



Big Data Analytics and National Data Sharing: A Case Study of Department of Science & Technology Government of India

**. Parveen Babbar
Deputy Librarian
Jawaharlal Nehru University, Delhi, India
E-mail: parveenbabbar@gmail.com**

Big Data

- ‘Massive amounts of digital information collected by companies and governments about human beings and our environment.’
- It is high volume, high velocity and high variety
- Expected 40,000 exabytes of data by 2020
- Government facing Vulnerabilities due to Large-scale cloud infrastructures, diversity of data sources and formats, the streaming nature of data acquisition and high volume inter-cloud migration.
- Karl Poppe said ‘Big Data explains everything but explain nothing’
- Nick Combs (EMC) says “It’s needle in a haystack or connecting the dots.”

Big Data

“Big Data is characterized not only by the enormous **volume** of data but also by the **diversity** and **heterogeneity** of the data and the **velocity** of its generation.” - **Farnam Jahanian** (NSF)

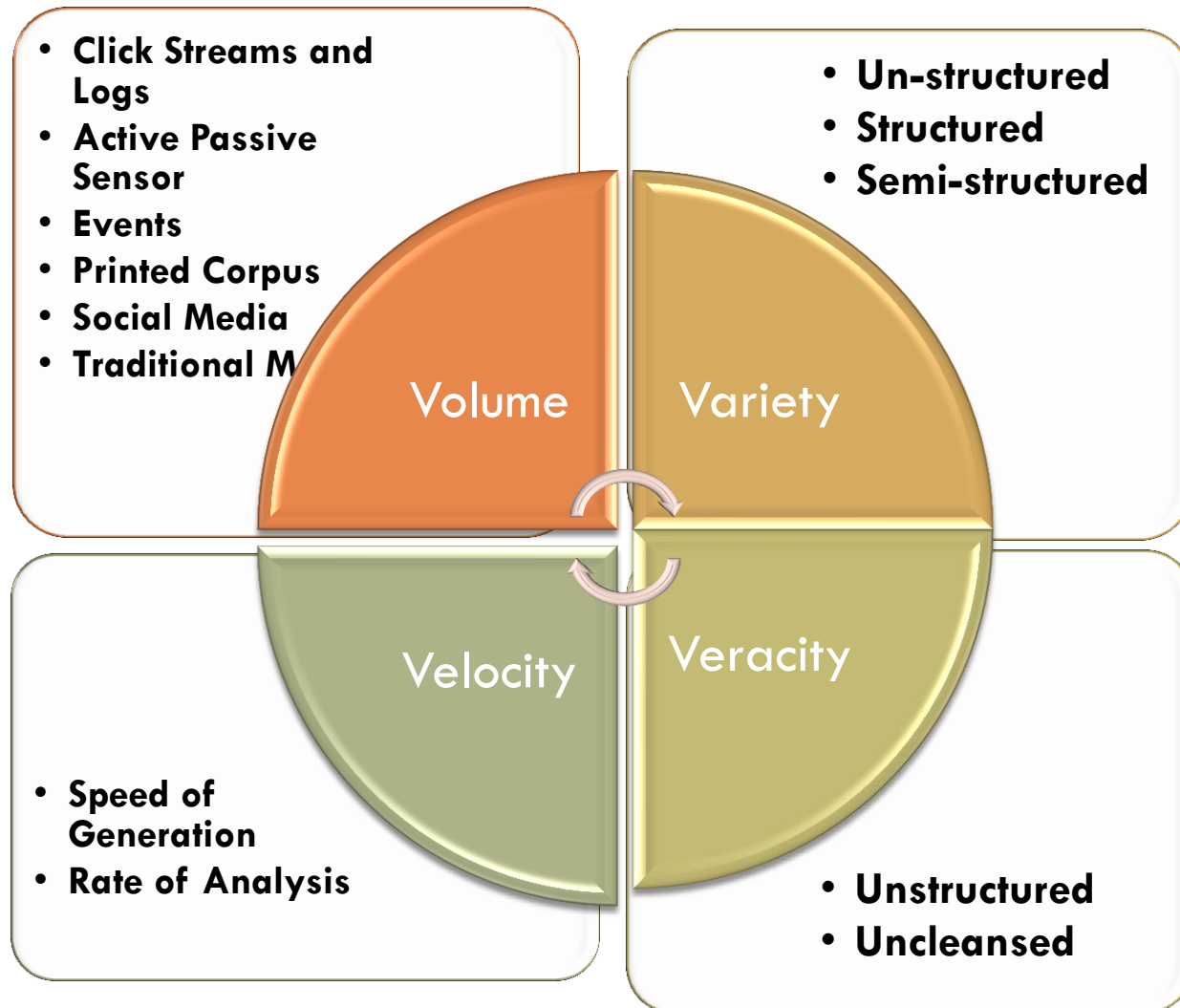
“it’s the **volume** and **velocity** and **variety** of data... to achieve new results for ...” **Nuala O’Connor Kelly** (GE)

