Vol.2011-CSEC-54 No.27
Vol.2011-SPT-1 No.27
2011/7/13

行動情報活用型クラウドサービス振興のためのデータ匿名化プラットフォーム「匿名化クラウド」のアーキテクチャ提案

永井康彦[†]　五十嵐亮基[†]　羅川広行[†]　松岡健[†]
藤田麻里央[†]　佐藤祥太郎[†]　美馬正司[†]

情報爆発時代の収集したパーソナル情報を利活用した新たなサービス創出の前提として，プライバシー確保のためにパーソナル情報を匿名化処理する技術の研究・開発が進められている．また，今後は共有匿名化ライブラリソフトを社会の新たなIT基盤であるクラウド基盤上に搭載することにより,匿名化処理のデータ保管・処理コスト削減，導入容易性や処理品質の確保，さらにサービス事業者間で協創する基盤となり，新たなサービス創出を啓発することが期待できる．本稿では，サービスに応じて適した匿名化処理を選択可能なクラウド上の共有匿名化処理ミドルウエア「匿名化クラウド」の機能アーキテクチャを提案する．

A Proposal on Architecture of “Anonymization Cloud” as Data Anonymization Platform for Promotion of Cloud Computing Services using Personal Activity Information

Yasuhiko Nagai[†], Ryoki Igarashi[†], Hiroyuki Serikawa[†],
Ken Matsuoka[†], Mario Fujita[†], Shotaro Sato[†],
and Tadashi Mima[†]

As a premise of the new service creation using personal activity information in the information explosion age, the research and development of the technique on anonymization processing of personal activity information for privacy protection. Especially, by using the shared anonymization library software on the cloud computing environment in future, it is can expect cost cut of data safekeeping and processing, introduction easiness, processing quality, and promotion of new service creation by collaboration on anonymization data and processings between service companies. In this paper, we propose the functional architecture of "Anonymization Cloud" as the shared middleware which can choose suitable anonymization processing depending on service.

1. はじめに

Web，非 Web 問わず膨大な数の多種多様な情報があふれ，収集・蓄積・解析・発信される情報爆発時代を迎えている中で，収集・蓄積した個人の生活や行動に関するパーソナル情報を利活用した新たなサービス（より適時で個人最適化されたレコメンドサービス等）の創出が期待されている．その前提としては，プライバシー確保が重要であり，情報の利活用と保護の両立を図るための技術として，暗号化やランダム化に並び，有効な技術としてパーソナル情報を匿名化処理する技術（個人の識別が困難になるよう処理する技術）の研究・開発が進められている [1)]．また，このパーソナル情報を利活用した新たなサービスや匿名化技術の有望な実用先として，社会の新たな IT 基盤になろうとしているクラウドコンピューティング環境がある．

匿名化技術としては，個人識別情報（各個人をそれ単体で一意に識別可能な情報，氏名や運転免許番号等）の削除や ID 化，準識別情報（それ単体では必ずしも個人は識別されないが，複数組合せることによって個人の識別に至る情報，年齢や性別等）の一般化等の個々のデータの範囲の匿名性（個人再特定不能性）を確保する匿名化処理が従来から一般的に活用されてきた．最近では，これらに加えて，他データや背景知識とのマッチングによって個人が再特定されたり個人の属性情報が推定されることを防ぐ，よりプライバシー保護を強化した集合データの範囲の匿名性，すなわち k-匿名性（パーソナル情報の表形式データにおいて，準識別情報の属性値の組合せが同じである行が少なくとも k 行存在することを保証）や l-多様性（パーソナル情報の k-匿名性を持つ表形式データにおいて，準識別情報の属性値の組合せが同じである k 行について，その属性値の多様性が l 個あることを保証）を確保する匿名化技術の研究や 2009 年頃から特定の匿名化処理のライブラリ・ツールが国内外で出現してきているところである [1),2)]．

ところで，これら匿名化技術は，適用されるアプリケーションやサービスで扱うパーソナル情報の情報項目の機密性や重要性，公開範囲等に応じて，多種の中から適切なものが選択されることとなる．このため，多様な匿名化処理を包含した共有匿名化ライブラリソフト・ツールを用意し，匿名化処理の利用者が適切な処理を選択して活用できるようにすることが有用である．また，通常，大規模データ処理となるパーソナル情報の匿名化処理のデータ保管・処理コスト削減のために，今後は共有匿名化ライブラリソフトをクラウド基盤上に搭載して共用することが考えられるが，単にコスト削減だけでなく，匿名化処理の導入容易性，匿名化処理の品質確保，さらに匿名化処理や匿名化データをサービス事業者間で協創する基盤として，新たなサービス創出を啓発することが期待できる．

[†]（株）日立コンサルティング
Hitachi Consulting Co., Ltd

ⓒ2011 Information Processing Society of Japan

Vol.2011-CSEC-54 No.27
Vol.2011-SPT-1 No.27
2011/7/13

そこで，本稿では，サービスに応じて適した匿名化処理が選択可能であり，マルチテナント環境で匿名化処理を実行できるデータ匿名化プラットフォームをクラウド上に実現するための第一ステップとして，クラウド上に新たな共有匿名化処理ミドルウエアを搭載したデータ匿名化プラットフォーム（「匿名化クラウド」と呼ぶ）の機能アーキテクチャを提案する [3].

2. 匿名化クラウドのアーキテクチャ開発の必要性

匿名化クラウドの開発において，クラウド利用の多様なアプリケーションに共通・有効活用できるようにするために，以下のような事項が要件となる.

(1) 汎用性・標準性の確保

クラウド利用の多様なアプリケーションに広く共通に活用できるよう，多様なクラウド基盤上で稼動できるようにするために，一般的・標準的クラウドアーキテクチャとの整合や相互接続性を考慮すること.

(2) クラウドコンピューティング技術の特長の有効利用

クラウドコンピューティングの技術的な動向を踏まえ，分散処理，KVS（Key-Value Store）型データベース等，クラウドコンピューティングの特長を有効利用すること.

