

業種業界を越えたデータ流通の信頼性を向上させるChain Data Lineage

Improving Reliability of Data Distribution Across Categories of Business and Industries with Chain Data Lineage

松原 正純 宮前 剛 伊藤 章 鎌倉 健

あらまし

昨今、企業の内外には様々なデータがあふれている。最近では、自社が持ちえない価値あるデータを企業間で流通して活用することによって、これまでになかった革新的なデジタルビジネスをCo-creation（共創）しようとする動きが広がってきている。しかし、このような世界を実現するためには、企業間でやり取りするデータが信頼できることが大前提となる。具体的には、データがどこからやってきて、どのように加工されてきたのか、個人データを含む場合には、本人が当該データ提供について同意済みであるか、といったことが確認できなければならない。富士通研究所では、このような課題を解決し、データを安心して扱えるようにするためのデータ流通・利活用技術に関する研究開発を推進している。

本稿では、まず、データ流通におけるデータの信頼性の課題について述べる。更に、富士通研究所が開発した、業種業界を越えてやり取りされるデータの信頼性を向上させる技術Chain Data Lineageを、活用事例とともに紹介する。

Abstract

Today, the inside and outside of companies are overflowing with various data. Recently, there is a growing movement to push the co-creation of innovative digital businesses by distributing and utilizing valuable data that cannot be obtained independently between different companies. The realization of a world like this, however, requires reliability of the data shared between companies. Specifically, it must be possible to verify where the data have come from and how they have been processed and, when personal data are included, whether or not individuals have given consent regarding the provision of the relevant data. Fujitsu Laboratories is moving ahead with research and development relating to data distribution and utilization technology that resolves these issues and enables the safe use of data. This paper first presents the issues relating to data reliability in data distribution. Then, it describes the Fujitsu-developed Chain Data Lineage technology, which improves the reliability of data exchanged across categories of business and industries, together with examples of its application.

1. まえがき

昨今、企業の内外には様々なデータがあふれている。例えば、顧客マスター、小売店のPOS (Point Of Sales) システムデータ、EC (Electronic Commerce) サイトでの商品購入履歴などは、従前から重要視されてきた。近年ではこれらに加えて、製造工場ラインの監視センサーデータなどのIoTデータ、気象情報や人口動態調査などのオープンデータ、個人の行動を表す自動車走行データといった個人データなど、ビジネスに利用可能なデータの種類や量がますます増えている。⁽¹⁾

企業のデジタル革新は、このようなデータを有効活用することによって実現される。つまり、データを組み合わせ、AI (人工知能) やデータ解析を駆使して新しい価値を見出すことによって、これまでになかった新しいビジネスを創出したり、従来のビジネスを抜本的に変えたりすることが可能となる。

更に近年、このデータ活用の動きは企業間にまで広がろうとしている。これは、自社が持ちえない価値あるデータを、企業間で流通させて活用することによって、これまでになかった革新的なデジタルビジネスをCo-creation (共創) しようというものである。

富士通研究所では、このようなデータドリブンなDigital Co-creationの世界の実現を目指している。⁽²⁾しかし、この世界を成り立たせるためには、扱うデータが信頼できることが大前提となる。そこで富士通研究所では、データを安心して扱えるようにするための、データ流通・利活用技術に関する研究開発を推進している。

本稿では、まずデータ流通におけるデータの信頼性の課題について述べる。更に、富士通研究所が開発した、業種業界を越えてやり取りされるデータの信頼性を向上させる技術Chain Data Lineageを、活用事例とともに紹介する。

2. 信頼性の高いデータ流通の実現に向けた課題

企業間のデータ流通においては、人の活動やビジ

ネスなどによって生み出されたデータを所有している企業と、そのデータを利用したい企業との間で交わされた契約に基づいて、対象データがやり取りされる。データを受け取った企業は、そのデータを使って、あるいは他データと組み合わせて、AIモデルを作成したり解析したりすることで自社のビジネスに活用する。更に、こうして新たに生まれたデータを第三者の企業に流通させることで、新たなビジネスが広がっていく。

