Proposal for a Data Sharing Disclosure Standard[1]

Incorporating a Data Sharing Disclosure Standard into the government's forthcoming data sharing proposals, and establishing a registry of data shares proposed and in place, is vital to

- Provide for transparent, accountable decision making over data sharing;
- Build citizen trust in government data sharing;
- Support efficiencies through better coordination;
- Support longer-term efforts to rationalise and simplify data sharing arrangements.

In this paper we make the case for such a standard, and outline possible options for its implementation.

Background

To provide public services government collects and holds significant amounts of data about citizens. Data such as tax records, benefit records, information on eligibility for services and information on contact with services all exist in different systems and at different levels of government. The current legal and regulatory environment that governs the sharing of this data between government agencies, and outside of government, is confusing and filled with inconsistencies. In many cases dataset or department specific 'legal gateways' exist to share data, but in others, data cannot be easily shared at all, or actual and perceived barriers to data sharing prevent it taking place.

The government has established a Data Sharing Open Policy Making process, to engage with stakeholders in government, industry and civil society, to develop proposals that would:

- Support data sharing for research;
- Support data sharing to address fraud and debt;
- Support data sharing to enable more tailored public services;
- Establish safeguards to govern data sharing.

Data sharing generally involves personal data, or data derived from personal data in ways that are not entirely anoymisable (e.g. data extracts for research). As such, data sharing involves restrictions on the re-use of data, unlike open data, which generally refers to non-personal data which is available for anyone to re-use.

More information on the data sharing policy discussions can be found at
http://www.datasharing.org.uk

---

[1] Prepared by Reuben Binns (rb5g11@soton.ac.uk) and Tim Davies (tdavies@cyber.law.harvard.edu).
Building on a previous proposal for an Open Register of Data Sharing by Government.

The case for a Data Sharing Disclosure Standard

Under a Data Sharing Disclosure Standard, any public bodies entering into arrangements to share datasets would have to publish, in a common format, details of those arrangements. This should contain, at a minimum[2]:

- Who the data controller is;
- Who the data is (or is to be) shared to;
- A description of the data in question;
- Details of the purpose of the share;
- The status of the share (i.e. applied for, planned, in place, elapsed)
- The dates within which the share is in place;
- The power under which the share is made;

By standardising these disclosures using a common data format, and common codes and identifiers:

- **There will be greater transparency of data sharing** - allowing citizens, civil society and the private sector to have greater trust in government data sharing arrangements, and to engage in debate over any arrangements which may give rise to concerns. Transparency also creates positive incentives inside government for improved decision making and scrutiny of data shares. With scrutiny from an empowered civil society, transparency also forms an important part of data sharing safeguards.

- **Perception barriers to data sharing can be overcome .** Being able to see the data sharing arrangements that other departments, authorities and public bodies have in force will support public officials to establish arrangements, where present misunderstandings or perceptions that data sharing in certain areas is not possible are limiting sharing that could be in the interests of citizens.

- **Data sharing arrangements can be made more efficient**. At present, departments such as HMRC or DWP have to process hundreds of similar data sharing requests from local authorities and other public bodies. A transparent register, including planned arrangements in the pipeline, would allow create consolidation and consistency across data sharing arrangements, bringing substantial efficiency savings to data sharing departments.

- **Continued rationalisation of data sharing regulations can be realised**. By requiring, over time, existing shares to also be disclosed, a data-driven picture can be built up of the current landscape of legal gateways, regulations and arrangements in use, and efforts made to simplify

---

[2] A full list of possible fields, justifications and details is provided in Appendix 2

Disclosure in practice: worked examples

The Data Disclosure Standard could be applied across all of the areas of the Data Sharing policy. The following section provides two worked examples of how it could operate:

<u>Data sharing for research</u>

An educational research firm applies to the Department for Education for access to an extract from the National Pupil Database.

Basic details of the application are entered into the DfE's Data Sharing Disclosure register whilst the application is under consideration. Third parties can be alerted to proposed shares, and through the register can see who to contact with any comments on them.

When the share is approved, further details are entered into the DfE's Data Sharing Disclosure Register. This information can be aggregated into a national register.

Anyone wanting in future to find out what sharing arrangements are in place for the National Pupil Database could go to an aggregated register and search it to get an accurate picture of shares in place.

<u>Data sharing for tailored public services</u>

A local authority wishes to use HMRC data to identify young people not in education, employment and training in order to offer them support services. It will do this by matching employment records against its list of 16 - 19 year old residents, and removing those in employment. The actual analysis will be carried out by a contracted-out youth support service.

The local authority applies to HMRC for an extract of data. They first check an aggregated record of data shares to see under which power other local authorities have made such arrangements, and submit a simple request aligned with existing arrangements. The share is approved and entered into the HMRC Data Sharing Disclosure register as an ongoing arrangement.

The local authority then record in their own Data Sharing Disclosures that the data has been shared for analysis by the contracted-out service, and update this record once the analysis is returned, and the copy held by the contracted-out service is destroyed, to note that this onwards arrangement has ended.

Getting the technology right

Just as a good data sharing policy will need to be backed by well worded legal powers, an effective Data Sharing Disclosure Standard will rely on carefully implemented technology. Through the Government Service Design Manual, and the UK's open data platforms, principles and infrastructures are in place that can support the creation of a streamlined service that helps public sector workers do their jobs, rather than adding administrative burdens. Technical implementation options are considered in Appendix 1.

A progressive process

We envision that steps can be taken towards a comprehensive Data Sharing Disclosure Standard over the coming years by:

- Requiring that all data shares under the proposed data sharing policy abide by the standard, and through this policy building the infrastructure to support disclosure;

- Encouraging all central government departments to move to disclosure of their existing data sharing arrangements (including those outside the new policy) via the standard;

- Incorporating a requirement for all data controllers in the public sector to make use of the standard in future reforms of relevant legislation.

