DATA MATCHING IN THE IDENTITY ECOSYSTEM

Increasing Identity Transactions by Enabling Service Providers

THE OPEN IDENTITY EXCHANGE |

REPORT WRITTEN BY **EMMA LINDLEY**
INNOVATE IDENTITY

Executive Summary

Organisations are recognising that moving services online has the potential to significantly reduce costs and improve service delivery. Within Central and Local Government moving a typical service from face to face to online delivery is estimated to drop the cost of the transaction from £15 to 17 pence[1].

To enable this transition online organisations must be assured of the identity of the person they are transacting with, particularly for transactions deemed to be higher risk. A market of Identity Providers is being created in the UK through the Government's Identity Assurance Programme. Identity Providers will allow citizens to assert a digital identity in transactions with both the public and private sectors. (**Fig. 1.)**

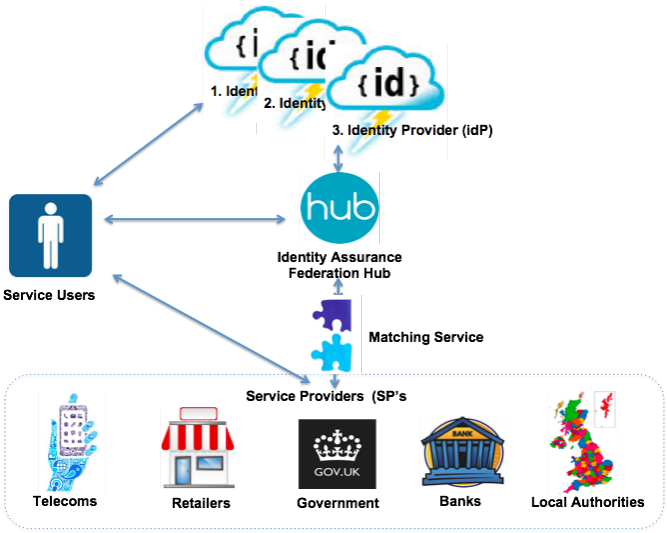


Fig 1. The Identity Ecosystem with a Data Matching Service

Adoption of an external digital identity asserted by the users Identity Provider requires the organisation to 'match' the identity details to customer records within its own systems. Previous Open Identity Exchange projects in Warwickshire and South Yorkshire have shown this to be a complex task because systems that hold customer records have different technical and data standards.

There is opportunity to reduce much of this complexity by developing data matching standards that any systems operator can deploy through a **'Matching Service'.** The service will allow a government standard digital identity to be linked to existing customer records at much lower cost than if each organisation develops its own bespoke approach.

[1] SOCITM 2013

Table of Contents

About Projects

In order to hasten adoption of the IDAP Framework there is a practical and strategic opportunity to leverage OIX domain expertise. OIX facilitates and coordinate the rapid formation and deployment of *White Papers*, *Discovery* and *Alpha Projects* in an agile manner.

These are defined as small scale, low risk assessments, analyses or tests of interoperable components that address the key challenges of the IDAP goal to create

This paper has been written as phase one of a discovery project into the matching service element of the identity ecosystem. It explains why it is required, what is already being done by organisations in the area of matching and, of that, what could potentially be reused. Additionally it explains the current challenges faced by organisations such as the service providers and attribute providers who will need to implement data matching in order to adopt and maximise the benefits of identity assurance.

It is estimated that UK organisations can save in the order of £1.5 billion by allowing customers to use their government standard Identity Provider service[2]. Matching Services are a requirement for any public or private sector service provider adopting external digital identities. Matching Service providers have a commercial opportunity to accelerate this adoption and cost savings for their clients.

