O-029

Change Credibility Evaluation-based Approach to Personal Data Update Support

Phan Thi Thanh Huyen‡ Shinji Itoh‡ Naoto Sato‡ Shuhei Nojiri‡ Takuro Mori‡ Hideyuki Kanuka‡

Abstract

This paper aims to improve the accuracy and up-to-dateness of user personal data in the service providers (SPs) which must ensure their clients are genuinely who they claim to be, as a user may provide an SP with misinformation or not update it on a personal data change. Our approach is to connect with many trusted and essential SPs to early detect changes on as many as possible user personal data attributes. The credibility of a change is then evaluated using many novel metrics such as the trust of the SP where the change occurs and the importance of the SP to users. Finally, the impacted SPs which have the right to know the latest values of the changed personal data attribute will be notified of the change. This approach helps removing the burden of a user on updating his personal data on many SPs and preventing some economic lost or illegal business activities due to the usage of inaccurate or outdated personal data attribute values.

1. Introduction

Number of services, especially online services, are growing quickly to cover more and more aspects of life. To manage who is using their services, many service providers (SPs) ask their clients to provide some personal data, for example name, address, and phone number, which are stored as the attributes of the corresponding accounts, or digital identities. A digital identity is the digital information representing a physical person in an ICT system [1]. As an individual may register for many services of independent SPs which do not share user identity, his personal data are duplicated and managed separately by different SPs.

Some SPs, for example financial services, governmental services, or healthcare services, must ensure their clients are genuinely who they claim to be. Although they have validated personal data of clients when clients register accounts and periodically over time using Know Your Customer (KYC) check, many personal data attributes are subject to change at anytime. Therefore, these SPs also require their clients to notify them for the personal data updated. However, clients may forget or don't want to notify all related SPs of the changes on their personal data because of the burden on managing many accounts of different SPs. Sometimes they do not know that they need to notify these SPs of the change. Clients may also notify SPs of misinformation. As a result of that, personal data managed in many SPs may be inaccurate or outdated. In this case, SPs often end up paying the cost of poor data integrity. Also, clients may lose the trust of SPs. They may also suffer from extra cost, for example undeliverable mails/goods because of forgetting to update address when moving.

To reduce the burden of an individual in updating their personal data in different SPs and to ensure the accuracy and up-to-dateness of personal data managed by the SPs in the important domains which requires that the digital identity of any client must match his actual identity in real-world, this paper introduces a personal data update support approach whose unique feature is to evaluate the credibility of a change on a user personal data using many novel metrics relating to the personal data update policy of the SP where the change occurs and the importance of the SP to users. If the change is evaluated as credible, the impacted SPs, which have the right to know the accurate and latest value of the changed personal data attribute, will be informed about the change.

The rest of the paper is organized as follows. Section 2 introduces the research questions and our proposed approach. Section 3 describes the design of the personal data update support system realizing the proposed approach. Section 4 presents the result of a case study of evaluating the credibility of some popular SPs in practice. Section 5 is about related work and Section 6 concludes the paper by reviewing the main achievements of our research.

2. Research questions and proposed approach

Our research handles the following questions.

RQ1. Which SPs should be updated/notified of a change on a user personal data attribute?

Different SPs have different requirements on the accuracy and the up-to-date status of the managed personal data. Some SPs, such as banking services, stock services, want to ensure that all of their user personal data accurate and up-to-date. Therefore, they request their users to notify them of changes on user personal data. On the other hand, some other SPs do not care the accuracy of the provided personal data except for the key attribute. For example, a social networking service (SNS) often verifies the validity of only email address but not the other information like name, date of birth, or home address.

Therefore, a change on personal data attribute should be notified to the right SPs, which are the SPs having an agreement with users on providing accurate and up-to-date

‡ Hitachi Ltd., Research and Development Group

personal data, but not all SPs which manage that personal data attribute.

