

# Identifiers and Usernames

By Ian Glazer

## Table of Contents

- ABSTRACT** ..... 1
- INTRODUCTION** ..... 2
  - WHAT ARE IDENTIFIERS AND USERNAMES? .....2
  - WHY CONSIDER IDENTIFIERS AND USERNAMES?.....2
  - TYPES OF IDENTIFIERS .....3
- TERMINOLOGY**..... 3
- ASPECTS OF USERNAMES**..... 4
  - SECRET.....4
  - PUBLIC.....5
  - MEMORABLE .....6
  - UNIQUE .....7
  - RECOVERABLE.....8
- CONCLUSIONS** ..... 9

## Abstract

An identifier is the way an identity management system or other entity refers to a digital identity. The identifier used by the system, however, likely differs from the identifier used directly by the user and will definitely differ from identifiers in another domain. This article reviews the concept of identifiers as they relate primarily to people, both from a user’s perspective and a system’s perspective, and their impact on the systems that use them.

# Introduction

## What are identifiers and usernames?

In the physical world, we use a variety of ways to identify a person or a thing. From serial numbers to mailing addresses to license plates to nicknames, humans select a specific thing from a collection of similar things via an identifier. In the online world, this behavior is no different. Computer systems and people who work with them use identifiers to distinguish between similar items. More formally, and in the context of identity management, we can think of an identifier as the way an identity management system refers to a digital identity.

However, the person associated with that digital identity may not use the same identifier that the system uses. In fact, it is highly likely that they do not. Likely the person uses a human-friendly identifier. For the sake of differentiation, let's call the way a person in control of a digital identity identifies themselves to a system a username.

## Why consider identifiers and usernames?

How systems refer to digital identities and how people refer to their digital identity in a system are crucially important. Identifiers and usernames are one of the most commonly used components of a digital identity management system. They have implications for usability, security, customer satisfaction, and system operations, and enable (or prevent) cross-system correlation and user account management. They have applicability in business to employee (B2E), business to business (B2B), business to customer (B2C), and business to business to customer (B2B2C) use cases.

Failing to consider identifiers, and especially usernames can have direct, negative impacts on the projects and systems you are working on.

## Types of Identifiers

Identifiers come in two varieties: internal and external. Internal identifiers are the means by which a system refers to a digital identity. Formats of internal identifiers can vary greatly. One common format of internal identifiers is a universal unique identifier (UUID.) Specified in the IETF RFC 4122, UUIDs come in 4 variants or versions.<sup>1</sup> Many systems use UUID version 4 (often referred to as UUID4), which are randomly generated identifiers. An example of a UUID4 is: d5372288-697b-42bf-928a-562aca0deeaf.

But not all internal identifiers are UUIDs. Systems can use other means of uniquely identifying a specific thing from a collection of similar things. Examples include identifiers that have specific meaning to the system but are meaningless outside of the system, such as the following identifier, "005o000000s4Hu."

The second variety of identifier is an external identifier. An external identifier is the means by which a person in control of a digital identity refers to that identity when interacting with a system. These include but are not limited to a telephone number, email address, nickname, or handle.

In some cases a system identifier can be used by both internal and external purposes. Since email addresses must be unique within an organization i.e. a company, or domain i.e. college or physicians, the member's 'net id' i.e. first part of their email address, will be used within corporate systems as a user's identifier. A net-id could be comprised of first initial, second initial, last name and a number that ensures uniqueness.

## Terminology

- Internal identifier: the way an identity management system refers to a digital identity
- External identifier: the means by which a person in control of a digital identity refers to that identity when interacting with a system
- Username: a common term used for an external identifier

# Aspects of Usernames

When considering what the format of usernames should be, an identity practitioner must consider the five guiding principles of usernames. The practitioner should consider username format in greenfield situations as well when new B2C or B2B2C services are being created, at the very least. Often, especially in B2B scenarios, usernames have formats established in previous generations of systems, and those formats take on an almost mythic quality. It is not reasonable to simply change username formats, and needless to say, changing username formats, especially in an enterprise B2B setting, is not an undertaking one should take lightly.