(3) 法・制度上の制約への対応

個人情報保護法等へのコンプライアンスやデータの安全性確保の観点から，パーソナル情報自身（元データ）は，アプリケーションの各種機関側で管理し，匿名化データをクラウド側で管理するハイブリッド型の情報管理といったセキュリティドメイン間隔離のある情報管理等，法的制約を考慮すること.

(4) 実現・実用段階への適応性

今後の新サービス創出に有効なパーソナル情報を匿名化して収集機関とは別の機関へ提供・流通させる場合（2 次利用）のデータ管理・匿名化処理だけでなく，既存の収集機関内で匿名化データを活用・管理する場合（1 次利用）の利用形態にも対応可能な実現・実用段階への適応ができること.

さらに，2009 年度の経済産業省の情報大航海プロジェクトにおいて，多様な匿名化処理を包含した共有匿名化ライブラリソフトである個人情報匿名化基盤 [1]を開発しており，この資産を有効活用して，これを拡張・発展させて効率的に開発することも追加要件としている.

これら要件を満足し，代表的クラウド基盤をベースとした匿名化クラウドの必要構成，必要機能，必要インタフェースを体系的に明確化するためには，匿名化クラウドのアーキテクチャを開発することがまず必要であると考えた.

3. アーキテクチャの開発アプローチと方法

匿名化クラウドのアーキテクチャ開発は，まず，匿名化クラウドのクラウド全体の中での位置づけや汎用性・標準性を確保するための外部インタフェース要件を明確化し，次に，匿名化クラウドのユースケースモデル（利用形態・動作フロー）から，必要な機能要件を導出，両者の結果を併せて，アーキテクチャの要件定義を行い，この要件を実現するアーキテクチャ設計をするという開発アプローチで，図 1 に示すような開発プロセスにより，実施した．各開発プロセスの概要・方法を以下に示す.

(1) クラウドアーキテクチャの動向調査

匿名化クラウドの位置づけ，関連標準 API 動向を把握するために，クラウドアーキテクチャの標準化団体での標準化動向および現状の代表的 CSP(Cloud Service Provider)のアーキテクチャ/API 動向を調査する．標準化動向調査は，匿名化クラウドが関係する PaaS/SaaS 及び関連 API を含む検討をしている代表的な標準化団体を，公開報告書，図書，ホームページ等の検索から安全サイドに網羅して，以下の 7 団体を調査対象として選定した.

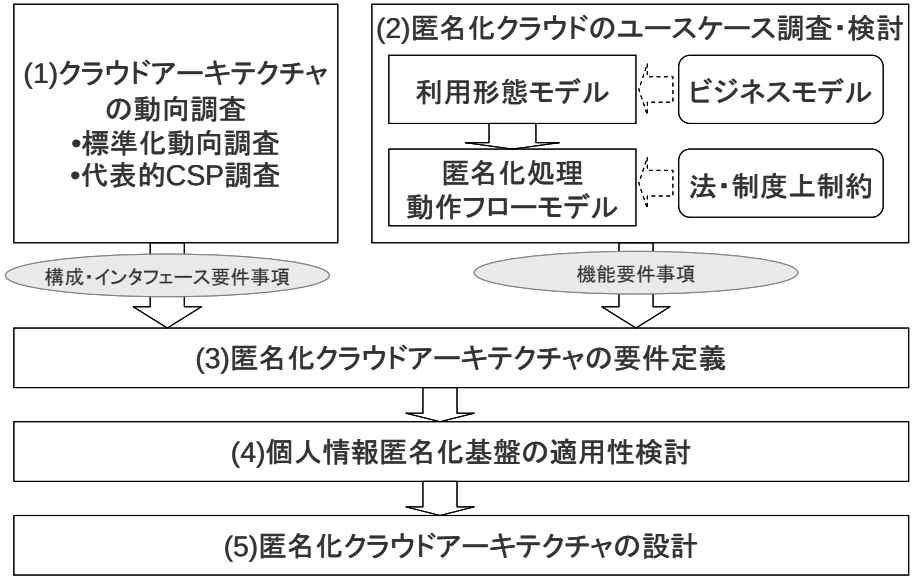

図 1　アーキテクチャ開発プロセス

ⓒ2011 Information Processing Society of Japan

Vol.2011-CSEC-54 No.27
Vol.2011-SPT-1 No.27
2011/7/13

① Open Cloud Manifesto
② CCIF（Cloud Computing Interoperability Forum）/UCI（Universal Cloud Interface）
③ CSA（Cloud Security Alliance）
④ DMTF（Distributed Management Task Force）
⑤ OMG（Object Management Group）
⑥ ISO/IEC JTC1 SC38：Distributed Applications Platforms and Services
⑦ OGC（Open Government Cloud consortium）

これら７団体の公開サイトから入手可能なドキュメントやホームページ情報より，まず７団体全体の概要調査を行い，次にその調査結果から匿名化クラウドのアーキテクチャ・位置づけや API 検討に有益な情報のある①，②，④の３つの標準化団体 [4)-7)] に絞り，詳細調査を実施した．

代表的 CSP の調査は，匿名化クラウドが多様な SaaS/AP から匿名化処理を横断的に共用してもらう共通基盤になるものであることから，PaaS を提供している代表的 CSP として，以下の３社の CSP を調査対象とした．

(a) Google：Google App Engine Platform [i)]
(b) Microsoft：Azure Services Platform [ii)]
(c) Salesforce：Force.com [iii)]

各社の PaaS に関して公開サイト/セミナー等から入手可能なドキュメント，ガイドや市販図書 [8)-10)]から，各社 PaaS が匿名化クラウドの基盤になることを想定し，現状 CSP の PaaS アーキテクチャ，API の観点だけでなく，匿名化クラウドの位置づけや開発環境・組込み方法も併せて調査した．