RQ2. How to know if a change on a user personal data attribute is credible, namely the new value of the changed attribute is accurate and up-to-date?

Not every change of personal data is credible or trusted. If an SP has strictly requirements on the up-to-dateness and accuracy of user personal data, for example validating the identity of users and their personal data by using identity verification documents, or using KYC services, the personal data managed in that SP and the changes on these data are highly credible. On the contrary, for the SPs with mild requirements, users can provide pseudonym information for some personal data attributes because they may not want to provide too much personal information except for the information which will be verified. For example, to use a SNS service, a user just needs to provide a valid email address verified by the SP. For other information such as name, job, or date of birth, he may provide pseudonym information. Therefore, the policies of SPs on the personal data verification will affect users on providing accurate and latest personal data values or not.

Also, the importance of the services provided by SP to a client also affects the credibility of the provided information. For the important services, clients are willing to provide accurate and latest personal data to avoid unexpected troubles due to incorrect information.

In addition, SPs may define their own list of trusted parties whose published information can be trusted. For example, a bank only trusts a home address if the address is registered with the city hall. In other words, a user must show a certification of the city hall for the new home address. Therefore, only declaration from the user is not enough, the certification on the change of the user personal data from of a trusted party of SPs is also necessary.

RQ3. How to know if a user personal data attribute is changed?

Except for the permanent personal data attributes like date of birth, the other attributes are changeable. Some temporary attributes like address, job can be changed frequently while the long-lived attributes like passport number are not changed regularly. In addition, changeable attributes may be changed suddenly due to unexpected events such as passport lost, credit card number stolen. SPs can use KYC services to check the validity of the managed user personal data periodically but not regularly due to the cost of verification. Therefore, SPs will request users to notify them of the changes on the registered personal data. However, users may forget to notify or notify SPs very late. So, the research question, how to notify the SPs, which have the right to know the change of user personal data, of the change as soon as possible?

Although some SPs are not updated or notified of the change on a personal data attribute late, there are always some SPs which are notified of the change very soon by users. These SPs may be a governmental SP where registration and update of some specific personal data attributes are citizen obligations. Other examples are e-commerce or logistics SPs where incorrect delivery address or expired credit cards will directly affect users through the stop of some necessary services or goods undelivered. Therefore, our approach to answer the above research question is to connect with many trusted and essential SPs to detect credible changes on many personal data attributes as soon as possible.

In short, the key idea of our approach is to connect to many trusted and essential SPs, to be able to early detect credible changes on as many as possible personal data attributes. The detected change will be evaluated its credibility using many metrics including the trust of the Source SP, which is the SP where the change is detected, and the importance of the Source SP to the user. A credible change will be notified to the right SPs which are impacted by the changes and have an agreement with users on the personal data update obligation.

3. Personal data update support system using change credibility evaluation

In this section, we present the structure and behavior of a personal data update support system realizing our proposed approach.

3.1 System structure

Figure 1 shows the structure of the proposed system which is composed of a system for managed SP information, **SP Information Management System**, and one or many instances of personal data management, **Personal Device**, where each instance manages the personal data of one user.

Personal Device includes the following components.

- **Personal Data Setting UI** is a UI for a user to specify the SPs whose services are registered and used by the user, and the Access IDs and Access Credentials for logging into these SPs. Personal Device can use these Access IDs and Access Credentials to retrieve the personal data registered in these SPs automatically.

- **Change Detector** is responsible for detecting changes on the user's personal data. It will regularly log into the user's SPs to obtain the latest values of user personal data, then compare with the values of the personal data stored in the local database (Personal Data DB) to detect the change.
- **Change Analyzer** analyzes the impact of the change by analyzing the credibility of the change and the relationship among the value of the personal data registered in different SPs to find which SPs and which personal data attributes of the user managed in these SPs should be updated if the change is credible.
- **SP Information Updater** processes Change Analysis Result sent from Change Analyzer by notifying impacted SPs of the change if the change is credible, and asking Change Analysis Result Delivery for storing Change Analysis Result.
- **Change Analysis Result Delivery** sends Change Analysis Result to SP Information Management System for storing.
- **Personal Data DB** stores personal data of the user and the related information such as the registering SPs and their Access IDs and Access Credentials.
- **Change History DB** stores history of changes on personal data of the user.

