Balancing Data Utility and Privacy Protection in the Socially Aware Data Cloud

Yuh-Jong Hu
hu@cs.nccu.edu.tw

Emerging Network Technology (ENT) Lab.
Department of Computer Science
National Chengchi University, Taipei, Taiwan

Nov-23-2013

Int. Conf. on EITA-New Media 2013, Taipei, Taiwan


Part I

Introduction

Motivations

1. Centralized closed social networking sites are *walled gardens* that limit themselves to the relationships between people with accounts on a single site.
2. Big data analytics has been proposed for online social network, but the related privacy protection issue does not arise much attention.
3. Statistical Disclosure Control (SDC) methods for microdata protection have been well-developed, but they are not used for data disclosure of online social network.
4. We intend to use Semantic Web technology to establish a privacy-aware policy Web architecture, and achieve the objective of balancing data utility and privacy protection.
5. This presentation extends our previous works for Int. Conf. on Web Intelligence, Mining and Semantics (WIMS'13), Madrid, Spain, 2013.


Motivations

1. Centralized closed social networking sites are *walled gardens* that limit themselves to the relationships between people with accounts on a single site.
2. Big data analytics has been proposed for online social network, but the related privacy protection issue does not arise much attention.
3. Statistical Disclosure Control (SDC) methods for microdata protection have been well-developed, but they are not used for data disclosure of online social network.
4. We intend to use Semantic Web technology to establish a privacy-aware policy Web architecture, and achieve the objective of balancing data utility and privacy protection.
5. This presentation extends our previous works for Int. Conf. on Web Intelligence, Mining and Semantics (WIMS'13), Madrid, Spain, 2013.


Motivations

1. Centralized closed social networking sites are *walled gardens* that limit themselves to the relationships between people with accounts on a single site.
2. Big data analytics has been proposed for online social network, but the related privacy protection issue does not arise much attention.
3. Statistical Disclosure Control (SDC) methods for microdata protection have been well-developed, but they are not used for data disclosure of online social network.
4. We intend to use Semantic Web technology to establish a privacy-aware policy Web architecture, and achieve the objective of balancing data utility and privacy protection.
5. This presentation extends our previous works for Int. Conf. on Web Intelligence, Mining and Semantics (WIMS'13), Madrid, Spain, 2013.


Motivations

1. Centralized closed social networking sites are *walled gardens* that limit themselves to the relationships between people with accounts on a single site.
2. Big data analytics has been proposed for online social network, but the related privacy protection issue does not arise much attention.
3. Statistical Disclosure Control (SDC) methods for microdata protection have been well-developed, but they are not used for data disclosure of online social network.
4. We intend to use Semantic Web technology to establish a privacy-aware policy Web architecture, and achieve the objective of balancing data utility and privacy protection.
5. This presentation extends our previous works for Int. Conf. on Web Intelligence, Mining and Semantics (WIMS'13), Madrid, Spain, 2013.


Motivations

1. Centralized closed social networking sites are *walled gardens* that limit themselves to the relationships between people with accounts on a single site.
2. Big data analytics has been proposed for online social network, but the related privacy protection issue does not arise much attention.
3. Statistical Disclosure Control (SDC) methods for microdata protection have been well-developed, but they are not used for data disclosure of online social network.
4. We intend to use Semantic Web technology to establish a privacy-aware policy Web architecture, and achieve the objective of balancing data utility and privacy protection.
5. This presentation extends our previous works for Int. Conf. on Web Intelligence, Mining and Semantics (WIMS'13), Madrid, Spain, 2013.


Research Goals

1. Argue why we choose decentralized but not centralized online social networking architectures for data analytics services?
2. How can we use the Hadoop and R integration paradigms in the socially aware data cloud with deep data analytics services?
3. How can we provide data analytics services through types of semantics-enabled policy enforcement on the decentralized policy-aware social Web?
4. How can we provide privacy-preserving data analytics through data anonymizing and selective data revelation techniques?


Research Goals

1. Argue why we choose decentralized but not centralized online social networking architectures for data analytics services?
2. How can we use the Hadoop and R integration paradigms in the socially aware data cloud with deep data analytics services?
3. How can we provide data analytics services through types of semantics-enabled policy enforcement on the decentralized policy-aware social Web?
4. How can we provide privacy-preserving data analytics through data anonymizing and selective data revelation techniques?


Research Goals

1. Argue why we choose decentralized but not centralized online social networking architectures for data analytics services?
2. How can we use the Hadoop and R integration paradigms in the socially aware data cloud with deep data analytics services?
3. How can we provide data analytics services through types of semantics-enabled policy enforcement on the decentralized policy-aware social Web?
4. How can we provide privacy-preserving data analytics through data anonymizing and selective data revelation techniques?


Research Goals

1. Argue why we choose decentralized but not centralized online social networking architectures for data analytics services?
2. How can we use the Hadoop and R integration paradigms in the socially aware data cloud with deep data analytics services?
3. How can we provide data analytics services through types of semantics-enabled policy enforcement on the decentralized policy-aware social Web?
4. How can we provide privacy-preserving data analytics through data anonymizing and selective data revelation techniques?


Contributions

1. Propose concepts of a semantic big data analysis pipeline for automated services of data analytics, protection, and interpretation.
2. Semantics-enabled policies, as a combination of ontologies and rules, are represented and enforced for data analytics services.
3. Data users can enable SDC selection techniques for selective data revelation on the semantics-enabled privacy-aware social Web.
4. Preliminary results on crafting a balance between data utility and protection.


Contributions

1. Propose concepts of a semantic big data analysis pipeline for automated services of data analytics, protection, and interpretation.
2. Semantics-enabled policies, as a combination of ontologies and rules, are represented and enforced for data analytics services.
3. Data users can enable SDC selection techniques for selective data revelation on the semantics-enabled privacy-aware social Web.
4. Preliminary results on crafting a balance between data utility and protection.

Contributions

1. Propose concepts of a semantic big data analysis pipeline for automated services of data analytics, protection, and interpretation.
2. Semantics-enabled policies, as a combination of ontologies and rules, are represented and enforced for data analytics services.
3. Data users can enable SDC selection techniques for selective data revelation on the semantics-enabled privacy-aware social Web.
4. Preliminary results on crafting a balance between data utility and protection.

Contributions

1. Propose concepts of a semantic big data analysis pipeline for automated services of data analytics, protection, and interpretation.
2. Semantics-enabled policies, as a combination of ontologies and rules, are represented and enforced for data analytics services.
3. Data users can enable SDC selection techniques for selective data revelation on the semantics-enabled privacy-aware social Web.
4. Preliminary results on crafting a balance between data utility and protection.

Part II

Background

Problems with Centralized Online Social Networks

- Information silos on one site is not usable in the others.
- A user is "stuck": migrating to another application is hard.
- Users cannot choose what Web applications do with their data.
- New application must acquire a critical mass of data from scratch.
- Do not allow users control over how their personal information is disseminated, which results in potential privacy problems.

–Decentralization: The Future of Online Social Networking

–A World Wide Web Without Walls, Krohn, M., et al.

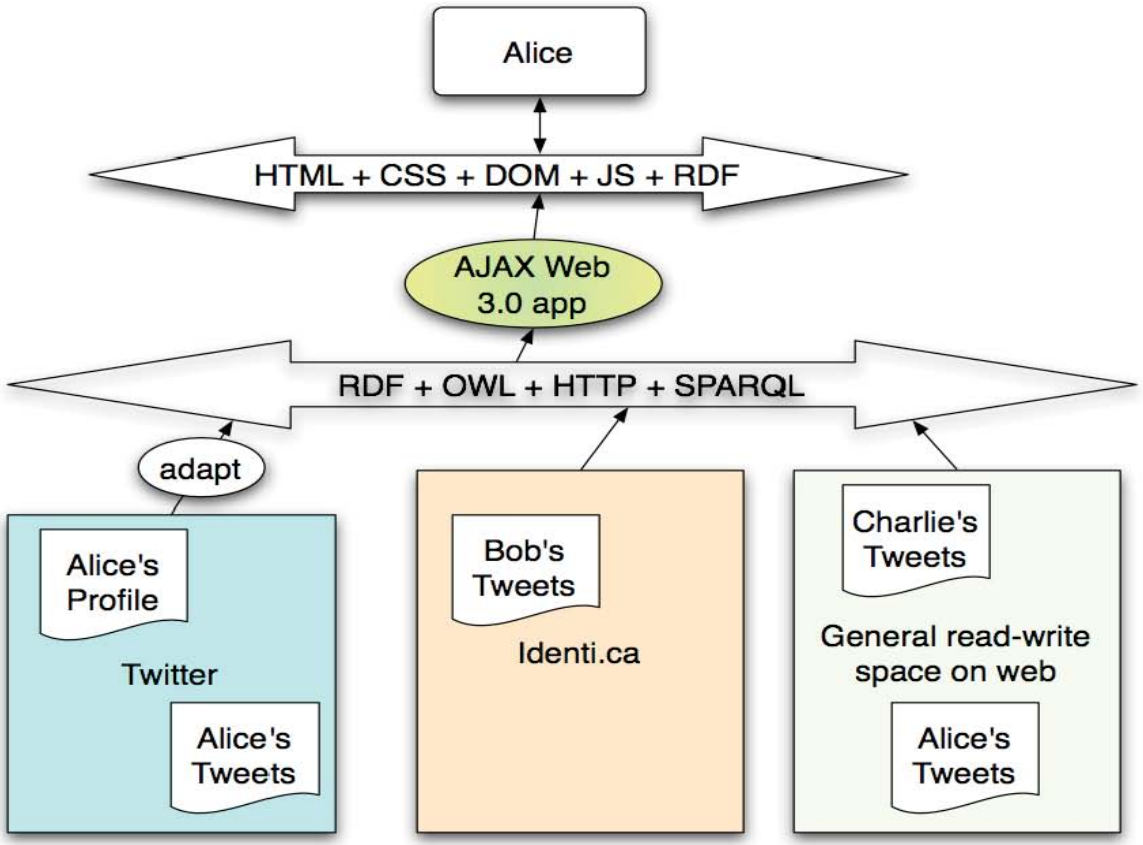
What are Decentralized Online Social Networks?

