# Account Linking

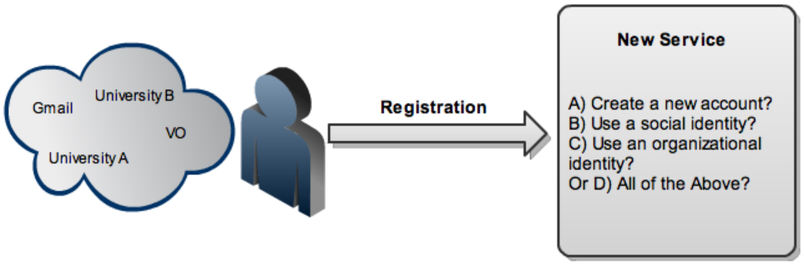
## Introduction

The concept of account linking has been around since at least the origins of federation, but despite its age, the understanding of account linking, the ramifications and technical requirements, is poor. So what is account linking? It is the act of associating one account, generally an existing local account, with a different account, such as a federated identity. Account linking happens regularly in an increasingly federated world and is usually the first action when an existing site tiptoes into federation. At a basic level, dealing with account linking means applications may need to deal with multiple identifiers for a single user. This may be a user-initiated action or it may be a programmatic action within a domain. Sites may also have to start dealing with the reconciliation of multiple identities in to a single account. As sites look at account linking and the policies and technologies required, a variety of questions need to be answered: is account linking done in the most efficient manner? How will the issues of privacy and assurance be handled? What about scenarios where it is a question of more than two accounts being linked? What site is authoritative when conflicting information exists in a linked account?

Account mapping, related to account linking, is on a technical level the mapping of an identifier from account A to account B when requirements go beyond beyond just the one identifier being used across multiple identities. This implies the additional problem space of dealing with conflicting information from the various sources of account information. The issues around account linking impact a variety of areas in identity and access management, including attribute aggregation, permission management, user experience, and more. The questions of where the work must be done - within an IdM system or within each individual application - are still under discussion. There are no common set of best practices.

A common identifier used to link accounts is email address. That is known to be a very weak and unreliable method for linking accounts, although it is widely used. Who "owns" a particular address, and what is it used for? Who is authoritative to assert an account identifier for a user? E.g. should a university be able to assert that a social identifier is a valid identifier for user john.doe@foo.edu?

This document looks to gather the use cases around account linking and provide guidance on how to cope with the complexities in this space.



## Use Case Library

### Account linking and Groups - SURFConext

- In a collaboration environment, users are taking advantage of the groups they have defined in LinkedIn, a business/social web site. Unfortunately, the identities of the users in the group do not necessarily map to accounts in the collaboration environment. The identifiers are different than what they may use in the SURFnet federation.

### Account linking and Learning Management Systems

- Universities want to share learning management systems. Behind the learning system are a variety of administrative systems (student records, instructor records, etc). It is difficult to integrate all the accounts in to one identity within a single institution; the complexities added when trying to integrate all the accounts at all of the institutions in to one identity that can follow a student as they change institutions are extreme.

### Account linking across transitions in one's relationships within an institution

- A campus has applicants using external identities (social or something else) to register an account with an institution during their application process. For some of these students, there will come a time when that external identity needs to transition to an institutional student identity and yet the history of the original account kept.
- Alternatively, but in the same style of relationship changing use case, a staff member has separated from institution. The campus has outsourced handling of federally mandated documents such as the W-2. The campus also has a policy to remove or lock down staff accounts at the end of employment. Still, the campus is required to provide a process to access those W-2 for 36 months, and so the institutional identity will need to be linked to another identity

### Account linking for the purpose of IdP choice

- An individual may have more than one account they want to use within a single service. When joining a collaborative organization, they may initially sign up with a social identity and transition to using an institutional identity.

- Depending on the context within which an individual is working, they may need to log in to a service with different identifiers. A researcher with affiliations to multiple institutions may need to log in to the NSF Fastlane service from one institution or another depending on what information she is accessing and which "hat" she is wearing when accessing the site. Different context require different identities.

### Account linking so that a 3rd party can serve permissions to a service provider

- account linking so a 3rd party could assert permissions to a SP. E.g. a VO asserting permission for a user, when the SP would not trust the campus to assert it.
- An individual creates an account in a cloud-based service using their institutional identity. At a future date, that institution enters in to an enterprise agreement with that same service such that all members of the institution may create accounts. The individual's account and the institutional identifier for an account that could be created under a different operating agreement need to be linked or transformed in some way. Does the enterprise account take over? Both parties might have an incentive (or not) to connect those relationships, though for privacy reasons they perhaps should not be linked.
- Also, in this same use case space, certain attributes will be loaded from the enterprise in to the cloud-service profile. The questions include: what attributes can be loaded, which ones can the user modify on their own in the box.net profile and which one are they not allowed to modify? The boarding process may be an aspect of the linking environment.

### Account linking in response to an invitation

- A VO adds someone to a group using their email address. That person may not have a presence in that project or IdM system yet, so an invite is automatically sent. The person has to respond to the invite by logging in from their IdP of choice and linking that account to the stub account created as part of the invitation process.

### Account linking and workflow

- The business processes at an institution require sign off on for certain contracts from a third party, resulting in the linking of an e-mail address to a new or existing id.

### Account linking of multiple federated identities

- An individual at Institution A associates their Institution A identity with a service provider. The individual later receives a joint-appointment at Institution B. This person also has a social identity. At any point in time, the individual may try to access the service provider with any one of these three (or more) federated identities and expect to access the same information at the service provider.