また，これら調査結果から標準・業界のクラウドアーキテクチャ動向との整合に配慮した匿名化クラウドへの構成・インタフェース要件考慮事項を導出した．

(2) 匿名化クラウドのユースケース調査・検討

データ匿名化処理の多様な機関（１次事業者，第３者機関，２次事業者）や多様な役割分担（統合型：一括自己処理，分離型：複数組織で分担処理，委託型：一括委託）の利用形態に適応できるよう，匿名化クラウドの利用形態モデルを検討・定義し，各利用形態モデルにつき，基本的な動作処理フローモデルを検討・定義する．その際，多様なサービス利用形態（統合・委託型，分離型）に応じて必要な匿名化機能を選択，追加可能とする情報大航海プロジェクトの個人情報匿名化基盤のプラグ＆プレイの機能モジュール構成の特長を，匿名化クラウドでも継承して実現，利用形態モデル定義では想定されるビジネスモデルの調査を行いそれへの適用性，動作処理フローモデル定義では関連する法・制度の調査を行ってそれを制約条件として検討・定義に反映する．また，この調査・検討結果とクラウド化により有効利用できる特徴（分散処理，マルチテナント，スケーラビリティ等）を併せ，匿名化クラウドへの機能要件考慮事項を導出した．

(3) 匿名化クラウドアーキテクチャの要件定義

上記(1)の結果から導出された構成・インタフェース要件事項，上記(2)の結果から導出された機能要件事項を統合して，匿名化クラウドアーキテクチャの要件（構成要件，機能要件，インタフェース要件）を定義した．

(4) 個人情報匿名化基盤の適用性検討

上記(3)で定義された構成・機能・インタフェース要件を基に，情報大航海プロジェクトで開発した個人情報匿名化基盤の流用可能部，改良・新規追加必要部を明確化した．

(5) 匿名化クラウドアーキテクチャの設計

上記(4)の適用性検討結果を考慮して，上記(3)の要件を実現する匿名化クラウドのアーキテクチャ（構成，機能，インタフェース）を設計した．また，その際には，実現性の観点から，匿名化クラウドの基盤と想定される代表的な現状 PaaS より利用する機能と新たに開発すべき機能を明確化した．

4. アーキテクチャ開発結果

4.1 標準化・代表的 CSP 動向調査結果概要

標準化動向調査結果と代表的 CSP 調査結果とをまとめたクラウドアーキテクチャ動向調査の結果まとめを表 1 に示す．表中の縦軸が調査の観点（アーキテクチャ，インタフェース），横軸が順次，標準化動向調査結果まとめ，代表的 CSP 調査結果まとめとなっており，参考に情報大航海プロジェクトの個人情報匿名化基盤（非クラウド）の現状仕様，最後にこれら結果まとめの標準化動向・現状 CSP 動向を踏まえ，匿名化クラウドアーキテクチャへの構成・インタフェース考慮事項を，匿名化クラウド方針として記載している．

i) Google, Google App Engine Platform は，米国 Google Inc.の米国およびその他の国における商標または登録商標です．
ii) Microsoft, Azure Services Platform は，米国 Microsoft Corporation の米国およびその他の国における商標または登録商標です．
iii) Salesforce, Force.com は，米国 Salesforce.com の米国およびその他の国における商標または登録商標です．