SP Information Management System includes the following components.

- **SP Information Setting UI** is a UI for setting change notification preference and change credibility preference of an SP.
- **SP Information Provider** provides SP information, such as its credibility or change notification preference
- **SP Credibility Calculator** calculates the credibility level of SPs using the change credibility preference of SPs and the Change Analysis Result DB. The credibility level of each SP is assigned an expiration date, so it can be reused within the expiration date instead of being recalculated.
- **Change Analysis Result Storage** stores Change Analysis Result sent form Personal Device to Change Analysis Result DB for further analysis purpose.
- **Change Analysis Result DB** stores Change Analysis Result of all changes on the personal data of all end-users.
- **SP Information DB** stores information of all SPs in collaboration with the proposed system, such as change notification/change credibility preference, and current credibility level.

3.2 Main processing flow

The main processing flow of the system is shown in Figure 2.

1. Change Detector periodically asks related SPs for the latest personal data values of the user.
2. Change Detector compares the latest personal data

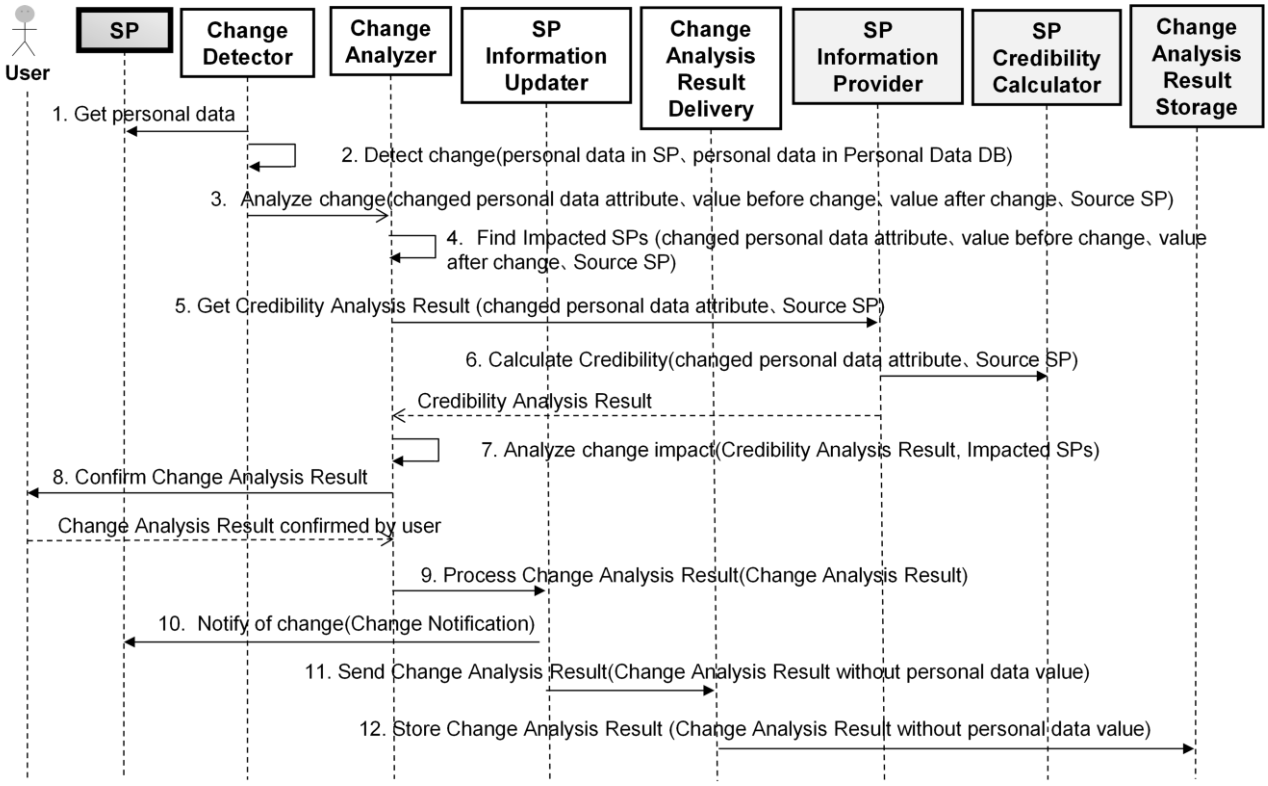

Figure 2　Main processing flow

values managed by SPs with the local values stored in Personal Data DB. If two values of the same personal data attribute are different, it concludes that there is a change on that personal data attribute. When a change is detected, Change Detector can ask Source SP, which is the SP where the change is detected, for more information of the changed personal data attribute, such as its previous value, the dates when the previous value and the current value were registered in Source SP, and Changer which denotes who makes the change, to prepare Change Detail. It also asks Source SP for a digitally signed certificate of change, which can be sent to Impacted SPs, the SPs which will be updated to be synchronized with the personal data value at Source SP, as the evidence of the source of change, if the digitally signed certificate service is available at Source SP.

