

The AGI response to:

A Consultation on Making Open Data Real: A Public Consultation

The Association for Geographic Information (AGI) is the UK national membership body for all with an interest in Geographic Information (GI). A summary of the background and activities of the association follows at the end of this document.

Compilation

This document has been created from several sources:

- A workshop held on 5 October, 2011 comprising members of the AGI Council, and other AGI members;
- AGI members' views (submitted through an online survey) and;
- Round tables on the Public Sector Mapping Agreement held between AGI members and Shareholder Executive in late 2010.

The response is structured as a series of answers to the questions posed by the consultation. These now follow.

Glossary of Key Terms

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

The AGI would like to comment on the definition of *Public Services*. Ensuring that contracts take account of Intellectual Property (IP) in data created to do the work is logical. It should avoid inadvertently introducing issues seen where contracting private companies to do work then means that the public sector customer is subsequently not able to use the data.

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

The AGI believes that *unless* there are privacy or security issues, there should be a *presumption* that data should be published, rather than that not. In other words, publication should be the *default* position.

There are some additional basic tests that can be applied. For example, is the data in question a core reference geography/data set?¹ If so it should be released. Is the data used by government body as part of a public task? If so, and assuming the aforesaid privacy or security issues do not arise, then the dataset should be made open. Is the data 'raw' as opposed to forming a 'downstream' value-added data product? If so and assuming privacy or security issues do not arise, then the dataset should be made open.

Beyond these basic tests, it is not always possible to determine in advance what data is of value; often value only becomes apparent when data has been released and end users start

to use it. Similarly raising concerns about data ‘quality’ can be presumptuous because the nature of use of Open Data by third parties cannot be necessarily predicted. Therefore the AGI believes that feedback from users will be helpful in deciding what other public data could be released in future.

Some public sector members of the AGI have expressed a view that there should be a caveat that data should be released if there is *demand* and it adds to *accountability* and *transparency*. This view does not presume that all public datasets should necessarily be supplied free of charge. There are concerns that releasing *all* public data for free is neither affordable nor sustainable because licence revenue from datasets generally supplements a shortfall in taxpayer revenue or central government funding.

Feedback from some AGI public sector members includes a need to recognise that there are likely to be some costs to the public sector of making data open and available. This is no doubt correct. Some public bodies will need government help in moving from the cost recovery of data publication to open data publication where difficulties occur. However, in general this should not be part of the decision as to whether data should be released or not. In addition, some AGI public sector members have stated that they are keen to retain an ability to charge for data unless they are directly related to transparency and accountability.

There is a clear preference from AGI members for data to be released “as is”, with “health warnings” about data quality, rather than not releasing it at all, or waiting for the data quality to be improved. Furthermore, if government is using this data as part of its public task then some quality level can presumably be expected. Furthermore AGI members would like any shortcomings with the data to be addressed once it has been made available. Once released, data users should be encouraged to publish reviews of the data (including the overall value of the data, its quality and content) which would be of great value to the data publishers. Releasing public data can also assist in the quality assurance process with third party users being trusted as a source of valuable feedback.

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?

As stated above, in terms of whether charges should apply, by default it should be assumed that the data should be released at no cost bearing in mind the criteria outlined in answer to question 2. That said there is certainly a case for public organisations to charge for data sets that are outside these criteria, those that are ‘polished’ and have value added.

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?

The AGI has no comment to make on this aspect.

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

The basic tests as expressed above should be used to ensure publication. Whilst over 7000 datasets are registered on data.gov.uk (and it is believed that many more public datasets exist with largely unexplored potential). It is likely that the vast majority of demand will be for a small number of well recognised datasets that fit well within the basic tests. Focusing the publication on these key datasets should reduce costs and complexity of publishing Open Data and the burden, real or otherwise, on publishers.

Additionally, in line with localism, the need for what data is released and how could be devolved to local level bodies with the views of the local community being accommodated to support transparency, accountability and community rights. Mechanisms used should aim to balance these needs and wants, especially given the economic pressures on the public sector. That said recognition of the potential downstream benefits and not just the initial 'costs' need to be recognised and acknowledged by publishers.

An enhanced right to data: how do we establish stronger rights for individuals, businesses and other actors to obtain, use and re-use data from public service providers?

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

Other countries (e.g. New Zealand) have dealt with this by setting an expectation that data will be released unless there are reasons of privacy or security that would over-ride this. The proposed approach appears to follow similar principles.

Public sector would want some recognition of proportionality – i.e. balancing the potential benefits of releasing data with the costs of doing so. In mitigation, data publishers should be encouraged to develop standard production processes as far as possible, and avoid performing custom work for each enquiry that is received. Some government departments have assessed the time spent dealing with requests and identified that 80% of requests related to questions that can be answered by 20% of datasets. The remaining 20% are more challenging and need to be dealt with individually, and after some analysis. There is significant scope for streamlining by developing standard processes – this could include standard datasets containing that meet the majority of enquiries.

