A Privacy-preserving Approach for Records Management in Cloud Computing

Eun Park and Benjamin Fung

School of Information Studies

McGill University

Digital transformation

- Information sharing
- Information integration
- Data mining

Conflict?

Privacy

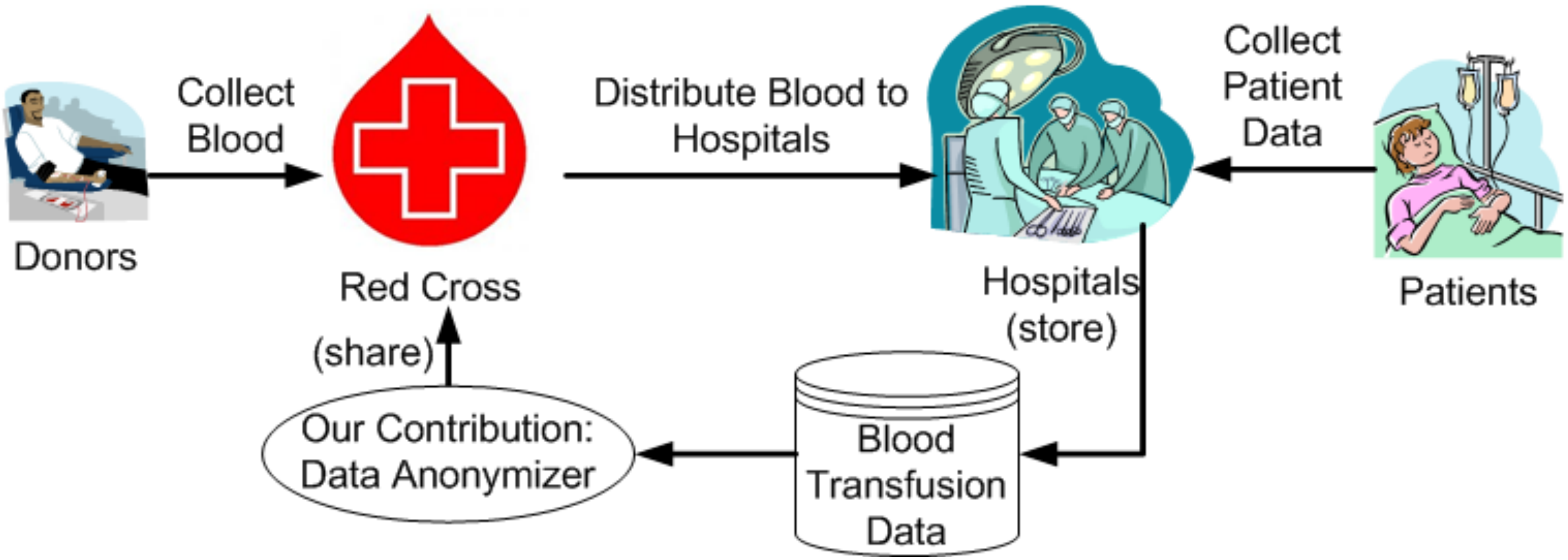
- Personal Health Information Protection Act (PHIPA), 2004

Can we share and integrate person-specific data to support effective data mining without compromising individual privacy?

Project Background

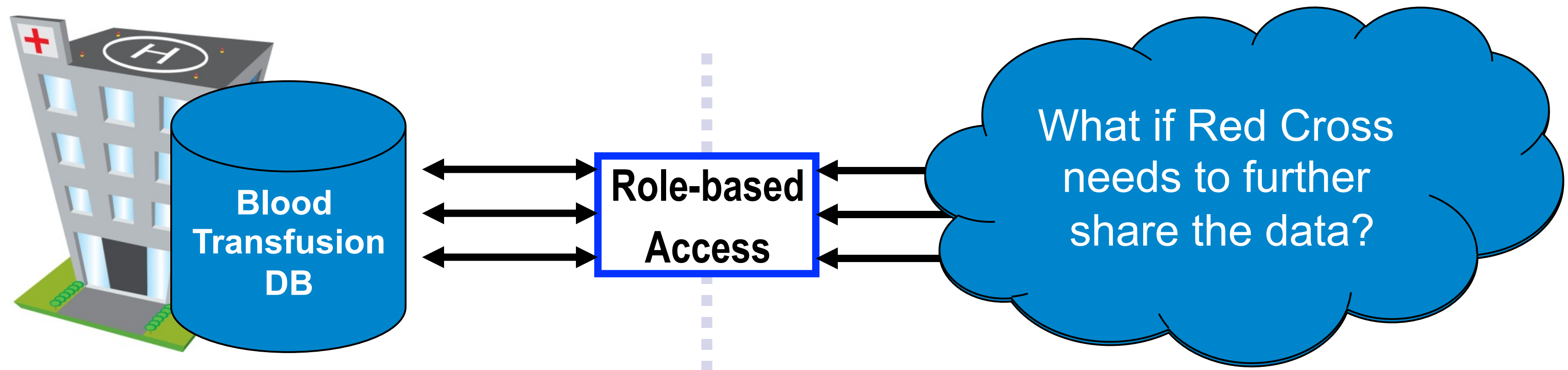
- Location: Hong Kong (Population - 7 million)
- Organizations:
  - Red Cross Blood Transfusion Service
  - ~30 public hospitals

