Comments from IT for Change, India to Article 19's Public Consultation on Principles of Privacy and Freedom of Expression

A. Definitions of key terms (Page 7)

Existing text:

"Personal Data is defined as any information relating to an identified or identifiable natural person ; an identifiable person is one who is identified directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic, or social identity".

Suggested Revision:

"Personal Data is defined as any information relating to an identified or identifiable natural person; an identifiable person is one who is identified directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic or social identity.

This includes identification directly or indirectly by :

a) interlinking of databases that are publicly available or developed privately ; or

b) the use of algorithms to analyse datasets; or

c) any other means"

Comment:

Interlinking of one database with another, causing the accretion of **any** data related to an identified or identifiable natural person can lead to re-identification of data subjects. A pioneering study in the US has demonstrated that 87% of the American population was identifiable by just their zip code, gender and date of birth. The researcher merely cross

393, 17th Main 35th Cross, Jayanagar 4th T Block, Bengaluru 560 041 Karnataka – India
Tel: +91-80-2665 4134, 2653 6890
Fax: +91-80-4146 1055
www.ITforChange.net

linked medical data released by the state, which included ZIP code, birth date, ethnicity and gender (but no name or address), with the voters list of the state which could be purchased.[1]

Two other famous cases that arose with respect to this in the US are discussed below.
Case 1: In August 2006, AOL released search queries of 6,50,000 of their users by replacing the user name with a unique ID. Soon after, the New York Times published an article, in which the authors claimed and proved that they were able to correctly identify certain users from the published search results.
Case 2: In October 2006, Netflix released records of millions of its users' movie ratings by replacing user names with unique identifiers, as part of a contest that invited members of the public to develop a movie recommendation algorithm that was ten percent more effective than its own in making accurate movie predictions. . Researchers at Texas University realized that through interlinking this supposedly 'anonymized database' with other datasets, they could identify the reviewer, to a high degree of accuracy.[2] As the number of data points increases, like in the case of hundreds of movie ratings per ID, the more unique and identifiable the data subject becomes.[3]

What is thought of as innocuous data like movie ratings, or adequately anonymized data, suddenly attain special significance with interlinking. As Danial J. Solove observes, “Viewed in isolation, each piece of our day-to-day information is not all that telling; viewed in combination, it begins to paint a portrait about our personalities.”[4]

B. Section 2: Freedom of expression and the right to privacy are mutually reinforcing rights (Page 11)

Existing text:

“Principle 8. Mandatory data retention

8.1 : Mandatory retention laws requiring internet and telecommunications service providers continuously to collect and preserve the content of users' communications, communications data and information about users' online activities and identity significantly interfere with the rights to freedom of expression and privacy

8.2 : The mandatory retention of communication data for any period beyond the originally

1 http://dataprivacylab.org/projects/identifiability/paper1.pdf
2 https://epic.org/privacy/reidentification/ohm_article.pdf
3 http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf
4 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=283924

stated purpose with a view to enabling access by law enforcement or intelligence agencies is disproportionate restriction on the rights to privacy and freedom of expression."

Suggested Revision:

"Principle 8 : Mandatory data retention

8.1 : Mandatory retention laws requiring internet and telecommunications service providers continuously to collect and preserve the content of users' communications, communications data and information about users' online activities and identity significantly interfere with the rights to freedom of expression and privacy.

8.2 : Private actors should not be mandated to retain content of users' communications, communications data and information about users' online activities and identity for any purpose, including enabling access by law enforcement or intelligence agencies.

8.3 : Only mandatory preservation of content of users' communications, communications data and information about users' online activities and identity may be allowed, subject to sufficient safeguards such as, but not limited to:

a) administrative, judicial and parliamentary oversight over the orders for preservation.

b) notice to target(s) of preservation orders either before or after preservation, as appropriate.

c) right / access to effective remedy against illegal data preservation orders.

d) limited period of data preservation, after which the data shall be destroyed."