The head of statistics for central government departments could be responsible for publishing data. The Government Statistical Service could facilitate this. For local authorities, performance data may achieve the same aims.

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?

The AGI believes the Information Commissioner's Office could perform an external policing role.

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

Perhaps not; there has been concern expressed that by releasing a wide range of datasets, it may make it possible for users to reverse-engineer information on individuals by combining multiple datasets (see the O'hara transparency review at <http://www.cabinetoffice.gov.uk/news/government-publishes-independent-transparency-and-privacy-review>). This aspect needs to be considered.

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?

It has already been stated that it is likely that the vast majority of demand is most likely to be for a small number of datasets. Additionally where public data exists the 'burden' of publication, bearing in mind the (ideal) use of electronic distribution and the acceptance that public data may be released 'as-is' and improved over time (with feedback from users), then the cost overhead should be proportionally very low and the benefit to the bodies within its scope significantly greater.

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

Active promotion should lead to greater use of Open Data and the increased likelihood of it becoming embedded in new ICT contracts. There is a role to play within government both as users and promoters of Open Data, as well as amongst third parties.

Setting transparency standards: what would standards that enforce this right to data among public service providers look like?

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

There is a plethora of existing standards for data interoperability, for example those defined by the EU INSPIRE legislation and associated Implementing Rules. The best way of ensuring compliance is to have user testing. Where non-compliance is demonstrated, this can be notified to the information supplier, for follow-up. This process would be overseen by the Information Commissioner's Office (ICO).

This pre-supposes that the existence of the data is known. Metadata is essential to data discovery. The provision of basic metadata (data about data) through a discovery metadata service, for example, that being set up by the UK Location programme, will enable searches to be made by key parameters such as subject area, keywords, geographic coverage, date etc.

To achieve interoperability, it will be vital to promote the use of Open Standards.

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

User experience will be varied and depend on the user, usage and the type of data. A user body could be set up (possibly under OPSI, the Office of Public Sector Information), and any standards for collecting user experience should be left to them. This might include a use case register, following ISO 19135 *Geographic information - Procedures for item registration*.

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

There is no need to establish a scheme for accreditation of information intermediaries, as this would inhibit usage and economic activity. However, should it be decided to so do, then it is important to follow or adapt existing accreditation schemes, for example ISO 19158 *Geographic information - Quality assurance of data supply*, rather than inventing new ones.

Corporate and personal responsibility: how would public service providers be held to account for delivering open data through a clear governance and leadership framework at political, organisational and individual level?

- 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?**

There may be several layers of responsibility which should be built on existing governance arrangements instead of re-inventing what already exists. For example, the organisation's Chief Information Officer (CIO) would be the first line of accountability. Subsequent layers could include the Information Commissioner's Office, which would perform an external policing role.

- 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?**

Publishing data could be part of the defined public task of the bodies within its scope and Open Data Key Performance Indicators could be part of their operating objectives. This can and should be managed by the organisation's Chief Information Officer (CIO)

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

A sanctions framework could be inappropriate, unwieldy and time-consuming to operate. Rather adherence to publishing data could as described in the previous answer.

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

There are boards in various areas including the UK Location Council providing cross-sector coordination for geographic information across various thematic headings. The AGI would support the UK Location Council or equivalent, future, body by being a route into Transparency Boards if desired.

Meaningful Open Data: how should we ensure collection and publication of the most useful data, through an approach that enables public service providers to understand the value of the data they hold and helps the public at large know what data is collected?

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

As mentioned above, it is difficult to determine what data will be valuable before it is released. However, clear guidance can be given around data which will not be published. This could include considerations of security, personal information, or even data that will soon be published.

In relation to data.gov.uk, the public sector should itself be using data.gov.uk because it will derive benefits. To reach its potential, it needs to be more readily searchable, (including spatially), and there needs to be more clarity on what is there and how unique it is. A mechanism to seek and process user feedback would be useful.

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

As previously mentioned, it is very difficult to estimate value in advance. If data is not released, based on the criteria outlined above, we will never understand what it could be used for. The initial assumption should be that all data is valuable and available for publication unless it breaches aforesaid rules.

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

The start point should be what is already being done, with other data being considered on a case by case basis as appropriate to defined public task. Data collection needs to be commensurate with the needs and activities of the public body and publication a by-product of the public bodies operating function.

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

The start point should be what is already being done, with other data being considered on a case by case basis. Statistics on usage should be used to cull data, unless the data is required for internal purposes. Data collection could be stopped where collection of data has no ongoing role in a public body’s statutory activities.

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?

Data does not always need to be high quality as long as it is supplied with “health warnings”. “Quality” can be subjective and depends on the purpose to which data is put. What is important is the level of authoritativeness. Data should be released with information on the measured quality (e.g. known accuracy, update cycle) and purpose for which it was captured, and how it was collected. Public bodies’ perception of data quality for their own policy purposes may not match that of third party users at either end of the quality scale. Whilst it is understandable that Open Data publishers may be apprehensive about its third party use and its appropriateness, holding back publication on this basis of potential risk would be detrimental. ‘Polishing’ data may only be required following user feedback that argues a case for doing so. Provision of such polished data with added value may not be free of charge.

