

IEEE SmartSys, 18 June 2018

# TSAR: a fully-distributed Trustless data ShARing platform

Hanqing Wu\*, Jiannong Cao\*, Shan Jiang\*, Ruosong Yang\*,  
Yanni Yang\*, Jianfei He†

\*The Hong Kong Polytechnic University, Hong Kong, China

†Huawei Technologies Co. Ltd., Shenzhen, China

# Table of Contents

- Background
- Related Works
- System Goal & Design Principles
- System & Module Design
- System Implementation & Evaluation
- Acknowledgments

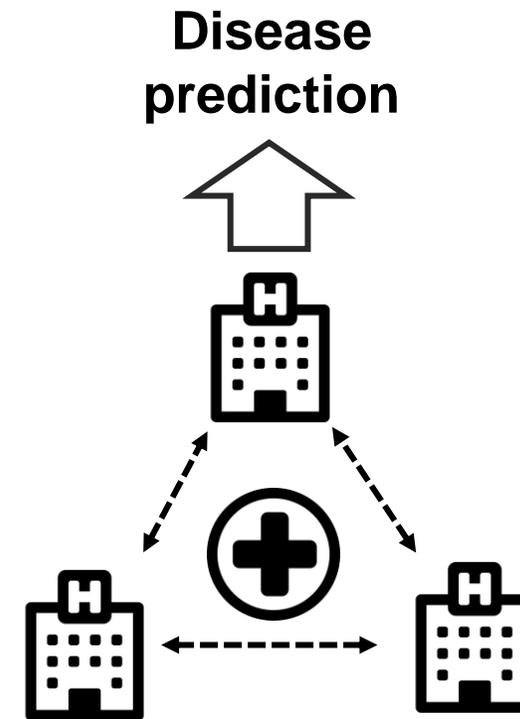
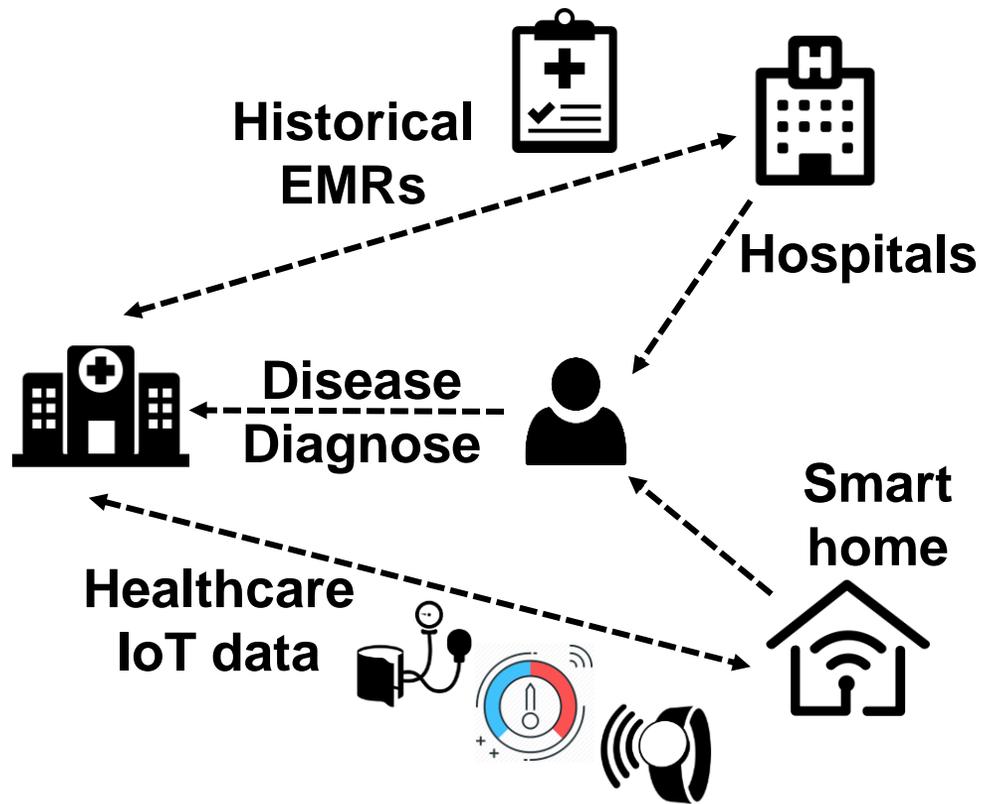
# Big Data Era