1. Background

During previous discovery projects in Warwickshire and South Yorkshire it was identified that there is a requirement in the identity ecosystem for a matching service for all Service Providers (SPs)

This would make it an essential and ubiquitous component for Service Providers, making it possible for them to adopt Identity Assurance. This in turn would enable new online transactions, and would assist central and local government in moving to full "Digital by Default", and private sector organisations to move services online and for all to realise significant service improvements and cost reductions. The matching service would also be essential for Attribute Providers, which can be referenced in the Open Identity Exchange white paper "Can Attribute Provision together with Identity Assurance, Transform Local Government Services?"

[2] The Economics of Identity – Ctrl Shift

Why Data Matching

- ***The potential*** for the UK organisations to reduce their identity assurance costs over the next decade from £1.65 bn to £150m based on "make once, use many times" electronic processes

- ***The commercial*** opportunity for providers of a matching service to span the 433 principal authorities in the UK: 27 county councils, 55 unitary authorities, 32 London boroughs, 36 Metropolitan boroughs, 201 districts, 32 Scottish unitary authorities, 22 Welsh unitary authorities, and 26 Northern Ireland districts plus hundreds of private sector organisations

2. What is a "Matching Service" and why is it required?

In this context a matching service can be described as "user recognition" or the process of finding a local identifier through matching, which is useful to the service provider (SP) when completing a transaction. This matching service functionality allows the SPs to say that the "John Smith" presenting himself for a transaction is the same "John Smith" that they hold already on file in a locally hosted database. Data matching is a requirement of the identity ecosystem. Without a matching service the identity ecosystem cannot function.

Matching Requirements

As part of the phase 1 of this project a workshop was completed to understand the view on data matching across a number of central and local government departments. HMRC, DWP, DVLA, Warwickshire County Council and Hackney Council. Some of these organisations have already completed data matching projects, some had not, but are keen to explore the benefits of identity assurance, and as such understand they will need to adopt a matching service.

This workshop identified that whilst the scale requirements of these organisations was different e.g. DWP would have many more data matching requests in terms of volume than a local authority, it was found that the level of complexity of the matching requirements was very similar. This leads us to believe that there could be some standardisation of requirements, or ability to configure a matching service that might allow matching services to be provided on scale. This scaled implementation would contain standard elements or configuration by the Service Provider rather than bespoke implementation each time, with the aim of bringing the unit cost of delivery of a matching service down through economies of scale.

Based on information taken from the organisations and stakeholders during the workshop the following information was

single customer view

"an aggregated, consistent and holistic representation of the data held by an organisation about its customers or users"

gleaned and scoped out. The data matching requirements include: how a single customer view might help; alternative approaches to a single customer view; current methods for matching in the identity ecosystem; how persistent identifiers are dealt with; matching cycles; commercial views and confidence scores.

It is possible that some of this information could be developed further into "best practice" principles for data matching in the future.

The document below outlines the findings and requirements outlined as part of that workshop.

3.Value of Single Customer View vs Individual Database Matching

Single Customer View

Often when users interact with an organisation that provides multiple services a new record is created each time. This leaves a data record on disparate databases across the organisation, and can make it difficult to have a single view on what products or services that user takes.

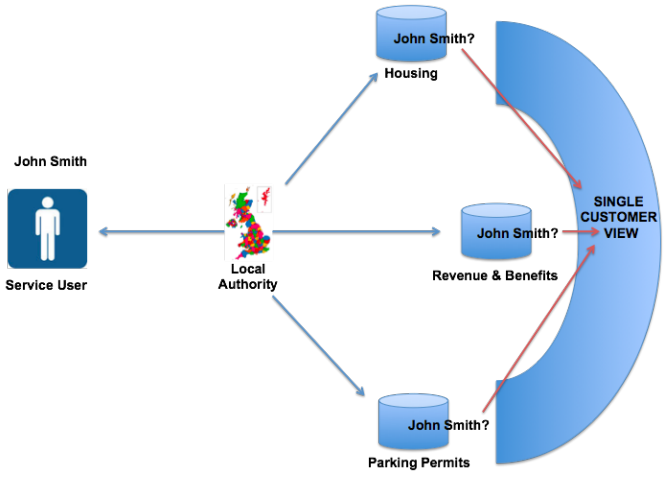
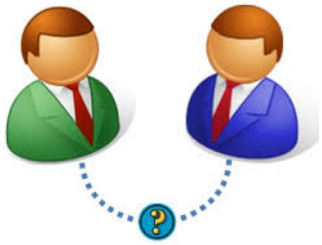


