AstraZeneca response to UK Government Open Data consultation

Response deadline : 27 October 2011

Purpose of consultation : to review reactions to policy proposals and perceived challenges to making large amounts of data collected in the course of public services readily available for citizens use

About AstraZeneca

Number of Employees (UK and Global): Global 61,000 of which 13% are in UK
Annual Turnover: Global turnover $33 billion

The importance of public sector information to AstraZeneca in the UK:

The public sector information which is of particular importance to AstraZeneca is that resulting from the treatment of patients within the NHS. This information is used to inform the development of new medicines and to assess the clinical effectiveness of these medicines in routine clinical practice. Within the UK marketing company it is crucial to gaining NICE approval and engaging with clinicians, payers and commissioners at a local level.

This note sets out:

- **AstraZeneca's views on principles for handling data generally**
- **Responses to specific questions in the Open Data consultation**
- **Responses to specific questions raised by the UK Government in relation to**
    - New Business opportunities
    - Economic benefits
    - Wider economic and social risks
    - Further research

Appendix : the Six Opportunities of Open Data and the policy challenges identified in the Open Data consultation document

- **AstraZeneca's views on principles for handling data generally**

As a scientific information based business, various factors play a role in how AZ handles data of varying sorts. Certain data will be fundamental to the maintaining of a competitive market edge which is the lifeblood of AstraZeneca's survival as a commercial company. Different considerations on data availability therefore apply to such data compared to data generated or held as part of public service provision by governments. Certain other types of data also warrant particular protection within a commercial setting, for example identifiable patient data used in developing medicines. AZ's Code of Conduct and supporting policies and standards reflect the need for such protection.

However, even within a commercial enterprise, there is much data of particular types that in the interests of the progress of medical science and research capabilities should be (and is) shared openly. An example would be clinical trial results data.

The type of data that results from public service provision in the UK and which is particularly relevant to AZ's business of drug development is patient data contained in NHS medical records. There are several principles that need to be borne in mind when considering

whether and how much of such data should be subject to the presumption of ready availability.

- **Safety** - greater availability of patient data from medical records could provide much better information on the specific outcomes of various treatment regimes, not only for the purposes of safety (for that patient and the wider population) but also for further research and refinement of particular drugs' safety profile and effectiveness

- **Better research capability** – greater availability of different types of patient information from databases and registries would enable both more complex research questions to be asked in the course of drug development

    - The Research Capability Programme is working towards this and AZ fully supports the objectives

- **Terminologies and meaningful interpretation** – quality of data capture within widely understood and adopted standard terminologies would improve data usage. There is a danger currently that misinterpretations can be made from poorly understood data in analyses that are made.

- **Visibility on breadth and type of data availability** – researchers' awareness of what data is available, by means of a publicly available catalogue that lists and describes the type and collection criteria, would be a useful addition to the availability of the data itself

- **Rights to privacy –** an individual's medical record is an integral part of who they are and how they live. Rights to manage this information in the way that the individual chooses should be reflected in an appropriate governance framework that applies to granting access to medical records. Whilst identifiable patient data is often the most useful for research, strict rules to ensure that individuals not wishing to share any of their information must be in place. At the same time, maximizing the data that is available, by building on the current rules that allow researchers to use anonymised datasets without recourse to the relevant individuals, would be the most helpful for medical research.

Responses to specific questions in the Open Data consultation (Page 6)

1. Do the definitions of the key terms go far enough or too far?

No specific comment

2. Where a decision is being taken about whether to make a dataset open, what tests should be applied?

Does making the dataset open, either in isolation or linked to other data, risk putting 'personal data' in the public domain?

To what extent can additional meta-information around the content, quality and completeness be made available alongside the dataset itself? Lower quality or lack of coverage or completeness should not be a bar to datasets being open but issues associated with datasets should be explicitly stated.

At what level will data be made available? Open access to raw data may pose risks to confidentiality, even when clear 'personal' data is removed. Aggregation of data may result in a loss of granularity which is crucial for analytical or research purposes. A balance between these will need to be established.

3. If the costs to publish or release data are not judged to represent value for money, to what extent should the requestor be required to pay for public services data, and under what circumstances?

Who will be the judge of what is considered 'value for money'? How is it intended that it will be assessed? If the data owner assesses value for money there is the risk that they may not be aware of how their dataset might be used outside the NHS.