- Numerous big data applications benefit human beings



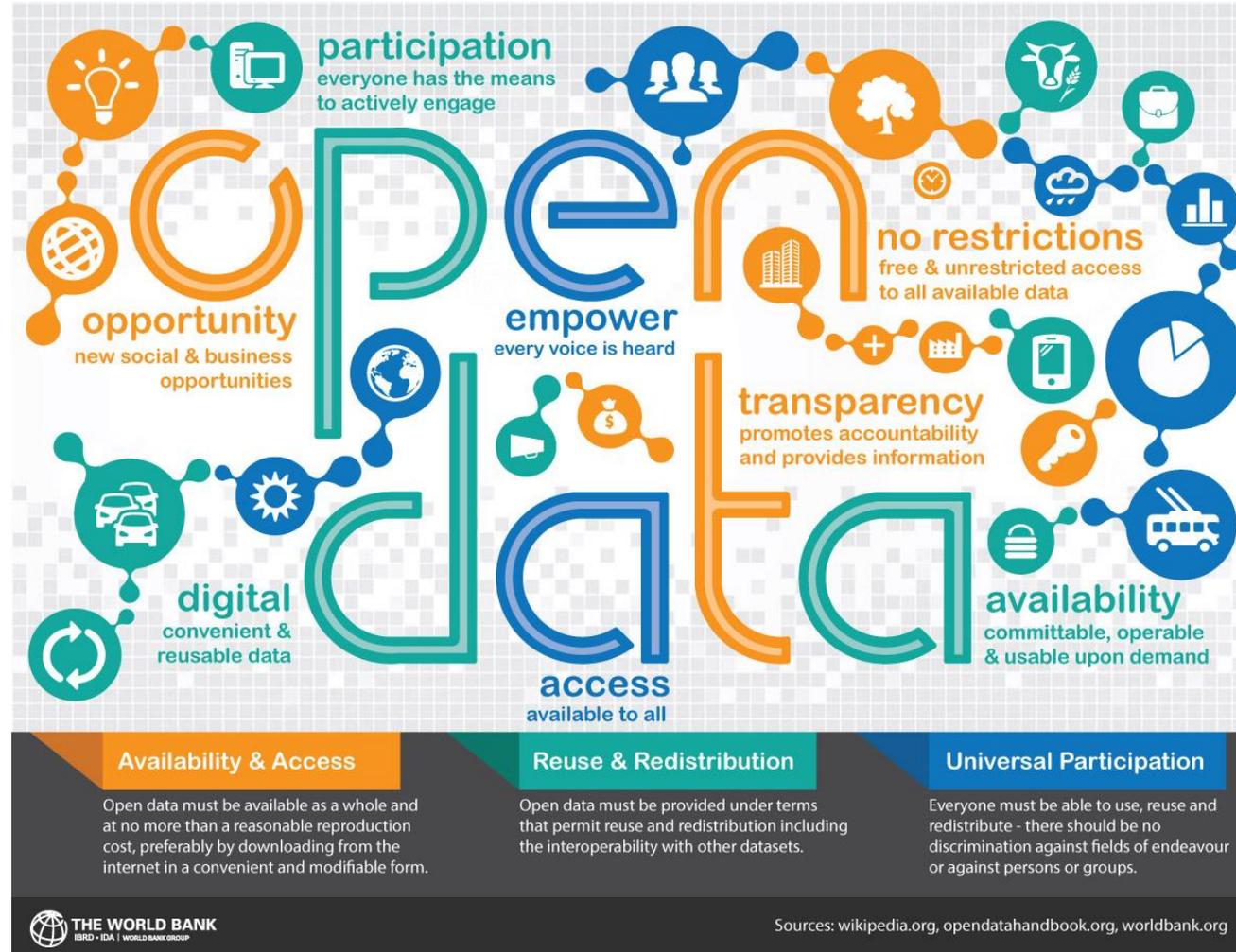
# Big Data Sharing

- Data Sharing and exchange



# Big Data Sharing

- Open Data



# Existing Solutions



WHY LINKED DATA? SERVICES ▾ OUR WORK

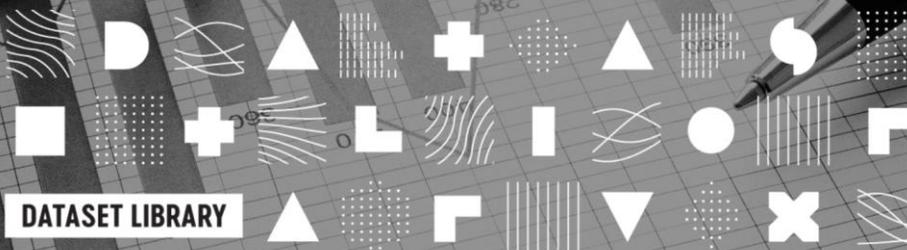
## DATA SHARING PLATFORM

Our Data Sharing Platform is a resilient, scalable and cloud-based solution for publishing data. It hosts data for a wide range of our public and private sector clients, and is available as a fully hosted and managed enterprise solution, or elements can be installed locally on your own infrastructure. It is widely used for publishing linked data on data.gov.uk, including data at: environment.data.gov.uk, location.data.gov.uk, landregistry.data.gov.uk, and many others.



### HKSTP DATA STUDIO

DATA DEVELOPERS DATA PUBLISHERS BLOG CONTACT US ABOUT FAQ



### DATASET LIBRARY

READY TO USE API FOR YOUR NEXT DATA INNOVATION

TYPE KEYWORD HERE



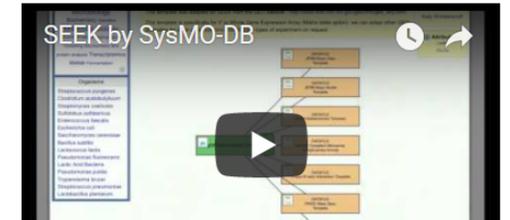
For finding, sharing and exchanging Data, Models, Simulations and Processes in Science.

HOME NEWS ABOUT SEEK ABOUT US FAQ GETTING SEEK PUBLICATIONS PRESENTATIONS REGISTRATION CONTACTING US

The SEEK platform is a web-based resource for sharing heterogeneous scientific research datasets, models or simulations, processes and research outcomes. It preserves associations between them, along with information about the people and organisations involved. Underpinning SEEK is the ISA infrastructure, a standard format for describing how individual experiments are aggregated into wider studies and investigations. Within SEEK, ISA has been extended and is configurable to allow the structure to be used outside of Biology. SEEK is incorporating semantic technology allowing sophisticated queries over the data, yet without getting in the way of your users.

[Read More](#)