しかし、その際に信頼のおけない低品質のデータを使用してしまうと、提供するサービスの質が低下するという問題が生じる。また近年、EU一般データ保護規則 (GDPR) をはじめとする個人情報保護する法律が施行されている。これに伴い、個人データを第三者の企業へ提供するためには、本人から提供の同意を得ることが必須となっている。したがって、個人データがどのように使われるのかを、その個人にとって分かりやすくしつつ、活発に活用する上での工夫も必要になっている。

今後、多くの企業や組織が個人情報を含んだデータの生成・加工に関わることで、多様かつ有用なデータが増えていく。一方、データが不特定多数の企業の手を渡って加工されることによって生じる信頼性の低下が懸念されており、データの成り立ちを把握するニーズが高まっている。

企業間で流通されるデータの大半は、他データを二次利用し、情報の追加・削除・変換などの加工・分析処理が加えられたものである。したがって、データ流通の信頼性を向上させるためには、どのようなデータを使ってどのように加工・分析したかという、来歴情報の管理が重要となる。

しかし従来は、データの加工・分析に関わる履歴情報は各企業に閉じて管理されており、データ流通先の企業にその情報が伝わることはなかった。また、各社が保有する個人データの利用目的、提供データ、提供先などに変更が生じる場合には、データ提供者から改めて同意を得る必要がある。更に、データ利用者は同意に基づいて個別に個人データを取得する必要がある。その結果、本人同意の再確認や個別の個人データ取得に多くの工数を要することになる。

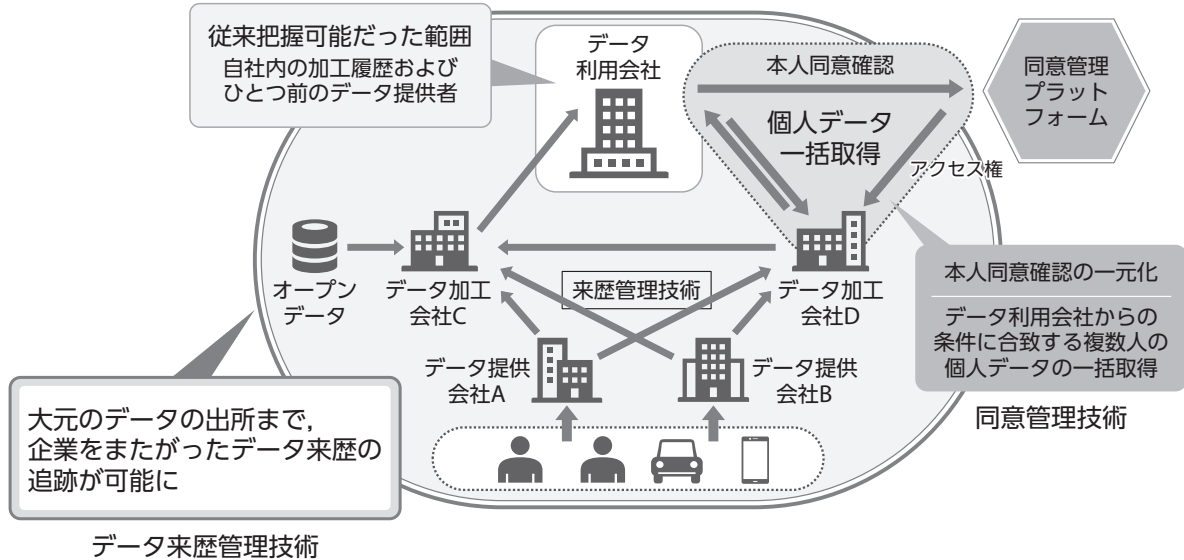


図-1 Chain Data Lineageの特徴

3. 開発技術

富士通研究所では、前章で述べたデータの信頼性に関する課題を解決するために、業種業界を越えてやり取りされるデータの信頼性を向上させるChain Data Lineage⁽³⁾を開発した。本技術は、データ来歴管理技術と同意管理技術から構成される(図-1)。本章では、それぞれの技術について詳述する。