- Desire properties:
  - Decouple application from data
  - Give end-users control over their own data
  - Infrastructure (or platform) providers with IaaS or PaaS
  - Social network service developers with SaaS for users (or data owners)
  - Minimize the trust footprint for users to enforce their own data usage and control policies across applications.

–Decentralization: The Future of Online Social Networking

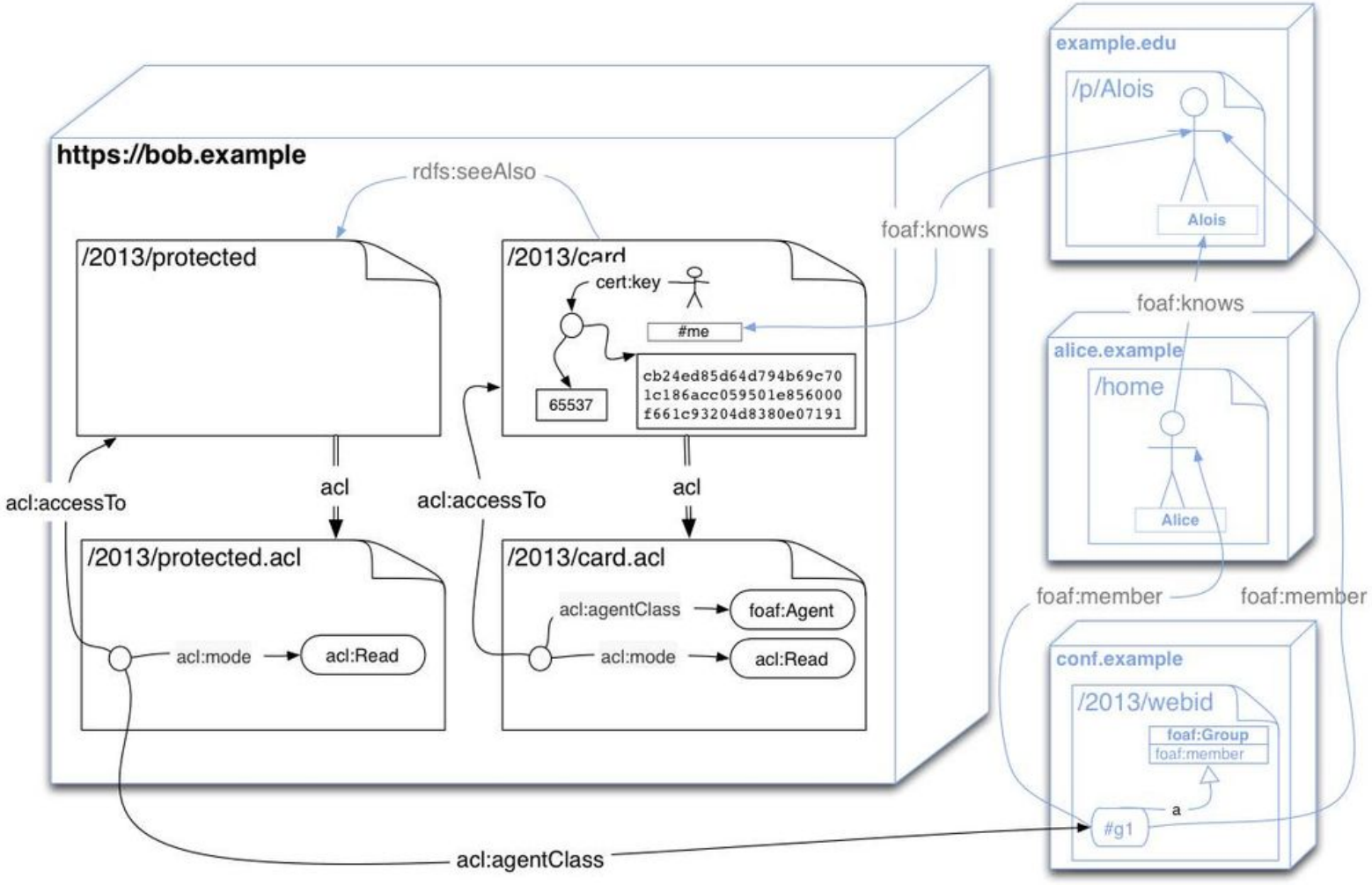
–A World Wide Web Without Walls, Krohn, M., et al.


Decentralized Online Social Network



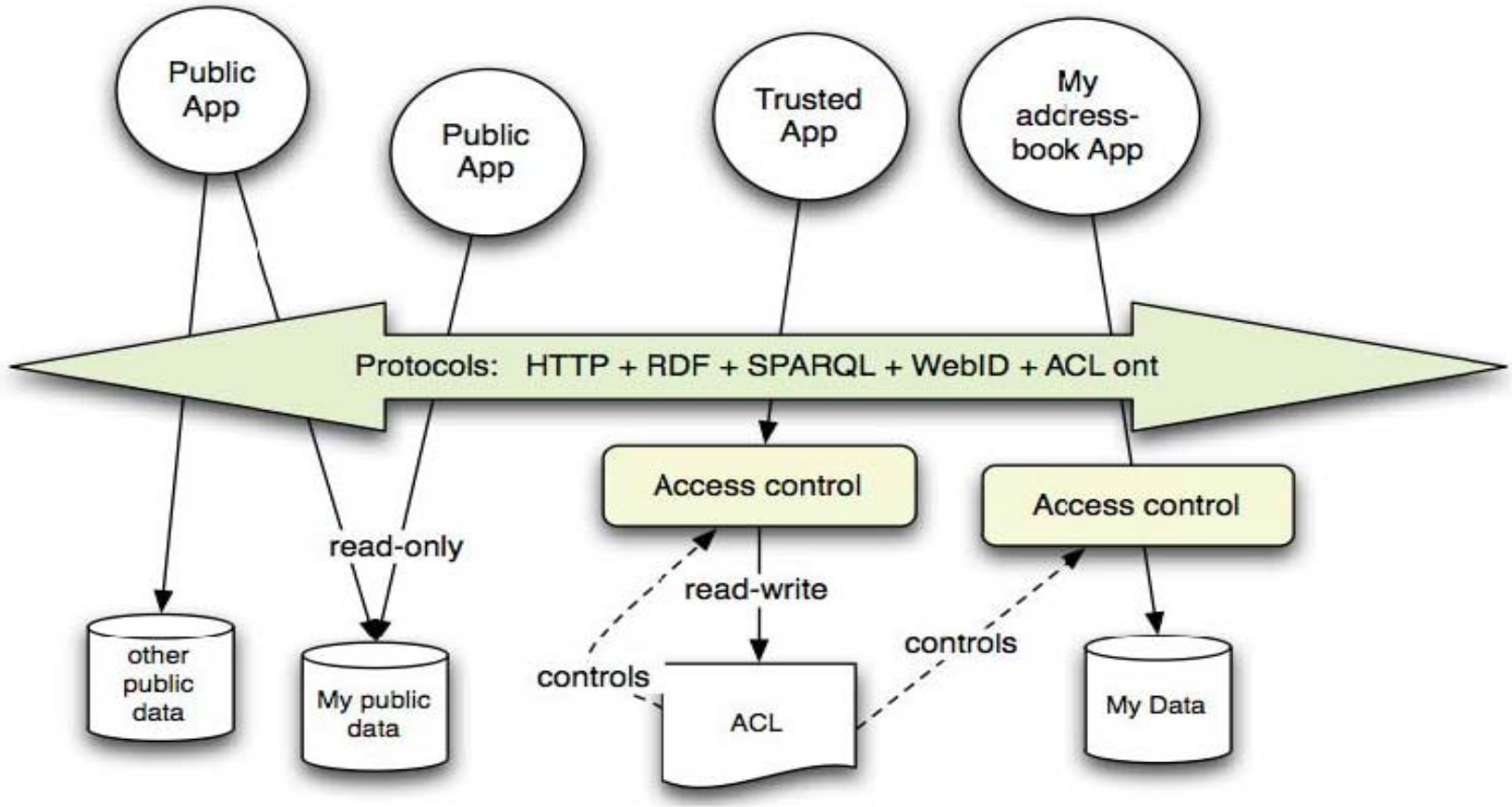
–Socially Aware Cloud Storage, Tim Berners-Lee