Comment:

We agree with the contention in 8.1 that mandatory retention laws covering content and meta data significantly interfere with freedom of expression and privacy. Compelling information and internet service providers to retain data, whether content or traffic data, will only enhance the state's ability to carry out mass, unencumbered surveillance, making individuals vulnerable to human rights violation. Data retention is also expensive and prone to security risks in the form of theft, fraud and accidental disclosure. We are thus against any retention regime for any purpose.[5]

Besides, a data protection regime must run in parallel to a data preservation regime to ensure that private companies retain data for only the specified purpose and the specified period, beyond which the data is destroyed. When a state does not have such a fundamental data protection regime (like India), any kind of data collection by private

5 http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf

companies, including under a preservation regime, for any purpose, can be very dangerous to society.[6]

Instead, if the state requires communication or service providers to provide access to communications, a data preservation regime should be preferred. In a data preservation regime, only the data of the suspect(s) is collected, and hence, it is a far more balanced approach to data collection. However, even a data preservation system must be subject to safeguards that adequately ensures and protects the right to privacy. Data preservation regimes should be codified in law and must be precise. The reasons for which data shall be preserved must be clearly stated.[7] In Slovenia, for example, an order for data retention can only be issued if there are reasonable grounds to believe that a crime has been committed or is going to be committed, and that there are no means other than data preservation through which the crime can be investigated.[8]

Data preservation regimes must incorporate an oversight mechanism, which is a combination of administrative, judicial and parliamentary oversight. The target must be given notice that s/he or they are being subjected to communication surveillance either before or after the fact, as appropriate, and must also be given the opportunity to seek redress if such communication surveillance is unlawful.[9] Further, data collected must be preserved for only a fixed period. For example, in the Budapest Convention on Cybercrime, the maximum period of preservation is ninety days.[10] **Principle 10 : Data disclosure by companies** addresses few of the concerns raised above, but it largely relies on private companies to ensure that privacy rights are adequately protected. While this may be an additional safeguard, it is the primary duty of the state to protect the right to privacy of individuals. Any data preservation regime must codify procedural safeguards that the state must abide by, to protect the privacy rights of individuals.

Thus, as the noted in Report of the Special Rapporteur on the ‘promotion and protection of the right to freedom of opinion and expression’, Frank La Rue, “provision of communications data by the private sector to states should be sufficiently regulated to

6 https://www.eff.org/issues/mandatory-data-retention
7 http://www.europarl.europa.eu/meetdocs/2014_2019/documents/libe/dv/7_conv_budapest_/7_conv_budapest_en.pdf
8 http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/police-cooperation/data-retention/docs/drd_task_2_report_final_en.pdf
9 http://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session27/Documents/A.HRC.27.37_en.pdf
10 http://www.europarl.europa.eu/meetdocs/2014_2019/documents/libe/dv/7_conv_budapest_/7_conv_budapest_en.pdf

ensure that individuals' human rights are prioritized at all times."[11]

C. Order in which Principles 8, 9 and 10 appear

Existing order

Principle 8: Mandatory Data Retention

Principle 9: Mandatory user registration and real name requirements

Principle 10: Data disclosure by companies

Suggested order:

Principle 8: Mandatory Data Retention

Principle 9: Data disclosures by companies

Principle 10: Mandatory user registration and real name requirements

D. Section 4 : Reconciling freedom of expression, data protection and privacy (Page 19)

Existing text:

"Principle 22: Data protection exemptions

22.1 : Exemption from and/or limitations on the application of data protection principles should include a broad exemption for the exercise of freedom of expression.

22.2 : At a minimum, existing exemptions and/or limitations in data protection for the protection of journalistic, literary, academic and artistic purposes and in discharge of a legal obligation to make information publicly available, such as the maintenance of archives for historical or other public interest purposes or under right to information laws, must be interpreted broadly so as to give meaningful effect to the right to freedom of expression and information.

Suggested Revision :

"Principle 22: Data Protection exemptions

22.1 : Exemption from and/or limitations on the application of data protection principles should include exemption for the exercise of freedom of expression.

22.2 : However, any exemptions and/or limitations in data protection, which gives effect to the exercise of freedom of expression, including, the protection of journalistic, literary,

11 http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf

academic and artistic purposes and in discharge of a legal obligation to make information publicly available, such as the maintenance of archives for historical or other public interest purposes or under right to information laws, must be interpreted to give meaningful effect to right to privacy of data subjects.

22.3 : Measures to give meaningful effect to the right to privacy of data subjects may include:

a)Making it mandatory for institutions including universities, private actors, and publishers to set up ethics and privacy bodies to review research proposals, and the final research outputs that use data gathered from human data subjects.

b) Ensuring de-identification of data subjects prior to publication

c) Requiring prior approval of the relevant public authority that controls access to data, or state authority set up for this purpose, including a data commissioner, to interlink sensitive personal data

d) Encryption of sensitive data prior to access with decryption keys provided to only select individuals.

e) Restricting access to sensitive raw data to a predefined area as prescribed by the relevant public authority or state authority set up for this purpose, including a data commissioner.

f) Requiring the destruction of the data after it is used for the stipulated purpose."