3. Change Detector sends Change Detail to Change Analyzer for analyzing the impact of the change.
4. Change Analyzer finds Impacted SPs by analyzing the relationships among the personal data values registered in different SPs and the personal data update obligation of the user at that SP.
5. Change Analyzer asks SP Information Provider of SP Information Management System for Credibility Analysis Result of Source SP by providing the name of the changed personal data attribute and Source SP.
6. SP Information Provider asks SP Credibility Calculation to calculate the credibility of Source SP regarding the changed personal data attribute if a valid credibility value of Source SP is unavailable in SP Information DB. Credibility Analysis Result including the credibility value of Source SP and other information such as the calculation date and the expiration date, is then returned to Change Analyzer.
7. Change Analyzer uses Credibility Analysis Result and Impacted SPs for analyzing impacts of the change, namely deciding which SPs and which personal data attributes stored in these SP should be changed. Output is a Change Analysis Result.
8. Change Analyzer asks the user to confirm Change Analysis Result. The user will give final decision on the validity of the change and Impacted SPs.
9. Change Analyzer sends the confirmed Change Analysis Result to SP Information Updater for further processing.
10. SP Information Updater informs Impacted SPs about the change or directly updates them using associated Access IDs and Access Credentials, depending on their settings of the change notification method. The timing of updates or notifications is depended on the settings of the change notification timing.
11. SP Information Updater asks Change Analysis Result Delivery to send Change Analysis Result with personal data values removed to the SP Information Management System.
12. Change Analysis Result Delivery asks SP Analysis Result Storage to store Change Analysis Result in SP Information Management System.

3.3 Credibility Calculation

Credibility of a change on a user personal attribute at Source SP is calculated based on the credibility of Source SP on the changed personal data attribute. In the scope of this research, we propose to use at least one of the following metrics for calculating the credibility of an SP.