Decentralized Online Social Network with Access Control



–WebAccessControl

Decentralized Online Social Network with Access Control



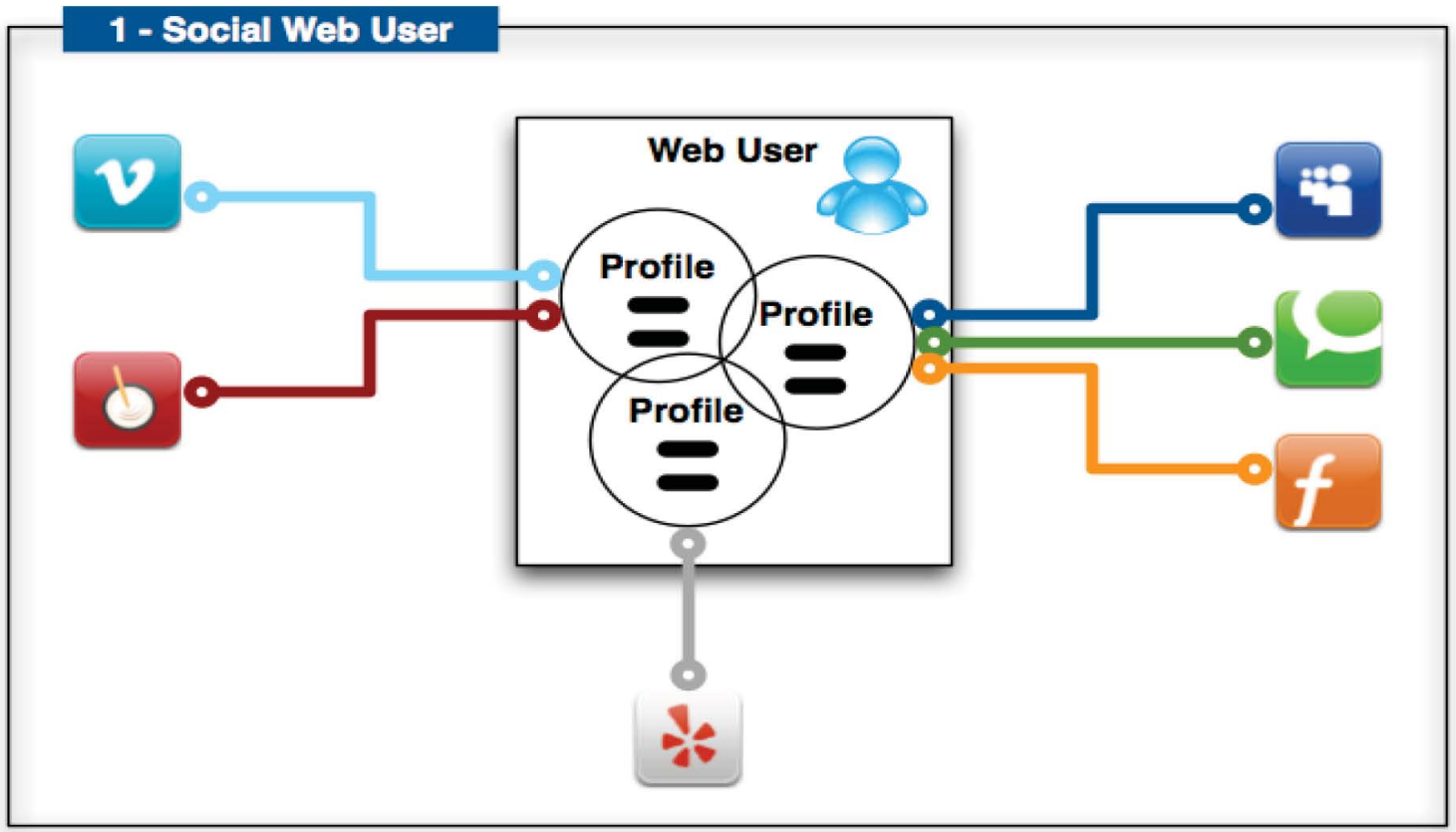
–Socially Aware Cloud Storage, Tim Berners-Lee

Open Privacy-Aware Social Web

- We need a new policy-oriented social Web architecture to support trust and privacy on the Web.
- A profile management service that could be run in the browser or via a third-party website would keep track of the distributed attributes and multiple profiles.
- Allow users to edit the attributes across multiple platforms and sites.
- A core feature or service of a social application is to make, maintain, and expand these connections.

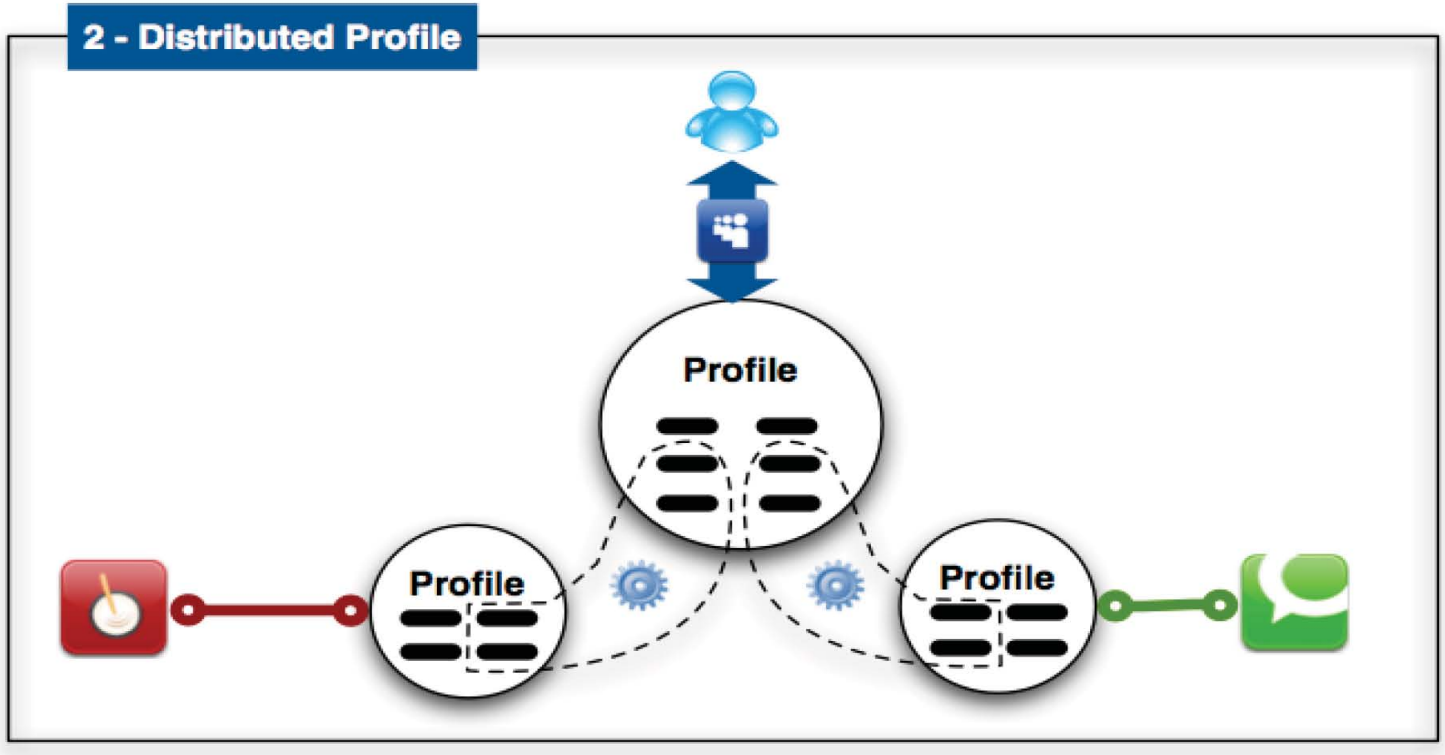
–A Standards-based, Open and Privacy-aware Social Web, W3C Incubator Group Report 6th Dec. 2010


Open Privacy-Aware Social Web (Conti.)



–A Standards-based, Open and Privacy-aware Social Web, W3C Incubator Group Report 6th Dec. 2010

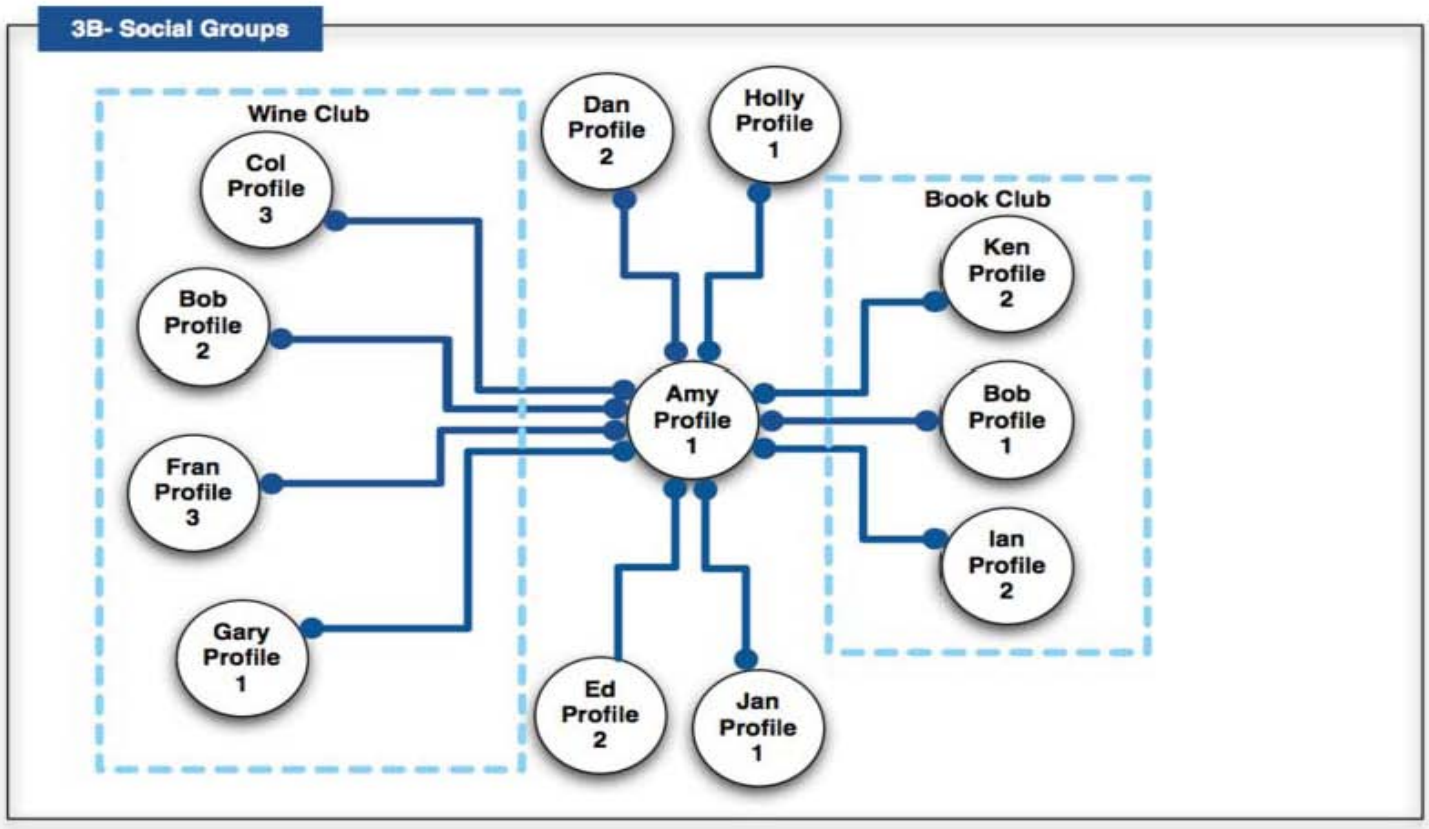
Open Privacy-Aware Social Web (Conti.)