なお、これらの技術は、ブロックチェーン技術を応用して開発された富士通研究所のVPX (Virtual Private digital eXchange) テクノロジー^{(4),(5)}をベースとする「FUJITSU Intelligent Data Service Virtuora DX データ流通・利活用サービス」⁽⁶⁾などに実装し、製品化する予定である。

3.1 来歴情報管理技術

データ流通システムにおけるデータの来歴情報とは、データの加工履歴や取引履歴を時系列に接続したグラフデータのことである。Chain Data Lineageでは、まず各社のデータ加工履歴とデータ取引システムの取引履歴のフォーマットを統一することによって、データの来歴情報全体を統合管理できるようにした(図-2)。また、流通・加工データの来歴情報の耐改ざん性を高めるために、取引履歴をブロックチェーン上で管理することとした。

更に、加工履歴を含めた全ての来歴情報に対して、ブロックチェーン技術の一部でもあるハッシュチェーン技術^(注1)を適用した。つまり、個々の来歴情報のハッシュ値を次の来歴情報の中に埋め込むことによって、過去に至る全ての来歴情報の改ざんを容易に検出できるようにして、来歴情報の信頼性を向上させた。同時に、実データのハッシュ値を来歴情報の中に埋め込むことによって、実データの改ざん、差し替えなども容易に検出できる、流通・加工データの原本性保証^(注2)も実現し、来歴管理技術に統合した。

これらの技術によって、流通データの原本性や来歴情報の正当性を保証するために必要な信頼点、第三者による証明などのTrusted Third Party (TTP)が不要になり、システムコストの大幅な引き下げが可能となった。また、信頼性が保証された来歴情報を実データに添付することによって、最終利用者の端末で来歴情報の正当性を検証できるようになった。