ⓒ2011 Information Processing Society of Japan

Vol.2011-CSEC-54 No.27
Vol.2011-SPT-1 No.27
2011/7/13

表 1　クラウドアーキテクチャ動向調査の結果まとめ

観点＼標準性			標準化団体		現状CSP (GAE,Azure,Force.com等)	参考: 個人情報匿名化基盤 (非ｸﾗｳﾄﾞ)	匿名化ｸﾗｳﾄﾞ方針
アーキテクチャ			ｸﾗｳﾄﾞ標準ｲﾝﾀﾌｪｰｽｻﾎﾟｰﾄ型(OCM,DMTF)	ｸﾗｳﾄﾞ統合ﾌﾞﾘｯｼﾞｴｰｼﾞｪﾝﾄ利用型(UCI)	•位置づけ PaaS内開発環境: ﾌﾚｰﾑﾜｰｸ, IDE		•位置づけ PaaS and/or SaaSに組込利用
インタフェース	各種ｸﾗｳﾄﾞ間連携	API		•API(ﾚﾍﾞﾙ4, ｶﾃｺﾞﾘ3): UCIとしてW3CｾﾏﾝﾃｨｯｸWeb参考ｵﾝﾄﾛｼﾞｰ言語ﾍﾞｰｽのRDF利用（どんなcloudAPIも抽象化し、semantic webとOWLにより提供CSPに係わらない一つのレイヤにAPI群を統合する）。 •データ検索/操作言語: SPARQL (SPARQL Protocol and RDF Query Language) •Agentｺﾏﾝﾄﾞｲﾝﾀﾌｪｰｽ: XMPP (eXtensible Messaging and Presence Protocol) •データ記述形式: RDF/OWL	・SOAP/REST		将来課題（時期尚早）
		ﾃﾞｰﾀ形式			・WSDL/HTML		将来課題（時期尚早）
	ｸﾗｳﾄﾞ-AP間	API	・APIﾚﾍﾞﾙ(全4ﾚﾍﾞﾙ) ﾚﾍﾞﾙ1:通信ﾌｫｰﾏｯﾄ&ﾌﾟﾛﾄｺﾙﾚﾍﾞﾙ(REST,SOAP) ﾚﾍﾞﾙ2:言語固有ﾂｰﾙｷｯﾄ利用してSOAP/REST処理 ﾚﾍﾞﾙ3:ｻｰﾋﾞｽ固有ﾂｰﾙｷｯﾄ利用してﾋﾞｼﾞﾈｽ・ｵﾌﾞｼﾞｪｸﾄ/ﾋﾞｼﾞﾈｽ・ﾌﾟﾛｾｽﾚﾍﾞﾙで処理 ﾚﾍﾞﾙ4:複数CSP共通ｲﾝﾀﾌｪｰｽ利用してﾋﾞｼﾞﾈｽ・ｵﾌﾞｼﾞｪｸﾄ/ﾋﾞｼﾞﾈｽ・ﾌﾟﾛｾｽﾚﾍﾞﾙで処理 ・APIｶﾃｺﾞﾘ(関連3/5全ｶﾃｺﾞﾘ, ｶﾃｺﾞﾘに応じAPI種別:ｻｰﾋﾞｽAPI/ｻﾎﾟｰﾄAPI全11種） ｶﾃｺﾞﾘ1:通常のﾌﾟﾛｸﾞﾗﾐﾝｸﾞ(C#, PHP, Java等, API3種） ｶﾃｺﾞﾘ2:ﾃﾞﾌﾟﾛｲﾒﾝﾄ・実装API(.Netｱｾﾝﾌﾞﾘ等, API5種） ｶﾃｺﾞﾘ3:ｸﾗｳﾄﾞ・ｻｰﾋﾞｽAPI(ｽﾄﾚｰｼﾞ, ﾃﾞｰﾀﾍﾞｰｽ, ﾒｯｾｰｼﾞｷｭｰｻｰﾋﾞｽ,ﾐﾄﾞﾙｳｪｱ共通API, API10種） ・ｾｷｭﾘﾃｨｺﾝﾄﾛｰﾙへの既存標準適用(OASIS,IEEE,ITU) •機能毎、ｻｰﾋﾞｽﾗｲﾌｻｲｸﾙのﾕｰｽｹｰｽ毎のAPI単位	・APIﾚﾍﾞﾙ(ﾚﾍﾞﾙ1-2) WebAPI:REST and/or SOAP 性能重視:REST AP開発容易性:SOAP (一部XMPP/RDF) 組込み用API: Python,Java, Ruby, PHP, C#, Apex, .NET ・API功ﾃｺﾞﾘ ｶﾃｺﾞﾘ1-2: Python,Java, Ruby, PHP, C#, Apex, .NET, SDK, VisualStudio ｶﾃｺﾞﾘ3:独自ｻｰﾋﾞｽAPI •API種別 独自定義種別	•APIﾚﾍﾞﾙ(ﾚﾍﾞﾙ1) WebAPI:HTTP(ﾓｼﾞｭｰﾙ連携用) 組込み用API:JavaAPI (DBMS/SQL通信：JDBC) •API功ﾃｺﾞﾘ ｶﾃｺﾞﾘ1:Java ｶﾃｺﾞﾘ3:独自ｻｰﾋﾞｽAPI •API種別（独自定義） WebAPI単位 入出力処理機能API 匿名化処理設定ﾌｧｲﾙXML(対象ﾃﾞｰﾀ, 匿名化処理定義） JavaAPI単位 ①ｺﾝﾄﾛｰﾙ機能API ②入出力処理機能API ③匿名化ﾓｼﾞｭｰﾙAPI(開発用IF) ※WebAPI化は②のみ	•APIﾚﾍﾞﾙ(ﾚﾍﾞﾙ1-2) WebAPI:REST 組込み用API: JavaAPI 課題：左記①、③IFのWebAPI化範囲(GUI/IF含む)の明確化 •API功ﾃｺﾞﾘ ｶﾃｺﾞﾘ1:Java ｶﾃｺﾞﾘ2:標準参照型/基盤改良型独自開発ﾊﾟｯｹｰｼﾞAPI ｶﾃｺﾞﾘ3:標準参照型/基盤改良型独自ｻｰﾋﾞｽAPI •API種別（標準参照型/基盤改良型独自定義）
		ﾃﾞｰﾀ形式	•下位ﾚﾍﾞﾙ:XML_上位ﾚﾍﾞﾙ:RDF-OWL(W3C/P3P) •ﾃﾞｰﾀ種別毎に項目規定		・XML, WSDL	•XML(PDXML)	•XML
	ｸﾗｳﾄﾞ-企業間	API	JavaEE (API:XML関連JAXP)等との統合		・SOAP/REST		•WebAPI: RESTorSOAP •組込み用API: JavaAPI
		ﾃﾞｰﾀ形式			・XML/HTML		•XML

OCM: Open Cloud Manifesto, DMTF: Distributed Management Task Force, UCI: Universal Cloud Interface, OWL: Web Ontology Language, RDF: Resource Description Framework, GAE: Google App Engine platform, IDE: Integrated Development Environment, REST: REpresentational State Transfer, SOAP: Simple Object Access Protocol

(1) 構成要件考慮事項

匿名化クラウドの位置づけ，すなわち構成要件考慮事項としては，標準・代表的 CSP のクラウド機能アーキテクチャから，PaaS 内アプリケーション開発環境に組み込まれ共有される匿名化基盤ライブラリソフトウエアの位置づけとなる．

(2) インタフェース要件考慮事項

インタフェース要件考慮事項としては，各インタフェース種別に対して，以下のとおりである．なお，考慮事項は，あるべき方向性として，現状の標準化状況から参照が有効な範囲は取り入れ，一方で実現性を踏まえた方向性として，現状 CSP 動向や個人情報匿名化基盤の流用性を考慮して規定している．

（a） **クラウド-AP 間インタフェース**

匿名化処理のように処理性能が要求されるものには REST が向いていることや，より広範な CSP で採用されているという普及の観点から，匿名化クラウドの WebAPI としては REST，組込み用 API としては，情報大航海プロジェクトの個人情報匿名化基盤の API の流用性と現状 CSP でも代表的な開発言語の一つとして活用されていることから JavaAPI，データ形式は，現状 CSP で代表的に採用されており，個人情報匿名化基盤の流用性も考慮して，XML 形式が適当であると考える．