Red Cross Blood Transfusion Service

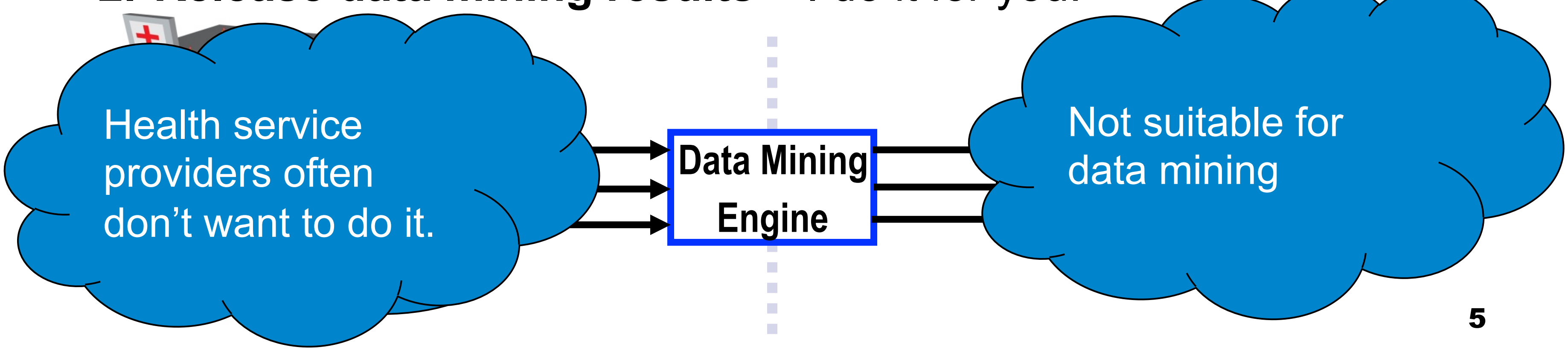


Options for Health Service Manager

1. **Role-based access control** – you are responsible for it!



2. **Release data mining results** – I do it for you!



Objectives

1. Identify the technical challenges of **hosting person-specific information on cloud through the lens of security and privacy.**
2. Evaluate the state-of-the-arts privacy-preserving techniques and their applicability in cloud.
3. Study the readiness of the Canadian health and government agencies to use cloud computing.
4. Develop a privacy-preserving Data-as-a-Service (DaaS) system for hosting person-specific information.
5. Make recommendations on privacy-preserving DaaS for government and health agencies.

Methodology

1. Examine the current legal guidelines of privacy management in the United States and Canada. Draw the criteria to evaluate privacy management.
2. Examine the available security and privacy-preserving techniques, mechanisms and tools.
3. Conduct case study to test how to implement the technique and tool and minimize the privacy risk with sample records and data.
4. Make suggestions on how to manage security and privacy risks in records and data management at government and health agencies.

Examining Three Guidelines: PIPEDA, Privacy Act, HIPPA

PIPEDA

- *Personal Information Protection and Electronic Documents Act*
- Received Royal Assent on April 13, 2000 and implemented on January 1, 2001.
- “Sets out ground rules for how private sector organizations can collect, use or disclose personal information in the course of commercial activities.”
- To balance an individual's privacy rights with the need of organizations in private sector
- To promote consumer trust in electronic commerce
- Obtaining consent and identifying the purpose for the collection of personal information

Privacy Act of 1974

- Enacted on September 27, 1975.
- To balance the government’s need to maintain information about individuals with the rights of individuals to be protected against unwarranted invasions of their privacy
- Provides the Government with a framework to conduct its day-to-day business when that business involves the collection or use of information about individuals.