Chain Data Lineageでは、高いセキュリティが求められる取引履歴をブロックチェーン上で管理する。その一方で、データ量の多い加工履歴を各企業

(注1) あるデータに対して、一方向性などを満たす特殊な数学的関数(ハッシュ関数)を繰り返し適用することによって、改ざんを困難にする技術。

(注2) データの故意または過失による虚偽入力、書き換え、消去、および混同を防止すること、あるいはそれらを検出できるようにすること。

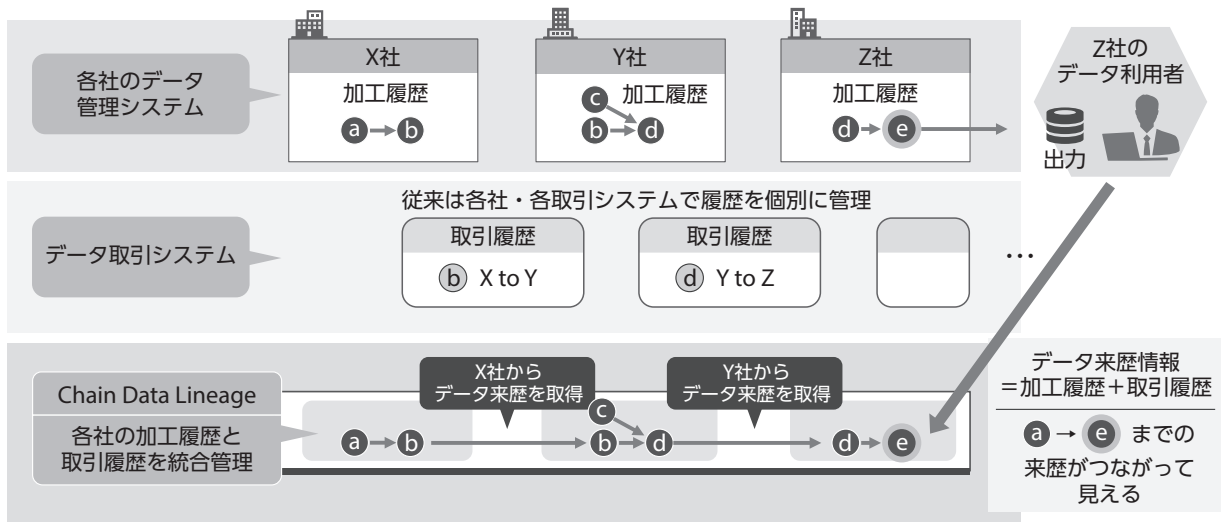


図-2 来歴管理技術

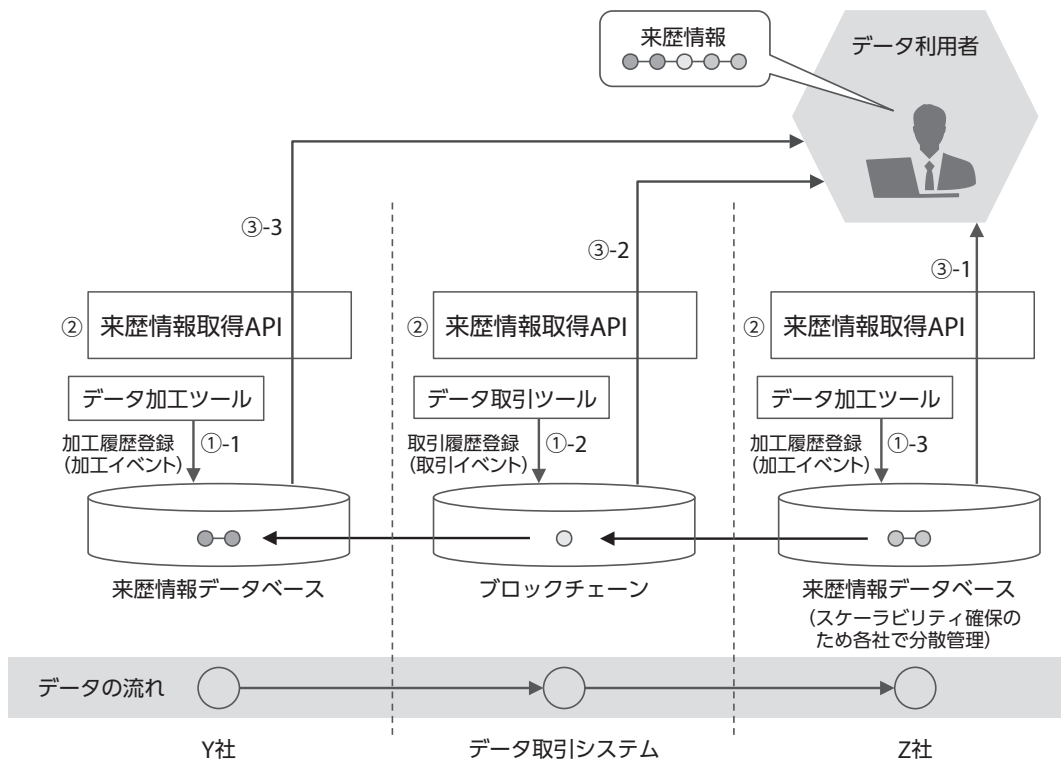


図-3 データ来歴情報の企業間での共有

の責任においてブロックチェーンの外側にあるデータベースで管理させることによって、セキュリティとスケーラビリティのバランスを取っている。データ来歴情報を企業間で共有する際のシステム構成図を図-3に示す。

データの来歴情報を記録するためには、データ取引システムやデータ加工ツールがそれぞれ取引・加工イベントを抽出し、自身が運用する来歴情報データベースに登録する (図-3 ①)。来歴情報データベースは、スケーラビリティを確保するために各

