

# Comments on the draft National Data Sharing and Accessibility Policy

*Centre for Internet and Society* | 2011-06-08

Pranesh Prakash

*A draft of the 'National Data Sharing and Accessibility Policy', which some hope will be the open data policy of India, was made available for public comments in early May. This is what the Centre for Internet and Society submitted.*

---

These are the comments that we at the Centre for Internet and Society submitted to the National Spatial Data Infrastructure on the draft National Data Sharing and Accessibility Policy.

## **Comments on the National Data Sharing and Accessibility Policy by the Centre for Internet and Society**

We would like to begin by noting our appreciation for the forward-thinking nature of the government that is displayed by its pursuit of a policy on sharing of governmental data and enabling its use by citizens. We believe such a policy is a necessity in all administratively and technologically mature democracies. In particular, we applaud the efforts to make this applicable through a negative list of data that shall not be shared rather than a positive list of data that shall be shared, hence making sharing the default position. However, we believe that there are many ways in which this policy can be made even better than it already is.

### **1. Name**

We believe that nomenclature of the policy must accurately reflect both the content of the policy as well as prevailing usage of terms. Given that 'accessibility' is generally used to mean accessibility for persons with disabilities, it is advisable to change the name of the policy.

#### **Recommendation:**

A. We would recommend calling this the "National Open Data Policy" to reflect the nomenclature already established for similar policies in other nations like the UK. In the alternative, it could be called a "National Public Sector Information Reuse Policy". If neither of those are acceptable, then it could be re-titled the "National Data Sharing and Access Policy".

## 2. Scope and Enforceability

It is unclear from the policy what all departments it covers, and whether it is enforceable.

### Recommendation:

A. This policy should cover the same scope as the Right to Information (RTI) Act: all 'public authorities' as defined under the RTI Act should be covered by this policy.

B. Its enforceability should be made clear by including provisions on consequences of non-compliance.

## 3. Categorization

The rationale for the three-fold categorization is unclear. In particular, it is unclear why the category of 'registered access' exists, and on what basis the categorization into 'open access' and 'registered access' is to be done. If the purpose of registration is to track usage, there are many better ways of doing so without requiring registration.

### Recommendation:

A. Having three categories of:

- Open data
- Partially restricted data
- Restricted data

B. Data that is classified as non-shareable (as per a reading of s.8 and s.9 of RTI Act as informed by the decisions of the Central Information Commission) should be classified as 'restricted'.

C. The rationale for classifying data as 'open' or 'partially restricted' should be how the data collection body is funded. If it depends primarily on public funds, then the data it outputs should necessarily be made fully open. If it is funded primarily through private fees, then the data may be classified as 'partially restricted'. 'Partially restricted' data may be restricted for non-commercial usage, with registration and/or a licence being required for commercial usage.

## 4. Licence

No licence has been prescribed in the policy for the data. Despite India not allowing for database rights, it still allows for copyright over original literary works, which includes original databases. All governmental works are copyrighted by default in India, just as they are in the UK. To ensure that this policy goes beyond merely providing access to data to ensure that people are able to use that data, it must provide for a conducive copyright licence.

**Recommendation:**

A. The licence that has been created by the UK government (another country in which all governmental works are copyrighted by default) may be referred to: <http://www.nationalarchives.gov.uk/doc/open-government-licence/>

B. However, the UK needed to draft its own licence because the concept of database rights are recognized in the EU, which is not an issue here in India. Thus, it would be preferable to use the Open Data Commons - Attribution licence:

<http://www.opendatacommons.org/licenses/by/>

The UK licence is compatible with both the above-mentioned licence as well as with the Creative Commons - Attribution licence, and includes many aspects that are common with Indian law, e.g., bits on usage of governmental emblems, etc.

## 5. Integrity of the data

Currently, there is no way of ensuring that the data that is put out by the data provider is indeed the data that has been downloaded by a citizen.

**Recommendation:**

It is imperative to require data providers to provide integrity checks (via an MD5 hash of the data files, for instance) to ensure that technological corruption of the data can be detected.

## 6. Authenticity of the data

Currently, there is no way of ensuring that the data that is put out by the data provider indeed comes from the data provider.

**Recommendation:**

It is preferable to require data providers to authenticate the data by using a digital signature.

## 7. Archival and versioning

The policy is silent on how long data must be made available.

**Recommendation:**

There must be a system of archival that is prescribed to enable citizens to access older data. Further, a versioning and nomenclature system is required alongside the metadata to ensure that citizens know the period that the data pertains to, and have access to the latest data by default.

## 8. Open standards

While the document does mention standards-compliance, it is preferable to require open standards to the greatest extent possible, and require that the

data that is put out be compliant with the Interoperability Framework for e-Governance (IFEG) that the government is currently in the process of drafting and finalizing.

**Recommendation:**

A. The policy should reference the National Open Standards Policy that was finalised by the Department of Information Technology in November 2010, as well as to the IFEG.

B. The data should be made available, insofar as possible, in structured documents with semantic markup, which allows for intelligent querying of the content of the document itself. Before settling upon a usage-specific semantic markup schema, well-established XML schemas should be examined for their suitability and used wherever appropriate. It must be ensured that the metadata are also in a standardized and documented format.

## 9. Citizen interaction

One of the most notable failings of other governments' data stores has been the fact that they don't have adequate interaction with the citizen projects that emerge from that data. For instance, it is sometimes seen that citizens may point out flaws in the data put out by the government. At other times, citizens may create very useful and interesting projects on the basis of the data made public by the government.

**Recommendation:**

A. The government's primary datastore (data.gov.in) should catalogue such citizen projects, including open and documented APIs that have been made available for easy access to that data.

B. Additionally the primary datastore should act as a conduit for citizen's comments and corrections to the data provider. Data providers should be required to take efforts to keep the data up-to-date.

C. Multiple forms of access should preferably be provided to data, to allow non-technical users interactive use of the data through the Web.

## 10. Principles, including 'Protection of Intellectual Property'

It is unclear why 'protection of intellectual property' is one of the guiding principles of this policy. Only those ideals which are promoted by this policy should be designated as 'principles'. This policy, insofar as we can see, has no relation whatsoever with protection of intellectual property. The government is not seeking to enforce copyright over the data through this policy. Indeed, it is seeking to encourage the use of public data. Indeed, the RTI Act makes it clear in s.9 that government copyright shall not act as a barrier to access to information.

Given that, it makes no sense to include ‘protection of intellectual property’ amongst the principles guiding this policy. Further, there are some other principles that may be removed without affecting the purpose or aim of this document: ‘legal conformity’ (this is a given since a policy wouldn’t wish to violate laws); ‘formal responsibility’ (‘accountability’ encapsulates this); ‘professionalism’ (‘accountability’ encapsulates this); ‘security’ (this policy isn’t about promoting security, though it needs to take into account security concerns).

**Recommendation:**

A. Remove ‘protection of intellectual property’, ‘legal conformity’, ‘formal responsibility’, ‘professionalism’, and ‘security’ from the list of principles in para 1.2.