Under the strongest form of a Data Sharing Disclosure standard, possession of relevant data without being listed as a recipient in a disclosure could be made into an offence, giving citizens and the state a strong guarantee of who legally has access to their personal information. Consideration should be given to mechanisms such as this that ensure the implementation of disclosure.

Appendix 1: Technical implementation options
There are a range of different ways in which Data Sharing Disclosures could be made sustainably available. Options include:

- **(1) Gazette notifications** - the London Gazette is a paper of record, and provides for publication of a wide range of official notices. The Gazette has manual and automatic routes for submission of notices, and provides digitally signed machine-readable copies of records as feeds of data under the Open Government License.

  Arrangements are in place for the publication of mandatory notices free of charge in the Gazette when provided in a machine readable format, and at a nominal cost when provided in other formats.

  A Gazette notice can be both human readable - accessible to citizens searching by organisation name or other information - and machine-readable, enriched with annotation of company identifiers, terms of data sharing and so-on. As the Gazette is a established platform, this route may offer the most robust approach to ensuring records of data sharing are available over the long-term.

  Publishing in the Gazette requires minimum technical skills for public bodies, and publication in the Gazette could also be open to private firms engaging in onwards sharing of data. However, the option of feeding machine-readable data into the Gazette may also stimulate development of value-added local tools for managing a register of data sharing, and feeding notifications to the Gazette.

- **(2) A distributed register -** a machine-readable data template (e.g. CSV field specification) for public bodies to make their Data Sharing Disclosure could be made available, and public bodies asked to publish data on their local websites.

  Public bodies would be obliged to register the location of their Data Sharing Disclosure Registers with Data.gov.uk, and to ensure they are maintained[3].

  A central software system could regularly check these distributed registers to ensure they are kept updated, and to optionally aggregate the data into a central store also.

  A similar model is currently used for public sector spending disclosure, with a standard suggested to local authorities and central government, and, for central government departments, a dashboard at http://data.gov.uk/data/openspending-report/index. The limited currency of spending data, as shown at that dashboard (75% of records outdated), highlights the need for a distributed register to be backed by engagement with

---

[3] Alternatively, the location of a Data Disclosure Register could be made a requirement for public bodies during the annual updating of their Data Controller information with the Information Commissioners Office.

public bodies to ensure records are regularly updated.

● **(3) A central register** - the UK government would develop a streamlined, web accessible, central register in which any public body can record their data sharing arrangements.

Public bodies would log into the system to manage their records, and the system would be able to give an at-a-glance view of data sharing arrangements across the public sector, supporting better management of key data sharing workflows.

Further work would be needed to scope out the most appropriate option, but Option (1) and (2) are likely to be preferred.

Appendix 2: Suggested fields

The following table sets out fields that a Data Sharing Disclosure Standard may require, along with details of each field. The fields that are essential are flagged 'Y' in the core column.

Core?	Data fields	Justification & notes
Y	Record ID	A unique identifier for this record. Potentially made as a UUID by being a composite of Data Controller ID + local ID.
Y	Status	[Application, Preparation, Active, Ended, Refused] A code list for possible statuses of this arrangement.
Y	Originator Name	The name by which the organisation sharing / to share the data is known.
Y	Originator ID	The ID of this data controller in the ICO register. The ICO register currently requests, but does not publish, company registration details. If this was published, the the ICO registration number would provide a link to connect disclosures to companies. Questions: What about overseas transfers? How could this be updated if changes to requirement for a data controllers register in future?
Y	Originator Contact Details	Contact details for an enquiry about this arrangement. This may not be the same as the data controller details, but should be the department or person dealing with the case.
Y	Recipient Name	The name by which the organisation receiving / to receive the data is known.
Y	Recipient ID	The ID of this data controller in the ICO register.
Y	Data description	A description of the data to be shared

?	Data URL(s)	The URL(s) of the dataset record on data.gov.uk. Data.gov.uk includes details of Unpublished Datasets as well as open data. See e.g. http://data.gov.uk/dataset/pension-credit
Y	Sharing purpose	A short text description of the purpose of sharing the data, usually drawn from an application for data sharing.
Y	Active from	The date from which the share is active
Y	Active until	The data until which the share is active
Y	Authorising power	A link to the legislation or regulation on legislation.gov.uk which is used as the basis for this sharing. A codelist of relevant powers can be built up.
Y	Flags: Privacy Impact Assessment	Yes / No flag for whether a Privacy Impact Assessment was completed.
Y	Flags: Business Case	Yes / No flag for whether a Business Case was completed
?	Flags: Algorithmic processing	Yes / No flag to indicate whether the proposed use of the data involves the use of algorithms to make judgements which will affect citizens eligibility for, or access to, public services. This is included as there is a case to be made for algorithmic transparency, and users may wish to identify and scrutinise data use with algorithmic components.
?	Flags: Onward sharing	Yes / No flag to indicate whether permission was granted for some onward sharing of this data.
	Related arrangements	Identifiers of related data sharing arrangements. For example, if this is onwards sharing of data under a previous arrangement, the identifier of that arrangement should be given here.
Y	Linked Documents	Links to relevant documents, such as Privacy Impact Assessments, Business Cases, Algorithms used in processing data and other documents. Each should be tagged with meta-data on it's type, and date of publication or revision, so it's possible to check what kinds of documents exist for an arrangement. These could be with URLs, or just records to indicate that documents are held, and will only be released subject to FOI requests.
	Special conditions	Details of any special conditions placed on re-use. E.g. must only be handled in secure facility. Could either be free-text, or based on a codelist.