社で分散管理される。全ての来歴情報データベースは来歴情報取得API (Application Programming Interface) を用意し、自身が管理している来歴情報をデータ取引に関与している参加企業に対して公開しておく (図-3 ②)。来歴情報を取得したい参加者は、自身に最も近い直前の取引履歴から、オリジナルデータの加工履歴の方向へ、各企業の来歴情報取得APIを順に呼び出していき、最終的にオリジナルの来歴情報にたどり着く (図-3 ③)。

富士通は、データ流通のためのブロックチェーン技術を応用したVPXシステムと、企業内でデータ加工履歴を管理するOSS (Open Source Software) のメタデータリポジトリの一つであるApache AtlasをChain Data Lineageに対応させ、接続実験を実施した。

ハッシュチェーンに基づいて来歴管理を実現する、Chain Data Lineageの普及における一つの大きな問題として、各社のデータ取引・加工ツールが来歴情報のハッシュチェーン形式に対応することが困難であることが挙げられる。そこで今後は、データ流通推進協議会 (DTA) などを通じて、来歴情報のフォーマットや来歴情報登録・取得APIを標準化していく。これによって、相互運用性を高め、各社がハッシュチェーン形式の来歴情報を導入しやすい環境を整える。また、内閣府の戦略的イノベーション創造プログラム (SIP) での実証実験などを通じて、データ流通の信頼性担保のために、ハッシュチェーン技術による来歴情報利用の重要性を周知する。更に政府に対しては、来歴情報の保管を義

務付けるための法整備を働きかけていく。技術的な観点では、ハッシュチェーンを生成するためのライブラリを整備することによって、データ取引・加工ツール側の開発の負担を低減できる仕組みを検討している。

3.2 同意管理技術

企業間で個人データを流通させる際の関連標準には、権限認可を行うOAuth⁽⁷⁾ や、認可に基づくアクセス管理プロトコルUMA (User-Managed Access)⁽⁸⁾ がある。しかし、これらはいずれも個人が所有するデータを前提に設計されている。そのため、例えば医療機関が既に所有している同意済みの診療情報を一括して別の医療機関に送信するなどの用途には、新たな設計が必要である。

図-4に、OAuthとその拡張であるUMAの典型的な構成を示す。UMAプロトコルは、データのオーナー単位でデータを認可して第三者に提供するモデルである。このモデルでは、それぞれのサーバやクライアントの間で発生する複数の通信を利用者の人数分繰り返すと、メッセージ数が膨大になってしまう。

この問題を解決するために、複数人分のデータを一括取得できるようにUMAプロトコルを拡張した。図-5に、主な処理シーケンスを示す。データフォーマットや属性指定によって複数のデータ群を仮想的に一人分のデータとして一括して扱い、このデータ群をトークンに割り付けることで、一括取得できるようにした。

