# Data Privacy, Security and Management

Dr. Dipankar Dasgupta
Department of Computer Science
Hill Professor in Cybersecurity
Director, Center for Information Assurance
The University of Memphis

# Digital Research Data

https://science.energy.gov/funding-opportunities/digital-data-management/#DigitalResearchData

- Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.
- *Digital research data* are *research data* that can be stored digitally; accessed and transferred electronically.
- The term *digital data* encompasses a wide variety of information stored in digital form including: experimental, observational, and simulation data; codes, software and algorithms; text; numeric information; images; video; audio; and associated metadata. It also encompasses information in a variety of different forms including raw, processed, and analyzed data, published and archived data.

# Research Data Examples

https://science.energy.gov/funding-opportunities/digital-data-management/faqs/

Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This 'recorded' material excludes physical objects (e.g., laboratory samples). Research data also do not include:

(A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and

(B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.

# Terminologies

**Data Preservation:**

*Data preservation* means providing for the usability of data beyond the lifetime of the research activity that generated them.

**Data Sharing:**

*Data sharing* means making data available to people other than those who have generated them. Examples of data sharing range from bilateral communications with colleagues, to providing free, unrestricted access to the public through, for example, a web-based platform.

**Validate:**

*Validate* means to support, corroborate, verify, or otherwise determine the legitimacy of the research findings. Validation of research findings could be accomplished by reproducing the original experiment or analyses; comparing and contrasting the results against those of a new experiment or analyses; or by some other means.

https://science.energy.gov/funding-opportunities/digital-data-management/

# The Office of Science affirms that the following principles related to the management of digital research data directly support fulfillment of its mission.

- Effective data management has the potential to increase the pace of scientific discovery and promote more efficient and effective use of government funding and resources. Data management planning should be an integral part of research planning.
- Sharing and preserving data are central to protecting the integrity of science by facilitating validation of results and to advancing science by broadening the value of research data to disciplines other than the originating one and to society at large. To the greatest extent and with the fewest constraints possible, and consistent with the requirements and other principles of this Statement, data sharing should make digital research data available to and useful for the scientific community, industry, and the public.
- Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning.

# DATA STORAGE & PRIVACY ISSUES

Data in
Hard drive

Backup
(Onsite/Off
site)