Cloud applications are a potential area of username confusion. For a multitenant application, usernames should be common, i.e., the username for a digital identity in one application is the same as that used in another application. However, in some cases a user sets up an account in a SaaS application and selects another username. If this application is subsequently interfaced to the identity management environment a transformation mechanism will be required. API gateways or identity provider services maintaining multiple usernames are options.

The five guiding principles identity practitioners should consider are that usernames:

- [Are not a secret](#)
- [Must be classified as public data](#)
- [Must be memorable](#)
- [Must be unique](#)
- [Must be recoverable](#)

## Secret

There is an instructive lesson in the United States' Social Security Number (SSN) as an anti-pattern for usernames.<sup>ii</sup>

SSN was meant as an internal identifier. Originally it was something the Social Security Administration would use to tie a human to their earned wages and eventually to their entitlements; it was something that they would use for their business processes. They shared this internal identifier with people and their employers to make business processes run. However, the use of this internal identifier grew. Businesses began to use SSN as a way for people to identify themselves to the business; in essence, business turned this internal identifier into

a username. This secondary use was based on the idea that only the person would know their SSN and thus, because it was secret, the holder of the SSN would be assumed to be the correct person. And this is where things went wrong.

The need for this specific secret permeated so many of our business processes throughout our economy. This need has created a massive amplifier for damage when data brokers and others have breaches.

The lesson of SSN is that usernames cannot be secrets. If you share an internal identifier with a party outside of your organization, you have turned that internal identifier into public information, and thus it cannot be a secret.

If you have a username or an internal identifier that has to be treated like a secret, then you do have an authentication mechanism on your hands, not a username. And this means that it needs to be treated akin to a password.

As a pointer to an advanced topic for a later date, consider this- biometrics, broadly speaking, cannot be secrets. A person cannot keep their fingerprints, facial geometry, or irises secret. Because of this, a system or process can use biometrics as external identifiers. But because they are “just” identifiers, some degree of authentication is required to ensure the person actually intends to present their biometric. This degree of uncertainty is why liveness detection and attention detection are so crucial. For example, it is insufficient to accept a fingerprint biometric without also checking that the finger is real, has blood pumping through it, and [isn't a fake made of gelatin](#).<sup>iii</sup>

## Public

It is not enough to make sure that usernames are not secrets. Identity practitioners must also classify usernames as public in one's data classification scheme. This action applies to employees, partners, and customers alike.

Classifying usernames as public does not mean attributes related to the individual are public. Such attributes cannot reasonably or safely be used in a username. Consider a simple four-level data classification system:

- Public: this data can be shared across organizational boundaries freely and with a low level of concern.
- Restricted: this data is essential to business process and likely cannot leave organizational boundaries. Only data subjects, employees, and contractors can have access to this data.
- Confidential: this data is crucial to business operations. Significant harm may occur if this data transits organizational boundaries.
- Secret: this data is extremely organizationally sensitive. Only a small select group of people and systems can have access.

In an ideal world, an airline or hotel loyalty number (another kind of identifier) is likely classified “Restricted.” Usernames must be classified as “Public.” Airline or hotel loyalty identifiers demonstrate the problem of an identifier that is “public” but contains attributes that have value.

In addition, classifying usernames as public reinforces the idea that identifiers cannot be secrets.

As a clarification, the recommendation is that the username should be classified as public data, in a data classification system. That does not mean that usernames should be publicized (e.g., listed on a public site) – that is a self-inflicted enumeration attack.

## Memorable

Part of the canon of US literature is Herman Melville’s Moby Dick. And its first sentence reads “Call me Ishmael.” Ishmael, the username, is not the most important part of that sentence – the “call me” part is. The power to name something is the power to control it. And by naming himself Ishmael takes control over himself, away from the Reader and away from the author.

In support of self-determination, people have to give themselves names, and in the digital world, this is crucially important. Usernames need to be self-generated in B2C and B2B2C settings, which is to say the person should have the power to create their preferred username. It is important to also consider self-generated usernames in B2B and B2E settings as well.