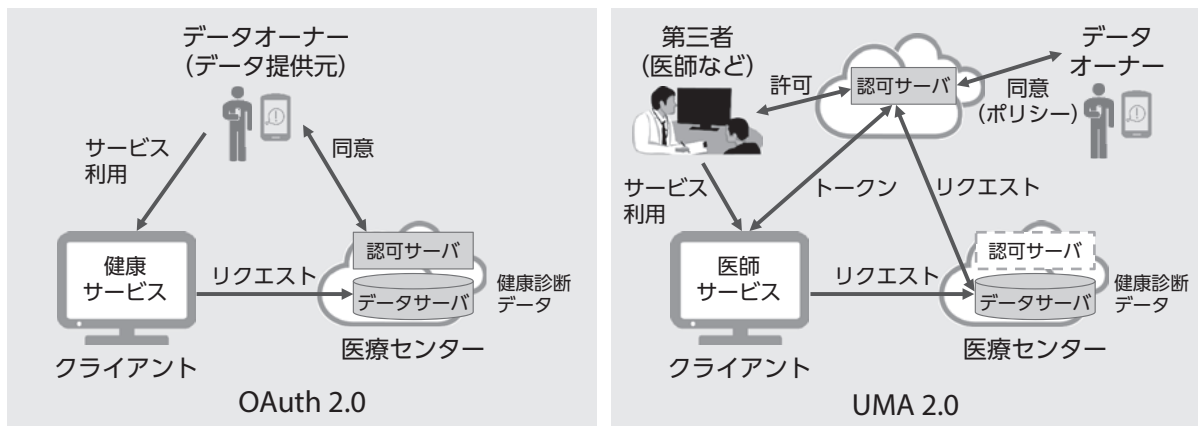


図-4 OAuth, UMAの構成例

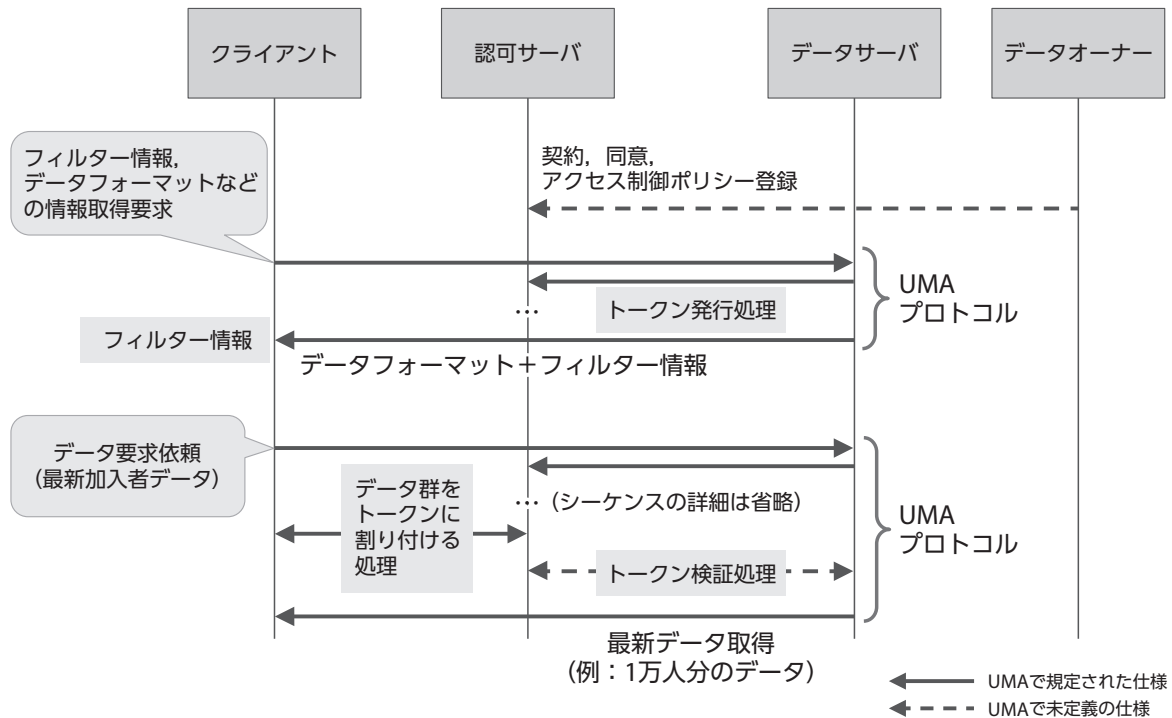


図-5 UMA 2.0を拡張した処理シーケンス

これによって、UMAプロトコルが本来備えるアクセス制御の安全性を活かしながら、取得する人数に依らず2回分のシーケンスで複数人分のデータを一括取得できるようになった。UMA 2.0では、認可サーバで同意の確認を行っているため、組織間でのデータの一括取得においても同意されたデータのみが転送される。本人同意の履歴はハッシュ値としてブロックチェーンに格納されるため、改ざんは困難である。

