to collate the experiences of trailblazing constabularies who are experimenting with other ways of releasing crime data and engaging with communities.
- Another would be to identify a specific problem or issue, and to engage with third parties via the Open Data Institute to address it.
- A third would be to collate experience in other sectors, perhaps through the Cabinet Office.
- A fourth is to continue the focus group analyses in which the Home Office is already engaged with stakeholders (see under (1) above).