## dataexchange

Data : Sorted



Why How Uses Install Media Docs Blog

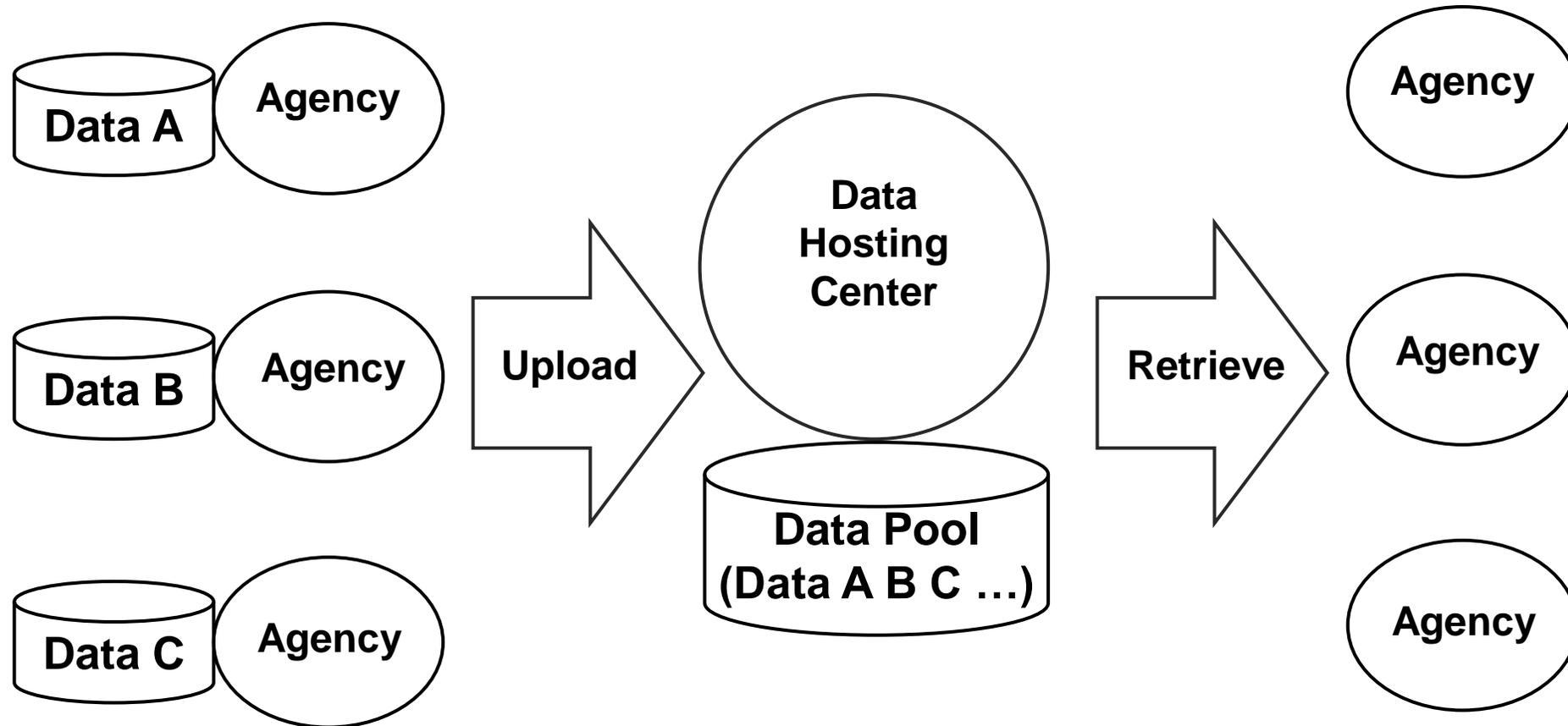
## IPFS is the Distributed Web

A peer-to-peer hypermedia protocol to make the web faster, safer, and more open.

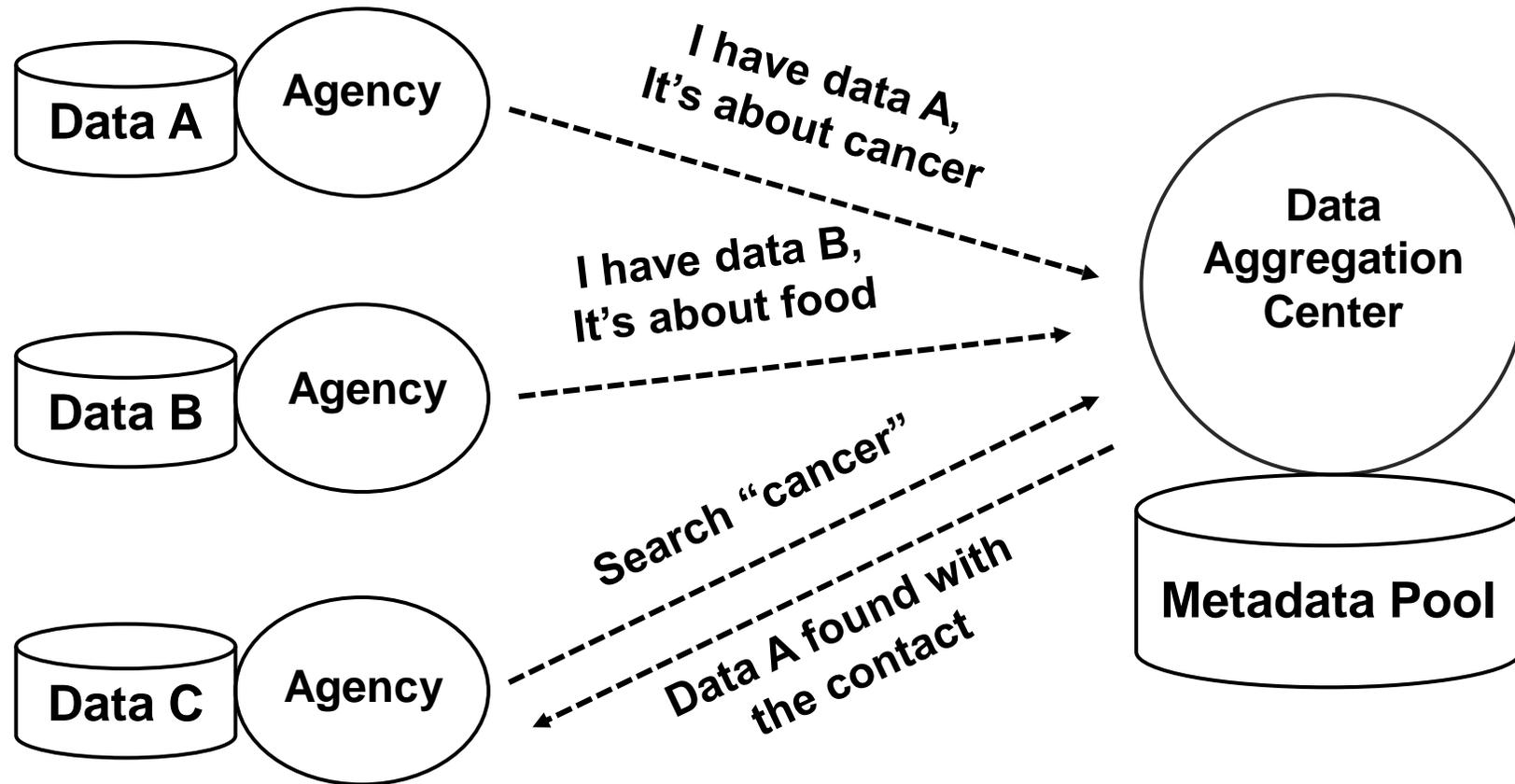
## Related Works

- Traditional Data Sharing and Transaction Models
- Data Hosting Center (DHC)
  - each agency will host, upload and publish its own data to the central database, controlled and maintained by the DHC
- Data Aggregation Center (DAC)
  - the center links data services through the API interface among agencies.