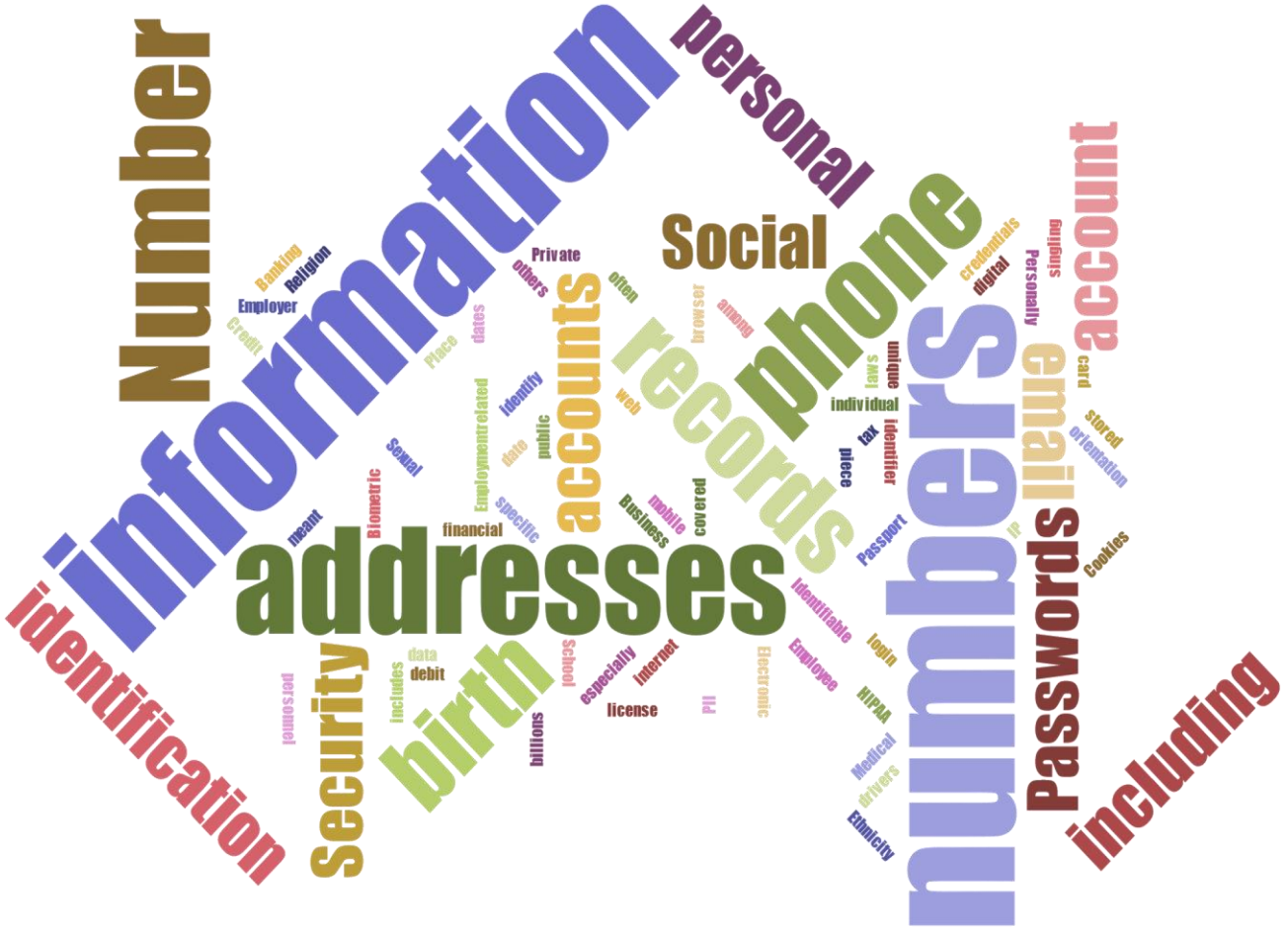

Network Attached
Storage

Data Centers

Cloud Storage


**Data is growing fast, as increasing Cloud Storage**

# Personally Identifiable Information (PII)



# Data Privacy Law

These laws[1] are based on Fair Information Practice that was first developed in the United States in the 1970s by the Department for Health, Education and Welfare (HEW). The basic principles of data protection are:

- For all data collected there should be a stated purpose.
- Information collected from an individual cannot be disclosed to other organizations or individuals unless specifically authorized by law or by consent of the individual
- Records kept on an individual should be accurate and up to date
- Data should be deleted when it is no longer needed for the stated purpose
- Transmission of personal information to locations where "equivalent" personal data protection cannot be assured is prohibited
- Some data is too sensitive to be collected, unless there are extreme circumstances (e.g., sexual orientation, religion)

[1]Wikipedia, “Information Privacy Law”, Last accessed on 20th April 2019. https://en.wikipedia.org/wiki/Information_privacy_law#United_States

# Data Privacy (NIST Guidelines[2])

NIST provided definite guideline about data privacy and protection. The basic principle of these guidelines are:

- **Privacy of personal information:**
  - Privacy of personal information involves the right to control when, where, how, to whom, and to what extent an individual shares their own personal information, as well as the right to access personal information given to others, to correct it, and to ensure it is safeguarded and disposed of appropriately.
- **Privacy of the person:**
  - This is the right to control the integrity of one's own body. It covers such things as physical requirements, health problems, and required medical devices.
- **Privacy of personal behavior:**
  - This is the right of individuals to keep any knowledge of their activities, and their choices, from being shared with others.
- **Privacy of personal communications:**
  - This is the right to communicate without undue surveillance, monitoring, or censorship.

[2] National Institute of Standards and Technology Interagency Report 7628, vol. 2 69 pages (August 2010)https://www.smartgrid.gov/files/Demand_Shifting_With_Thermal_Mass_in_Light_Heavy_Mass_Commer_201009.pdf

# Data Privacy & Security Policy: NIH

National Institute of Health NIH proposed some security policy principles to guide organizations in developing and implementing appropriate security plan. According to NIH any Precision Medicine Initiatives (PMI) should, at a minimum:

- Seek to preserve data integrity, so that participants, researchers, and physicians and other healthcare providers, can depend on the data.
- Identify key risks, and develop evaluation and management plans that address those risks, while still enabling science and research to advance.
- Provide participants and other relevant parties with clear expectations and transparent security processes.