A Single User with a Distributed Social Graph

–A Standards-based, Open and Privacy-aware Social Web, W3C Incubator Group Report 6th Dec. 2010

Open Privacy-Aware Social Web (Conti.)



Distributed Social Graphs and Groups

–A Standards-based, Open and Privacy-aware Social Web, W3C Incubator Group Report 6th Dec. 2010

Why Integrating R and Hadoop?

- Open-source *R* statistical computing packages are used for heavyweight data analytics, but only works for in-memory data in a stand alone computer.
- Hadoop with MapReduce framework allows for distributed processing of large data sets, but only works for lightweight data analytics.
- Integrating R and Hadoop bring distributed (or parallel) processing capability of heavyweight data analytics.
- Possible R + Hadoop: RHadoop, RHipe, RHive, Ricardo, etc.


Why Integrating R and Hadoop?

- Open-source *R* statistical computing packages are used for heavyweight data analytics, but only works for in-memory data in a stand alone computer.
- Hadoop with MapReduce framework allows for distributed processing of large data sets, but only works for lightweight data analytics.
- Integrating R and Hadoop bring distributed (or parallel) processing capability of heavyweight data analytics.
- Possible R + Hadoop: RHadoop, RHipe, RHive, Ricardo, etc.


Why Integrating R and Hadoop?

- Open-source *R* statistical computing packages are used for heavyweight data analytics, but only works for in-memory data in a stand alone computer.
- Hadoop with MapReduce framework allows for distributed processing of large data sets, but only works for lightweight data analytics.
- Integrating R and Hadoop bring distributed (or parallel) processing capability of heavyweight data analytics.
- Possible R + Hadoop: RHadoop, RHipe, RHive, Ricardo, etc.


Why Integrating R and Hadoop?

- Open-source *R* statistical computing packages are used for heavyweight data analytics, but only works for in-memory data in a stand alone computer.
- Hadoop with MapReduce framework allows for distributed processing of large data sets, but only works for lightweight data analytics.
- Integrating R and Hadoop bring distributed (or parallel) processing capability of heavyweight data analytics.
- Possible R + Hadoop: RHadoop, RHipe, RHive, Ricardo, etc.


How to use R/Hadoop for Social Web of Data?

- We intend to leverage the Semantic Web's open technologies for social web of data, e.g., RDF(S)-based for FOAF, access control linked data, and Web-ID for single-signon service.
- Translate from existing centralized social network graph data in a JSON format to RDF(S) Turtle outputs, and integrate with other emerging Turtle linked data of decentralized social networks.
- SPARQL queries linked data outputs from various RDF(S) (or non-RDF(S) JSON) data stores for R.
- Why use RDF(S) linked data stored in the Hadoop HDFS for big data format, and furthermore queried outputs through SPARQL for R with deep data analytics?


How to use R/Hadoop for Social Web of Data?

- We intend to leverage the Semantic Web's open technologies for social web of data, e.g., RDF(S)-based for FOAF, access control linked data, and Web-ID for single-signon service.
- Translate from existing centralized social network graph data in a JSON format to RDF(S) Turtle outputs, and integrate with other emerging Turtle linked data of decentralized social networks.
- SPARQL queries linked data outputs from various RDF(S) (or non-RDF(S) JSON) data stores for R.
- Why use RDF(S) linked data stored in the Hadoop HDFS for big data format, and furthermore queried outputs through SPARQL for R with deep data analytics?


How to use R/Hadoop for Social Web of Data?

- We intend to leverage the Semantic Web's open technologies for social web of data, e.g., RDF(S)-based for FOAF, access control linked data, and Web-ID for single-signon service.
- Translate from existing centralized social network graph data in a JSON format to RDF(S) Turtle outputs, and integrate with other emerging Turtle linked data of decentralized social networks.
- SPARQL queries linked data outputs from various RDF(S) (or non-RDF(S) JSON) data stores for R.
- Why use RDF(S) linked data stored in the Hadoop HDFS for big data format, and furthermore queried outputs through SPARQL for R with deep data analytics?


How to use R/Hadoop for Social Web of Data?

- We intend to leverage the Semantic Web's open technologies for social web of data, e.g., RDF(S)-based for FOAF, access control linked data, and Web-ID for single-signon service.
- Translate from existing centralized social network graph data in a JSON format to RDF(S) Turtle outputs, and integrate with other emerging Turtle linked data of decentralized social networks.
- SPARQL queries linked data outputs from various RDF(S) (or non-RDF(S) JSON) data stores for R.
- Why use RDF(S) linked data stored in the Hadoop HDFS for big data format, and furthermore queried outputs through SPARQL for R with deep data analytics?


Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analytics and protection, and rules are used for enforcing the principles of data analytics and privacy protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation with selective data revelation:
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through SPARQL queries.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and data users' usage context.
   - Data Releasing Policy (DRP) describes available SDC methods for data analytics with selective data revelation to preserve privacy.


Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analytics and protection, and rules are used for enforcing the principles of data analytics and privacy protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation with selective data revelation:
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through SPARQL queries.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and data users' usage context.
   - Data Releasing Policy (DRP) describes available SDC methods for data analytics with selective data revelation to preserve privacy.


Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analytics and protection, and rules are used for enforcing the principles of data analytics and privacy protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation with selective data revelation:
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through SPARQL queries.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and data users' usage context.
   - Data Releasing Policy (DRP) describes available SDC methods for data analytics with selective data revelation to preserve privacy.


Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analytics and protection, and rules are used for enforcing the principles of data analytics and privacy protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation with selective data revelation:
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through SPARQL queries.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and data users' usage context.
   - Data Releasing Policy (DRP) describes available SDC methods for data analytics with selective data revelation to preserve privacy.


Semantics-enabled Policies

1. Semantics-enabled policies are composed of ontologies and rules, where ontologies are used for describing the concepts of data analytics and protection, and rules are used for enforcing the principles of data analytics and privacy protection.
2. Semantics-enabled policies, ACP, DHP, and DRP are respectively correspond to, query restriction, data manipulation, and output perturbation with selective data revelation:
   - Access Control Policy (ACP) provides restricted Pattern-Based Queries (PBQs) through SPARQL queries.
   - Data Handling Policy (DHP) provides data usage conditions matching between data owners' privacy preferences and data users' usage context.
   - Data Releasing Policy (DRP) describes available SDC methods for data analytics with selective data revelation to preserve privacy.


Part III

Privacy-Preserving Data Analytics

A Big Data Analytics Life cycle

- A pipeline process with five stages:
  1. Data acquisition and recording
  2. Information extraction and cleaning
  3. Data integration, aggregation, and representation
  4. Query processing, data modeling, and analysis
  5. Interpretation
- Privacy protection is one of the common criteria for all of the related pipeline stage services.
- Apply semantics-enabled policy management for seamlessly bound pipeline services together to achieve balancing data utility and privacy protection on the decentralized social Web.

–A. Labrinidis et al. Challenges and opportunities with big data, Computing Research Consortium (CSR), 2012.


A Big Data Analytics Life cycle

- A pipeline process with five stages:
  1. Data acquisition and recording
  2. Information extraction and cleaning
  3. Data integration, aggregation, and representation
  4. Query processing, data modeling, and analysis
  5. Interpretation
- Privacy protection is one of the common criteria for all of the related pipeline stage services.
- Apply semantics-enabled policy management for seamlessly bound pipeline services together to achieve balancing data utility and privacy protection on the decentralized social Web.

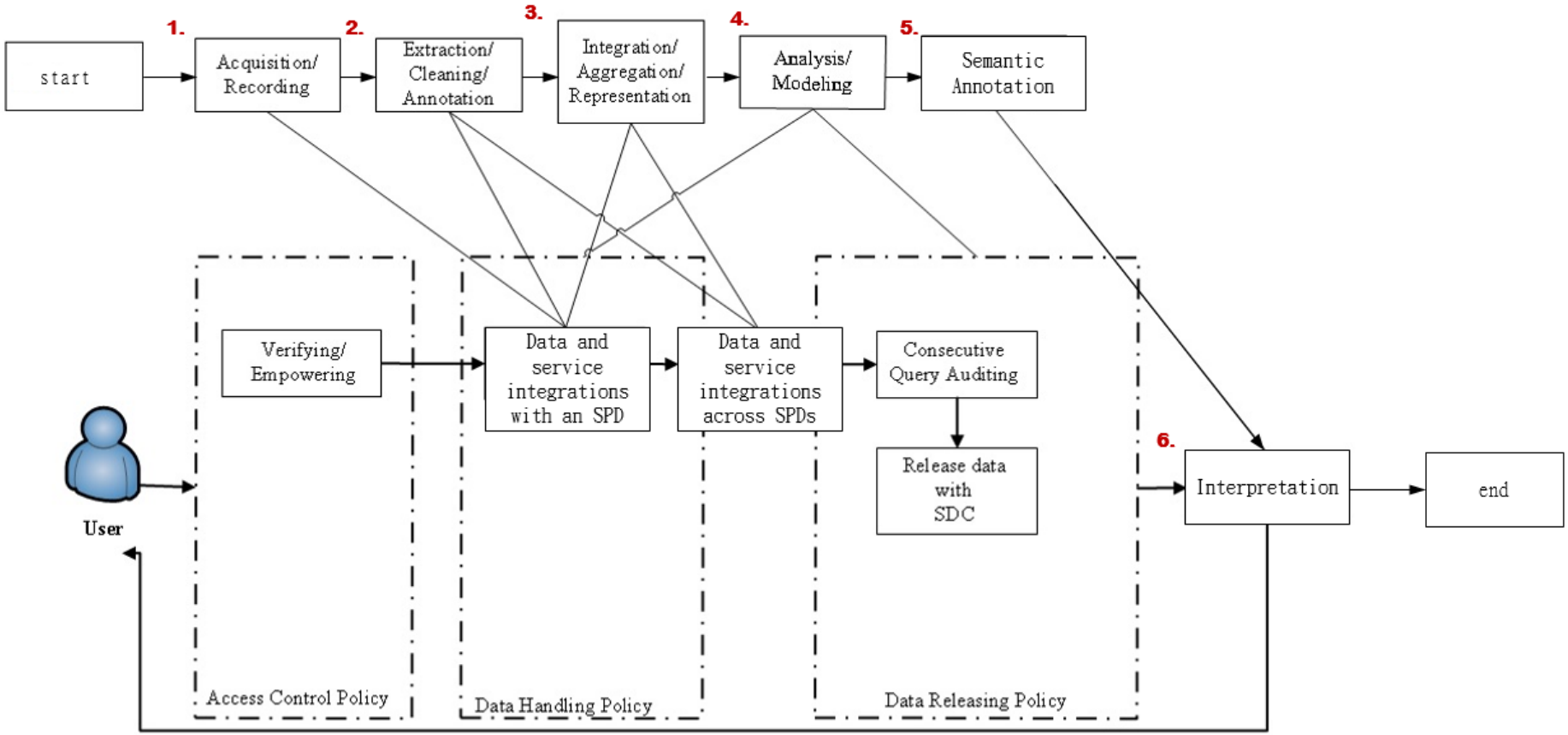
–A. Labrinidis et al. Challenges and opportunities with big data, Computing Research Consortium (CSR), 2012.