The current system of access to anonymised health data from the Information Centre whereby the charges are for the time taken to process the request, including any analyses requested, does not seem unreasonable in theory. However if there are repeated requests for data that was not considered of value at initial consideration then there should be a review to consider whether the dataset should be in scope. In addition it is key that data is provided in a timely manner.

4. How do we get the right balance in relation to the range of organisations (providers of public services) our policy proposals apply to? What threshold would be appropriate to determine the range of public services in scope and what key criteria should inform this?

No specific comment

5. What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?

Providing initial support to enable public organisations to organise and curate datasets so that their ongoing release does not impose a disproportionate burden on providers. The aspiration could be that the resource required for maintenance and delivery of the open dataset should be the same or lower than those currently providing responses to Freedom of Information requests. This might require significant investment in a flexible self-service model for data access for those who do not have the capacity or analytical ability to use the full datasets.

Responses to specific questions in the Open Data consultation (Page 25)

1. How would we establish a stronger presumption in favour of publication than that which currently exists?

Ensuring there is a clear mechanism for obtaining public service information (website access probably) and demonstrating that it works. Clarity on what sort of information is accessible and from where (through an efficient search engine possibly) should be available. Generators of the information should be trained to understand which information recorded will be automatically accessible.

2. Is providing an independent body, such as the Information Commissioner, with enhanced powers and scope the most effective option for safeguarding a right to access and a right to data?

An Information Commissioner could oversee the implementation of a framework such as that described in the previous answer. Independence would be a good way of balancing the interests of the public as well as the data owners and generators.

3. Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda?

Current provisions would need to be road tested for practical application and understanding of the rules in relation to a new open data framework. In particular any potential conflict between UK law and EU law on privacy rights needs to be resolved and articulated in the new framework.

4. What might the resource implications of an enhanced right to data be for those bodies within its scope? How do we ensure that any additional burden is proportionate to this aim?

There will undoubtedly be cost and resource implications in organising and/or re-organising quantities of data for entities subject to the rules of a new framework – involving investment in new information management systems or redesign of existing mechanisms. Release /mandatory accessibility of data (and therefore supporting systems) could be on a staggered basis over a number of months/years to avoid immediate and significant cost impacts. Priority release could perhaps be linked to data clearly linked to activity likely to underpin economic growth.

5. How will we ensure that Open Data standards are embedded in new ICT contracts?

Mandatory terms of the contract setting out a commitment to adhere to the standards and to supply all necessary systems and training, together with an appropriate audit framework to check compliance, articulate sanctions and publicise violations.

Responses to specific questions in the Open Data consultation (Page 28)

1. What is the best way to achieve compliance on high and common standards to allow usability and interoperability?

Contractual requirements specifying standards to be adhered to. In view of the cost and resource implications for meeting such standards to facilitate data access it may be necessary to provide a government funded incentive to quicken investment by (non-government) entities that will be required to comply with the standards. A related idea is the investment by the US government in new IT systems for physicians practices that could demonstrate "meaningful use" of the data being collected and organised. A quality standards system like the ISO might be an accreditation that service providers could apply for.

2. Is there a role for government to establish consistent standards for collecting user experience across public services?

Yes – perhaps by giving information on prevailing standards and information on providers meeting and using those standards to underpin and reward best practice.

For example: Government should encourage the use of accepted and validated instruments, eg EQ5D for Quality of Life.

3. Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?

Any scheme proposed should promote rather than restrict access to data and support appropriate quality standards in data management, analytics and interpretation.

Responses to specific questions in the Open Data consultation (Page 30)

1. **How would we ensure that public service providers in their day to day decision-making honour a commitment to Open Data, while respecting privacy and security considerations**.

Development of a culture, framework and training that supports the presumption of open data with periodic "road testing" of data requests to check how quickly and efficiently the data is provided, together with periodic auditing of systems and publicising of best practice and continuous improvement methods.

Any recorded decision could answer a mandatory question on how the open data requirements have been met – itself accessible for scrutiny.

2. **What could personal responsibility at Board-level do to ensure the right to data is being met include? Should the same person be responsible for ensuring that personal data is properly protected and that privacy issues are met?**

Clear accountabilities for oversight of the open data system and clear sanctions for violations or failure to meet standards. A well understood system with well articulated rights and exceptions could mean that this could be the same person without a conflict of interest.