For more information see the following link

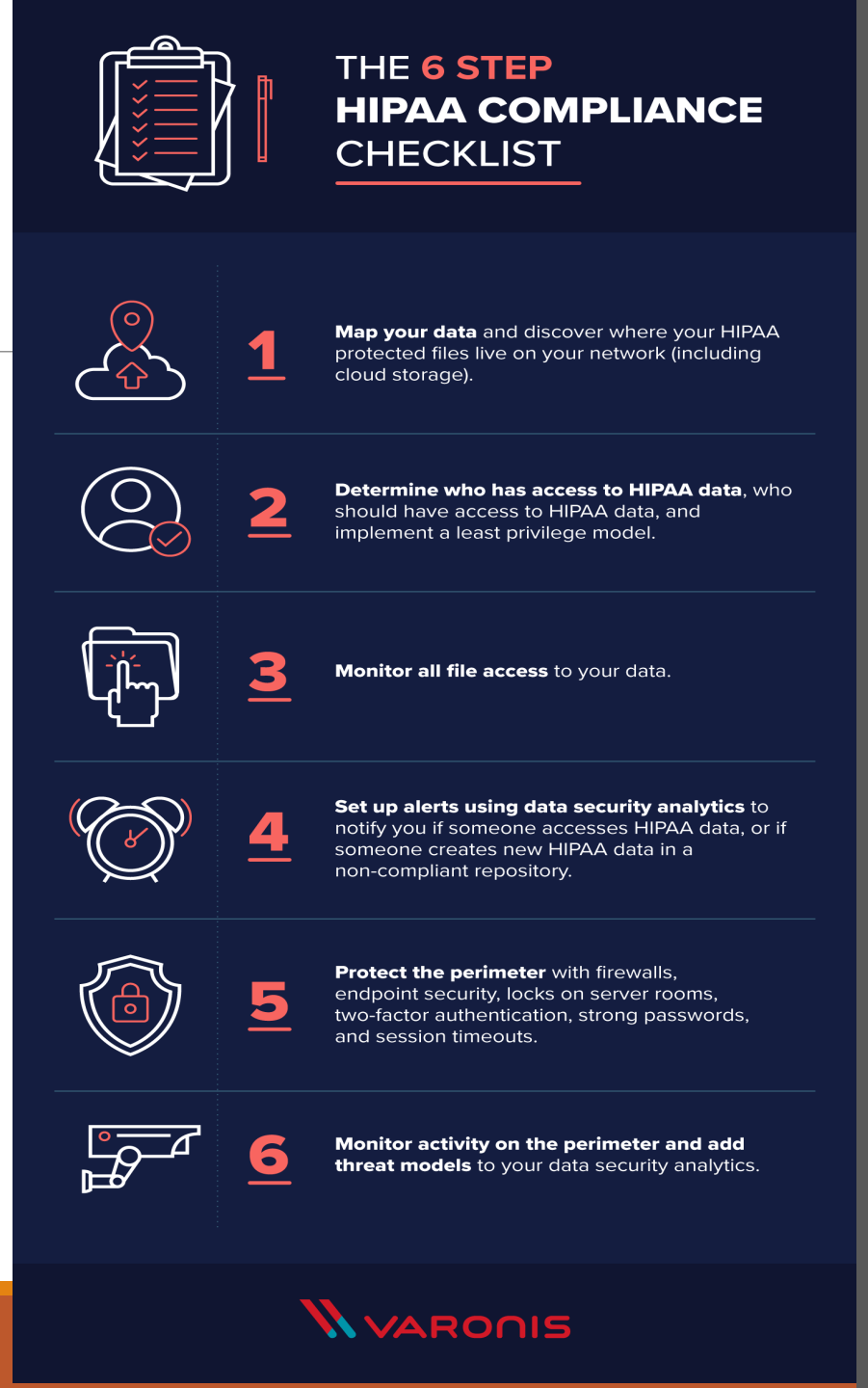
https://allofus.nih.gov/about/program-overview/precision-medicine-initiative-data-security-policy-principles-and-framework-overview/data-security-policy-principles

# Data Privacy: HIPAA

**HIPAA (Health Information Privacy and Portability Act)[3]**

One of the most prominent US data protection and privacy laws at the federal level is HIPAA—a data privacy regulation that was put in place to safeguard patient personal health information.