### Account linking between social identities and institutional identities

- An individual is a member of a VO. As a member of the VO, they have registered both their organizational identity and their social identity with the collaboration management platform. When the individual transitions from Institution A to Institution B, they use the social identity to link in the new account.

### Account linking and LOA

See the LIGO use case

### Account linking and licensing

- Institution A may pay a fixed annual fee for access to part of a resource while Institution B has a pay-per-view agreement for a larger, but overlapping, selection. That pattern of licenses may well already exist for academics who are also clinicians, for example. It means both that the same user sees **different** views of the same resource depending on which hat he is wearing and that the two institutions (and the service provider!) may have different commercial views on which identity should be used when accessing the dual-licensed parts of the resource.

## Examples of services where account linking is happening now

- Educause
- NSF Fastlane
  - NSF is using standard account linking pattern: the first time an individual come in with their federated id, the NSF ask them for their existing, local-to-the-NSF account and password, and the linking happens behind the scenes. Then the individual may log in with the new account and have all the standard information. Educause is also doing account linking this way. This method still leaves dangling issues: is the old account removed? What happens when the individual changes to a third institution? How are the two federated accounts linked?
- various NIH apps
- Google's experiences with its account system migration, merging previously silod account spaces into one, is a high-profile large scale example in the wild.
- At Indiana, because of the way email domains are constructed, they are requiring users to link their multiple institutional identities back to one another. This does not happen automatically for business reasons. This kind of institutional account linking is required when vendors allow a user to change their primary email address.
- At the University of Chicago Medical Center, the Medical Center continues to manage their own identities. They are a part of U. Chicago, so users may use their Medical Center identities without ever needing an additional U. Chicago identifiers. The account linking happens on a more fundamental level.
- Trusted Attribute Aggregation Service (TAAS) acts as a secure service to link multiple IdP or attribute authorities together. It does so by using persistent identifiers without requiring the service that is performing the linking to know anything about the user at all. The TAAS stores the

attribute types that the IdPs return as part of the account. It can then work as a proxy IdP service that authenticates the user at an IdP and retrieves the user attributes that are requested by the SP from multiple AAs. A demo running for firefox browsers is available online at http://sec.cs.kent.ac.uk/demos/taas.html. Feedback more than welcome.

## Issues still to explore

- LoA issues
    - When a higher level of assurance is required, how should account linking be handled? What if the accounts are at different LoA? What if they are at the same, but have different information (e.g. name, contact info, etc.)?
- Campus policy issues
- Privacy
    - Can accounts be linked and still preserve the privacy of the individual?
- consumer versus institution accounts
    - What should the guidance be when discussing the linking of social versus institutional accounts?
- a schema that encompasses linked account info

## Community thoughts

- Who is authoritative in data conflict situations?
  Each attribute authority (AA) should be authoritative for data from its own domain. However, when multiple AAs are returning conflicting data it does not necessarily mean that any of the data is wrong. It is quite possible that there may be two sets of attributes that conflict if the individual works for two universities at the same time with two different roles but both sets will still be valid. When conflicts such as this occur it is up to the user to decide which data set they wish to present
  to the SP as both sets may be valid. The underlying problem is more about how we provision IdPs and attribute authorities with information for which they should not be authoritative.

- Can accounts be linked and still preserve the privacy of the individual?
  Yes absolutely, so long as the user is in control of the linking process and has the ability to decide which accounts he/she chooses to release to each requesting party. e.g. I may have linked two accounts my work account and a dating site. If i'm then trying to access a site for work I mightn't want to release the fact that I've signed upto a dating site to all and sundry and I also might not want to send work attributes to the dating site.
  An example service that tries to do this is briefly described below:
  The technology already exists and is well established to receive a persistent but otherwise unidentifiable string to identify a user between a single AA and SP, if we authenticate to multiple AAs as part of the same session then we have multiple persistent ids that can be used to identify the same user at multiple AAs. If the linking service is attacked it doesn't matter if the PIDs are stolen as they are only valid between the service and the AA and cannot be used by an external attacker and no personal information needs to be stored on the service. When an SP requires authn /attributes it redirects the user to the service with the linked accounts which authenticates the user (requesting no attributes). The service can then query all (or a selection) of the linked IdPs for attributes including a token identifying user via the stored PID. The IdP can then decide whether or not it trusts this linking service to authenticate/aggregate and if it does create an encrypted assertion (encrypted to the SP not the linking service) to be sent to the requesting SP. This does of course require a higher level of trust in the linking service than in a standard SP but unless collusion occurs between multiple parties the use of PKI can ensure that no attribute data is ever visable except to the intended recipient SP.

- When accounts at different LoAs are linked, is the resulting information at a higher, lower, or varied LoA overall?
    - We need to differentiate between the login LoA and the registration LoA. The LoA of the actual authentication will then be the highest level of assurance for all the data returned as that is how sure at the time we are that the user is who they say they are. However if the registration process of the other AAs is lower than this threshold i.e. the authenticating IdP used a smartcard login and the second linked account registers its users via a web form then it cannot be presented as the same LoA value it must be downgraded to the registration LoA as that is how sure we are that the data is accurate.
    - If access within the SP is determined based on the user's affiliation with an specific Credential Provider, then a user that has logged in from multiple IDPs could have a higher level of access because of the accumulation of privileges based on affiliation with multiple Credential Providers. If the user's affiliation with a particular Credential Provider is terminated the SP may never learn of this because the user will simply stop authenticating against that IdP. Therefore, the SP may never remove the associated permissions, yet the user still has access because of a valid account at at least one other IdP.