3. **Would we need to have a sanctions framework to enforce a right to data?**

Yes.

4. **What other sectors would benefit from having a dedicated Sector Transparency Board?**

No specific comment.

Responses to specific questions in the Open Data consultation (Page 31)

1. **How should public services make use of data inventories? What is the optimal way to develop and operate this?**

Access to a comprehensive picture of what data is readily available from where should be one aim of the use of data inventories- aggregating those currently available and developing linked inventories perhaps to facilitate anticipated research questions. Inventories at both the individual provider level and aggregated and accessible via a search facility of some kind should be fully exploited.

2. How should data be prioritised for inclusion in an inventory? How is value to be established?

*If the aim of open data is to encourage innovation and economic growth perhaps the data that is clearly used by businesses to that end should take priority***.**

3. In what areas would you expect government to collect and publish data routinely?

Education investment and achievement at individual establishment level, health investment and achievement at individual practice and acute trust level including health outcomes achieved, public infrastructure investment and services at local provider level, comparative data with other countries at as granular a level as possible.

4. What data is collected "unnecessarily" ? How should these datasets be identified? Should collection be stopped?

No specific comment

5. Should the data that government releases always be of high quality? How do we define quality? To what extent should public service providers 'polish' the data they publish, if at all?

There will always be questions of data quality and appropriate interpretation. However, releasing data in its raw form will encourage interpretation and debate and the development of interpretational standards. A practice of polishing data will delay and restrict access, encourage narrow interpretation and be contrary to the spirit of pluralist data use for the greater good.

The three principles in the following quote capture the issue well:

"Three basic principles govern my work. First, data is a public good and therefore should be out there. Second that if it is available by Freedom of Information then citizens or residents shouldn't have to go through any bureaucratic nightmares to get it, and [third] that we should have a presumption of openness – extremely important in restoring public confidence in public institutions…"

Director of Digital Projects, Greater London Authority
26 NAO, A Short Guide to Structured Cost Reduction 2010
27 Figures from MOJ and UCL Constitution Unit

Responses to specific questions in the Open Data consultation (Page 33)

1. **How should government approach the release of existing data for policy and research purposes: should this be held in a central portal or held on departmental portals?**

A central portal would make it much easier to access. Contributors should post the data in one place. Having to make multiple requests for data collected in various locations is not helping access much.

2. **What factors should inform prioritisation of datasets for publication, at national, local or sector level?**

Linkage to key business sectors that will spur innovation and economic growth would be one factor to consider. The Life Sciences sector and patient health data is a key example.

3. **Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?**

Both are important, however increasing awareness of what is or could be available for researchers by prioritizing broader data sets (ie datasets that have not been available before) would spur more innovative research and thinking more readily. More detail across the broader collection would then naturally follow. The government should also consider the

ease with which some datasets could be made available; for some datasets the underlying data which underlies the currently published data is already available in an electronic form to internal users (eg NHS prescribing).

Responses to specific questions in the Open Data consultation (Page 36)

1. Is there a role for government to stimulate innovation in the use of Open Data? If so, what is the best way to achieve this?

Examination of data currently used in innovative industries and exploration with those industries as to how this could be improved and developed to underpin faster progress in those industries would be one idea. For example interlinked datasets to answer critical questions in disease understanding and development to assist in drug development projects in the life sciences.

- **Responses to specific questions raised by the UK Government in relation to**

<u>New Business opportunities</u>

1. **How can more public data grow your existing business and/or expand your business into new markets?**

Greater availability of public health data, in terms of completeness, quality and integration or linkage would enable our business to ask more complex questions around safety, effectiveness and health outcomes of our medicines. Availability of more data would enable these questions to be answered earlier in a medicines lifecycle. The ability to assess societal outcomes of health, such as use of social services and effect on employment and sick leave would enable better health economic cases to be made in countries which consider these in reimbursement decisions, growing our global business.

2. **What sort of data would you need to have access to for this growth opportunity to be realised?**

    - *Greater coverage of primary care health data*
    - *Dispensed medications (acute hospital trust and community pharmacy)*
    - *Easier access to registry and other public health data which is not collected directly for the clinical treatment of patients*
    - *Social Services usage*
    - *Employment, including sick leave.*

In order for the benefits to be realised these sources would need to be able to be linked to other datasets, rather than provided solely as stand-alone datasets. Access to national data rather than accessing data from each of the 4 nations would be desirable.