- **SP mutual trust (T1):** We assume that each SP will have a self-definition of trusted parties for each personal data attribute. Each SP can assign a trusted party a credit denoting their trust level to this SP. The higher the total credits an SP receives, the higher the credibility of that SP.
- **Update obligation trust (T2):** An SP which has a clear definition of the obligations of users on notifying it of personal data changes and the consequences for not obeying is more credible.
- **SP verification trust (T3):** An SP is more credible if it verifies the user to ensure their clients are genuinely who they claim to be, and validates information provided by user.
- **Accumulated trust (T4)**: Credibility of an SP is evaluated based on the validity of the changes which have already happened on that SP. The more the number of valid changes made in an SP is, the higher the credibility of that SP. T4 can be calculated using the Change Analysis Result DB which stores the validity analysis result of all changes analyzed by the proposed system.
- **User trust (T5)**: We assume that the trust of a user to an SP can be expressed through the importance of that SP to him. If an SP is important to users, they often supply accurate personal data to make the provided service work smoothly. So, when a change on a personal data happens, they should update that SP as soon as possible. Therefore, the earlier a user notifies an SP of the change, the higher the credibility of that SP to that user. T5 can be calculated based on the order of the changes in the same change transaction stored in Change Analysis Result DB.

Assigning a weight for each metric, **the overall credibility C** of SP on a personal data attribute is calculated as the weighted average of these metrics as follows.

$$C = \frac{T1*weight1 + T2*weight\ 2 + T3*weight3 + T4\ *weight4 + T5*\ weight5}{weight1 + weight2 + weight3 + weight4 + weight5}$$

4. Case study

In this section, we present a case study for evaluating the credibility of some popular SPs in practice using the proposed approach in Section 3.3. As the personal data update support system explained in Section 3 is not implemented yet, this case study evaluates the credibility of these SPs using 3 metrics, *SP mutual trust*, *Update obligation trust* and *SP verification trust*.

The study is conducted by searching the websites of common SPs for information relating to procedure for updating a personal data attribute. In this case study, we pay attention to the documents required by SPs for changing home address, for example identity verification documents, the information about personal data update obligation of users, and the procedure for verifying user and validating user personal data. Based on the required documents for updating home address of an SP, we can refer its trusted parties and their priorities which are used to calculate its *SP mutual trust,* T1. Similarity, information about personal data update obligation is used to calculate *Update obligation trust,* T2. The more formal and clearer the statements about personal data update obligation and the more serious consequences for not obeying the update obligation are, the higher *Update obligation trust* of that SP is. Finally, how an SP validates the user personal data is used to calculate *SP verification trust,* T3.

4.1 Result

Table 1 shows the result of investigating 8 popular SPs, including city hall, police station, bank A, bank B, telephone company C, telephone company D, e-commerce company E, and SNS company F.

T1 of an SP is calculated as the average of the trust priorities given by the other SPs to it. The trusted SPs of an SP are referred from the SPs issuing the documents required by that SP for updating address, including identity verification documents. The identity verification documents are basically divided into three groups including type A, for example my number card or driver license, type B, for example insurance card or pension book, and type C, for example electric, water, or telephone bills. Identity verification often requires 1 documents of type A, or 2 documents of type B, or 1 document of type B and 1 document of type C, in order of preference. Based on this classification, the trust priority an SP is assigned 1 for SPs issuing type A documents, 0.5 for SPs issuing type B documents, and 0.25 for SPs issuing type C documents, and 0 for SPs not mentioned. City hall is a special case. Because its required documents include a document issued by only city hall, city hall is assigned 1 while other SPs issuing type A documents, such as police station, are assigned 0.75. For simplicity, Table 1 shows the priorities of the trusted SPs in this case study only.

T2 of an SP is identified based on the information about the personal data update obligation of its users and the consequences of not updating. T2 is assigned 1 if the update obligation is defined by law, 0.75 if the update obligation is specified in the contract between users and SP or in the important notices for changing information in the contract which might cause the provided service stopped and documents/goods undelivered, 0.5 if the update obligation and the risk of documents/goods undelivered are given in SP's website, 0.25 if update obligation is not mentioned by SP but users may recognize the risks of documents/goods undelivered if not updating address, and 0 if no update obligation is mentioned and no consequence of not updating address happens.