# Data Hosting Center

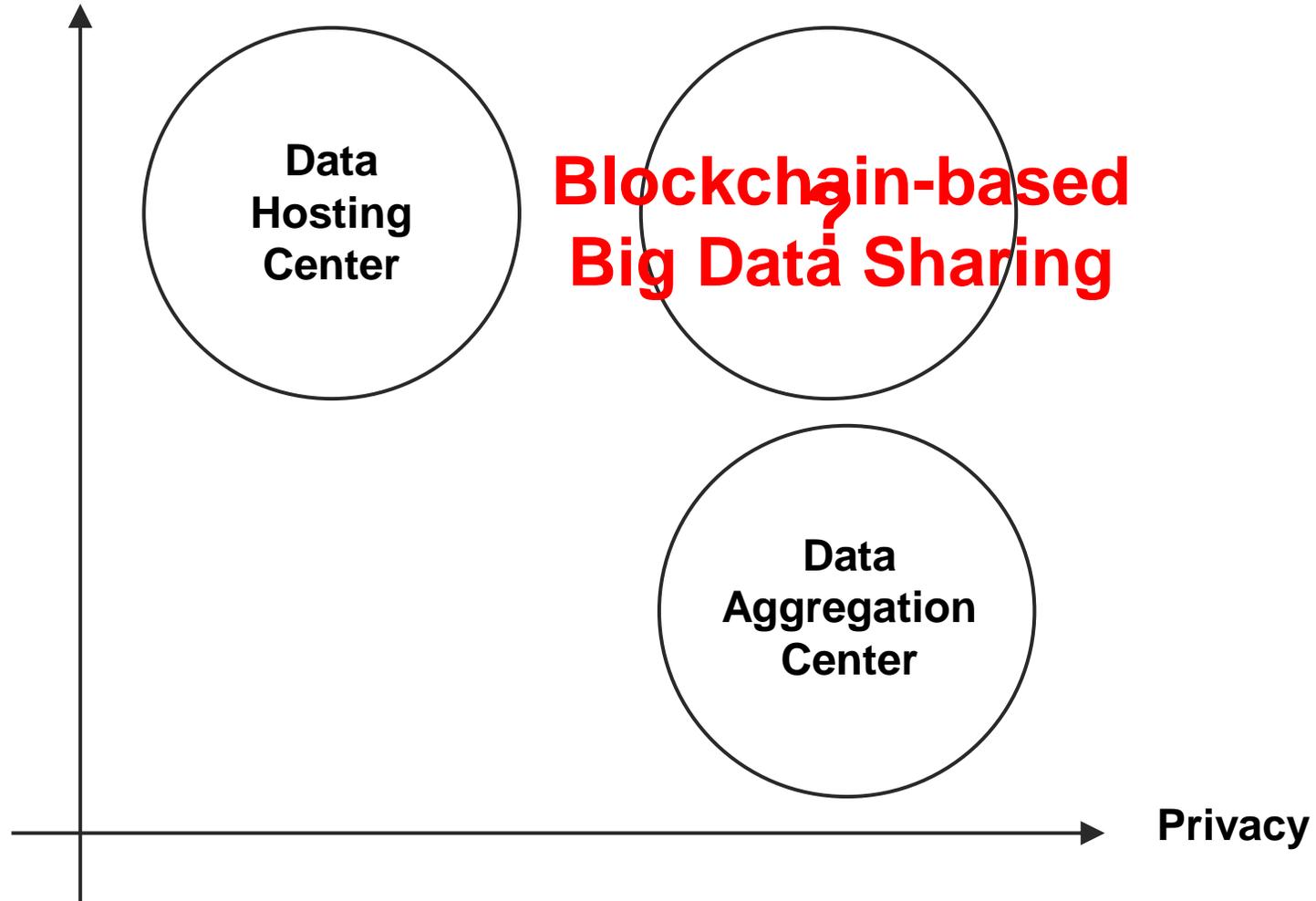


# Data Aggregation Center



# Blockchain as a Solution

Authenticity



# Blockchain

## Blockchain

A technology of  
**distributed ledger**  
for **trustless** data  
storage with  
**auditability**