（b） **クラウド-企業間インタフェース**

現状 CSP で代表的に採用されており，個人情報匿名化基盤の流用性も考慮して，WebAPI は REST，組込み API は JavaAPI，データ形式は XML 形式が適当と考える．

（c） **クラウド-企業間インタフェース**

本インタフェースは，複数の異種クラウド間を統合管理・利用する際のものであり，標準化動向も現状仕様規定までなく RDF 試行レベルであること，匿名化クラウドの技術開発上の優先度は低いことから，現時点でサポートすることは時期尚早であり，将来課題とすることが適当と考える．

4.2 ユースケース調査・検討結果概要

4.2.1 利用形態モデル

情報大航海プロジェクトの個人情報匿名化基盤の機能モジュール構成を図 2 に示す．個人情報匿名化基盤は，入出力や各機能モジュールの制御等を行う共通基盤実行機能をベースに，①単純匿名化機能モジュール，②匿名化保証機能モジュール，③統合匿名化機能モジュール，④データ統合機能モジュール，⑤匿名化保証管理機能モジュールの５つのプラグ＆プレイ機能モジュール群から構成され，５つの中からサービスや利用形態に応じて選択利用できることを特長とする．

利用形態は，匿名化対象データが一種類の場合か複数種類の場合か，匿名化処理を１次事業者，委託第３者機関，第３者評価機関でどのように分担するかで多様なパターンとなる．このようなサービスや利用形態に応じて匿名化処理関連の必要機能を適合できる特長は，マルチテナントで共有される環境となるクラウドにおいて，より有効活用されるものであり，匿名化クラウドにおいても継承すべき特徴と考える．

可能性のある利用形態モデルパターンは，大別して，一種類のデータを匿名化処理する単純データ匿名化の場合と，複数種類のデータを結合して匿名化処理する複合データ匿名化の場合があり，各々の場合に統合型，分離型（１），分離型（２），分離型（３），委託型の５つのパターンが存在する．単純データ匿名化の場合の例を，図 3 に示す．

また，匿名化処理の適用が有効とされるビジネスモデルを調査したところ，以下の３パターンのモデルが代表的であることが分かった．

(a) 利用者への適切な情報の提供（広告・宣伝・アドバイスなど）

ⓒ2011 Information Processing Society of Japan

Vol.2011-CSEC-54 No.27
Vol.2011-SPT-1 No.27
2011/7/13

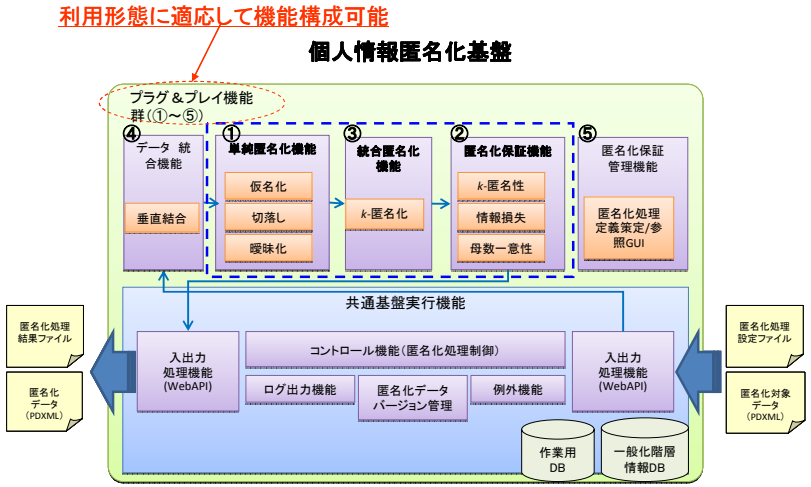

図 2　情報大航海プロジェクトの個人情報匿名化基盤の機能モジュール構成

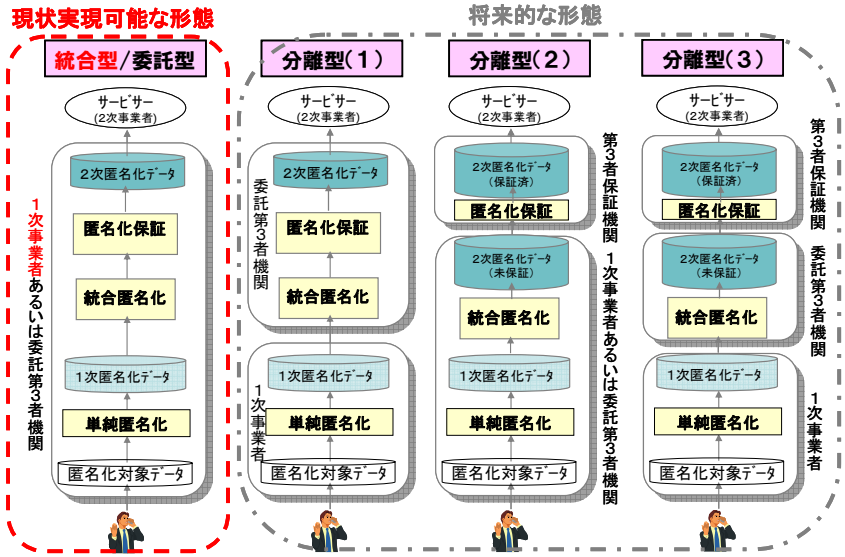

図 3　単純データ匿名化の場合の利用形態モデルパターン

(b)　２次利用サービス事業者による情報活用（商品開発・マーケッティングなど）
(c)　利用者による第３者情報の活用（情報検索・情報共有など）

いずれも匿名化処理の利用形態モデルの統合型で対応可能なものである．すなわち，匿名化クラウドにおいて，統合型の実現が現状ビジネスモデルへの適用性から優先度が高いと言える．匿名化処理の他の利用形態は，現状ないサービス形態であり，将来の新たなサービス形態の候補として位置づけられる．