ただし、この拡張によって、標的型メールなどで不正なリンクを送りデータを取得する攻撃に対する脆弱性が生じる。この対策として、トークンの検証機能も追加した。

4. 活用事例

本章では、Chain Data Lineageに関する二つの活用事例について述べる。

4.1 テレマティクス自動車保険

保険業界では現在、リスクに応じたきめ細かい保険料算定が施される傾向にある。一方、ICTやセン

サー技術の進展によって、詳細なドライバーの運転履歴を収集できるようになってきている。テレマティクス自動車保険は、自動車の運転履歴を収集分析することにより、ドライバー個人に応じた保険料を算出し、リスクを最適化した保険を提供する。これによって、保険会社とドライバーの双方にメリットが生まれる。今後は、更に最適化するために、レンタカーやカーシェアリングなど、自分の車以外の車を運転したときのデータを集めたり、運転時の天候などの条件を組み合わせたりすることによって、より高度な分析が可能になると考えられる。

このように保険業界では、今まで使われていなかったデータを用いたり、高度な分析のために新規のデータ分析事業者の加工データを用いたりする場合がある。Chain Data Lineageでは、データの来歴情報をデータ使用者が確認できるため、データが信頼できるかどうかを事前に判断できる。

運転履歴に基づいた自動車保険の仕組みについて、図-6に示す。この仕組みでは、ドライバーモデルデータを算出するまでの情報が、来歴情報として保険会社のデータ利用者に示される。どの自動車メーカーから提供された運転履歴データか、それが

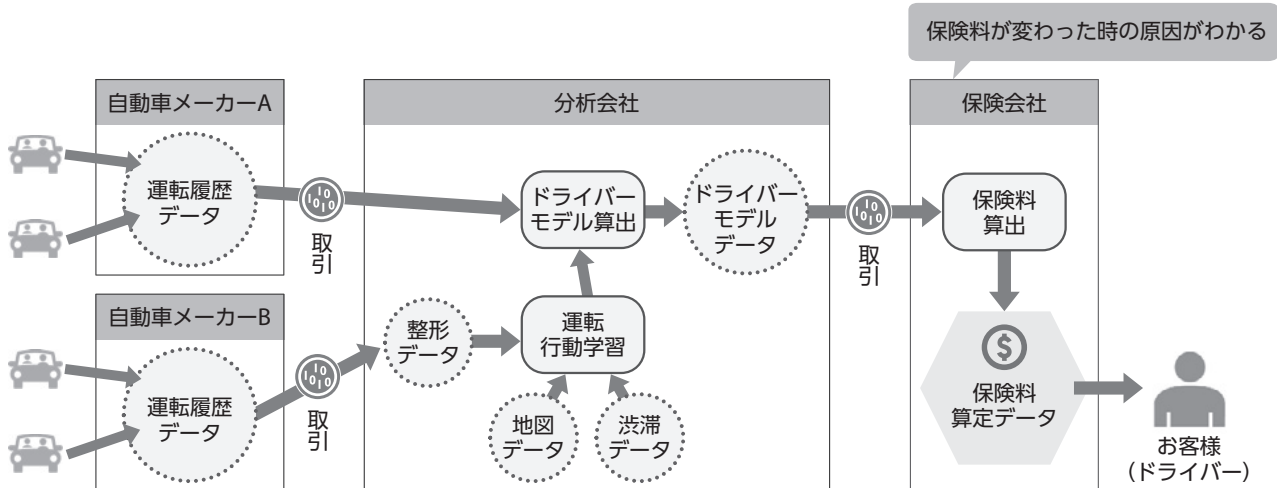


図-6 運転履歴に基づく自動車保険の仕組み

どの分析会社で分析されたのか、またドライバーモデルデータがどのように算出されたかといった情報が、この来歴情報に含まれる。従来は、保険会社から見て一つ手前であるドライバーモデルデータの情報しか分らなかった。しかし、この来歴情報を確認することによって、データの素性が分かり、信頼性が大幅に向上する。