Healthcare providers maintains the 6 steps shown in the figure to make their product HIPAA compliant.




[3] Cindy Ng, Data Privacy: Definition, Explanation and Guidehttps://www.varonis.com/blog/data-privacy/

# New California Consumer Privacy Act

**Your personal information is being sold to businesses you don't even know exist.**

**The California Consumer Privacy Act will give you important new consumer privacy rights to take back control of your personal information, including:**

Your life is not their business.

https://www.caprivacy.org/about

# California Consumer Privacy Act (cont.)

CA ACT has **the following three major goals:**

**You will have the right to know what information large corporations are collecting about you**.

**You will have the right to tell a business not to share or sell your personal information..**

**You will have the right to protections against businesses which do not uphold the value of your privacy**...

The California Consumer Privacy Act will give you these important new rights on January 1, 2020.

**It's your personal information. Take back control!**

If you've come to our site from a search engine like Google or a social platform like Facebook, your information is possibly being collected, tracked, and shared by them. **Learn more** and/or **opt-out** of some sharing on CAPrivacy.org.

# EU General Data Protection Regulation (GDPR)

GDPR is legislation that will update and unify data privacy laws across the **European Union.** GDPR was approved by the EU Parliament on April 14, 2016 and goes into effect on May 25, 2018.

Enacted in May 2018, the GDPR aims to protect EU citizen personal data. There are many action items that companies in scope need to take to be become compliant, including but not limited to:

- Explicit opt-in consent
- The right to request their data
- The right to delete their data

**GDPR is set to affect many, and not just those in the EU. If an organization collects data online, they may soon have to make some tough choices.**

Under GDPR, companies may not legally process any person's personally identifiable information without meeting at least one of six conditions.

# GDPR guidelines:

1. Express consent of the data subject.
2. Processing is necessary for the performance of a contract with the data subject or to take steps to enter into a contract.
3. Processing is necessary for compliance with a legal obligation.
4. Processing is necessary to protect the vital interests of a data subject or another person.
5. Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.
6. Processing is necessary for the purposes of legitimate interests pursued by the controller or a third party, except where such interests are overridden by the interests, rights or freedoms of the data subject.

# GDPR compliance

**Organizations that must be GDPR compliant should take the time to visit the European Union's GDPR website to familiarize themselves with the requirements and penalties imposed by the regulation.**

For organizations that do business with or collect data on citizens of the European Union, there are three potential courses of action:

Stop all business activities related to the EU.

Find a way to do business without actually collecting any data.

Work toward compliance with the regulations.

https://searchdatabackup.techtarget.com/tip/Being-GDPR-compliant-is-not-just-a-concern-for-the-EU

# Analytics an uneasy balance between data collection and privacy

**In the age of GDPR and privacy regulations, attention must be paid to user privacy. Data management tools that employ AI as part of analytics can help achieve that.**

Advanced analytics, BI and AI are booming and can potentially offer great business benefits, but these technologies are extremely data hungry. Meanwhile, GDPR and other privacy regulations are forcing companies to re-evaluate how much data they collect and what they do with it.

**The corporate data warehouse has been stretched beyond its natural limits. Data now comes from such a variety of sources that traditional approaches are breaking down.**

**Data management has evolved from centralized data accessible by only the IT department to a flood of data stored in data in cloud data centers, mobile devices, App providers, etc.**

**Questions raises, Can privacy and Data analytics coexist?**

--We use AI to dynamically detect where information is private and where governance is needed.
**Emily Washington** Senior vice president of product management, Infogix Inc.

# NIST Data Privacy Workshop

The NIST Privacy Framework is currently under development. NIST envisions that it will be a voluntary tool for organizations to better identify, assess, manage, and communicate about privacy risks so that individuals can enjoy the benefits of innovative technologies with greater confidence and trust.

**Drafting the NIST Privacy Framework: Workshop #2**, on May 13-14, 2019, in Atlanta, Georgia!

# THANKS!