4.2.2 匿名化処理動作フローモデル

匿名化クラウドの機能要件を導出するために，その基本的な動作フローモデルを定義する際，個人情報保護法等のプライバシー保護，匿名化処理に係わる法・業界制度を考慮することが制約条件となる．そこで，国内外の関連する法・業界制度を調査して，匿名化クラウドにおける制約条件を以下のとおり特定した．

(a) 匿名化クラウド自身で対策要の条件

個人情報（対応表含む）と匿名化情報の分離

- 個人情報と匿名化情報を物理的あるいは論理的に分離
- 匿名化情報と識別情報の対応表を分離（同じクラウド内に置かない）

(b) 匿名化クラウドを含むクラウド基盤で対策要の条件

匿名化情報の保管・管理

- 蓄積場所のコントロール（自国内，特定国内）
- 委託先，CSP 等の特権ユーザへの匿名化情報への不要な閲覧を許さないアクセス制御
- 匿名化情報の記録管理（匿名化処理のログ，バージョン管理機能は既に個人情報匿名化基盤に存在）

この内，上記（a）の条件が，匿名化クラウド自身，アーキテクチャ設計の際の対応検討事項となる．さらに，“単純匿名化情報（切り落し，ID 化等）は，プライバシーリスクが高いため，クラウド上には保存しないこと”を方針とした．そして，これら条件・方針を満たすために以下の対策をアーキテクチャ設計に反映することとした．

- 対策１：
  個人情報⇔単純匿名化情報（１次匿名化データ）⇔統合匿名化情報（２次匿名化データ）の各々相互間を，論理的あるいは物理的に分離し，個人情報及び単純匿名化情報はオンプレミス側で保存し，統合匿名化したものをクラウド上に保存．
- 対策２：
  相互対応表は，オンプレミス側の個人情報管理領域で保管．
- 対策３：
  匿名化情報の利用者の個人情報へのアクセス禁止（アクセス制御）．

ⓒ2011 Information Processing Society of Japan

Vol.2011-CSEC-54 No.27
Vol.2011-SPT-1 No.27
2011/7/13

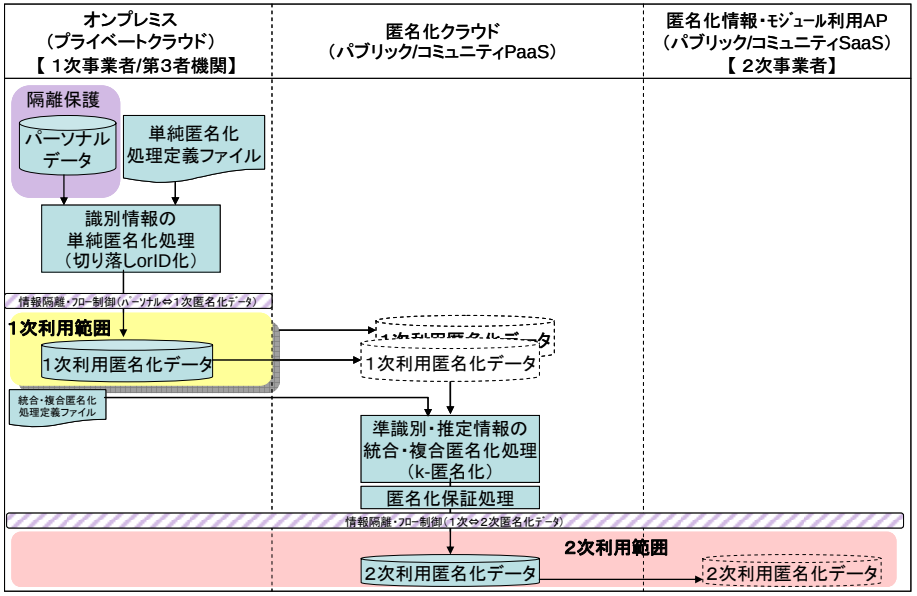

図 4　匿名化クラウドの運用時の動作フローモデル（統合型／委託型）

上記の対策および 4.2.1 で述べた利用形態モデルを考慮し，匿名化クラウドへの機能要件考慮事項導出のため，匿名化クラウド（匿名化モジュール）の基本的な動作フローモデルを定義した．動作フローモデルのパターンは，利用フェーズ（環境構築時，匿名化処理運用時），匿名化処理のタイプ（単純データ匿名化，複合データ匿名化），利用形態（統合型，分離型（１），分離型（２），分離型（３），委託型）の組合せとなる．運用時の統合型/委託型の動作フローモデルの例を，図 4 に示す．

4.2.3 アーキテクチャ機能的考慮事項

利用形態モデル定義結果，匿名化処理動作フローモデル定義結果から，匿名化クラウドアーキテクチャへの機能的考慮事項をまとめると，以下のとおりとなった．

(1) 情報大航海プロジェクトの個人情報匿名化基盤の機能群・構成の継承

利用形態モデル，動作フローモデルの匿名化処理単位・配置を実現するためには，個人情報匿名化基盤の機能モジュール単位・構成をクラウド化しても継承することが必要である．

(2) クラウド化に伴い必要となる新機能群

(a) データ匿名化/フロー制御機能

- 個人情報（対応表含む）⇔単純匿名化情報（１次匿名化データ）⇔統合匿名化情報（２次匿名化データ）間で DB 独立化（物理的，論理的分離）．
- 匿名化情報の利用者の個人情報へのアクセス禁止．

(b) 単純匿名化機能セット（識別情報切落とし，ID 化）のダウンロード機能

- オンプレミス側での個人情報⇔単純匿名化情報間データ隔離化のため，匿名化クラウドより単純匿名化機能セットをオンプレミス側にダウンロードして利用する機能．

(c) 匿名化クラウド実現機能

- クラウド基盤機能を利用して，個人情報匿名化基盤のマルチテナント化，スケーラビリティ化や１次事業者/第３者機関と２次事業者間の匿名化データや制御ファイルのセキュアな通信・認証機能を実現する機能．