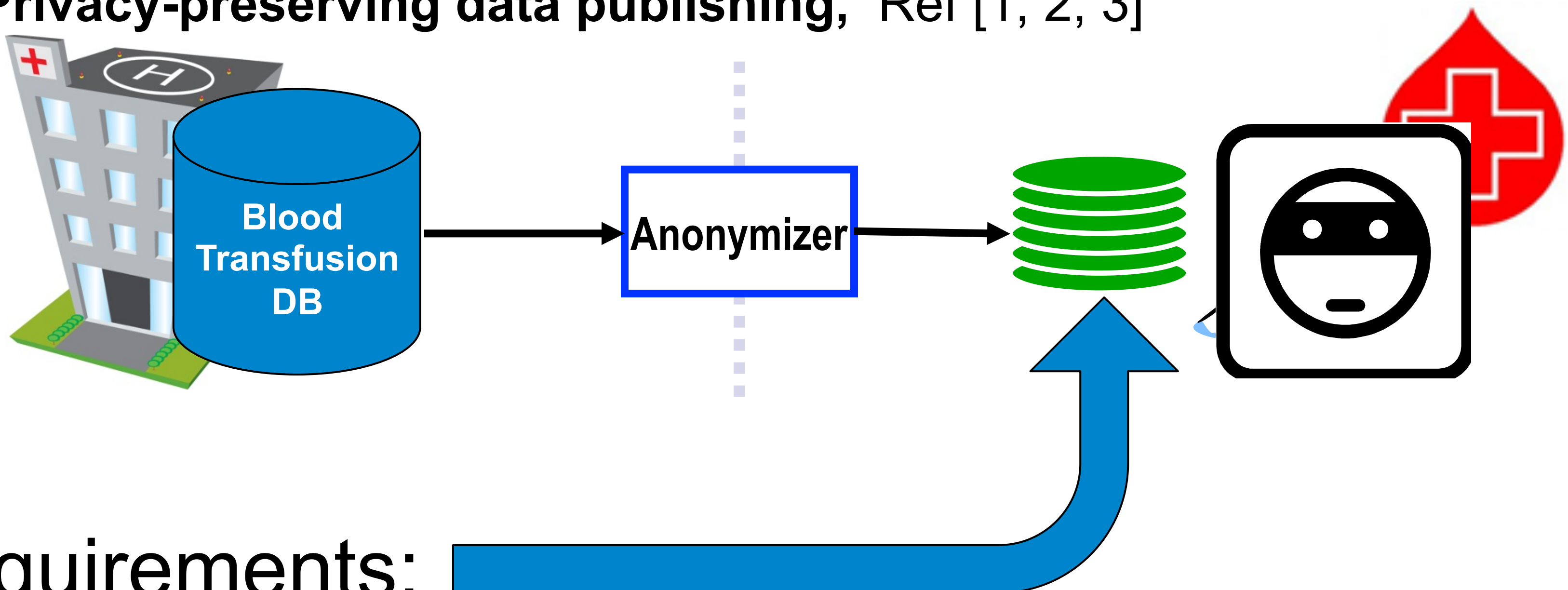
HIPPA

- *Health Insurance Portability and Accountability Act* of 1996
- Enacted August 21, 1996.
- To assure that individuals' health information is properly protected
- To make it easier for people to keep health insurance, protect the confidentiality and security of healthcare information and help the healthcare industry control administrative costs
- Applies to health providers who transmits health information in electronic form