T3 is identified by the way an SP validates users and their personal data. T3 is assigned 1 if both identity verification and change verification are done, 0.75 if identity verification is always conducted but change verification is done for some specific cases only, 0.5 if only identity verification is done, and 0 if no verification is done.

Finally, the overall credibility C is calculated as the weight average of T1, T2, and T3 where weight of T1 is 1 and weights of both T2 and T3 are 2. The values of C show that for the home address, SPs in the government domains are the most credible, followed by the SPs in the banking domain, and telecommunication domain. E-commerce domain has low credibility and SNS is not credible at all.

Table 1 Case study result

	Metric / SP	T1 (Trusted SP Priority Average)			T2	T3	C
		Trusted SP Priority		Average			
1	City hall	City hall	1	0.75	1	1	0.95
		Police station	0.75				
		Bank, telephone company	0.25				
2	Police station	City hall, police station, telephone company	1	0.72	1	1	0.94
3	Bank A	City hall, police station	1	0.06	0.5	1	0.61
		Telephone company	0.25				
4	Bank B	City hall, police station	1	0.06	0.25	0.75	0.41
		Telephone company	0.25				
5	Telephone company C	City hall, police station	1	0.19	0.75	0.5	0.44
		Telephone company	0.25				
6	Telephone company D	City hall, police station	1	0.19	0.75	0.5	0.44
		Telephone company	0.25				
7	E-commerce company E	Not mention	0	0	0.25	0	0.13
8	SNS company F	Not mention	0	0	0	0	0

4.2 Discussion

The result of the case study helps answering the research questions relating to a change on home address of a user.

- **RQ1: Which SPs should be updated/notified of a change?** Although all of these SPs have information of home address of users, values of T2 are different from SP to SP. For the city hall, police station, and telephone companies, the home address update obligation is compulsory due to law or contract constraint. For the banks and e-commerce company, the update is recommended for user due to the risk of document/goods undelivered. For SNS company, no need to update because of no update obligation or risk for not updating.
- **RQ2: How to know if a change is credible?** Assuming that the credibility threshold is 0.75, by comparing the values of the overall credibility C with the credibility threshold, a change on home address at city hall or police station is evaluated as credible. For a change at other SP, we need the judgement of the user for its credibility.
- **RQ3: How to know if a user personal data attribute is changed?** From the values of T1, city hall and police station are the most trusted by the other SPs who prefer the identity verification documents issued by city hall or police station. Also, because the update obligation of users for city hall and police station is defined in law, there is a high possibility that users will update their home address in city hall and police station first. Therefore, a regular check on the change of user personal data, in particular home address, in city hall and police station will help to detect the change on the user home address soon.

Also, through collecting information of SPs for the case study, we recognize that information about personal data update obligation and the consequences for not updating in SPs are not easy to find. As a result of that, users might not know their personal data update obligation, and hence do not notify SPs of their personal data changes. Our proposed system can support users by providing end-users about their update obligation and warning them about the consequences for not updating. SPs can also improve their credibility by improving the process for identity verification and user personal data validation and providing users with clear transparent and easily understandable information about their obligations and rights in providing and updating their personal data.

This case study in specific and an investigation for the credibility of real-world SPs on different personal data attributes in general are necessary preparation before developing a real system. This preparation helps to decide which SPs the proposed system should connect to so that it can detect credible changes on as many as possible user personal data attributes to avoid the oversight of a credible change on certain personal data attribute and can provide users with more accurate Change Analysis Result to avoid extra processing of users.

5. Related work

User personal data are managed as the attributes of user identities managed by an identity access and management system of an SP. There are many identity access and management products such as Microsoft Azure Active Directory, IBM Security Verify, and Okta Workforce Identity tools [2] which support both Single Sign-On (SSO) inside an organization and cross-domain SSO with identity federation solution [3]. In both cases, SPs in the same organization or joining into the same identity federation must trust on each other and have agreed in advance of how to share identity information including personal data attributes, among them. In other words, a change on a personal data attribute in a partner SP is assumed to be credible and updating other SPs of the changes or not will depend on their agreement before. On the other hand, our system is a supplement to the existing identity access and management products by handling the real-world context where SPs may be independent from each other and a personal data change on an SP may be incredible as users may provide misinformation.