4.3 アーキテクチャ要件定義概要

4.1 節の動向調査から導出された匿名化クラウドアーキテクチャへの構成・インタフェース考慮事項，4.2 節のユースケース調査・検討から導出された機能的考慮事項を基に統合・整理して，匿名化クラウドアーキテクチャの構成，機能，インタフェースの要件を以下のように定義した．

4.3.1 構成要件

(1) 匿名化クラウド（クラウド向け匿名化基盤ライブラリソフト）の位置づけ

多様な SaaS/AP 向けの共通ミドルウエアであること，適用領域が非ミッションクリティカルであり差別化 AP 実現用であることから，標準クラウドアーキテクチャの機能レイヤとしては PaaS の位置づけであり，さらに，代表的 PaaS/CSP の機能構成の中の PaaS 内 AP 開発環境に組み込まれる共有匿名化処理ソフトウエアの位置づけとなる．

(2) 匿名化処理機能のモジュール構成

多様なサービスや利用形態に適応可能とすべく，情報大航海の個人情報匿名化基盤の特長でもあるモジュール構成を継承する．

(3) PaaS 基盤上での構成

PaaS 基盤上に匿名化処理を実施する複数の１次事業者（オンプレミス），委託第３

ⓒ2011 Information Processing Society of Japan

Vol.2011-CSEC-54 No.27
Vol.2011-SPT-1 No.27
2011/7/13

者機関，第３者保証機関等の組織各々を一つのインスタンスとするマルチテナント匿名化基盤インスタンスを生成する．インスタンス間でデータ等干渉しないよう独立化するために, Java 実行環境/VM 単位に, 個人情報匿名化基盤の DI(Dependency Injection) コンテナ相当ごと割当てて匿名化基盤インスタンスを生成する．

4.3.2 機能要件

匿名化クラウドの機能として以下のような機能群を設ける．

(1) 情報大航海プロジェクトの個人情報匿名化基盤の継承機能群

(a) 共通実行基盤機能

個人情報匿名化基盤が，プラグ＆プレイ機能群の機能を用いて匿名化を実行するために必要となる共通の処理を行う機能で，コントロール機能，入出力処理機能，ログ出力機能，例外機能，匿名化データバージョン管理機能から構成される．

(b) プラグ＆プレイ機能群

個人情報匿名化基盤の利用形態に応じて必要な機能を選択して実行できる機能で，データ統合機能，単純匿名化機能，統合匿名化機能，匿名化保証機能，匿名化保証管理機能から構成される．

ただし，個人情報匿名化基盤はクラウド化しても基本的な部分は流用可能であるが，クラウドの特長活用のための改良が必要である．例えば，機能モジュールでは，単純匿名化機能，統合匿名化機能につき，クラウドの特長である分散並列処理やスケールアウト性を活用できるよう分散処理フレームワーク利用型の機能モジュールへの改良が必要である．

(2) クラウド化に伴い必要となる新機能群（標準 PaaS 基盤機能利用）

(a) 分散処理基盤/仮想化

- マルチテナント匿名化基盤インスタンスの生成・実行管理基盤機能
  匿名化基盤インスタンスを各利用 AP 単位に登録・生成する機能．
- マルチテナント匿名化基盤インスタンスの並列・スケーラビリティ処理機能
  匿名化基盤インスタンスの並列分散処理を実行管理する機能．

(b) 分散ストレージ

- 匿名化処理向けデータストア（分散 KVS）
  匿名化基盤インスタンス生成・実行管理ミドルウエアが分散データベースを利用し，並列処理するデータベース機能．
- データ隔離化/フロー制御機能
  パーソナル情報⇔単純匿名化情報（１次匿名化データ）⇔統合匿名化情報（２次匿名化データ）の各々相互間を，論理的あるいは物理的に分離し，パーソナル情報及び単純匿名化情報はオンプレミス側で保存し，統合匿名化したものをクラウド上に保存する，また匿名化情報の利用者の個人情報へのアクセスを禁止（アクセス制御）する機能．

(c) Web フロント/ネットワーク

- 匿名化データ/制御ファイルのセキュア通信/認証機能
  匿名化データや制御ファイルのセキュアな通信・認証を実現する機能．
- 単純匿名化機能セット（識別情報全削除，ID 化）のダウンロード機能

4.3.3 インタフェース要件

(a) 匿名化クラウドーAP 間インタフェース
データ形式：XML，API：WebAPI/REST，GUI：WebGUI

(b) 匿名化クラウドー企業（オンプレミス）間インタフェース
データ形式：XML，API：WebAPI/REST，GUI：WebGUI

(c) 匿名化クラウド，オンプレミス AP 組込み用インタフェース
データ形式：XML，API：JavaAPI

4.4 アーキテクチャ開発結果概要

前章の匿名化クラウドアーキテクチャの要件定義に基づき，これを実現する匿名化クラウドアーキテクチャ（機能構成）を，図 5 に示す．匿名化クラウドは，PaaS 基盤上の匿名化基盤ライブラリソフト（共有匿名化処理ミドルウエア）として実現され，個人情報匿名化基盤相当の各利用 AP 単位に生成される匿名化基盤インスタンスと，匿名化基盤インスタンス群の PaaS 上での生成・分散並列実行・セキュア通信/認証等マルチテナント/スケーラビリティ性を管理する匿名化基盤インスタンス生成・実行管理ミドルウエアから構成されるものとなる．

匿名化基盤インスタンス生成・実行管理ミドルウエアは，大規模並列分散処理基盤，分散データストア，統合開発環境，共通ミドルウエアサービスライブラリの PaaS 標準機能を基盤機能として利用して，匿名化基盤インスタンス並列分散実行管理機能を核とする，匿名化基盤インスタンス登録・生成機能，匿名化処理関連分散データストア（分散 KVS 利用），匿名化データセキュア通信機能から成る．