Fig 2. Multiple Service Organisation and Single Customer View

individual database matching



"data matching is completed in real time against each customer record in each database that could hold information about the individual inside an organisation"

A "single customer view" can be described as an aggregated, consistent and holistic representation of the data held by an organisation about its customers or users. This single view aggregates data from disparate systems allowing real time matching. The advantage to an organisation of attaining this unified view comes from the ability it gives to analyse past behaviour, and to then better target and personalise future customer interactions. A single customer view is also considered especially relevant where organisations engage with customers through multiple channels, since customers expect those interactions to reflect a consistent understanding of their history, preferences and to give customer insight. There are additional benefits to building a single customer view that include the increased ability to identify and then prevent fraud and misuse of public services quicker than when records are kept in disparate databases.

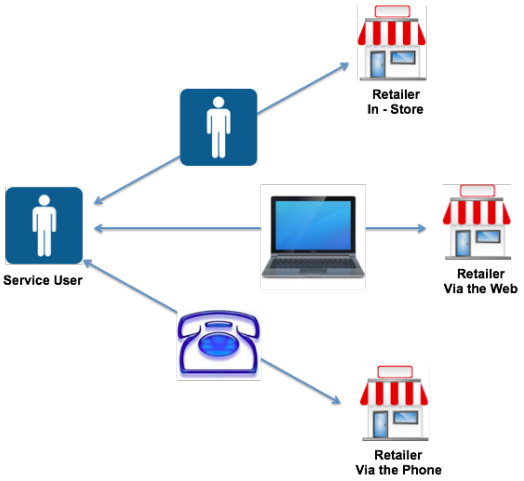
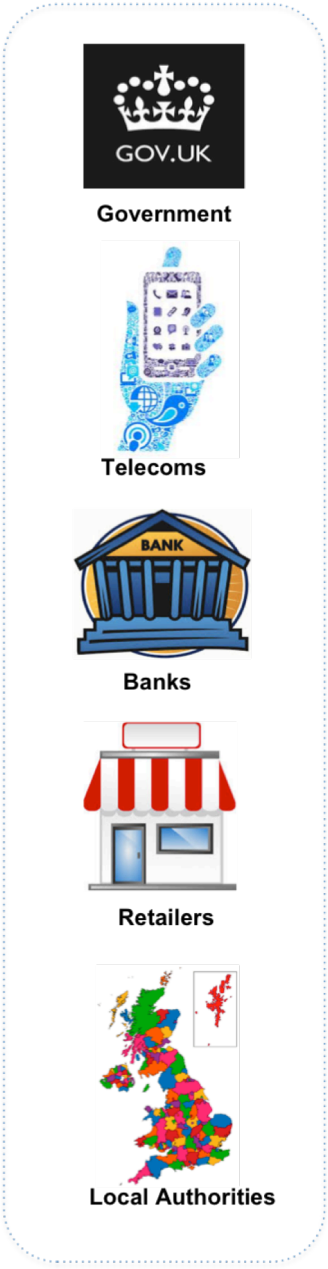


Fig 3. Multi-Channel Interaction with the User

There are multiple identified benefits to building a single customer view, however the view was these projects are sometimes considered complex and high cost.

Individual Database Matching

As technology has improved individual data matching is another method of matching. In this model the data matching is completed in real time against each customer record in each database that could hold information about the individual inside an organisation.



