1C-04

データ二次利用を促進する為の利用権限を継承するデータ流通基盤技術

加藤　孝浩*　　坂本　久†　　石田　和生‡　　稲垣　嘉信§
NECソリューションイノベータ株式会社||

1. データ流通の現状と課題

現在は、様々なＩｏＴサービスや分析サービスが世の中のデータを大量に収集・分析し、新たなデータを生み出している。しかし、それらは他の目的で活用されることが極めて少ない。企業が保有するデータの５０％は、蓄積されたまま眠っているという統計もある[1]。

我が国でも、この様なデータを流通し、利活用する機運が高まっているが、実現には、データ流通や提供条件等の透明化・明確化が必要である等、いくつか課題がある[2]。

本稿では、これら課題を解決し、データを二次／三次利用（以下、ｎ次利用）させ、新たな価値を創造するため、データ毎に保護し、適切な利用者にのみ利用させるための権限を持たせて、それらをｎ次利用データへ継承するアクセス権管理技術と、技術の利用を容易にするための基盤アーキテクチャを提案する。

2. データ流通基盤に必要な機能

データの保護を目的としたアクセス権管理と、データの利用制限を目的としたｎ次利用管理する基盤（以下、データ流通基盤）に必要である主な機能は、次の三つである。

１）データ自体の保護とアクセス権管理
２）データ利用履歴の管理
３）データ利用権限の伝播

2-1. データ自体の保護とアクセス権管理

流通させるデータそのものを暗号化して保護する。

暗号化に用いる鍵は、データにアクセスする事を許可されない第三者に決して漏洩させない様、データ流通元が厳重かつ安全に管理する。

しかし、厳重すぎると、適切なデータ利用者の利便性が失われる。本機能は、データの鍵を秘匿しながら、データは公開するという、矛盾した要件を満たす必要がある。

そのため本機能では、データの保護にＤＲＭ（デジタル権限管理）を応用した方式を用いる。

2-2. データ利用履歴の管理

データの参照や加工データの生成を記録する。

データ利用履歴管理の目的は、加工したデータの出典を明らかにすることで、データの信憑性を示したり、ｎ次利用したデータが新たに創造した価値を、データ流通側に還元したりすることにある。

この利用履歴管理には、分散型台帳管理を用いる。複数の当事者によって管理することで、非中央集権で、可用性・耐改ざん性を高める。

2-3. データ利用権限の伝播

参照元データから、加工されたｎ次利用データに、利用権限を伝播させる。

例えば、参照したデータに、利用者Ａへの利用を制限する権限が設定されている場合、ｎ次利用した加工データは、利用者Ａへの利用制限を引き継ぐ。さらに、加工データに、利用権限を追加して設定することもできる。

つまり、ｎ次利用の加工データには、ｎ－１次以前の利用権限とｎ次利用データ自身の利用権限が設定されており、ｎ次利用データ自身の利用権限は、ｎ＋１次利用以降へ引き継がれる。（図１参照）

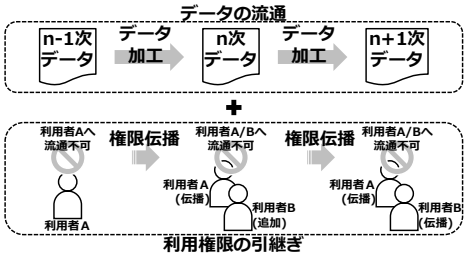

図１ データ利用権限の伝播

A data distribution platform that inherits usage rights to promote secondary use of data.

* Takahiro Katou
† Hisashi Sakamoto
‡ Kazuo Ishida
§ Yoshinobu Inagaki
|| NEC Solution Innovators, Ltd.

3. データ流通基盤の構成

データ流通基盤の構成は次の通りである。（図２参照）

ⅰ）認証・認可部
ⅱ）セキュアコンテナ部
ⅲ）利用履歴管理部
ⅳ）データ公開管理部
ⅴ）データストレージ部

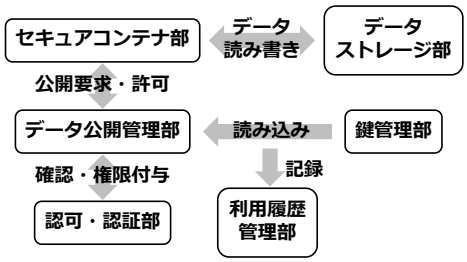

図２ データ流通基盤の構成

ⅰ）ⅱ）は、「１）データ自体の保護とアクセス権管理」にかかわる技術、ⅲ）は、「２）データ利用履歴の管理」にかかわる技術、ⅳ）ⅴ）は、「３）データ利用権限の伝播」にかかわる技術である。

3-1. 認証・認可部

認証は、誰であるかを確認し、認可は、権限を与える。例えば、流通側が「利用者は誰であるか」を確認し、データ参照する権限を与える、という手順である。

これにはOpenID Connect[3], OAuth2.0[4]という標準的な技術を用いる。利用側で認証し、発行するトークンと呼ばれる一意の識別子を、流通側が受け取り認可することで、流通されているデータを参照するための鍵を得ることができる。

3-2. セキュアコンテナ部

前述の鍵は、保護されているデータを復号するために、利用側に渡す必要がある。

これには、セキュアなコンテナ技術を用いる。一般的なコンテナ技術との違いは、コンテナ内のプロセスとコンテナ実行環境は、互いに不可侵なことにある。

つまり利用側は、自身のリソースで動作しているコンテナ環境でありながら、その中を知ることはできず、データの鍵を秘匿しながら、データは公開するという、要件を満たすことができる。

3-3. 利用履歴管理部

利用履歴は、いつ、だれが、どのデータを参照したのかを記録し管理する。例えば、データの参照に必要な鍵を「流通側が利用側に渡す／利用側で受け取る」というタイミングで記録する。記録には分散台帳管理の技術を用い、可用性・耐改ざん性の高い管理とする。

3-4. データ公開管理部

前述の認証・認可部およびセキュアコンテナ部と連携し、流通されているデータの公開を管理する。セキュアコンテナ部からの要求を受け付けるインタフェースを持ち、認証・認可のフローを制御する。

3-5. データストレージ部

流通するデータを格納するストレージは、既存のストレージサービスを用いることができ、特定のストレージサービスに依存しない。流通するデータ自体は、暗号化で保護しているため、流通させることに特化した保護は、必要ではない。

4. データ流通基盤の活用

今後は、同アーキテクチャを実装したデータ流通基盤を用いて、データの流通が連鎖的に付加価値を創造するモデルが、成立することを検証する。

検証で扱う実際のデータとして、環境ＤＮＡ[5]分析と呼ばれる技術で取得された生物情報を、対象とすることを検討している。

複数の研究機関の間で生成・共有される一次情報を基本に、その情報を、他の組織が二次利用し、情報の見える化サービス等を通じて、生態系の持続や環境保全といった価値を創造することを目指す。

本稿で提案したデータ流通基盤を活用することで、環境ＤＮＡ情報がサービスの枠を超え、安全に利活用されると考える。

5. 参考文献

[1] DATA’S DARK SIDE, https://www.veritas.com/dark-data（参照 2020/1/8）
[2] 平成 29 年版 情報通信白書, https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/pdf/n2200000.pdf（参照 2020/1/8）
[3] OpenID Foundation, https://openid.net/connect/（参照 2020/1/8）
[4] The OAuth 2.0 Authorization Framework, https://tools.ietf.org/html/rfc6749, https://tools.ietf.org/html/rfc6750 （参照 2020/1/8）
[5] 環境ＤＮＡとは, http://ednasociety.org/edna（参照 2020/1/8）