Veracity: data in doubt and describe 'contradictory data,' or noisy data. **Arvind Krishna** (IBM)

Challenges

- Size of available material with Big Science having Big budgets, Big staffs, Big machines, Big laboratories
- Availability of limited tools and methods for locating and accessing the material
- Availability of tools and methods for meaningful content analysis of the material
- Limited time and resources
- The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization.
- The growing globalization of data flows via big data increases the risk that countries can lose control of their own data

Big Data and Four V's



4Vs

- Data at Rest – Terabytes to Exabyte's of existing data to process
- Data in Motion- Streaming data, milliseconds to seconds to respond
- Data in many forms – Structured, unstructured text, multimedia
- Data in Doubt – Uncertainty due to data inconsistency and incompleteness, ambiguities, latency, deception, model approximations.

Why Big Data

- **Big Data** analysis and data analytics promise new opportunities to gain valuable insights and benefits – new predictive modes of analysis;
- But, it will also enable **expanded surveillance**, increasing the risk of unauthorized use and disclosure, on a scale previously unimaginable.
- Performing data analytics on context-free data will only yield correlations (which at times, will be spurious);

Why Big Data

- Innovation and creation of new IT products and services
- Acceleration in the pace of discovery in almost every SE discipline
- Solve the country most pressing/vital challenges

Department of S&T, India

भारत सरकार
GOVERNMENT OF INDIA

विज्ञान और प्रौद्योगिकी मंत्रालय
MINISTRY OF SCIENCE AND TECHNOLOGY



विज्ञान एवं प्रौद्योगिकी विभाग
DEPARTMENT OF
SCIENCE & TECHNOLOGY

सरयमेव जयते



Home / About DST / Scientific Programmes / S&T Policies / Geo Information Services / Administration & Finance / Autonomous S&T Institutions / What's New

Compendium On Specific S&T Interventions By State S&T Councils



Agriculture

Energy

Construction

Land and Water



Cottage Industry &
Traditional Crafts

Health and Food

Rural Industry

Transportation

What's New

- PIRE - Funding Opportunity Announcement
- Call for Proposals under Indo-French (DST-Inria/CNRS MoU) Programme in ICST
- DST-JSPS Call for Proposal - 2016
- National Awards for Science & Technology Communication 2016

विज्ञान एवं प्रौद्योगिकी विभाग
National Awards for Science & Technology Communication 2016

मेरा देश बदल रहा है
आगे बढ़ रहा है

DEVELOPING AND TRANSFORMING INDIA

Progress in Pictures | Then & Now | Voice of India | Articles | Gallery

Rate My Government | Governance Quiz | Live Interaction

DST

Committed towards

- ▣ **Managing and processing**
- ▣ **Storage**
- ▣ **Archiving and Distribution**
- ▣ **Analysis**
- ▣ **Visualization**
- ▣ **Commercial cloud computing services**

Further

- ▣ **Creating Transparency**
- ▣ **Innovating new administrative/governance models, products, and services**
- ▣ **Discovering through experimentation for the needs, expose variability, and improve performance**
- ▣ **Segmenting populations to customize better actions and services**
- ▣ **Replacing/supporting human decision making with automated algorithms**

Information for all

- **Open Data platform:** <http://data.gov.in> facilitates proactive release of datasets in open format by the ministries/departments for use, reuse and redistribution. Provides open and easy access to information for citizens.
- **Engagement through social media:** Pro-actively engage through social media and web based platforms to inform and interact with citizens.
- **MyGov.in:** This website crowd sources ideas from the public for design of programmes such as better traffic management, using big data for making cities smarter etc.
- **Online messaging:** Online messaging to citizens on special occasions/programs would be facilitated through emails and SMS.

Securing the Big Data

0 Secure Storage and Infrastructure

- 0 Use of technologies such as Hadoop and MapReduce be Secured

0 Secure Data Management

- 0 Techniques for Secure Query Processing

0 Big Data for Security

- 0 Analysis of Security Data (e.g., Malware analysis)

0 Regulations, Compliance Governance

- 0 Storing, retaining, managing, transferring and analyzing Big Data
- 0 Compliance with the regulations
- 0 Privacy of the individuals both for the raw data but also for data integration and analytics
- 0 Defining the clear Roles and Responsibilities

0 Secure Internet of Things

Some other initiatives

- **Broadband for All –**
 - Rural: 2,50,000 village Panchayats to be covered under the National Optical Fibre Network (NOFN) by Dec 2016.
 - Urban: Virtual Network Operators would be leveraged for service delivery and communication infrastructure in new urban developments and buildings would be mandated.
- National Information Infrastructure (NII): NII is integrating the network and cloud infrastructure in the country to provide high speed connectivity and cloud platform to various government departments up to the panchayat level.

- Mobile coverage in a phased manner to the remaining 55,619 villages in the country with the cost of around Rs.16,000 Cr (appx. US\$ 2.3 billion) during 2014-18.
- Common Service Centres (CSCs) will be strengthened and increased from the current 135,000 to 250,000. CSCs – one in each Gram Panchayat, would be made viable & multi-functional end-points for delivery of government services.
- Post Offices as Multi-service centres: 150,000 Post Offices to be converted into multi service centres.