Privacy

**Decentralization**: removal of  
third party

**Transparency**: visible to  
everyone

**Immutability**: once stored,  
cannot be changed

Authenticity

# Blockchain Approach to Big Data Sharing

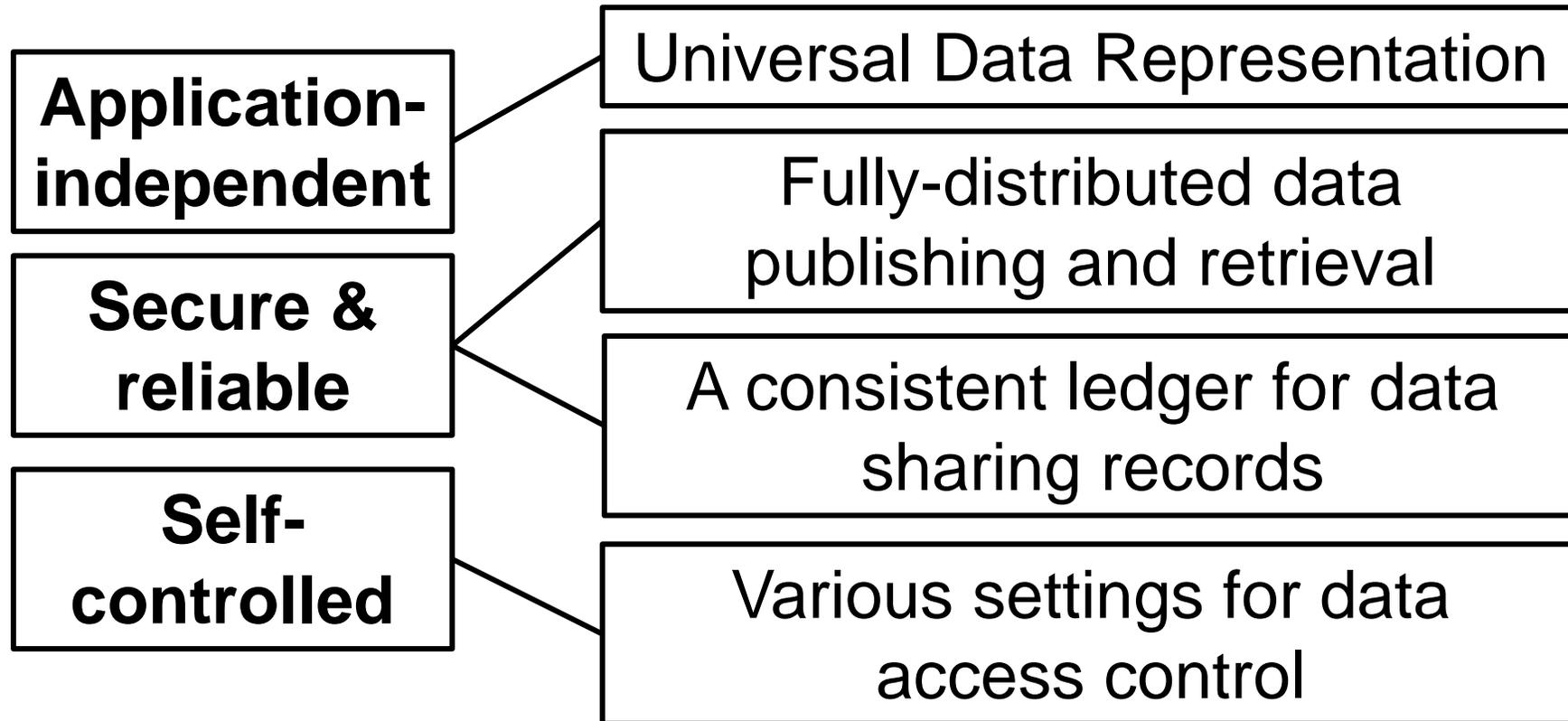
**AI3  
FOR BIG DATA SHARING**

**ACCOUNT  
ARCHIVE  
PUBLISH  
FAQ**

 **HUAWEI** 

THE HONG KONG  
POLYTECHNIC UNIVERSITY  
香港理工大學

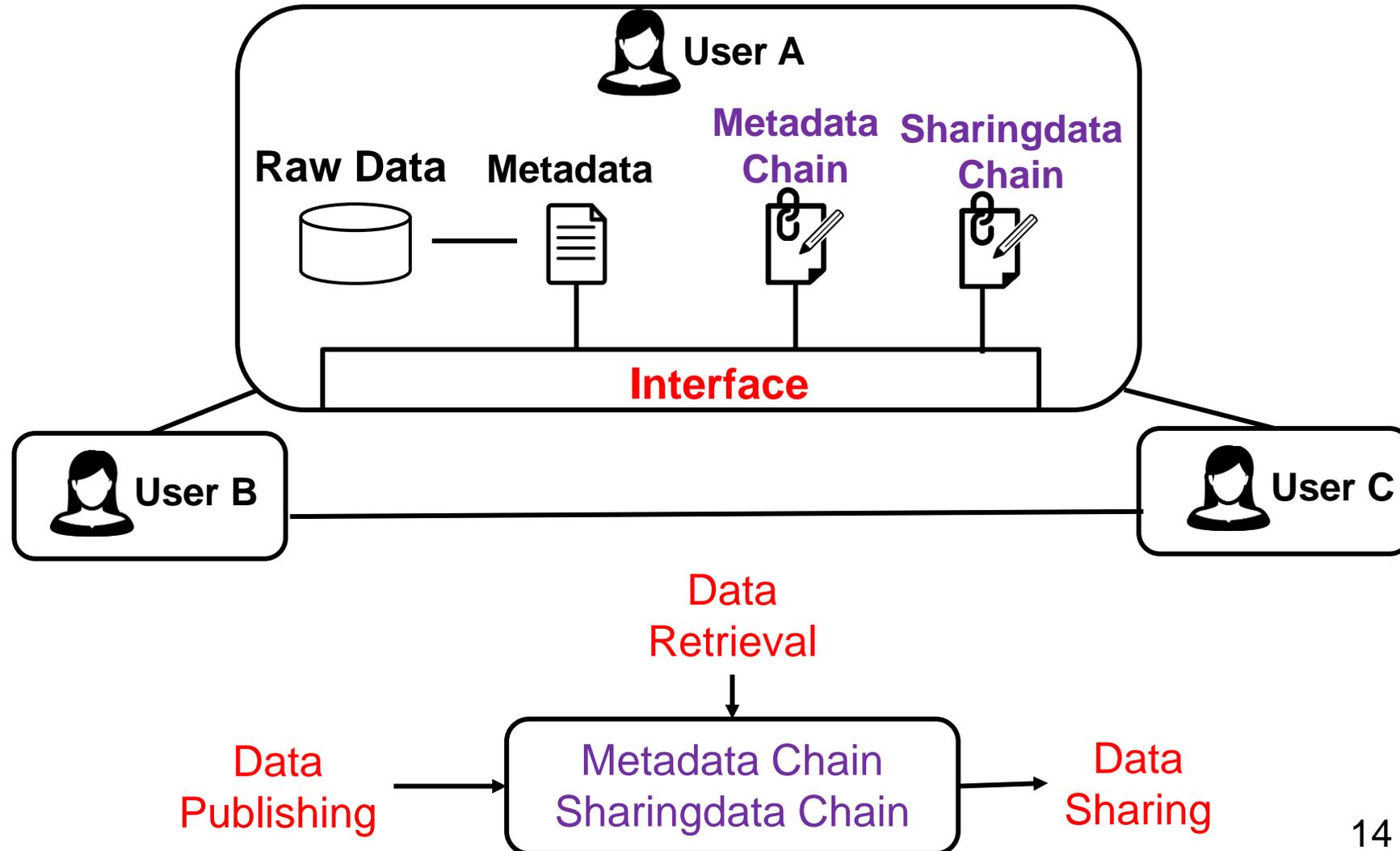
# TSAR Goal & Design Principles



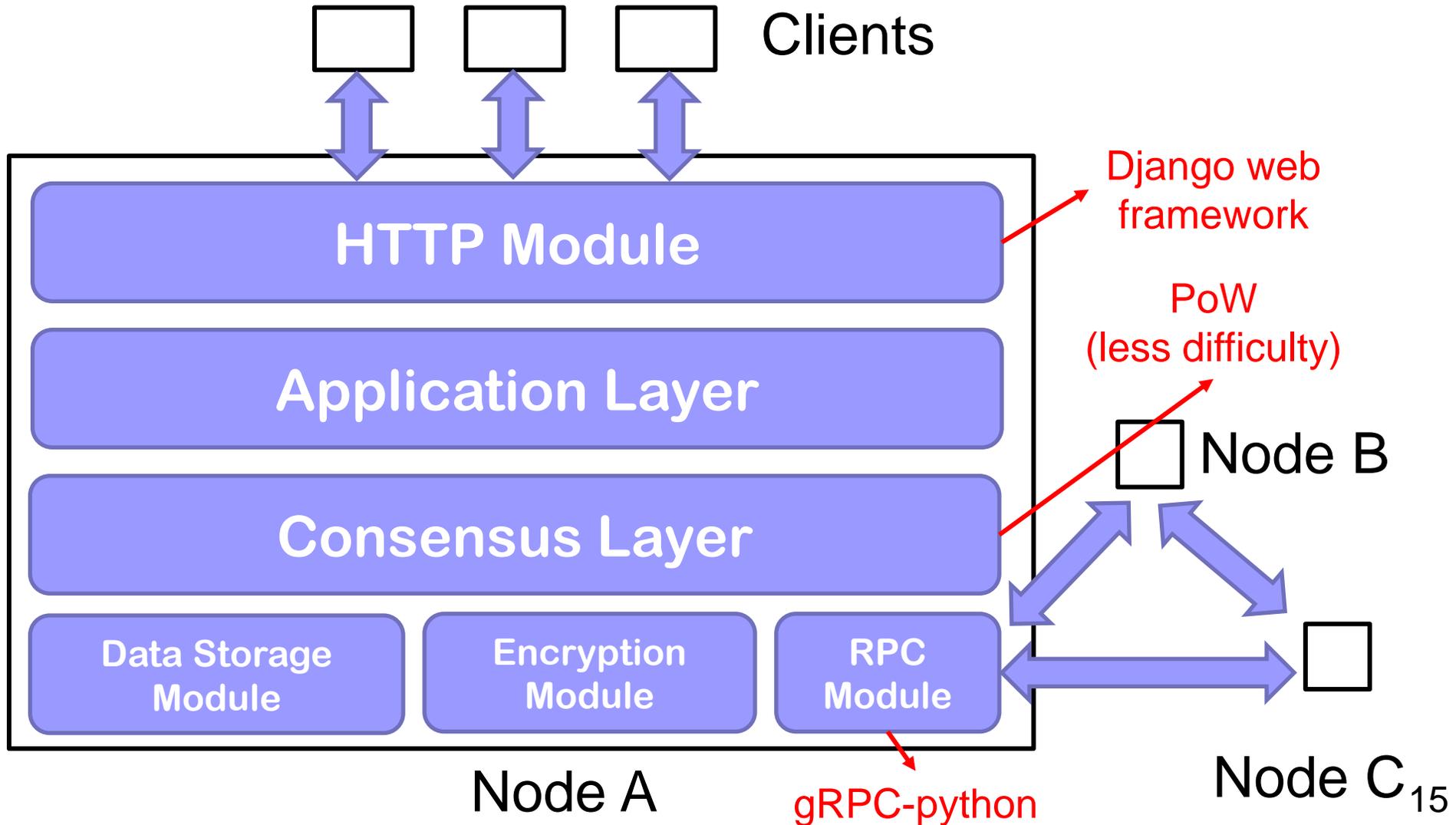
**Goal**

**Design Principles**

# TSAR System Architecture

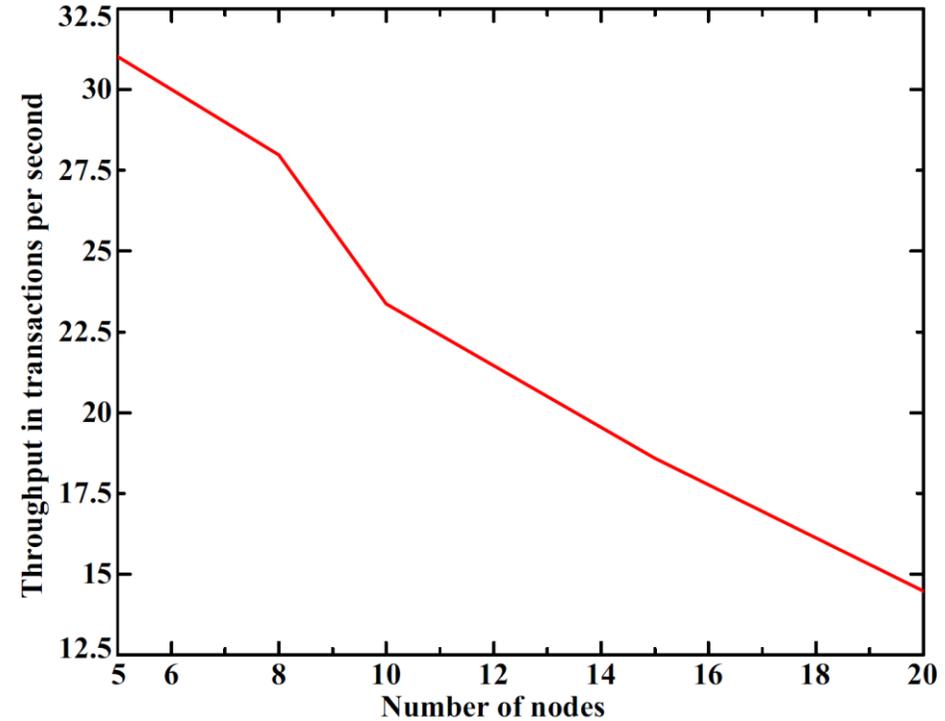
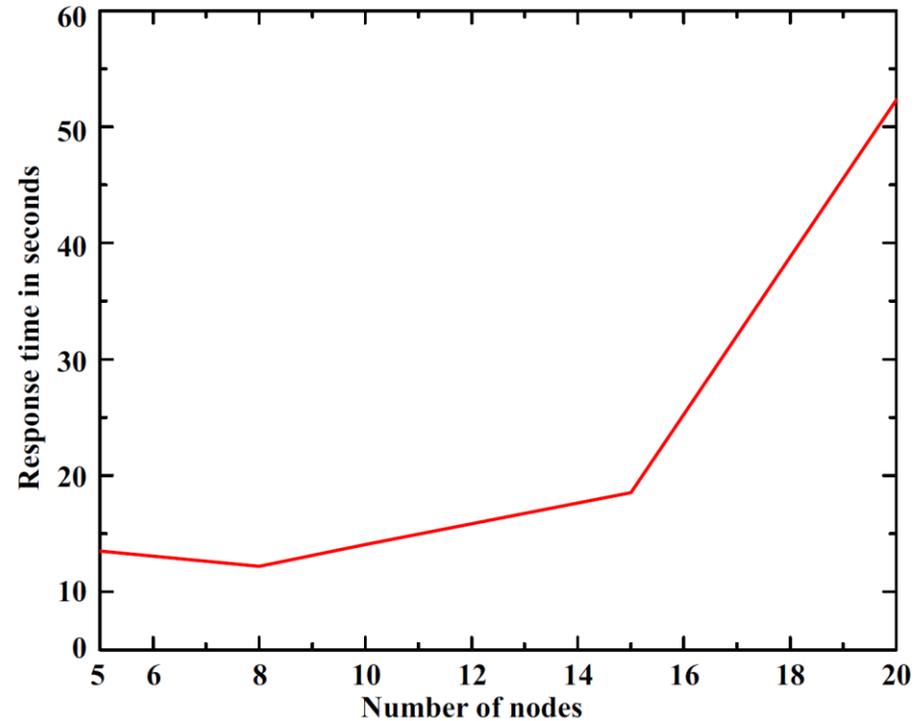


# TSAR Implementation



# TSAR Evaluation

- The system is deployed and tested on cloud (Alibaba two cores 4G 1M bandwidth)



# Acknowledgments

- This work is supported by Huawei Technologies Co. Ltd. with project code P15-0540 and RGC CRF with project number CityU C1008-16G.

# References

- [1] S. Erevelles, N. Fukawa, and L. Swayne, “Big data consumer analytics and the transformation of marketing,” *Journal of Business Research*, vol. 69, no. 2, pp. 897–904, 2016.