State-of-the-arts privacy-preserving techniques for different scenarios

Scenario #1: Single Provider, Single Release

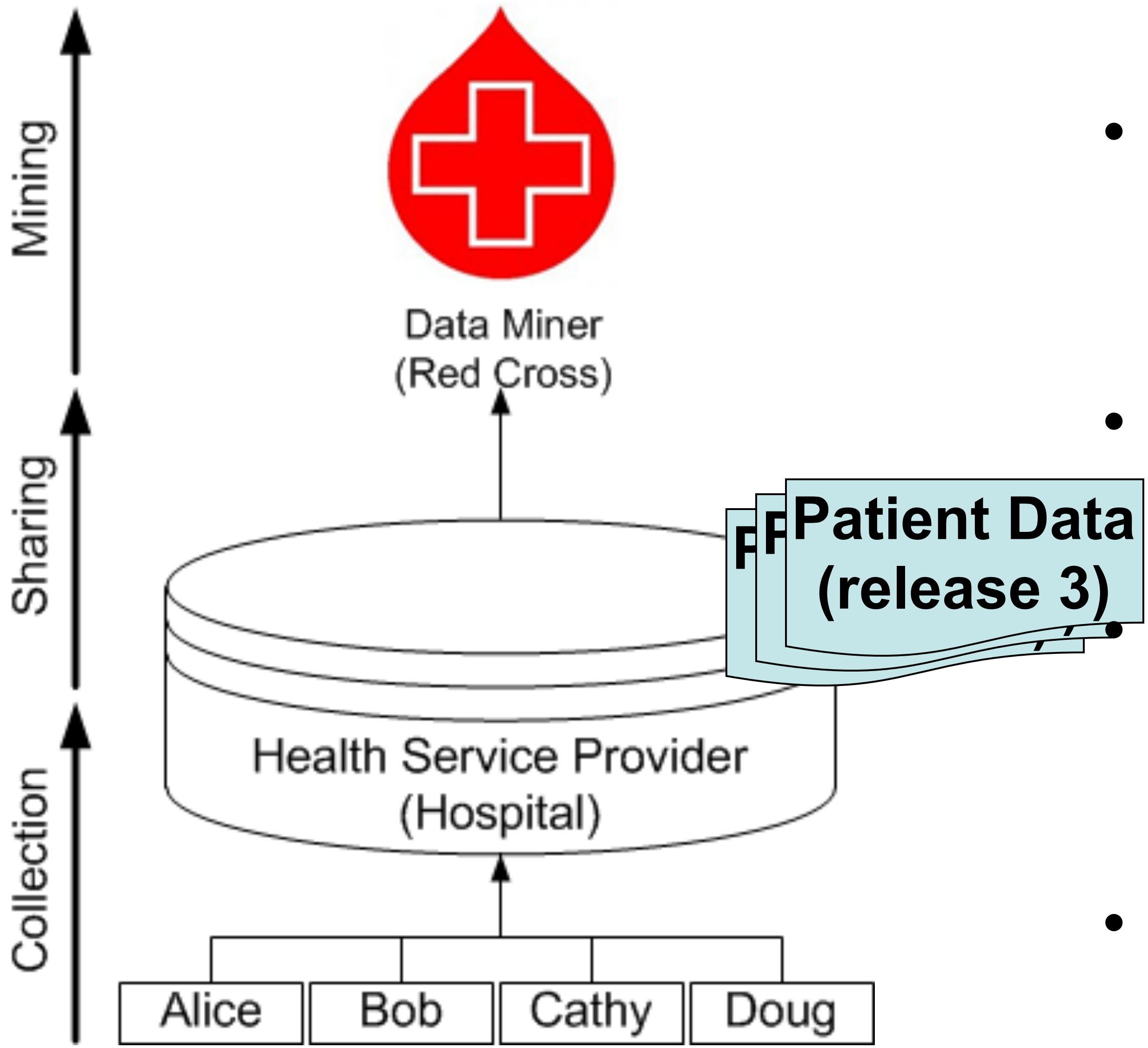
Privacy-preserving data publishing, Ref [1, 2, 3]



Requirements:

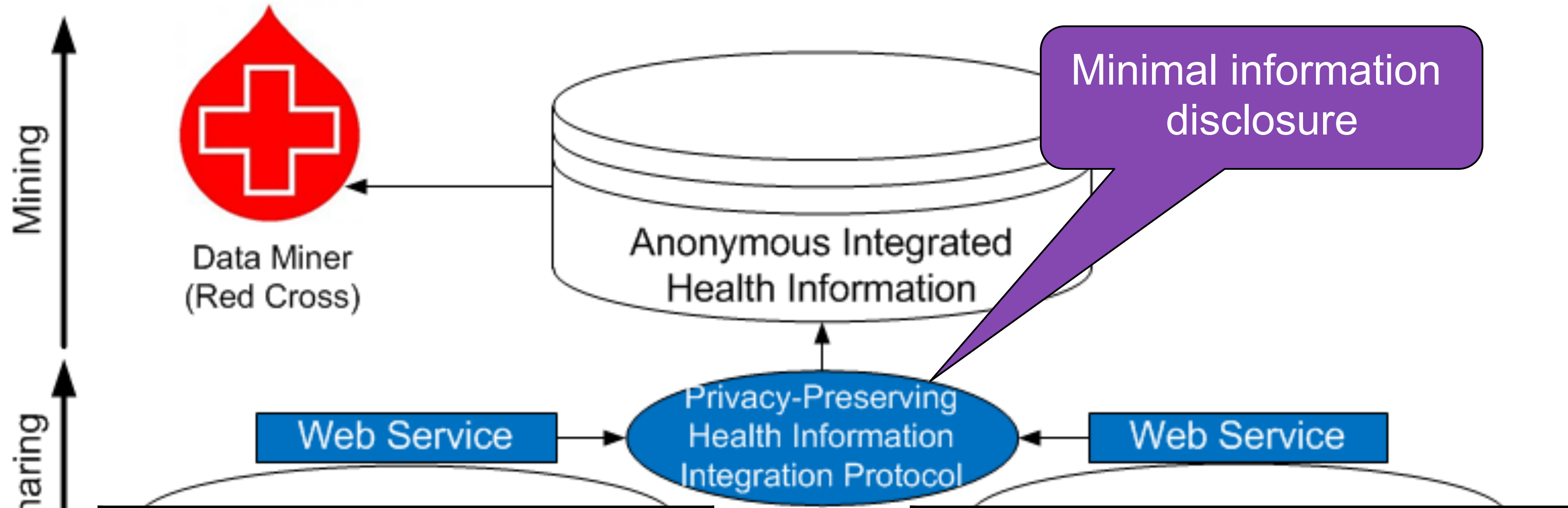
1. Prevent attacker from inferring sensitive information.
2. Keep the useful information for data analysis.

Scenario #2: Sequential Release



- Data are released sequentially in multiple versions.
- New information become available.
- Tailored view for each data sharing purpose.
- Ref [4, 5, 6]

Scenario #3: Collaborative Data Integration

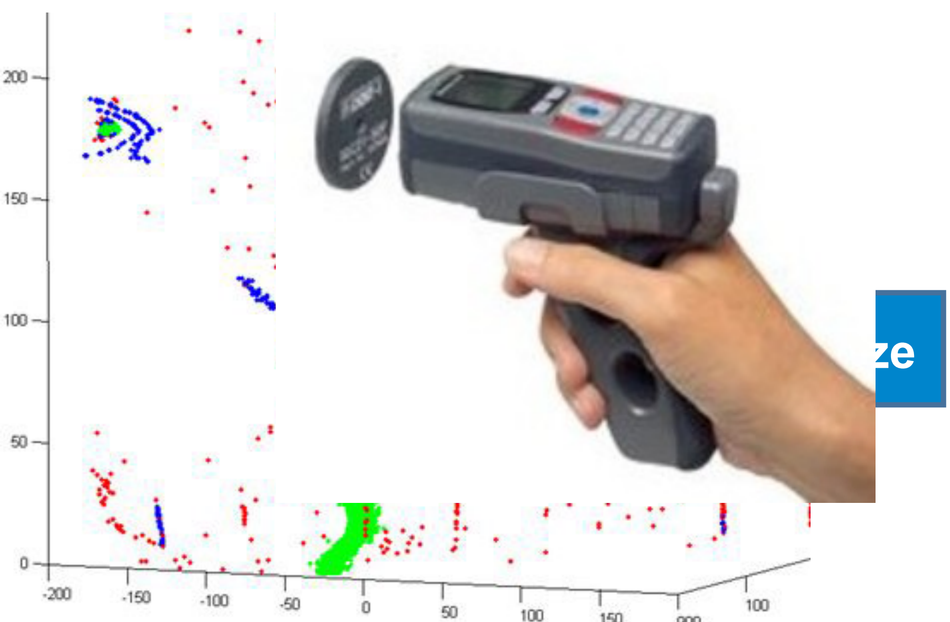


Hospital A		
Patient	Blood Group	Wgt.
Alice	A	88
Bob	B	60
Cathy	AB	88
Doug	O	60

Hospital B		
Patient	Blood Group	Wgt.
Emily	A	76
Frank	A	76
Gary	B	88
Harry	O	60

Ref: Horizontally-partitioned data [7], Vertically-partitioned data [8], Game theory [9]