また，PaaS 基盤とのインタフェースでは，複数の代表的 CSP/PaaS 上で動作できることが実用性・普及の観点から有効なことから，各 CSP/PaaS 基盤が提供する API と匿名化クラウド内共通 JavaAPI との間の変換処理をする API マッパーを設ける．

ⓒ2011 Information Processing Society of Japan

Vol.2011-CSEC-54 No.27
Vol.2011-SPT-1 No.27
2011/7/13

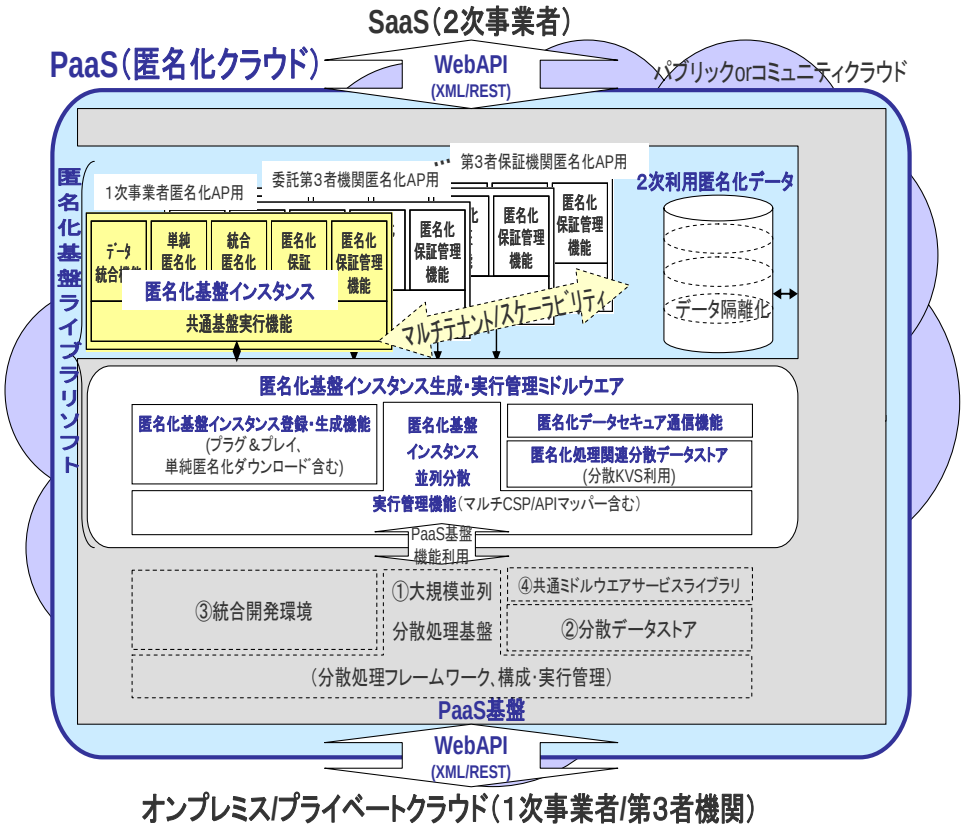

図 5　匿名化クラウドの機能アーキテクチャ

5. おわりに

本稿では，サービスに応じて適した匿名化処理を選択可能なクラウド上の共有匿名化処理ミドルウエア「匿名化クラウド」の機能アーキテクチャを提案した．今後，本アーキテクチャに基づき，匿名化クラウドの試作開発・検証を行う予定である．その際には，現在急速に拡充されてきている PaaS 等の分散処理基盤の最新動向を考慮して設計することが必要と考える．

謝辞　本研究は，経済産業省 平成 22 年度産業技術研究開発委託費による「次世代高信頼・省エネ型 IT 基盤技術開発事業（行動情報活用型クラウドサービス振興のためのデータ匿名化プラットフォーム技術開発事業）」の一環として行われた．この場を借りて，関係各位に感謝の意を表する．

参考文献

1) 経済産業省：情報大航海プロジェクト　個人情報匿名化基盤, http://www.meti.go.jp/policy/it_policy/daikoukai/igvp/cp2_jp/common/024/010/post-9.html（参照 2011-5-26）.
2) Aggarwal,C.and Yu,P.:Privacy-Preserving Data Mining: Models and Algorithms, Springer-Verlag（2008）
3) 経済産業省：平成２２年度次世代高信頼・省エネ型 IT 基盤技術開発事業報告書「行動情報活用型クラウドサービス振興のためのデータ匿名化プラットフォーム技術開発事業」, http://www.meti.go.jp/policy/mono_info_service/joho/cloud/2010/index.html（参照 2011-5-26）.
4) Open Cloud Manifesto: “Cloud Computing Use Cases White Paper Version 4.0”,（2010.7.2）http://opencloudmanifesto.org/Cloud_Computing_Use_Cases_Whitepaper-4_0.pdf （参照 2011-3-2）
5) CCIF/UCI（Cloud Computing Interoperability Forum）:“UCI Requirments”,（http://code.google.com/p/unifiedcloud/wiki/UCI_Requirements）（参照 2011-3-2）
6) DMTF（Distributed Management Task Force）:“Architecture for Managing Clouds Version1.0”,（2010.6.18） http://www.dmtf.org/sites/default/files/standards/documents/DSP-IS0101_1.0.0.pdf）（参照 2011-3-2）.
7) 経済産業省：「クラウドコンピュ-ティングと日本の競争力に関する研究会」報告書,（2010.3.26） http://www.meti.go.jp/press/20100816001/20100816001.html （参照 2011-3-2）.
8) 中田敦，他（共著）:「クラウド大全」，日経 BP 社，（2010.4.27）.
9) 清野 克行：「クラウド・アーキテクチャの設計と解析―分散システムの基礎から大規模データストアまで」，秀和システム，（2010.8.24）.
10) IPA（独立行政法人情報処理推進機構）:「クラウド・コンピューティング社会の基盤に関する研究会」報告書，（2010.3.24）.

ⓒ2011 Information Processing Society of Japan