A Big Data Analytics Life cycle

- A pipeline process with five stages:
  1. Data acquisition and recording
  2. Information extraction and cleaning
  3. Data integration, aggregation, and representation
  4. Query processing, data modeling, and analysis
  5. Interpretation
- Privacy protection is one of the common criteria for all of the related pipeline stage services.
- Apply semantics-enabled policy management for seamlessly bound pipeline services together to achieve balancing data utility and privacy protection on the decentralized social Web.

–A. Labrinidis et al. Challenges and opportunities with big data, Computing Research Consortium (CSR), 2012.


Automated Big Data Analysis Pipeline



Semantics of Super-Peer Domain (SPD) Cloud

1. Semantics of a super-peer data cloud is described as the policy ontology, including modular concepts of SPD.
2. Semantics-enabled policies perform data and services integration within an SPD.
3. Semantics-enabled policies are unified to fulfill data and services integration across SPDs
4. Proposed various ontologies and rules for socially aware data cloud will be downgraded from OWL-based to RDF(S)-based, but this not yet complete!
5. Why use RDF(S)-based linked data and SPARQL query services for big data in the socially aware data cloud?

Semantics of Super-Peer Domain (SPD) Cloud

1. Semantics of a super-peer data cloud is described as the policy ontology, including modular concepts of SPD.
2. Semantics-enabled policies perform data and services integration within an SPD.
3. Semantics-enabled policies are unified to fulfill data and services integration across SPDs
4. Proposed various ontologies and rules for socially aware data cloud will be downgraded from OWL-based to RDF(S)-based, but this not yet complete!
5. Why use RDF(S)-based linked data and SPARQL query services for big data in the socially aware data cloud?

Semantics of Super-Peer Domain (SPD) Cloud

1. Semantics of a super-peer data cloud is described as the policy ontology, including modular concepts of SPD.
2. Semantics-enabled policies perform data and services integration within an SPD.
3. Semantics-enabled policies are unified to fulfill data and services integration across SPDs
4. Proposed various ontologies and rules for socially aware data cloud will be downgraded from OWL-based to RDF(S)-based, but this not yet complete!
5. Why use RDF(S)-based linked data and SPARQL query services for big data in the socially aware data cloud?

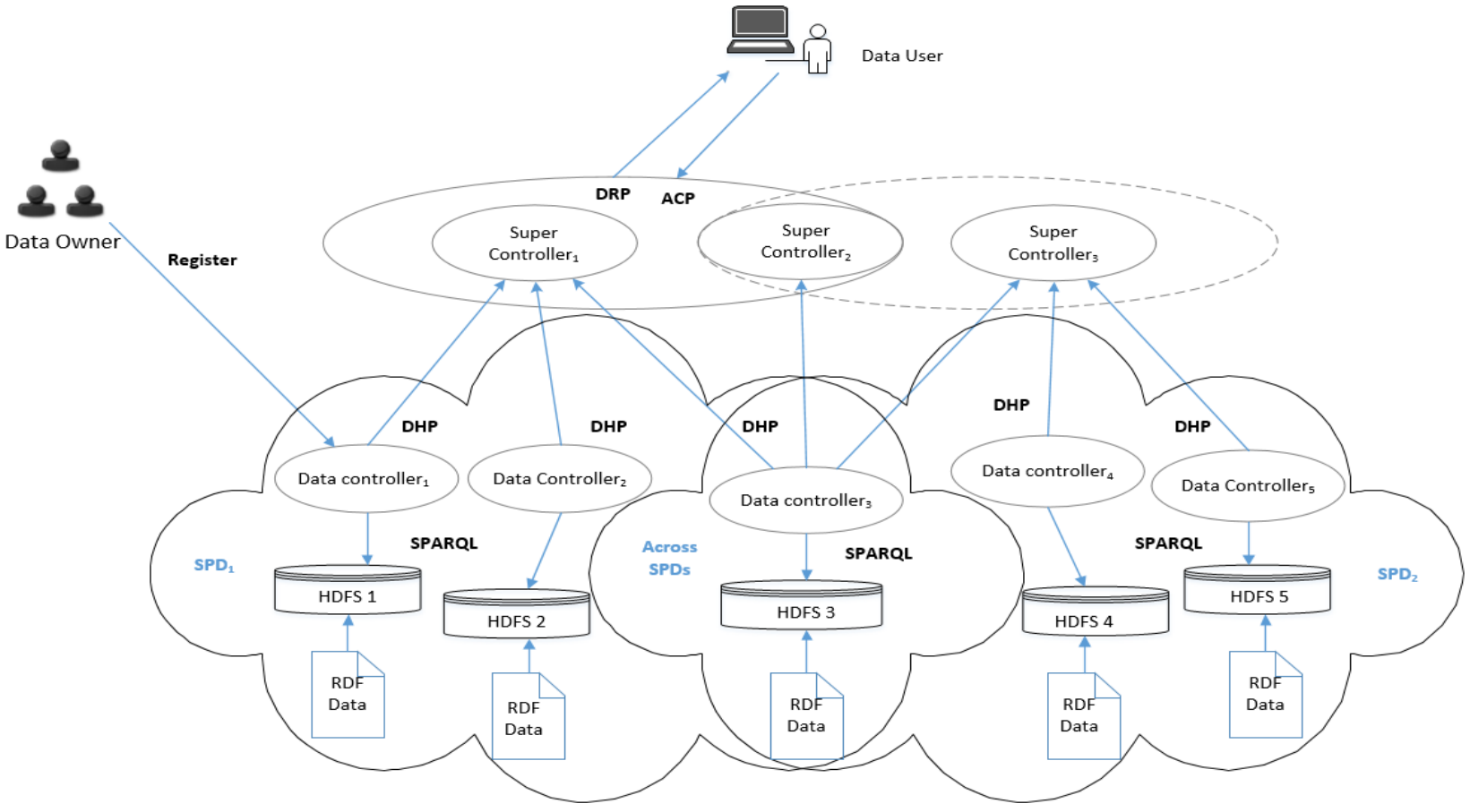
Semantics of Super-Peer Domain (SPD) Cloud

1. Semantics of a super-peer data cloud is described as the policy ontology, including modular concepts of SPD.
2. Semantics-enabled policies perform data and services integration within an SPD.
3. Semantics-enabled policies are unified to fulfill data and services integration across SPDs
4. Proposed various ontologies and rules for socially aware data cloud will be downgraded from OWL-based to RDF(S)-based, but this not yet complete!
5. Why use RDF(S)-based linked data and SPARQL query services for big data in the socially aware data cloud?

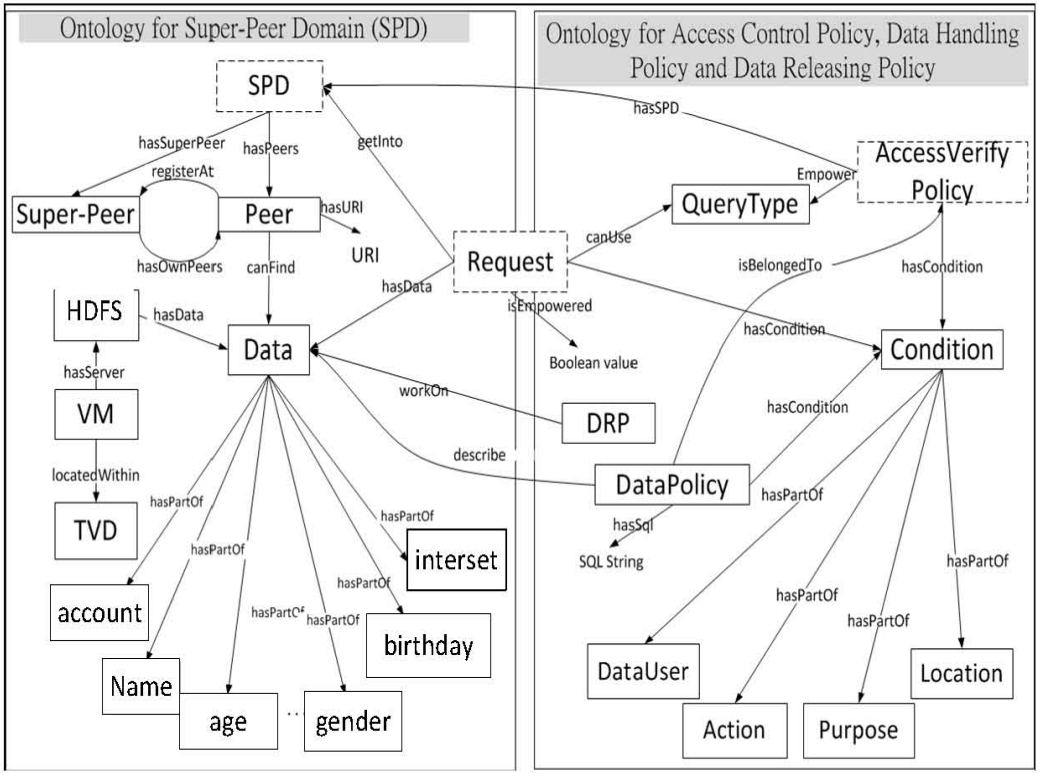
Semantics of Super-Peer Domain (SPD) Cloud

1. Semantics of a super-peer data cloud is described as the policy ontology, including modular concepts of SPD.
2. Semantics-enabled policies perform data and services integration within an SPD.
3. Semantics-enabled policies are unified to fulfill data and services integration across SPDs
4. Proposed various ontologies and rules for socially aware data cloud will be downgraded from OWL-based to RDF(S)-based, but this not yet complete!
5. Why use RDF(S)-based linked data and SPARQL query services for big data in the socially aware data cloud?