Scenario #4: RFID Trajectory Data Release



Ref: [10][11]

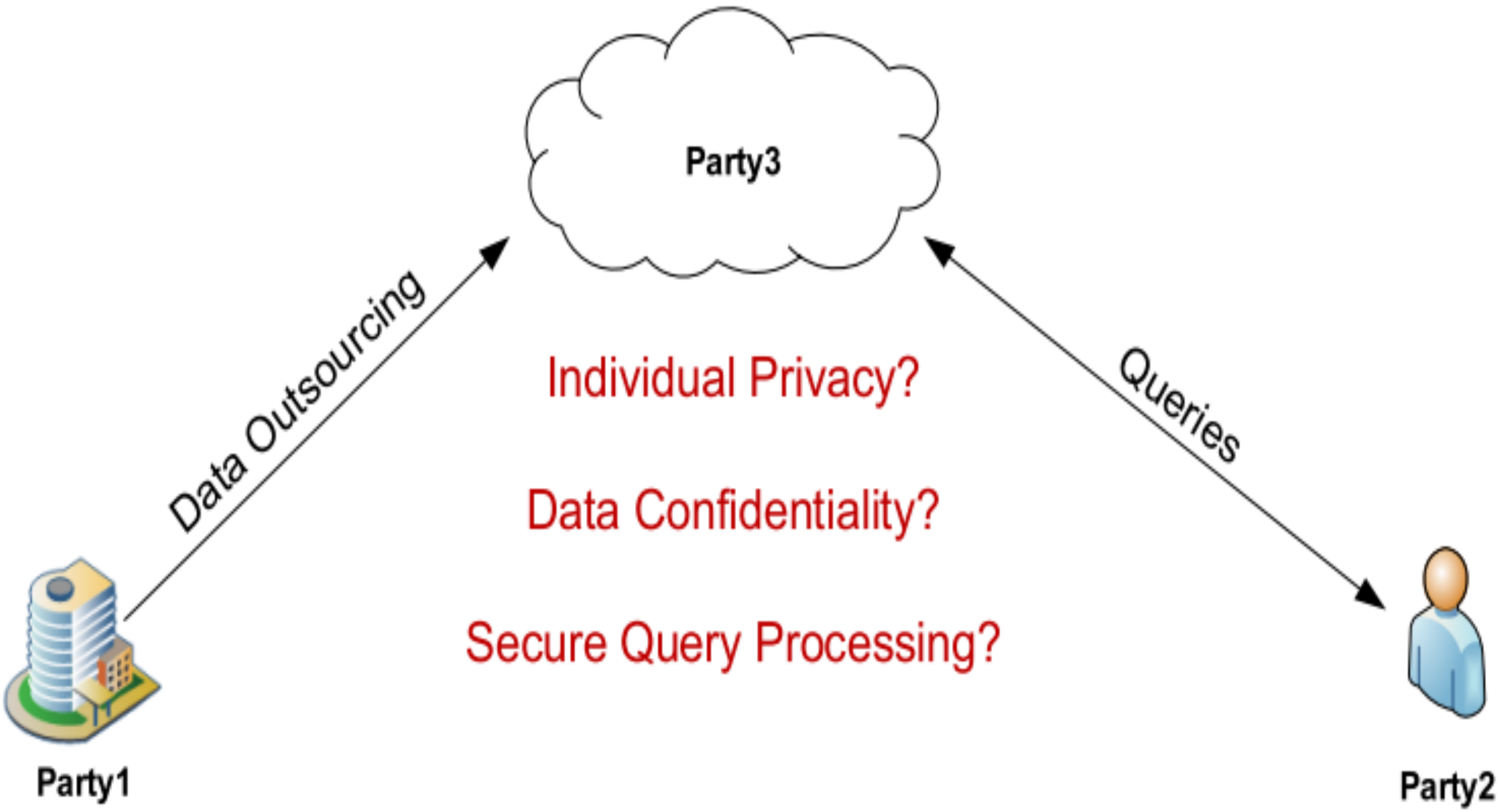
Patient-specific trajectory table		
EPC	**Trajectory**	**Department**
100	a1 → d2 → b3	Maternity
101	b3 → e4	Cardiology
102	b3 → c7 → e8	Radiology
103	d2 → f6	A&E
104	d2 → c5 → f6	ICU

Scenario 5:

Hosting private data on the cloud

Confidentiality-Preserving Query Processing on Anonymized Data in the Cloud

- Data confidentiality, privacy of personal information, and secure access to the data are major concerns.