Comment:

We agree with the rationale that data, especially Big Data, is very useful and beneficial, and that it is necessary for certain academic and journalistic purposes. The question then is not about whether access to data should be provided, "but rather how the benefits can be captured in a way that respects fundamental principles of ethics and privacy[12]". Consider, for example, the 2014 study conducted by Cornell University and Facebook in which changes in mood/ behavior of data subjects were studied by modifying their Facebook feed without taking their consent; or the release of genomic databases as part of a research study without taking into account the risk of reidentification of participants. States must regulate the use and processing of Big Data for academic purposes, in order to protect the privacy of the data subject. This may be done through: security controls that regulate access to data; and privacy controls that regulate how the data can be used .[13]

In the US, the Common Rule for the Protection of Human Subjects requires institutions

12 https://cyber.law.harvard.edu/node/99428
13 https://cyber.law.harvard.edu/node/99428

receiving federal funding involving human subjects to set up an Institutional Review Board (IRB), which will determine the ethical veracity of the project. Any research can commence only after IRB approval. .[14] Sometimes journals require proof of IRB approval before publication. IRB-like regulatory structures are important in Big Data research that have human subjects, as they can institute ex-ante safeguards against privacy violations. Additionally, such boards should extend their regulations to privately funded research as well.[15]

States may also adopt differentiated, tiered access to data. Access to sensitive data sets must be allowed only to specific individuals with purpose-limitation safeguards; and they must not be opened up to the public at large.[16] In Finland, when a researcher wants to access data stored with public authorities, the researcher must demonstrate compliance with data protection principles. If approval is granted, data is provided in encrypted form, to which only the researcher can have the decryption key.[17] Furthermore, for sensitive data like health data, states may regulate how the data maybe interlinked. In Singapore, any interlinking with the Ministry of Health's database requires prior approval of the Ministry. To further protect the privacy of the data subjects, staff conducting research involving interlinking can only conduct the research within the lab, and will also be constantly monitored.

Considering the risks in data release, especially sensitive personal data, access for academic, literary or journalistic purposes must be carefully counterweighted against privacy rights.

E. Section 5: Reconciling the right to information, data protection and the right to privacy (Page 21)

Existing text:

“Principle 25 : Official Records

25.2 : There should be a presumption that :

a) Court orders should be made public as anonymity orders can adequately protect the

14 https://www.nsf.gov/bfa/dias/policy/docs/45cfr690.pdf

15 http://www.forbes.com/sites/kalevleetaru/2016/06/17/are-research-ethics-obsolete-in-the-era-of-big-data/#5db717f01cb9

16 https://cyber.law.harvard.edu/node/99428

17 http://www.keepeek.com/Digital-Asset-Management/oecd/social-issues-migration-health/strengthening-health-information-infrastructure-for-health-care-quality-governance_9789264193505-en#page117

right to privacy where necessary;

b) Health records, because of their inherently sensitive nature, should be kept private unless there is strong countervailing public interest in publishing such information in individual cases;

c) Public records about children, whether medical or pertaining to social programmes, should not be published other than in an anonymized format."

Suggested Revision:

"Principle 25 : Official Records

25.2 : There should be a presumption that :

a) Court orders should be made public as anonymity orders can adequately protect the right to privacy where necessary;

b) Health records, because of their inherently sensitive nature, should be kept private unless there is a strong countervailing public interest in publishing such information in individual cases;

c) Access to public records on children, whether medical or pertaining to social programmes, should be subject to the fulfillment of conditions that protect the privacy of the children. This may include:

i) De-identification of data subjects through techniques such as : aggregation, privacy-aware methods for producing contingency tables, synthetic data, data visualizations, interactive mechanisms, multiparty computations etc., as deemed appropriate by the authority that controls the public record.

ii) Allowing access to sensitive raw data only after court order specifying the type of information that maybe accessed.

iii) Specifying persons or entities entitled to access the data.