Semantics of Super-Peer Domain (SPD) Cloud (Conti.)



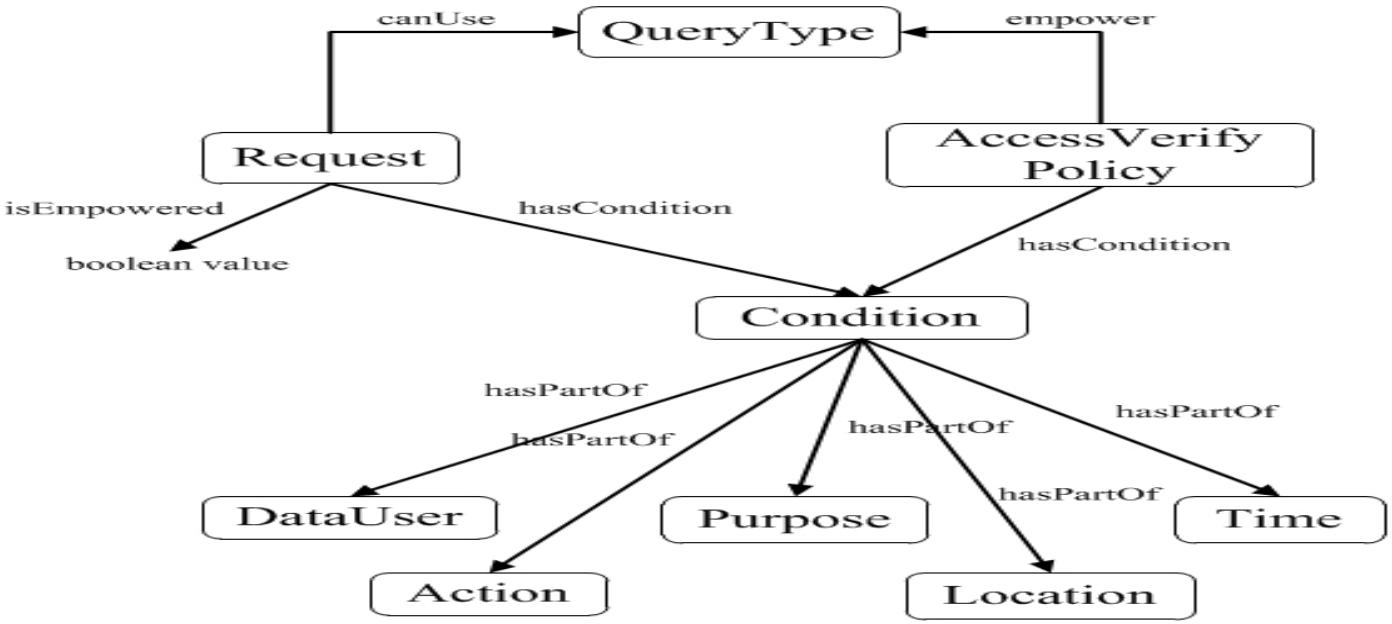
Policy Ontology for Super-Peer Domain Cloud



Ontology for Access Control Policy (ACP)

Definition of ACP Ontology

ACP describes the concept of data usage access control in the super-peer of an SPD.



Rule for Access Control Policy (ACP)

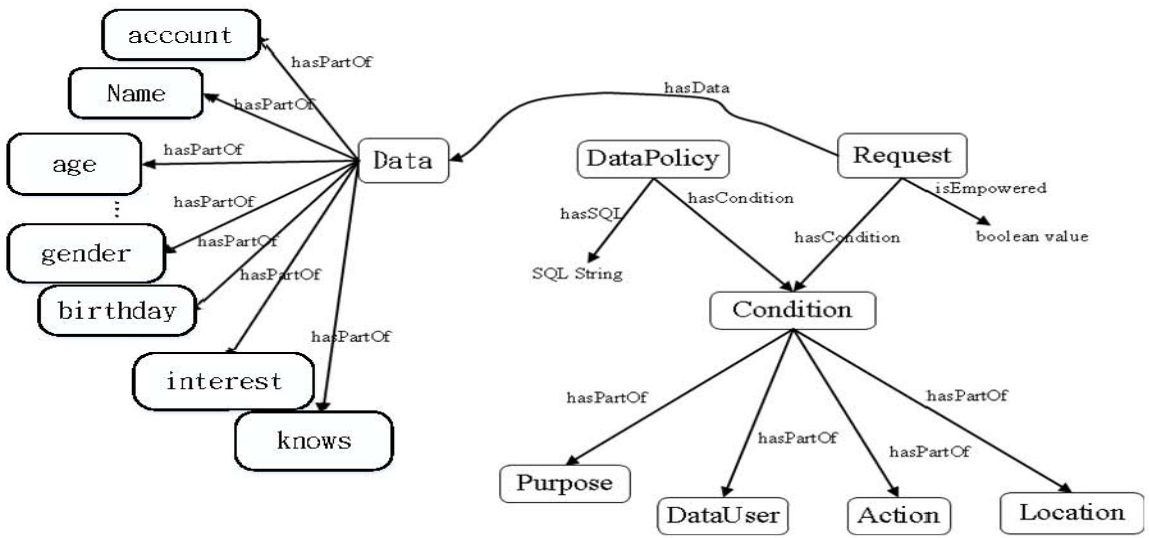
Specification of ACP Rule

$$\texttt{Request}(?r) \wedge \texttt{hasCondition}(?r, ?c) \wedge \texttt{Condition}(?c)$$
$$\wedge \texttt{hasCondition}(?avp, ?ac) \wedge \texttt{Condition}(?ac)$$
$$\wedge \texttt{AccessVerifyPolicy}(?avp) \wedge \texttt{sameAs}(?ac, ?c)$$
$$\wedge \texttt{empower}(?avp, ?qt) \wedge \texttt{QueryType}(?qt)$$
$$\longrightarrow \texttt{isEmpowered}(?r, 1) \wedge \texttt{hasQueryType}(?r, ?qt) \longleftarrow (1)$$

Ontology for Data Handling Policy (DHP)

Definition of DHP Ontology

DHP describes the concept of semantic metadata markup services and decides which data owners' privacy preferences match which data users' usage context.



Rule for Data Handling Policy (DHP)