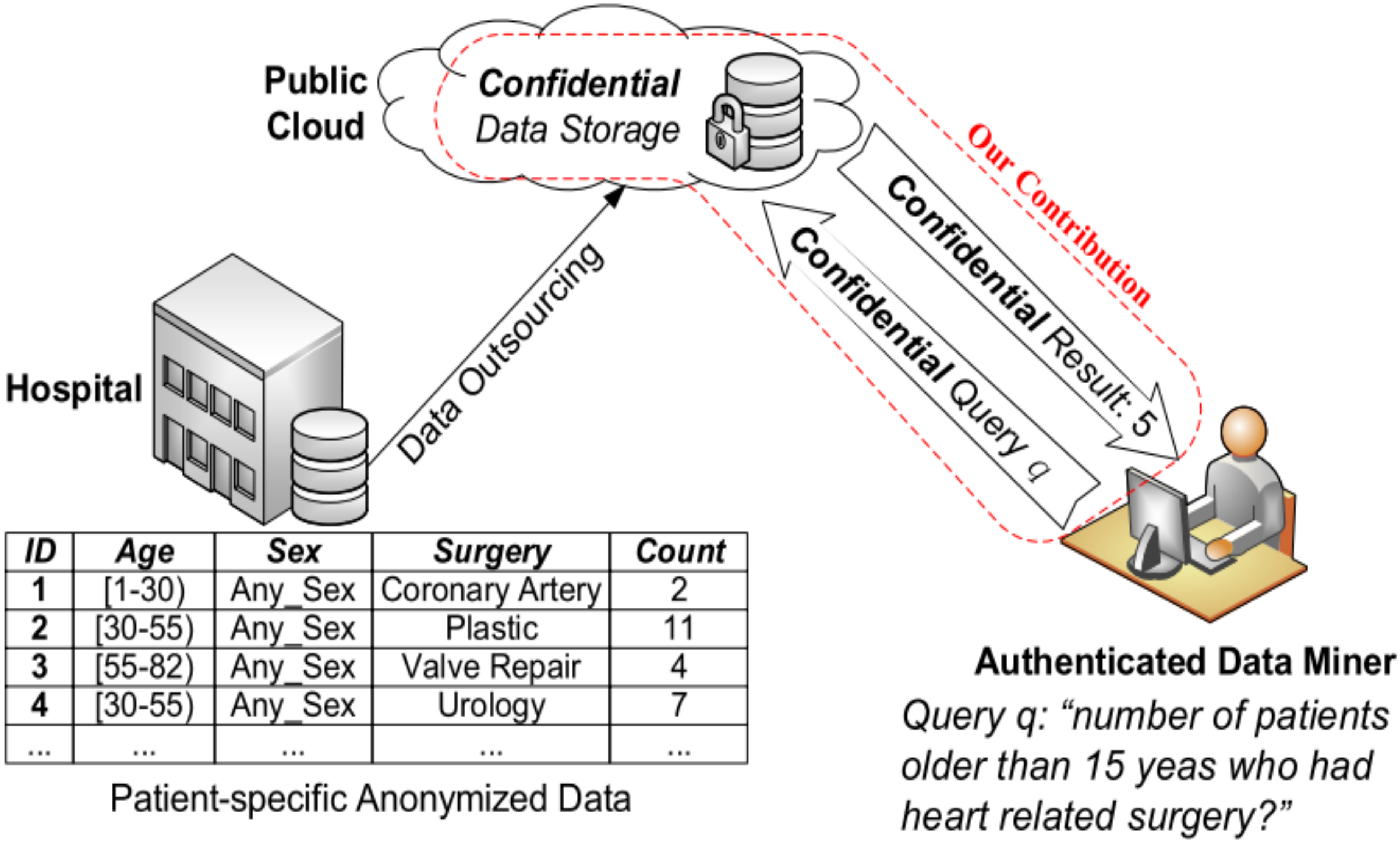
Objectives

1. To support privacy-preserving data outsourcing on the cloud
2. To provide data miners (researchers) access to the privacy-preserved patient information

Security and privacy requirements:

1. Assuming the cloud is not trusted.
   - Prevent the cloud from accessing the raw patient data → confidential data storage
   - Protect the confidentiality of the data miner's queries → confidential query
2. Assuming the data miner is not trusted
   - Prevent a data miner from inferring sensitive information or linking a target patient from the record → confidential results

Outsourcing Scenario



ID	Age	Sex	Surgery	Count
1	[1-30)	Any_Sex	Coronary Artery	2
2	[30-55)	Any_Sex	Plastic	11
3	[55-82)	Any_Sex	Valve Repair	4
4	[30-55)	Any_Sex	Urology	7
...	...	...	...	...