Comment:

Privacy concerns cannot be deemed to be adequately met merely by deleting "personally identifiable information" from a database, or substituting it with a unique ID, as reasoned in the comment to the definition of the key term '**Personal data**'. Any person who has knowledge of certain personal information of an individual in a dataset may, by combining this knowledge with other information (interlinking), be able to discover the latter's personal data, including sensitive health information.[18] The target may even be a child. In fact in 1997, researcher Latanya Sweeny, in order to show the dangers of release of insufficiently

18 http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf

de-identified data, sent the Governor of Massachusetts, in the US, William Weld, his health record from the anonymized health data of state employees released by the Massachusetts Group Insurance Commission, by linking it with the electoral roll.[19] Once a database is vulnerable to re-identification, anyone who has access to that dataset may carry out re-identification. It is extremely important that adequate mechanisms to prevent any re-identification of sensitive databases, including restricted access to such datasets, are in place.[20]The presumption that anonymized data is privacy compliant does not hold.

Access to public records on children is necessary for academic or journalistic purposes, to hold the government accountable, to improve welfare services to children etc. However, since these records contain very sensitive data, access to these databases must be conditional to the privacy rights of the children. For instance, states may have statutes that list or describe persons or entities that have access to departmental records on children. Depending on the circumstances and the kind of records, persons or entities that may be permitted to access these records may include, medical professionals, law enforcement and court officials, researchers and persons who are subjects of these records.[21]

A few states in the US safeguard privacy of the data subjects in juvenile records by allowing access to non- de- identified, confidential information only after a court order. However, if the data is de- identified, non confidential, and in a general statistical format, court orders may not be required.[22]

Data released in an aggregated statistical form may not sufficiently safeguard the privacy of the data subject. This is because as the number of data points associated with the data subject increases, the uniqueness and probability of identification also increases. Therefore, authorities who control these databases must use new alternatives to traditional deidentification techniques, including: “contingency tables, synthetic data, data visualizations, interactive mechanisms, and multiparty computations”.[23]

19 http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/
20 http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf
21 https://www.childwelfare.gov/pubPDFs/confide.pdf#page=2&view=Persons%20or%20Entities%20Allowed%20Access%20to%20Records
22 http://juvenilerecords.jlc.org/juvenilerecords/documents/publications/national-review.pdf
23 https://cyber.harvard.edu/node/99428

F. Section 6 : Remedies and Sanctions (Page 22)

Existing text:

"Principle 26 : General Principles

26.1 : Redress mechanisms for privacy violations should be easy to use, quick and effective. Self regulatory or voluntary redress mechanisms, alternate dispute resolution schemes such as ombudspersons, and non pecuniary remedies should be preferred to court action."

Suggested Revision:

"Principle 26: General Principles.

26.1: Redress mechanisms for privacy violations should be easy to use, quick and effective. Self regulatory or voluntary redress mechanisms, alternate dispute resolution schemes such as ombudspersons may be used. However, the courts must remain the primary forum to redress privacy violations."

Comment:

The Group of Experts, constituted by the Planning Commission of India to devise a privacy and data protection framework for India, noted the following,

> "Alongside the National Privacy Principles, self regulating bodies will have the option of developing industry specific privacy standards that would be in conformity with the National Privacy Principles, which should be approved by a Privacy Commissioner."

At present, in India, we have no general data protection principles, nor is the right to privacy an explicit right in the Constitution. People have had to rely on the court to interpret constitutionally guaranteed fundamental rights, including the right to life[24], to redress breach of privacy. In such a scenario, to rely primarily on self regulation, and secondly, only approach courts when other remedies are exhausted is not an appropriate or a justified framework.[25] As the Group of Experts, cited above, suggests, self regulatory bodies must function **alongside a** nationally recognized data protection framework.

Furthermore, self regulation has not always been an effective alternative. Online Privacy Alliance (OPA), a self regulatory body set up in the US in the mid 1990s as a result of a

24 **R. Rajagopalan v. State of Tamil Nadu,**1995 AIR 264, 1994 SCC (6) 632

25 http://planningcommission.nic.in/reports/genrep/rep_privacy.pdf

threat by the FTC to regulate directly, issued guidelines for privacy policies. Apart from the fact that these guidelines heavily favored industry and did not regulate the collection of sensitive data, very few firms actually joined the group. Self regulatory systems are skewed heavily in favor of industry as there is no obligation to take public interest into account. Without any external enforcement mechanism, these bodies have also been unable to ensure implementation or compliance. The Network Advertising Initiative, another self regulation initiative in the US that came after the downfall of the OPA, was also a failure because of its inability to ensure compliance.[26]

While independent regulatory standards may be useful for specialized industries, it is important that they meet at least a common minimum threshold of data protection, a contention that must be judged by an independent regulator like a data protection commissioner/officer.[27] States will not be able to redress privacy violations, if in the absence of a 'state declared' data protection regime, and an institutional framework to implement it, self regulatory or voluntary redress mechanisms are primarily relied upon. Industry specific self regulatory systems must hence be part of a co-regulatory system.