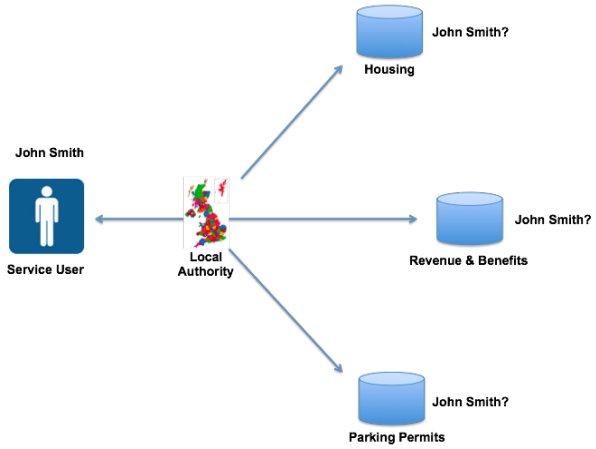


Fig 4. Individual Data Matching

The benefits of this model are that there is no single customer view project required. This model relies on technology, so uptime availability and functionality is critical. Some real-time matching relies on the Internet which can sometimes be affected by network availability and latency for the return of a match result.

4. Data Matching in the Context of Identity Assurance

In the context of identity assurance the following process is used for the Identity Provider to assure the identity, the identity is then asserted to the Service Provider which is where the data match needs to take place.

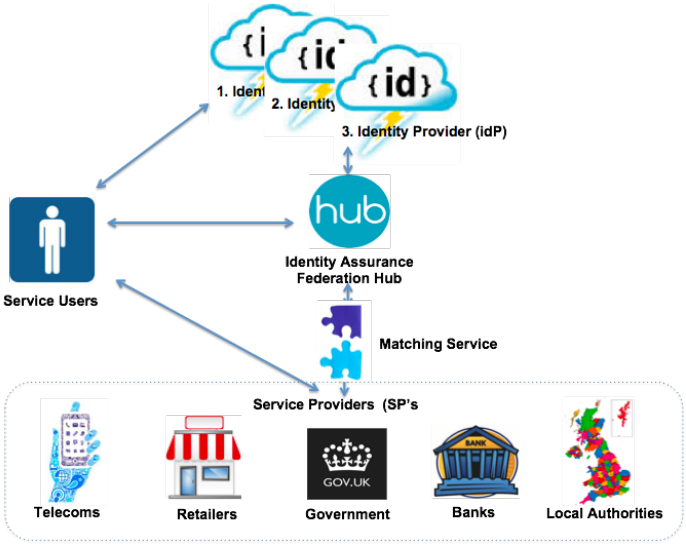


Fig 5. Identity Assertion and the Matching Service

Persistent Identifiers

- *Unique to user*
- *Unique to an identity provider*
- *Passed at registration*

1. The user wants to access the SP's service and the service requires an assured identity to do so.
2. The user is redirected to the hub service and selects and IdP. The IdP is then responsible for verification of the user's identity. The SP doesn't know (or need to know) if a user needs to register or not. The SP simply asks the user to authenticate and the hub handles the rest. N.B. during any subsequent registrations the user will not need to repeat this process, they will simply re-authenticate without going through registration again.
2. The matching data set (MDS) can contain first name, middle name, last name, address, postcode, gender and D.O.B. The MDS may also contain historical values with to/from date ranges
3. A verified flag is included to denote that the element has been assured by the provider as existing in the real world and belonging to the asserting entity.

Persistent Identifiers

A persistent identifier (PID) can be defined as an identifier that is unique to a user and an identity provider. It's generated by the identity provider and is present in assertions in order to identify the user that the assertion refers to within that single IdP.

In the process of identity assurance Identity Providers are also required to create and pass a PID. The PID is a unique identifier representing the asserted identity's account and is passed at registration and on subsequent authentications. For privacy purposes the PID is hashed before being passed used at the SP's matching service. Matching can be done on the hashed PID after the first customer transaction with the SP to shortcut the matching process.