3. **Do you know of any specific datasets Government collects that would fulfil your data requirements?**

Patient level primary care datasets such as that currently in GPRD, but expanded to include other eHealth Record provider systems (eg EMIS)

Cancer, Cardiac, Diabetes, Renal and other registries
HES
Minimum Mental Health Dataset
Prescribing/dispensing data in primary and secondary care
Acute Trust eHealth clinical data (from trusts who have successfully implemented a full eHealth Record)
Personal Demographics service
Choose and Book
ONS record of births and deaths
DWP records of employment and sickness

Economic benefits

1. **What is your central estimate of the increase in turnover that you could realise with increased access to public data?**

   Access to a greater breadth and depth of data will enable studies on the safety, effectiveness and cost-effectiveness of medicines to be undertaken much earlier. This should in turn result in earlier, appropriate usage of new medicines. These data would also impact re-imbursement decisions in other markets. Financial models forecast that bringing forward a positive regulatory or re-imbursement decision by 3 months may result in increased sales in the region of 25% of peak year sales.

2. **Do you think a venture of this sort could create new jobs? If yes, what is your central estimate for the number it could generate?**

   Healthcare data is complex and requires a deep understanding of clinical, organisational and external factors in addition to specialist analytical skills. Recognition of the quality of UK health data and the challenges around appropriate and effective usage of this data has already created some new jobs in the UK within AstraZeneca. Non-Pharma companies that specialise in the use of healthcare data would be likely to expand within the UK to exploit newly available or expanded datasets and the increased requests for projects from Pharma companies that do not have internal resource.

3. **What would be your projected timescale, in months, for turning raw public data into a viable business opportunity?**

   We anticipate that information derived from raw data could be included in business decisions, HTA submissions and customer interactions within 6 months of acquisition. More complex research studies, which would be subject to external ethical and scientific review, could have an impact within 12 months of approval.

Wider economic and social risks

1. **What wider public benefits might be realised alongside anticipated commercial benefits?**

Increased access to broader and deeper healthcare data could contribute to improvements in Public Health. Earlier completion of safety and effectiveness studies could identify medicines which should me more widely adopted or discontinued. Greater understanding of all types of interventions and outcomes in a wide range of diseases could lead to easier identification of best practice in clinical care.

2. **What are the potential costs to industry associated with Government opening up more public data?**

   There would be costs associated with employing and training staff and in acquiring hardware and software to maximise usage of newly available or expanded datasets.

3. **Are there any datasets which pose particular risks to current industry?**

   Any datasets which might allow individuals to be identified, either in isolation or after linkage to other datasets, have the potential to compromise patient confidentiality. This, in addition to the legal consequences, could lead to a loss of public faith in increased access to healthcare data.

Further research

1. **Could you direct us to any published evidence or case studies that would be useful for our enquiry?**

AstraZeneca has used linked NHS data relating to usage and outcomes of current therapies in sucessful UK and other national Health technology assessments of ticagrelor. The abililty to do similar research in less prevalent diseases will be limited unless a greater proportion of primary care patient level data is made available.

2. **Would you be willing to allow us to follow up with any further queries? If yes, who should we contact within your organisation?**

Jane Juniper
R&D Policy Lead

Cathy Emmas
UK Medical & Regulatory Affairs

3. **Is there anyone else that you suggest we speak to (internally or externally)?**

Other governments approach to data transparency, effect on the commercial world and the dynamics that drive it (eg the intellectual property framework), awareness of cultural differences and attitudes to the power that comes from being better informed – this would inform the thinking on the UK's positioning in this initiative and the likely effects over time in other regions.

AstraZeneca
October 2011

Appendix

“Six Opportunities of Open Data” are identified in the Open data consultation document

- Accountability
- Choice
- Productivity
- Quality and Outcomes
- Social Growth
- Economic Growth

Policy challenges identified are in relation to:

- Establishing stronger rights to use data
- Implementing suitable standards of transparency
- The corporate and personal responsibilities for delivering data
- Making available the most meaningful and useful data
- Using government and the public sector as examples in open working
- A role for government in stimulating innovation and enterprise from data use