4.2 製薬メーカーでの利用

製薬においては、様々な診療事例を収集・分析する必要がある。前節のテレマティクス自動車保険の事例と同様に、様々な関係者がデータの受け渡しをする過程にChain Data Lineageが適用できる。病院などで生成された診療データが大学やデータ分析会社などで分析され、その結果が製薬メーカーにおいて利用される。製薬メーカーは、データが生成された時点からの来歴情報をChain Data Lineageによって確認でき、信頼性を確認した上でデータを製薬に利用できる。また、診療データは患者のプライバシーに関する情報であるため、提供に当たっては患者の同意を得る必要がある。これにもChain Data Lineageの同意管理技術が適用できる。

5. むすび

本稿では、多数の組織をまたがって生成・加工されるデータの信頼性を確保するための来歴情報

管理、本人同意管理を実現する技術Chain Data Lineageについて述べた。これらの技術によって、人々は大量のデータを安心して使用でき、トラストなデータ社会の構築につながる。

今後は、データ利活用・分析を伴う具体的なサービスに沿ったChain Data Lineageの活用事例を増やしていく。また、本技術を本格的に浸透させるために、データ利活用の標準化も推進する。

これによって、より多くのデータに本技術が適用され、多数の企業がそれぞれの専門分野を活かして分析・加工を施し、それらを、安心して人々が利用できる環境作りを目指していく。

本稿に掲載されている会社名・製品名は、各社所有の商標もしくは登録商標を含みます。

参考文献

- (1) 総務省：平成24年版 情報通信白書。
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h24/>
- (2) 松原正純ほか：価値創出のサイクルを回すデータマネジメント. FUJITSU, Vol.69, No.5, p.30-36(2018).
<https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol69-5/paper05.pdf>
- (3) 富士通：業種業界を超えたデータ流通の信頼性を向上する技術を開発。
<https://pr.fujitsu.com/jp/news/2018/09/20-1.html>

- (4) 富士通：ブロックチェーンの応用による安心・安全なデータ流通ネットワークを実現するソフトウェアを開発。

<https://pr.fujitsu.com/jp/news/2017/06/5.html>

- (5) 江尻祐介ほか：業種・業界の枠を超えたデータ流通・利活用サービス。FUJITSU, Vol.69, No.6, p.9-15 (2018)。

<https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol69-6/paper02.pdf>

- (6) 富士通：Virtuora DXデータ流通・利活用サービス。

<https://www.fujitsu.com/jp/products/network/carrier-router/dataexchange/virtuora-dx/saas/>

- (7) Internet Engineering Task Force (IETF)：RFC 6750 The OAuth 2.0 Authorization Framework: Bearer Token Usage.

<https://tools.ietf.org/html/rfc6750>

- (8) Kantara Initiative:User-Managed Access (UMA) 2.0 Grant for OAuth 2.0 Authorization.

<https://docs.kantarainitiative.org/uma/wg/rec-oauth-uma-grant-2.0.html>



伊藤 章 (いとう あきら)

(株) 富士通研究所
スーパーミドルウェア・ユニット
データ流通・利活用の研究開発に従事。



鎌倉 健 (かまくら けん)

(株) 富士通研究所
セキュリティ研究所
データセキュリティやブロックチェーン
応用の研究開発に従事。

著者紹介



松原 正純 (まつばら まさずみ)

(株) 富士通研究所
スーパーミドルウェア・ユニット
データ利活用基盤技術の研究開発に
従事。



宮前 剛 (みやまえ たけし)

(株) 富士通研究所
セキュリティ研究所
ブロックチェーン、データ利活用基盤
技術の研究開発に従事。