Patient-specific Anonymized Data

Description of the Diagram

1. Hospital first anonymizes the data, encrypts the data, and uploads it to the cloud. The cloud only sees the encrypted data.
2. When a data miner needs to get obtain data, the data miner submits a query to the hospital. The hospital authenticate the data miner's identity, and returns an encrypted query to the data miner.
3. The data miner submits the encrypted query to the cloud. The cloud processes it and answers the encrypted answer to back to data miner. The data miner decrypted the result.

- Properties
  - The cloud can answer all the queries, but the cloud itself does not know anything about the data nor the queries.
  - Though the data miner can see the decrypted result, the decrypted result is still anonymous, and the data miner cannot link a target patient to specific record.

Next Step

- Examine the basic requirements of the three legal guidelines in order to draw the criteria to evaluate privacy management.
- Conduct a case study to test how to implement the technique and tool with sample records and data at a government site or health related institution.
- Make suggestions on how to manage security and privacy risks in records and data management at government and health agencies.

Timeline

- Literature review
  - May 2014 – July 2014
- Evaluation of tools
  - July 2014 – December 2014
- Development and system evaluation: case study
  - November 2014 – September 2015
- Finalize by December 2015

References

1. B. C. M. Fung, K. Wang, A. W.-C. Fu, and P. S. Yu. Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques, ser. Data Mining and Knowledge Discovery. 376 pages, Chapman & Hall/CRC, August 2010. [ISBN: 9781420091489]
2. N. Mohammed, X. Jiang, R. Chen, B. C. M. Fung, and L. Ohno-Machado. Privacy-preserving heterogeneous health data sharing. Journal of the American Medical Informatics Association (JAMIA), 20(3):462-469, May 2013. BMJ.
3. N. Mohammed, R. Chen, B. C. M. Fung, and P. S. Yu. Differentially private data release for data mining. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), pages 493-501, San Diego, CA: ACM Press, August 2011.

References

4. X. Xiao, & Y. Tao (2007, June). M-invariance: towards privacy preserving re-publication of dynamic datasets. In Proceedings of the 2007 ACM SIGMOD international conference on Management of data (pp. 689-700). ACM.
5. K. Wang and B. C. M. Fung. Anonymizing sequential releases. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), pages 414-423, Philadelphia, PA: ACM Press, August 2006.
6. B. C. M. Fung, K. Wang, A. W. C. Fu, and J. Pei. Anonymity for continuous data publishing. In Proceedings of the 11th International Conference on Extending Database Technology (EDBT), pages 264-275, Nantes, France: ACM Press, March 2008.
7. D. Alhadidi, N. Mohammed, B. C. M. Fung, and M. Debbabi. Secure distributed framework for achieving ε-differential privacy. In Proceedings of the 12th Privacy Enhancing Technologies Symposium (PETS), LNCS 7834, pages 120-139, Vigo, Spain: Springer-Verlag, July 2012.

References

8. N. Mohammed, D. Alhadidi, B. C. M. Fung, and M. Debbabi. Secure two-party differentially private data release for vertically-partitioned data. IEEE Transactions on Dependable and Secure Computing (TDSC), 11(1):59-71, January/February 2014. IEEE Computer Society.
9. N. Mohammed, B. C. M. Fung, and M. Debbabi. Anonymity meets game theory: secure data integration with malicious participants. Very Large Data Bases Journal (VLDBJ), 20(4):567-588, August 2011. Springer.
10. R. Chen, B. C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang. Privacy-preserving trajectory data publishing by local suppression. Information Sciences (INS): Special Issue on Data Mining for Information Security, 231:83-97, May 2013. Elsevier.
11. R. Chen, B. C. M. Fung, B. C. Desai, and N. M. Sossou. Differentially private transit data publication: a case study on the Montreal transportation system. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), pages 213-221, Beijing, China: ACM Press, August 2012.