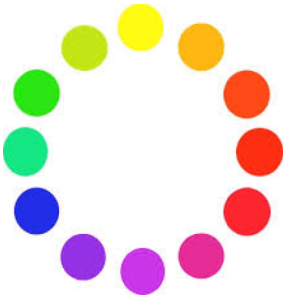
5. Matching Cycles

Matching cycle numbers refer to the sequence of attempts made to find a matching user for a particular transaction.

Cycle 0

Cycle 0 is the first data matching cycle and is used to determine whether the entity has been previously matched to a local identifier.

"matching cycles"



Matching cycle numbers refer to the sequence of attempts made to find a matching user for a particular transaction

The hashed PID and a local identifier (LID) are stored in a local matching store (LMS). The matching service searches for the hashed PID, if a match is found then the LID is retrieved and passed to the SP service, the following cycles are then not required.

Cycle 1

Is required when the details have not been matched previously at Cycle 0. The matching service checks to see if there's a match for the user's ID in the transaction's local store, using the matching data set (MDS) to try and achieve a match. Cycle 1 matches the MDS against organisational records; the objective is to achieve a sufficient match confidence (see section 6 confidence scoring) in line with service provider requirements. Matching routines will consist of a set of wide to narrow rules resulting in either a single or multiple matches.

Typically a set of rules would initially retrieve a set of records with surname and date of birth, and then filter on historical surname, followed by forenames, middle names then address. The method of scoring and weighting the outcomes would depend on service provider requirements. It is recommended that an extensive synonym list is used when matching forename.

Cycle 2

Is used after Cycle 0 and Cycle 1 have not been able to match the data to a high level of confidence, and it works by asking for additional attributes from third parties, examples of these third parties could include credit referencing agencies or other third party data providers. This process enriches the attributes allowing for a better chance of a match to take place. If the data is matched at this phase the match is approved, if not it goes to cycle 3.

Cycle 3

Is used to collect additional information from the user to arbitrate when multiple possible matches have been found. It works by asking the user for some additional personal information that the user knows and sends this back to the matching service. The user may enter a known fact (e.g. their National Insurance Number - NINO) to enable the matching service to differentiate between the other potential matches. Other additional attributes might include further

address information and contact information (e.g. phone or email), which can help further differentiate a candidate record.

If after cycle 3 a match cannot be made to a required confidence a temporary local identifier can be assigned or the department can create a new account based upon the MDS. This process would also occur if the individual after cycle 1 if the individual has never been encountered before.

Once a match is achieved the matching service can store the hashed PID and local identifier in the local matching store; because it is stored locally it removes the need to complete cycles 1 to 3 in future matching cycles.

6. Confidence Scoring

Confidence scoring is the score given to the data by which a match is called a match e.g. 100% match confidence would denote that all elements of data have been matched fully, a 80% match confidence might mean the forename, surname and address have matched but the D.O.B is showing a mismatch. Mismatches in data can be due to the user typing the asserted information incorrectly or that the data held at the locally hosted file is incorrect.

Data at the local host can be incorrect for many reasons for example, if someone has moved house but not informed the holder of the locally hosted file (e.g. the local authority), this means the new address asserted would mismatch against the locally hosted file. Sometimes the supplier configures the match confidence. Given attitudes to risk will differ for each organisation it makes sense for suppliers to make this a flexible and configurable by the organisation. This way they can agree their internal attitudes to risk set the system accordingly.

7. Matching Data Set

In the context of the identity assurance principles and the protection of users' privacy, each identity provider will be responsible for securely and separately holding data about the users that have registered with them. Each government department service will only

have access to the matching dataset it needs. If the SP requires verified data in addition to the MDS to complete the transaction in question, this will need to be sourced through attribute exchange and not the identity assurance service. The proposed approach to attribute exchange can be found in the sister white paper to this document which can be found on the Open Identity Exchange website called "Can Attribute Provision together with Identity Assurance, Transform Local Government Services?"