Data Analytics Applications

- Annoymization
- Aggregation
- Interpretation
- Processing
- Modeling
- Time Patterns
- Spatial Signatures
- Flows and presence
- Operations
- Data Mining
- Strategy

Data.gov.in



GOVERNMENT OF INDIA

A Digital India Initiative

Skip to navigation

Skip to main content



data.gov.in

Open Government Data (OGD) Platform India

Type search keyword



LOG IN | REGISTER



AT A GLANCE

26,367
RESOURCES

3,835
CATALOGS

101
DEPARTMENTS

7.13 M
TIMES VIEWED

2.83 M
TIMES DOWNLOADED

111
CHIEF DATA
OFFICERS

421
APIs

832
VISUALIZATIONS

VISUALIZATION

CATALOG



Annual Health Survey: Clinical, Anthropometric & Bio-chemical (CAB) Survey



HIGH VALUE DATASETS

TRANSPORT
TIMETABLES

GOVERNMENT
BUDGET

COMPANY
REGISTER

NATIONAL
STATISTICS

LEGISLATION

ENVIRONMENT

AGRICULTURE

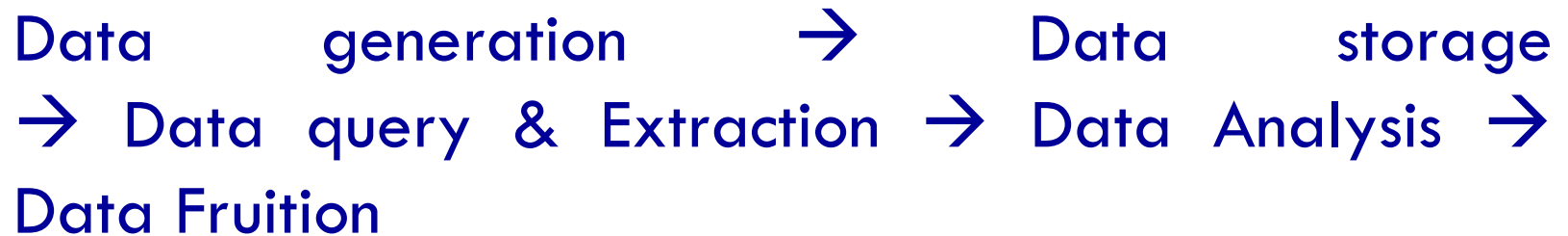
HEALTH
PERFORMANCE



MORE

Big data analytics

- Multi-level value chain of Big Data Analytics



- “Infinite” possibilities for multiple combinations
: so standardization to be applied at multiple
levels



- **A 12-digit unique identity for every Indian individual, including children and infants.**
- Establishes uniqueness of every individual on the basis of demographic and biometric information.
- Aadhaar will provide a universal identity infrastructure which can be used by any identity-based application (like ration card, passport, etc.)
- Coverage: **925 million and growing**
- Direct Benefit Transfer:
 - LPG
 - Scholarships
 - PDS scheme

E-TTAL

- ettal is a web portal for dissemination of e-Transactions statistics of National and State level e-Governance Projects including Mission Mode Projects.
- It receives transaction statistics from web based applications periodically on near real time basis.
- eTaal presents quick analysis of transaction counts in tabular and graphical form to give quick view of transactions done by various e-Governance projects.

National e-Transaction Count

Since 1st Jan, 2016

Since 1st Jul, 2016

3,96,59,91,931

17,82,90,393

Total Number of e-Services Integrated

3,051

- Promotes healthy competition amongst:
 - States/Union Territories
 - Departments and
 - Mission Mode Projects / e-Governance initiatives

E-Bhasha

- All government websites: Will move from bilingual to multilingual
- Digital India website: Available in 12 languages...Work in progress to cover all 22 constitutionally recognized languages

Digital Locker

- DigiLocker is a service launched by Government of India in February 2015 to provide a secure dedicated personal electronic space for storing the documents of resident Indian citizens. The storage space (maximum 1GB) is linked to the Aadhaar number of the user.
- **Advantages:**
 - Help citizens to digitally store their important documents like PAN card, passport, mark sheets and degree certificates.
 - Provide secure access to Government issued documents.
 - Eliminates the use of physical documents and enables sharing of verified electronic documents across government agencies.
 - A dedicated personal storage space in the cloud to citizens, linked to citizens Aadhaar number.
 - Reduce the administrative overhead of government departments and agencies created due to paper work.
 - Easy for citizens to receive services by saving time and effort as their documents will now be available anytime, anywhere and can be shared electronically.

Big data and public policy concerns?

- Correlation of disparate data such as healthcare, financial, demographic and location data.
- Tracking consumer behavior and sharing them with 3rd party without proper authorization for targeting and other purposes.
- Big data storage in the cloud across multiple geo boundaries
- Lack of transparency: who has access to which data, which data is collected and for what reason.

