PHS-PRES: Multi-sided Protection For Population Health Data

Chenliu Yang, Jiahui Hu, Yunman Fan, An Fang
Institute of Medical Information,
Chinese Academy of Medical Sciences
Beijing P.R.China, 100020
yang.chenliu, hu.jiahui, fan.yunman, fang.an {@ imicams.ac.cn }

Abstract – This poster presents the challenges of population health data long-term protection, put forward the security system and multi-sided approaches to defend these threats.

Keywords – PHS-PRES, Long-term Preservation, Multi-sided Protection, Security Strategy, Safeguard Procedures

Conference Topics – Exploring the New Horizons; Building the Capacity & Capability.

I. Introduction

This poster offers the population health data preservation from the view of safeguard measures on long-term **Pres**ervation system for **P**opulation **H**ealth **S**ciences (PHS-PRES) platform constructed by **I**nstitute of **M**edical **I**nformation, **C**hinese **A**cademy of **M**edical **S**ciences (IMICAMS).

II. Values of Health Data

Population health data include clinical diagnosis, pharmaceutical research, public health, life omics, disease surveillance, population management and etc. They are valuable and significant on business, academic and society development.

1) Assist decision-making, health data are national basic resources and references to support policy formulation.

2) Promote scientific research, high-quality health data are capable to prop up innovation on scientific and technical development.

3) Accelerate medical development, provide helpful materials to inspection and diagnosis, and improve medical technology.

4) Deal with public emergency, through analysis of similar and relevant cases, provide treatment reference and suggestion.

III. Challenges of Data Archiving

On account of the values and particularity of health data, it is necessary to recognize the current and subsequent threats and risks of health data.

1) Constraint absence, no standard or guideline for staff management and health data daily maintenance, lack of supervision and detection mechanism.

2) Unauthorized disclosure, sensitive or private health data are available or exposed to unauthorized organizations, individuals, or entities.

3) Data loss, health data contents become incompatible, unreadable, damaged, missed within a decade or even long time of their appearance.

4) Medium failure, including breakdown, errors, and obsolescence of storage hardware and software, health data are not available.

5) Security issues, natural disaster to physical entities, internal and external attacks or destruction of health data, network, and system.

IV. Approaches for digital protection

PHS-PRES multi-sided protection framework is established for the purpose of figuring out these challenges mentioned above, represented by management criterion, security strategy, and safeguard details, shown as figure 1.

17th International Conference on Digital Preservation
iPRES 2021, Beijing,China.

DOI: 10.1145/nnnnnnn.nnnnnnn

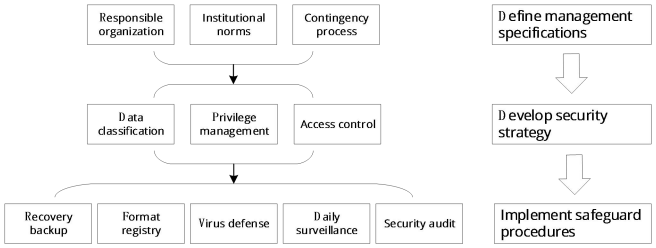

Figure 1 PHS-PRES Multi-sided Protection Framework

A. *Define management specifications*

1) Responsible organization, set up professional security team, design structure, scale and form, demarcate department, positon and privilege, and cooperate with related field or individuals to support health data security protection, escape intentional or unaware damage operation.

2) Institutional norms, according to policy and legislation, release the guideline of health data supervision, specify demands of archived data risk analysis, risk management, penalty policy, and log review, to discover and handle the hidden problems which may lead serious impact.

2) Contingency process, set up a series relevant solutions and suggestions to handle the directed problems immediately under emergency situation, avoid or reduce possibility of health data damage at the great extent.

B. *Develop security strategy*

1) Data classification, the first step of archiving is to distinguish data preservation level, classify data object based on category, attributes, distribution, access individual, access mode, utilization frequency and related secure elements. Commonly, high level means stricter safeguard, taking appropriate measures to protect data of each level will be efficient.

2) Privilege management, just provide the least privilege or forbid device access activities except for authorized staff, monitor operating status of PHS-PRES relevant equipment, once the unauthorized invade occurs, data will be transferred to other devices timely.

3) Access control, through HTTPS, VPN to build encryption path of chain, encrypt the data object by appropriate and security digest algorithm, as well as employ data or medium encryption technology in archiving. Unauthorized account is restricted to access or disclosure protected object, safeguard confidentiality of high level health data.

C. *Implement safeguard procedures*

1) Recovery backup, with the purpose of PHS-PRES disaster recovery (DR) and services continuity, dual-live system to make sure the platform normal operation, and support several backups of health data in different mediums and locations, to recover changed or damaged data rapidly without data loss.

2) Format registry, based on accredited format database, record original format of deposit data object, follow the latest updates of format and relevant software under rapid technology development, provide migration recommendation in case of obsolescence issues.

3) Virus defense, deploy antivirus hard software, divide safety areas in PHS-PRES network, not only focus on traditional cyber-attack types and purpose (steal data, break system), but also pay attention to the increased invade purpose (contaminate data, disturb analysis result).

4) Daily surveillance, realize "exante", "interim" warning and dispose timely under real-time monitoring, make sure PHS-PRES is capable to supervise and discover storage failure, unauthorized access, malicious operation, and other threating actions, dispose breakdown and errors in time, optimize detection mechanism.

5) Security audit, design inspection checklist, verify the integrity, coherence, accessibility of health data regularly, and check normal flow rate, error state, and security incident based on log information in period. Besides, improving the responsibility confirmation, performance improvement, and security evaluation to promote supervision abilities.

V. ACHIEVEMENTS

Currently, PHS-PRES multi-sided protection system has already been built and applied, provide safeguard services to all archived resources, and have good effect to some extents.

VI. FURTHER WORK

With the guide of Chinese "*Cyber Security Law*" [1], "*Data Security Management Regulation*" [2], "*Personal Information Protection Act (Draft)*" [3], based on the layered construction and classified protection, from the perspective of data-centric security system, PHS-PRES is still working on closed loop protection throughout data collection,

acceptance, ingestion, management and access, attempt to create an approximately comprehensive protection system for population health data in future.

VII. ACKNOWLEDGMENTS

This work is supported by Chinese Academy of Medical Sciences (CAMS) project, i.e., The Long-term Preservation of National Population Health Data.

REFERENCES

[1] Cyber Security Law (2016), Cyberspace Administration of China. http://www.cac.gov.cn/2016-11/07/c_1119867116.htm

[2] Data Security Management Regulation (2019), Ministry of Justice of the People's Republic of China. http://www.chinalaw.gov.cn/government_public/content/2019-05/28/657_235862.html

[3] Personal Information Protection Act (Draft) (2020), The National People's Congress of the People's Republic of China. http://www.npc.gov.cn/npc/c30834/202010/569490b5b76a49c292e64c416da8c994.shtml