8. Matching Service Availability

As a second part of the research for this paper a workshop was also held with a number of potential matching service providers, these ranged from solution providers who already had an existing solution, to those newer entrants to the market. All providers were given the background to the project, the opportunity and also some of the challenges faced by organisations wishing to adopt matching services (e.g. those outlined in section 9 Barriers to Adoption). The potential suppliers were given the opportunity to give feedback for this paper. The aim was to find out what was available in the market place and if the challenges could be met by the potential suppliers in the industry.

Not all potential suppliers responded, but of those that did there was a real appetite to embrace both the opportunity and challenges faced within this document. It was clear from some of the responses that there is significant knowledge in the market place already, which could be demonstrated by live implementations of a matching service. The responses also indicated they would be favourable to looking at ways in making the solution commercially viable through a flexible approach to the commercial model.

9. Barriers to Adoption

It is possible therefore that one, some or many of these suppliers will be in a position to provide their services at scale to the mass market, and by taking advantage of these economies of scale supplying their matching service offering at a price point that

works for both large organisations with big budgets, but also medium and smaller organisations with smaller budgets. If there were a matching service provided at a viable price point it is envisaged that Service Providers and Attribute Providers could participate more quickly and easily in the identity assurance ecosystem. The second challenge in the marketplace today is in relation to matching service integration. Under the identity assurance ecosystem an "out of the box" solution is required to allow integration with the matching data set through something called a "matching service adapter"(MSA). The MSA is designed to simplify data matching for the SP by providing a different technical interface (JSON) as an alternative to the standard interface (SAML). The matching service adapter is not currently available as a commodity "off the shelf" item and therefore would need to be developed by industry.

10. Commercial Opportunity

Any public or private sector organisation who wants to go "Digital by Default", deliver better customer service levels and to reduce the cost of service delivery will be exposed to the risk of fraud and misuse of those services. There are 433 principal authorities alone who will want to deliver all levels of services online, not to mention the countless private sector organisations. Identity assurance is critical to the delivery of online services to prevent this fraud and misuse.

Additionally the newly formed identity ecosystem has opened up new opportunities such as those for "attribute providers". All organisations who want to use identity assurance or to provide attributes into the identity ecosystem will need a matching service. This represents a significant opportunity for commercial suppliers of matching services

- *433 principle authorities*
- *Attribute providers*
- *Private sector organisations*

Conclusion and Recommendations

Data matching services are an essential part of the identity ecosystem. There is a large-scale commercial opportunity for those matching service suppliers that have the experience and skills to provide a matching service into the identity ecosystem.

However commercial organisations need to look at how they could standardise component parts, gain economies of scale and use creative commercial models to manage organisations with both large and small budgets. Considerations for commercial models could include transaction based pricing to allow for the differences in volume of transactions. It could also be considered that the supplier does not configure the matching service technology, and technology dashboards could be provided along with training to empower users to manage their requirements and reduce the cost of implementation.

For organisations that want to move more services online but manage their risk the identity assurance programme can provide the key. However matching services will be critical in this move into identity assurance and "Digital by Default" therefore these organisations need to be thinking about how they will likely adopt matching services and the other elements that might go with them, such as a single customer view.

This project has generated much interest from the identity community due to its core function to the identity ecosystem. Therefore it is recommended that as an output from this project an OIX working group is set up to agree the next stage of this project. It is critical that the next stage is a practical activity to test component parts of how matching will work in a real scenario and to look at how a set of standards or best practice could be developed to enable organisations to move swiftly to identity assurance adoption.

The output for phase 2 of this project should be directed to working out how matching services can technically work and be delivered to the identity ecosystem, but also how commercially they can be made viable and easily procured such as through the government CloudStore and other related procurement vehicles.

-End-