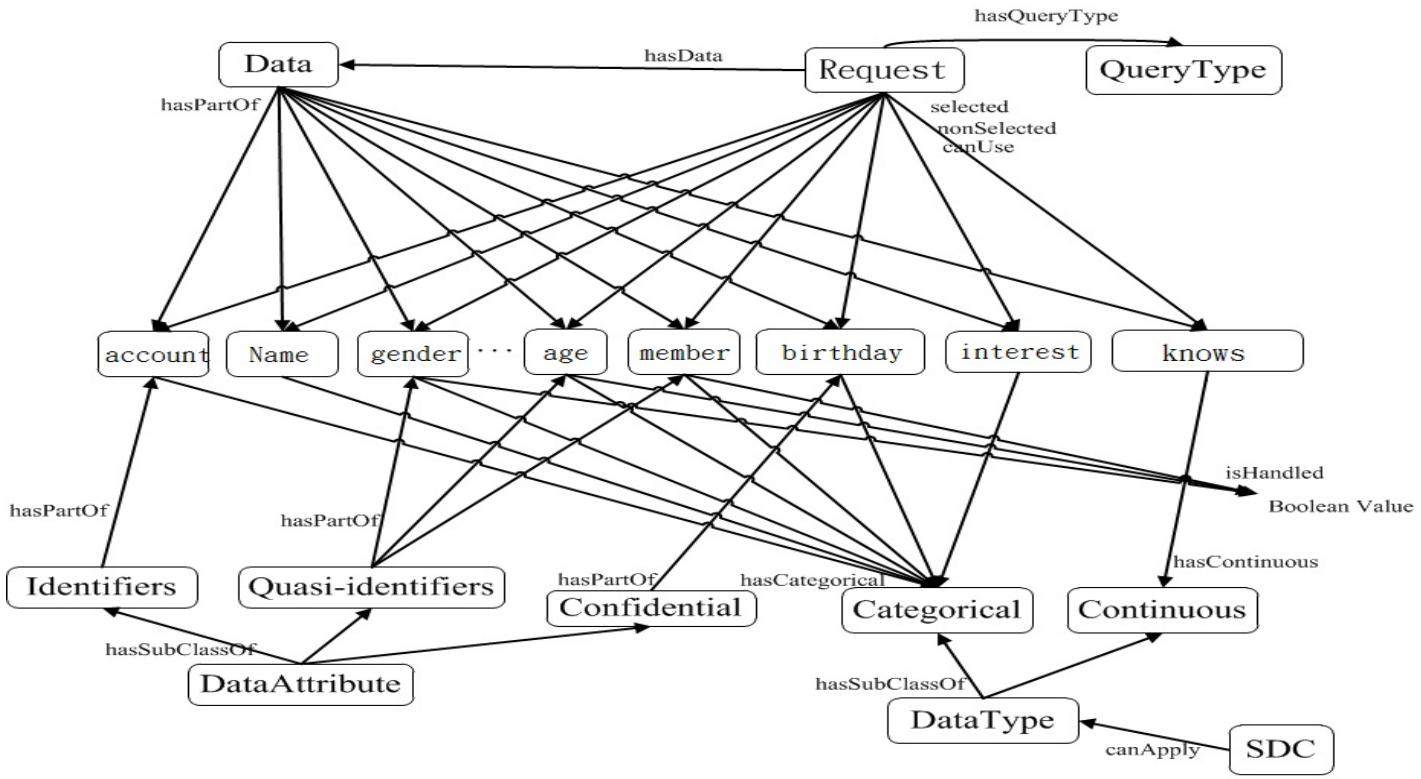
Specification of DHP Rule

$$\texttt{Request}(?r) \wedge \texttt{isEmpowered}(?r, 1) \wedge \texttt{hasCondition}(?r, ?c) \wedge \texttt{Condition}(?c) \wedge \texttt{DataPolicy}(?dp) \wedge \texttt{Condition}(?dc) \wedge \texttt{hasCondition}(?dp, ?dc) \wedge \texttt{sameAs}(?c, ?dc) \wedge \texttt{hasSPARQL}(?dp, ?s) \longrightarrow \texttt{sqwrl} : \texttt{select}(?s) \longleftarrow (2)$$

Ontology for Data Releasing Policy (DRP)

Definition of DRP Ontology

DRP describes the concept for which part of PII attributes are allowed to disclose for analysis and still ensures the privacy principles.



Ontology for Data Releasing Policy (DRP)(Conti.)

Definition of DRP Ontology

- hasData.Request(), hasData^{-}.Data().
- hasQueryType.Request(), hasQueryType^{-}.QueryType(PBQs).
- hasPartOf.Data(), hasPartOf^{-}.ID(), hasPartOf^{-}.Name(), ...
- hasPartOf^{-}.ZIP(), hasPartOf^{-}.Cholesterol().
- hasSubClassOf.DataAttribute(),
- hasSubClassOf^{-}.Identifiers(),
- hasSubClassOf^{-}.Quasi − identifiers(),
- hasSubClassOf^{-}.Confidential().
- hasPartOf.Identifiers(), hasPartOf^{-}.ID(id.), ...
- hasPartOf.Confidential(), hasPartOf^{-}.Disease().


Ontology for Data Releasing Policy (DRP)(Conti.)

Definition of DRP Ontology

- hasSubClassOf.DataType(),
- hasSubClassOf^{-}.Categorical(),
- hasSubClassOf^{-}.Continuous().
- hasContinuous.Cholesterol(), hasContinuous^{-}.Continuous().
- hasCategorical.ID(), $\text{hasCategorical}^{-}$.Categorical().
  ...
- hasCategorical.Doctor(), $\text{hasCategorical}^{-}$.Categorical().
- canApply.SDC(generalization), canApply^{-}.Categorical().
  ...
- $\text{canApply.SDC}(\text{top} - \text{coding})$, canApply^{-}.Continuous().


Rules for Data Handling Policy (DHP)

Specification of DHP Rules

$$\texttt{Request}(?r) \wedge \texttt{hasData}(?r, ?d) \wedge \texttt{Data}(?d)$$
$$\wedge \texttt{hasPartOf}(?d, ?pod) \wedge \texttt{hasQueryType}(?r, \texttt{PBQ})$$
$$\wedge \texttt{sqwrl} : \texttt{makeSet}(?rs, ?pod) \wedge \texttt{sqwrl} : \texttt{groupBy}(?rs, ?r)$$
$$\wedge \texttt{Quasi} - \texttt{identifiers}(?qui) \wedge \texttt{hasPartOf}(?qui, ?qpod)$$
$$\wedge \texttt{sqwrl} : \texttt{groupBy}(?qs, ?qui) \wedge \texttt{sqwrl} : \texttt{contains}(?rs, ?qs)$$
$$\wedge \texttt{Confidential}(?c) \wedge \texttt{hasPartOf}(?c, ?dc)$$
$$\longrightarrow \texttt{sqwrl} : \texttt{selectDistinct}(?qui, ?gpod) \longleftarrow (3)$$

Rules for Data Handling Policy (DHP)(Conti.)

Specification of DHP Rules

$$\texttt{Request}(?r) \wedge \texttt{hasData}(?r, ?d) \wedge \texttt{Data}(?d)$$
$$\wedge \texttt{hasPartOf}(?d, ?b) \wedge \texttt{selected}(?r, ?b)$$
$$\wedge \texttt{hasContinuous}(?b, ?con) \wedge \texttt{Continuous}(?con)$$
$$\wedge \texttt{SDC}(?sdc) \wedge \texttt{canApply}(?sdc, ?con)$$
$$\longrightarrow \texttt{sqwrl} : \texttt{select}(?b, ?sdc) \longleftarrow (4)$$

Specification of DHP Rules

$$\texttt{Request}(?r) \wedge \texttt{hasData}(?r, ?d) \wedge \texttt{Data}(?d)$$
$$\wedge \texttt{hasPartOf}(?d, ?b) \wedge \texttt{selected}(?r, ?b)$$
$$\wedge \texttt{hasCategorical}(?b, ?cat) \wedge \texttt{Categorical}(?con)$$
$$\wedge \texttt{SDC}(?sdc) \wedge \texttt{canApply}(?sdc, ?cat)$$
$$\longrightarrow \texttt{sqwrl} : \texttt{select}(?b, ?sdc) \longleftarrow (5)$$


Rules for Data Handling Policy (DHP)(Conti.)

Specification of DHP Rules

$$\texttt{Request}(?r) \wedge \texttt{hasData}(?r, ?d) \wedge \texttt{Data}(?d)$$
$$\wedge \texttt{hasPartOf}(?d, ?b) \wedge \texttt{selected}(?r, ?b)$$
$$\wedge \texttt{hasContinuous}(?b, ?con) \wedge \texttt{Continuous}(?con)$$
$$\wedge \texttt{SDC}(?sdc) \wedge \texttt{canApply}(?sdc, ?con)$$
$$\longrightarrow \texttt{sqwrl} : \texttt{select}(?b, ?sdc) \longleftarrow (4)$$

Specification of DHP Rules

$$\texttt{Request}(?r) \wedge \texttt{hasData}(?r, ?d) \wedge \texttt{Data}(?d)$$
$$\wedge \texttt{hasPartOf}(?d, ?b) \wedge \texttt{selected}(?r, ?b)$$
$$\wedge \texttt{hasCategorical}(?b, ?cat) \wedge \texttt{Categorical}(?con)$$
$$\wedge \texttt{SDC}(?sdc) \wedge \texttt{canApply}(?sdc, ?cat)$$
$$\longrightarrow \texttt{sqwrl} : \texttt{select}(?b, ?sdc) \longleftarrow (5)$$

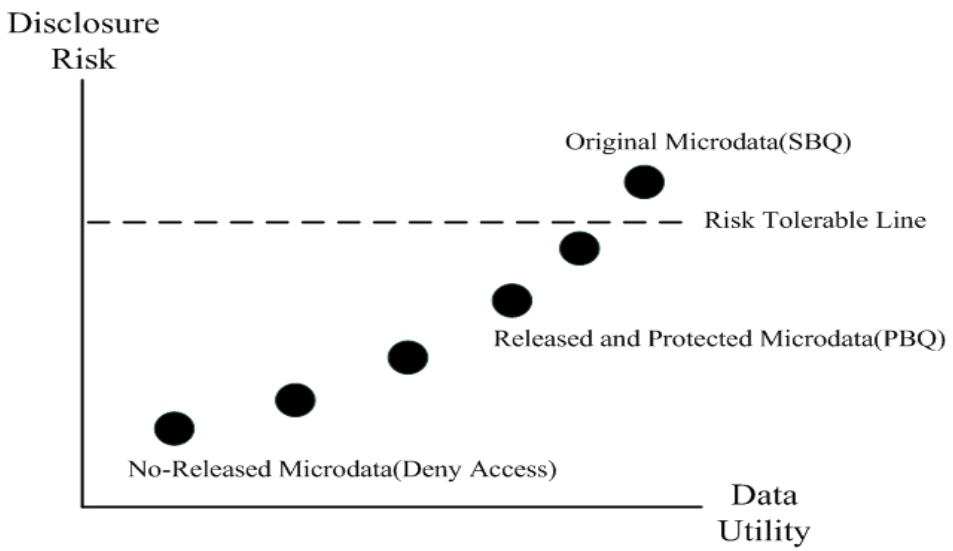

Rules for Data Handling Policy (DHP)(Conti.)