- [2] W. Raghupathi and V. Raghupathi, “Big data analytics in healthcare: promise and potential,” *Health information science and systems*, vol. 2, no. 1, p. 3, 2014.
- [3] M. Batty, “Big data, smart cities and city planning,” *Dialogues in Human Geography*, vol. 3, no. 3, pp. 274–279, 2013.
- [4] W. Li, J. Cao, J. Guan, M. L. Yiu, and S. Zhou, “Efficient retrieval of bounded-cost informative routes,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2182–2196, 2017.
- [5] X. Liu, K. Li, G. Min, Y. Shen, A. X. Liu, and W. Qu, “Completely pinpointing the missing rfid tags in a time-efficient way,” *IEEE Transactions on Computers*, vol. 64, no. 1, pp. 87–96, 2015.
- [6] S. Jiang, J. Cao, Y. Liu, J. Chen, and X. Liu, “Programming large-scale multi-robot system with timing constraints,” in *Computer Communication and Networks (ICCCN), 2016 25th International Conference on*. IEEE, 2016, pp. 1–9.20
- [7] D. Lazer, R. Kennedy, G. King, and A. Vespignani, “The parable of google flu: traps in big data analysis,” *Science*, vol. 343, no. 6176, pp. 1203–1205, 2014.

# References

- [8] C.-K. Chu, S. S. Chow, W.-G. Tzeng, J. Zhou, and R. H. Deng, “Keyaggregate cryptosystem for scalable data sharing in cloud storage,” *IEEE transactions on parallel and distributed systems*, vol. 25, no. 2, pp. 468477, 2014.
- [9] B. Cui, Z. Liu, and L. Wang, “Key-aggregate searchable encryption (kase) for group data sharing via cloud storage,” *IEEE Transactions on computers*, vol. 65, no. 8, pp. 2374–2385, 2016.
- [10] R. A. Poldrack and K. J. Gorgolewski, “Making big data open: data sharing in neuroimaging,” *Nature neuroscience*, vol. 17, no. 11, p. 1510, 2014.
- [11] S. Jiang, J. Cao, H. Wu, Y. Yang, M. Ma, and J. He, “Blochie: a blockchain-based platform for healthcare information exchange,” in *Smart Computing (SMARTCOMP), 2018 IEEE International Conference on (to appear)*. IEEE, 2018, pp. 1–8.
- [12] J. Benet, “Ipfs-content addressed, versioned, p2p file system,” arXiv preprint arXiv:1407.3561, 2014.
- [13] “grpc: A high performance, open-source universal rpc framework.” <https://grpc.io/>, accessed: 2018-01-25.
- [14] “Device for and method of one-way cryptographic hashing,” <https://www.google.com/patents/US6829355>, accessed: 2018-01-25.