Following the Open Government Licence (OGL), public authorities may not take responsibility for the data in any litigation, i.e. *“The Information Provider is not liable for any errors or omissions in the Information and shall not be liable for any loss, injury or damage of any kind caused by its use. The Information Provider does not guarantee the continued supply of the Information.”*ⁱⁱⁱ Hence, the data’s authority may be diminished (for example, various websites are good for general use but would not stand up in court as being accurate and reliable). That said, this does not necessarily differ from public data currently available and where licensing that data is charged for. Thus lack of, or limited, liability should not be an obstacle to data release.

Government sets the example: in what ways could we make the internal workings of government and the public sector as open as possible?

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 has numerous advantages, including those of easy recognition by and promotion to users, a common interface to work with, common access to a plethora of published data and a potentially centralised user feedback management process. That said a single portal can have complications when specifying, implementing and managing it. Such complexities and the possible time it takes to resolve them should not make government shy of using departmental portals if it is pragmatic to do so.

Essential to data discovery, and thus exploitation, is metadata. Metadata needs to be in a form that is common and readable, whether open data holdings reside on a central or departmental portal. Its use also needs to be mandated. AGI has developed the GEMINI metadata standard which has been adopted by the UK Location Programme for the UK's INSPIRE implementation. It is a solid example of the design, implementation and use of an effective metadata standard in growing use.

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

Prioritisation needs the use of metrics to determine what users most want. 'Demand' can be assessed from existing public data usage as well as indications, for example, from FOI requests. Public bodies could also be proactive, looking at, for example the "80-20 rule" where the focus for release would be on the core datasets used by the majority of current users.

Over time an Open Data User's Forum, for example, could provide very useful feedback of changing user requirements. Such requirements may offer distinct benefits to public bodies publishing the data.

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

This is not easy to answer; pragmatically the priority should be to publish currently available data within the parameters outlined above (i.e. public task related, core reference data, etc). Broader ranges of data may be available and 'easy' to publish because these are likely to already exist for internal use and the risk of disclosure of personal or sensitive information may be lower than releasing more detailed data. The AGI recognises that Open data is not about creating new data, it is about existing data which are created as part of a public body's activity.

Innovation with Open Data: to what extent is there a role for government to stimulate enterprise and market making in the use of open data?

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?

The AGI believes that government can help, but it is important that it plays to its strengths. It is not the government's role to engage in commercial innovation, but it can be effective at stimulating it in other sectors. Principally, this should include ensuring that data access and licensing is simple, consistent and easy to understand and apply.

It is possible to learn from international experience. For example, Australia has established the Collaborative Research Council (Spatial Information), which delivers research relating to technologies that may subsequently be commercialised. In America, Open311 (open311.org) enables Application Programming Interfaces (APIs) to be created from a standard, open-access, read/write model for citizens to report non-emergency issues. This enables companies to build software based on the open model and has created a competitive market.

If government also wishes to consider specific methods to stimulate enterprise and market making in the use of open data, it could, for example consider:

1. Sponsored information exchanges on best practice.
2. Pump-priming via grants and prizes (a current example being Ordnance Survey's Geovation).
3. Benchmarking public services on data release (although AGI public sector members have expressed concern that this would add burden at a time when powers are being devolved and bodies self regulate).
4. Simplified request process for information release.

About the AGI

The Association for Geographic Information (AGI) is the UK national membership body for all with an interest in geographic information (GI). Established in 1989, AGI membership is made up from both individuals and organisations. The latter includes government departments and agencies, local authorities, other national organisations, educational institutions, utilities, commercial businesses and geospatial software companies and data suppliers. In total over 2000 named members. The mission of the AGI is to *maximise the use of GI for the benefit of the citizen, commerce and good governance*.

The AGI is a member of the UK Location Council and provides the secretariat for the British Standards Institution (BSI) Committee for Geographic Information.

The AGI:

- Provides a focus for the GI community.
- Promotes value and benefit of GI.
- Represents the GI community to policy makers.
- Fosters appropriate GI standards.
- Encourages best practice and innovation.
- Promotes training, education and continuing professional development (CPD).
- Provides wide ranging member benefits.

Disclaimer

As a diverse membership organization with many interests and viewpoints, the AGI is an important voice to identify the wide range of issues, points and concerns that may arise when public policy consultations take place. This also means that the views expressed in this collective response do not necessarily reflect the views of all AGI members.

This response has been approved by the AGI Council and is submitted on behalf of the AGI by:

Chris Holcroft
Director and Chief Executive Officer
AGI

ⁱ <http://www.communities.gov.uk/documents/communities/pdf/locationstrategy.pdf>

ⁱⁱ <http://www.nationalarchives.gov.uk/doc/open-government-licence/>