Besides personal data management, data privacy also receives a lot of attention. [4] evaluates the credibility of a Web/SNS based on its data privacy policy. Similarity, our system evaluates the credibility of a change through evaluating the credibility of the SP where the change occurs, but we use many novel metrics relating to the change update policy rather than the privacy policy. However, we may consider integrating the privacy policy as a new metric into the proposed system because the privacy policy of an SP may affect to the trust of a user to this SP, and therefore affecting to their attitude towards providing correct and up-to-date personal data.

[5] gives an overview of the approaches to information credibility assessment including information propagation based approach which study how low-credibility information spreads and information classification based approach which use learning algorithms to classify information based on their credibility level. However, these studies focus on the online information which is diffused through the (Social) web by means of Web 2.0 technologies while our research focuses on the offline information in traditional environment characterized by interpersonal and persuasive communication, in particular personal data provided to SPs by the personal data owner only.

6. Conclusion

This paper focuses on updating personal data attributes of user identities in real-world context where different SPs have different requirements on the accuracy and up-to-dateness of user personal data and users may provide some SPs with inaccurate and outdated personal data. We clarify and handle three research questions including how to detect a change on user personal data, how to know a

change is credible, and how to know which SPs should be notified of change. Our approach is to connect with many trusted and essential SPs, to early detect changes on as many as possible user personal data attributes. The credibility of the detected change is then evaluated using at least one of the five novel metrics, including *SP mutual trust* which is the trust of SPs on each other, *Update obligation trust* which relates to the legally binding agreement on updating personal data among SPs and users, *SP verification trust* which involves how SPs validate user personal data, *Accumulated trust* which relates to the validity of all changes on an SP in the past, and *User trust* which represents the importance of an SP to users. Finally, the related SPs which are impacted by the changes and have the right to know the accurate and latest value of some user personal data will be notified of the change.

Toward realizing the above approach, we have designed a credibility evaluation based personal data update support system. In addition, we have conducted a case study for evaluating the credibility of some popular SPs in practice using three metrics: *SP mutual trust*, *Update obligation trust*, and *SP verification trust*. The result of the case study shows that for the home address, SPs in the government domains are the most credible, followed by the SPs in the banking domain, and telecommunication domain.

We expect that the proposed approach helps to remove the burdens of users on updating personal data in the SPs which requires accurate and latest user personal data. It also helps SPs preventing some economic lost or illegal business activities due to the usage of inaccurate or outdated personal data attribute values and contributes to easing the collaboration among SPs.

Refences

[1] J. Jensen, “Identity Management Lifecycle - Exemplifying the need for Holistic Identity Assurance Frameworks”, Lecture Notes in Computer Science, Vol. 7804, pp. 343-352 (2013).
[2] M. McDade, “Top 10 Identity and Access Management Solutions”, Expert Insight https://expertinsights.com/insights/top-10-identity-and-access-management-solutions/ (Accessed 2022/03/24).
[3] Okta, Inc., “Federated Identity Management vs. Single Sign-On: What’s the Difference?”, https://www.okta.com/identity-101/federated-identity-vs-sso/ (Accessed 2021/03/24).
[4] Y.Ichifuji and N.Sonehar, “Credibility Estimation of Web/SNS Site Using Privacy Policy”, IEICE Transactions on Information and Systems, Vol. J96-D, No. 6, pp.1493-1502 (2013).
[5] G. Livraga and M. Viviani, “Data Confidentiality and Information Credibility in Online Ecosystems”, In Proceedings of the 11th International Conference on Management of Digital EcoSystems (MEDES '19), pp.191-198 (2019).