# Front Page

Application-Independent Information Infrastructure

## Next Generation Distributed Big Data Sharing Platform

REGISTER NOW

Log In Your AI3 Account

Your Email  
t1@t.com

Password  
•

SIGN IN

©2017. AI3 Development Group. v1.4.5

# My Data

AI3 System

Search for shared data, domain and others

t1@t.com [LOG OUT](#)

PUBLISH

Data

Discover

My Data

Starred

Domain

Domains

Profile

Profile

## My Data

Loaded	Publisher	Keywords
Name		
wechat click data 002	Derek Ma	-
wechat click data 003	Derek Ma	-
wechat click data 005	Derek Ma	-
wechat click data 006	Derek Ma	wechat, weixin, click
WeChat Click Data	IMCL	clicks

# My Profile

AI3 System

Search for shared data, domain and others

t1@t.com [LOG OUT](#)

**PUBLISH**

Data

Discover

My Data

Starred

Domain

Domains

Profile

Profile

## Manage Your Profile

### Account Information

Information loaded

Your Email  
**t1@t.com**

Your Name  
**Derek**

Your Mobile Number(numeric only)  
**65704753** 8

Your Industry  
**COMP**

**MODIFY**

### Change Your Password

Your Current Password

New Password You Want

**CHANGE**

# Publish - DB

AI3 System t1@t.com LOG OUT

Search for shared data, domain and others

**PUBLISH**

Data

Discover

My Data

Starred

Domain

Domains

Profile

Profile

## Publish a New Data

Structured Data - Database    Unstructured Data - Local File    Import - with Metadata

### Information about Database?

Database IP Address \_\_\_\_\_

Database Name \_\_\_\_\_

Table Name \_\_\_\_\_

Port  
3030

Username of the Database \_\_\_\_\_

Password \_\_\_\_\_

### Can we know more about the data?

Identifier \_\_\_\_\_

Name \_\_\_\_\_

Contributor \_\_\_\_\_

Publisher \_\_\_\_\_

Description about the data shared \_\_\_\_\_

Keywords \_\_\_\_\_

Categories \_\_\_\_\_

[DOWNLOAD METADATA .XML FILE](#)    **PUBLISH**

# Publish - File

AI3 System

Search for shared data, domain and others

t1@t.com LOG OUT

**PUBLISH**

Data

Discover

My Data

Starred

Domain

Domains

Profile

Profile

## Publish a New Data

Structured Data - Database   Unstructured Data - Local File   Import - with Metadata

### Where is the data?

Currently, only single .csv file is supported.

**SELECT FILE**  Drop file here...

We will not upload this file, only the absolute path and attributes will be obtained.

### Can we know more about it?

user\_id, identifier, name, contributor, publisher, update\_time, format, description, keywords, access\_control, source\_link, sample\_link, categories, price

Identifier

Name

Contributor

Publisher

Format

Description to the data shared

Keywords

Categories

**DOWNLOAD METADATA .XML FILE**   **PUBLISH**

# Discover

AI3 System

Search for shared data, domain and others

t1@t.com [LOG OUT](#)

**PUBLISH**

Data

Discover

My Data

Starred

Domain

Domains

Profile

Profile

## Discover

Loaded	Publisher	Keywords
Name		
wechat click data 001	Peiyuan	-
wechat click data 002	Derek Ma	-
wechat click data 003	Derek Ma	-
wechat click data 004	Hanqing	-
wechat click data 005	Derek Ma	-
wechat click data 006	Derek Ma	wechat, weixin, click
WeChat Click Data	IMCL	clicks

# Discover with Details

AI3 System

Search for shared data, domain and others

t1@t.com LOG OUT

PUBLISH

Data

Discover

My Data

Starred

Domain

Domains

Profile

Profile

## Discover > WeChat Click Data

Loaded

### Basic about Data

**WeChatClickData**

Name: WeChat Click Data  
 Description:  
 Contributor: Organization1  
 Format: Database  
 Publisher: IMCL  
 Categories: internet  
 Keywords: clicks  
 Update Time: 2017-04-12 12:35:44.0

ACTION1
ACTION2

### Data Preview

date	title	url	reposts	read_count	timestamp
2015-06-20	??????	http://mp.weixin.qq.com/s?__biz=MzA4NTQ4NzgwMg==&mid=208573600&idx=6&sn=812196da3b4adfa11a82421457b67998&scene=6#rd	2	585	1434737647
2015-06-20	????????	http://mp.weixin.qq.com/s?__biz=MzAwNzAzMTQ4Mw==&mid=207629262&idx=1&sn=f4c553b2468796187d8f2ad7ed59d6ab&scene=6#rd	3	426	1434737647
2015-06-20	2?=????200????????	http://mp.weixin.qq.com/s?__biz=MzA4OTc4MDYwMA==&mid=203420929&idx=7&sn=34a4b90303bd4c3194ad93320eea1677&scene=6#rd	1	133	1434737647

# Domain Manage

AI3 System

Search for shared data, domain and others

t1@t.com LOG OUT

**PUBLISH**

Data

Discover

My Data

Starred

Domain

Domains

Profile

Profile

## Manage Your Domains

### Current Domains

#### PolyU

All data can be shared among all PolyU students and staffs. Including data of education, academic projects and so on...

LEAVE ACTION2

#### Dept of Computing

Research, homework and testing data shared by Dept of Computing at HK PolyU

LEAVE ACTION2

### Recommended Domains

#### IMCL Group

Internet and Mobile Computing Group at HK PolyU Computing

REQUEST ACTION2

#### Huawei Open Data

Public open data provided by Huawei

JOIN ACTION2

### Add to a New Domain

New Domain ID

ADD

thank  
you!