G. Section 6 : Remedies and Sanctions (Page 23)

Existing text:

"Principle 29 : Prior restraint , super injunctions, mandatory pre-moderation and notice prior to publication.

29.3 : A legal requirement to give notice prior to publication to an individual whose right to privacy might be engaged so as to enable him or her seek an injunction is incompatible with the protection of the right to freedom of expression."

Suggested Revision:

"Principle 29 : Prior restraint , super injunctions, mandatory pre-moderation and notice prior to publication.

29.3: A legal requirement to give notice prior to publication to an individual whose right to privacy might be engaged so as to enable him or her seek an injunction is incompatible with the protection of the right to freedom of expression. However, this shall not override any provision for prior notice and consent, in any data protection law, to be given to the data subject prior disclosure of personal data to a third party, or use of personal data

26 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1758078
27 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1758078

beyond the original stated and consented purpose.

Comment:

Prior notice, consent, purpose specification and use requirement are important aspects of data protection laws across the world. If the data is used[28] or disclosed to a third[29] party for a purpose other than that specified in the original notice, fresh consent from the data subject is required.[30] The data subject may choose to opt out of further use or dissemination[31]. If the data is used/ processed in disregard of the data subject's objection, an interim injunction to prevent such privacy harm may be the only recourse available.

We contend that notice prior to publication is an essential aspect of data protection, and if the data subject institutes an interim injunction post such notice, s/he is in her/ his legal capacity to do so.

H. Section 6 : Remedies and Sanctions (Page 23)

Existing text:

"Principle 30 : Injunction

30.1 : Interim injunctions prohibiting the publication or further publication of private information (i.e. interim non-disclosure orders) should only be permitted by an order of a court in the most exceptional cases where all of the following conditions are met:

a) the applicant can show that he or she would suffer irreparable damage, which could not be compensated by subsequent remedies should publication or further publication take place ;

b) The court is satisfied that the applicant is likely to establish that publication or further publication should not be allowed ;

c) the court must have particular regard to the impact on freedom of expression, and where the proceedings relate to journalistic, literary or artistic material, the extent to which the material has or is about to become available to the public or the extent to which it is, or would be in public interest for the material to be published.

d) the court must have regard to the protection of the right to privacy.

---

28 OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm#part1

29 Diclosure to any third party is considered publication.

30 https://ico.org.uk/for-organisations/guide-to-data-protection/principle-1-fair-and-lawful/

31 https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-11-19_Big_Data_EN.pdf

Suggested Revision:

30.1 Interim injunctions prohibiting the publication or further publication of **personal data** (i.e. interim non-disclosure orders) should only be permitted by an order of a court in the most exceptional cases where **all or any** of the following conditions are met as **assessed by the court**:

a) the applicant can show that he or she would suffer irreparable damage, which could not be compensated by subsequent remedies should publication or further publication take place ;

b) The court is satisfied that the applicant is likely to establish that publication or further publication should not be allowed ;

c) the court must have particular regard to the impact on freedom of expression, and where the proceedings relate to journalistic, literary or artistic material, the extent to which the material has or is about to become available to the public or the extent to which it is, or would be in public interest for the material to be published

d) the court must have regard to the protection of the right to privacy.

Comment:

Firstly, we recommend that, since the phrase 'private information' has not been defined under the 'Definition of key terms', it should be substituted by the phrase personal data.

Secondly, we agree that it is important that judicial processes are not used as a tool to effectively curb the freedom of speech and expression, especially journalistic, literary and artistic freedoms. However, the theory of irreparable damage sits oddly with the discourse on Big Data. Instead of the traditional privacy concern about present damage/ harm, data protection jurisprudence focuses on the fact that data collection, processing and dissemination pose a risk of future harm. The harm may not be directly visible, but can alter the way we behave. For example, fear of surveillance could have a chilling effect on our free speech, but this may not be recognized as a personal irreparable damage.[32] Therefore, the threshold of irreparable damage may be redundant from the point of view of Big Data. It is hence that we recommend that on fulfillment of any, and not all the sub-clauses that the court grant injunction.

32 https://www.law.upenn.edu/journals/lawreview/articles/volume154/issue3/Solove154U.Pa.L.Rev.477(2006).pdf