Specification of DHP Rules

$$\texttt{Request}(?r) \wedge \texttt{hasData}(?r, ?d) \wedge \texttt{Data}(?d)$$
$$\wedge \texttt{hasPartOf}(?d, ?b) \wedge \texttt{select}(?r, ?b) \wedge \texttt{isHandled}(?b, 1)$$
$$\wedge \texttt{hasPartOf}(?d, ?a) \wedge \texttt{notSelected}(?r, ?a)$$
$$\longrightarrow \texttt{canUse}(?r, ?a) \wedge \texttt{canUse}(?r, ?b) \longleftarrow (6)$$

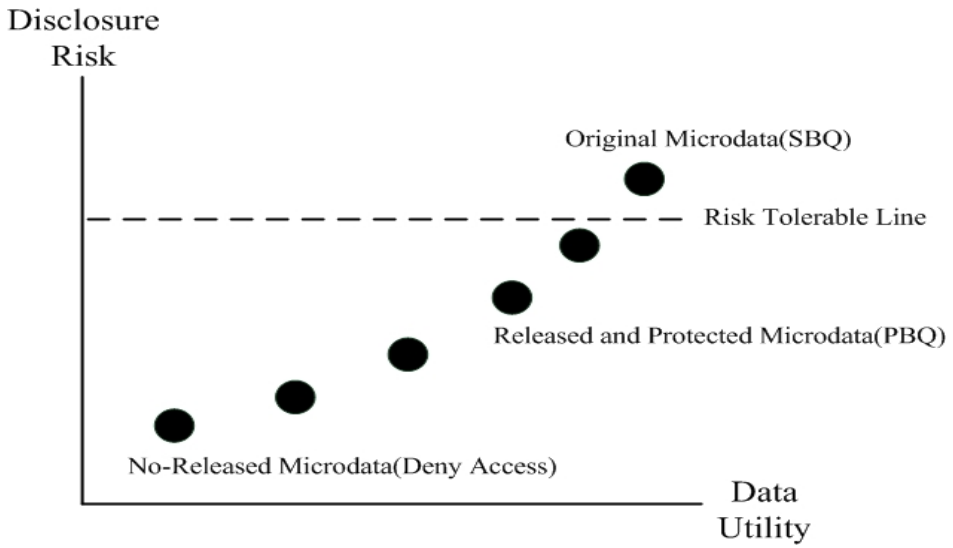
Semantic Data Analysis and Protection

- Improve the situation, where SDC enforcement is obliged to original data providers and a data analytics user lacks the flexibility to choose suitable SDC methods.
- Seek a balance between a data owner's right for privacy protection and a data user's need for data analytics through transparency of SDC methods releasing.
- Semantics-enabled Data Releasing Policy (DRP) calls for which SDC methods and ensures maximum data utility with a tolerable data disclosure risk.



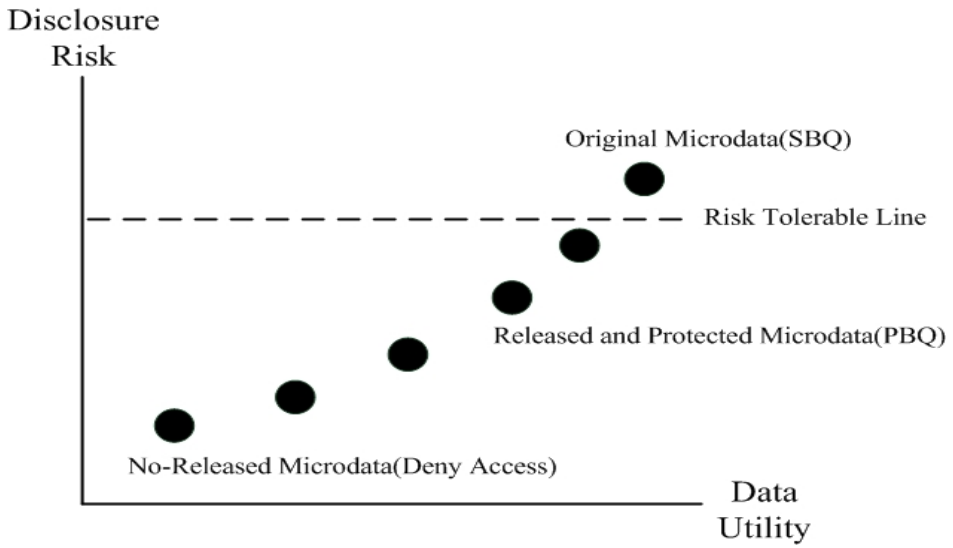
Semantic Data Analysis and Protection

- Improve the situation, where SDC enforcement is obliged to original data providers and a data analytics user lacks the flexibility to choose suitable SDC methods.
- Seek a balance between a data owner's right for privacy protection and a data user's need for data analytics through transparency of SDC methods releasing.
- Semantics-enabled Data Releasing Policy (DRP) calls for which SDC methods and ensures maximum data utility with a tolerable data disclosure risk.

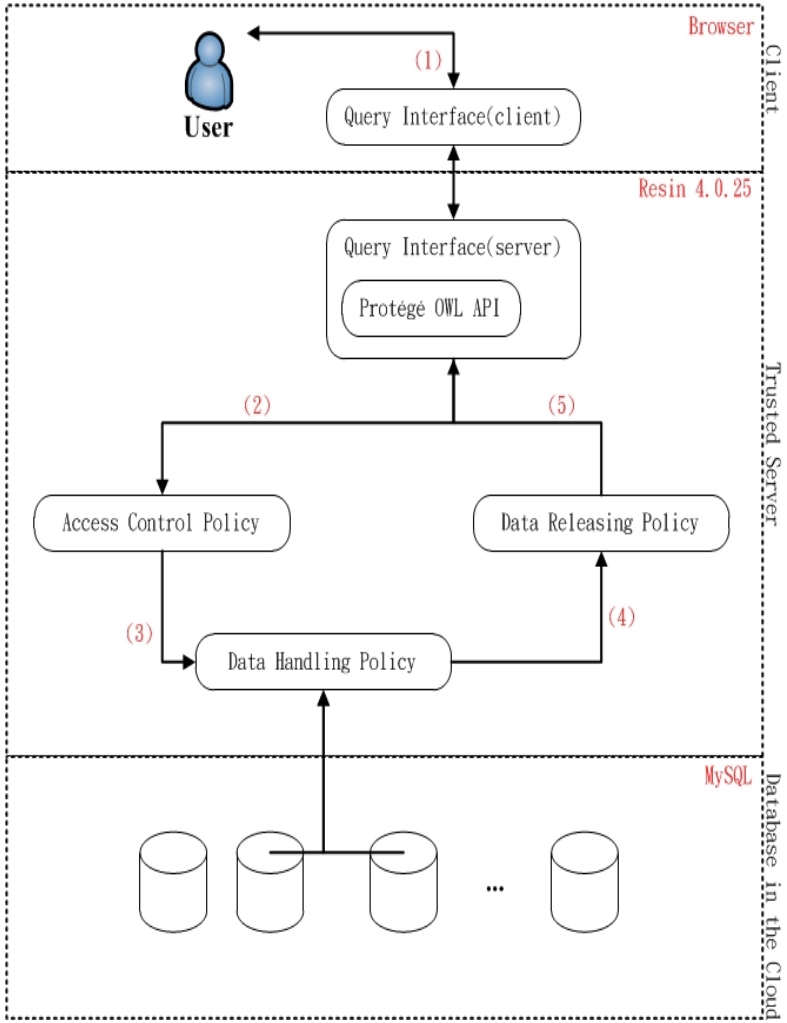


Semantic Data Analysis and Protection

- Improve the situation, where SDC enforcement is obliged to original data providers and a data analytics user lacks the flexibility to choose suitable SDC methods.
- Seek a balance between a data owner's right for privacy protection and a data user's need for data analytics through transparency of SDC methods releasing.
- Semantics-enabled Data Releasing Policy (DRP) calls for which SDC methods and ensures maximum data utility with a tolerable data disclosure risk.



A Three-Tier SDC Prototyping System



Conclusion and Future Works

- Preliminary Results:
  1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which ensure the privacy protection principles.
  2. A RHdoop platform is establishing to provide privacy-preserving deep analytics of socially aware big data.
  3. Supporting a simple balance between data utility and protection through semantics-enabled policies call for SDC methods.
- Future Works:
  1. Semantics-enabled policies, e.g., ACP, DHP, and DRP, are fully represented and enforced as RDF(S)-based linked data analytics for privacy-preserving on the decentralized social Web.
  2. The ultimate goal is to craft an optimized balance between data utility and protection through the automated big data analysis life cycle.


Conclusion and Future Works

- Preliminary Results:
  1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which ensure the privacy protection principles.
  2. A RHdoop platform is establishing to provide privacy-preserving deep analytics of socially aware big data.
  3. Supporting a simple balance between data utility and protection through semantics-enabled policies call for SDC methods.
- Future Works:
  1. Semantics-enabled policies, e.g., ACP, DHP, and DRP, are fully represented and enforced as RDF(S)-based linked data analytics for privacy-preserving on the decentralized social Web.
  2. The ultimate goal is to craft an optimized balance between data utility and protection through the automated big data analysis life cycle.


Conclusion and Future Works

- Preliminary Results:
  1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which ensure the privacy protection principles.
  2. A RHdoop platform is establishing to provide privacy-preserving deep analytics of socially aware big data.
  3. Supporting a simple balance between data utility and protection through semantics-enabled policies call for SDC methods.
- Future Works:
  1. Semantics-enabled policies, e.g., ACP, DHP, and DRP, are fully represented and enforced as RDF(S)-based linked data analytics for privacy-preserving on the decentralized social Web.
  2. The ultimate goal is to craft an optimized balance between data utility and protection through the automated big data analysis life cycle.


Conclusion and Future Works

- Preliminary Results:
  1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which ensure the privacy protection principles.
  2. A RHdoop platform is establishing to provide privacy-preserving deep analytics of socially aware big data.
  3. Supporting a simple balance between data utility and protection through semantics-enabled policies call for SDC methods.
- Future Works:
  1. Semantics-enabled policies, e.g., ACP, DHP, and DRP, are fully represented and enforced as RDF(S)-based linked data analytics for privacy-preserving on the decentralized social Web.
  2. The ultimate goal is to craft an optimized balance between data utility and protection through the automated big data analysis life cycle.


Conclusion and Future Works

- Preliminary Results:
  1. Semantics-enabled policies, ACP, DHP, and DRP, are proposed and verified through query restriction, manipulation, and output perturbation, which ensure the privacy protection principles.
  2. A RHdoop platform is establishing to provide privacy-preserving deep analytics of socially aware big data.
  3. Supporting a simple balance between data utility and protection through semantics-enabled policies call for SDC methods.
- Future Works:
  1. Semantics-enabled policies, e.g., ACP, DHP, and DRP, are fully represented and enforced as RDF(S)-based linked data analytics for privacy-preserving on the decentralized social Web.
  2. The ultimate goal is to craft an optimized balance between data utility and protection through the automated big data analysis life cycle.