Many enterprises have a standard username format and they bring that preference to B2C and B2B use cases. A classic username format is First Initial, Last Name. For example, Sally Smith would get a username of “ssmith” and if that wasn’t unique a random number would be added. This habit-based username format, although reasonably effective, doesn’t support a desire for self-determination which is so crucial in B2C use cases.

Failure to support memorable usernames means increased account recovery calls, more on-screen help, and more customer support needs. And it also leads to duplicate identifiers because people often forget the identifier they used to register.

When building a username scheme, one needs to provide choice to the user. If asked for email as username and then on the next screen the user says, ‘do not use email to talk to me’, then there is significant cognitive dissonance. In order to provide choice, consider supporting multiple username schemes such as email addresses and user-created nicknames. Supporting multiple schemes adds a level of complexity, but the user empowerment that brings with it engenders self-determination and customer satisfaction.

## Unique

Usernames need to be unique. Internal identifiers need to be unique. Neither statement should be controversial, but there is nuance here.

It is not enough to say a username must be unique; one must consider the scope of uniqueness. Is the username unique:

- at the individual service level?
- at the tenant level (if you are multi-tenant)?
- within a namespace with a service or set of services?
- globally across all of your services?
- universally?

Is there a clear picture of the scope being designed for? Even if there is, that picture may change; practitioners need to consider if the future might include merging internal systems or have to support various merger and acquisition activities in the future.

Also, uniqueness has implications depending on the type of identifier. Usernames and internal identifiers do not have to have the same scope of uniqueness. For example, while an internal identifier needs to be globally unique, a username might be unique only in a subset of systems in the enterprise. Internal identifiers have to be unique at the service-scope, e.g., unique in a specific enterprise service. To mitigate potential data subject reidentification, then those identifiers ought to be globally-scoped unique. Meanwhile, a person might use their email address to log into multiple systems - a service-level scoped unique username.

In addition, do not, in the same system, make the username and the internal identifier the same value. In some regards, this was one of the mistakes the US made with the Social Security Number.<sup>iv</sup> Practitioners should not make them the same value if only because changing either later can be enormously challenging. Furthermore, a common username scheme of choice is an email address, and these can change over a person's life based on life events such as marriage and divorce. Accommodating such changes to the username in a scheme where the username and internal identifier are the same requires that all systems with the "old" username/internal identifier need to be aware of the change and updated; in a complex environment, that task may be nearly impossible.

A final consideration is username reuse. Yahoo email allows people to use email addresses that were once used by someone else. Phone numbers are regularly reused. In this case, the username may still be unique but the person in possession of that username has changed. This transitional period is a difficult situation to be in if for no other reason than the new possessor of the email or phone looks like an attacker in many cases.

## Recoverable

Usernames need to be recoverable, which is to say, that there needs to be a way to get a person back to their digital identity. Recovery means re-attaching the person to the digital identity; it does not necessarily mean they will use the same username over again.

In this regard, recovery is more than just reminding the person of what email address they used to log in. Consider telling a person that the email address they used was their old work email address that they cannot access. That leaves the person little recourse but to call the help desk... or move on to a new service.



Recovery is a re-association and to do this safely, it often requires re-proofing the individual is who they claim to be. Especially in B2C scenarios, such a re-proofing process requires considerable thought as it has significant security and customer satisfaction implications.

## Conclusions

Identifiers are necessary to an identity system, with internal and external identifiers serving different purposes. While the two types of identifiers can be the same, the IAM practitioner should consider this with caution. External identifiers, also known as usernames, should consider these five guiding principles:

- Usernames should not be considered a secret.
- Usernames must be classified as public data.
- Usernames must be memorable.
- Usernames must be unique.
- Usernames must be recoverable.

Each principle has implications for the identity practitioner to consider as they develop an identity management system. Constructing a username framework is part of the 'identity orchestration' task.

---

<sup>i</sup> Leach, P., Mealling, M., and R. Salz, "A Universally Unique Identifier (UUID) URN Namespace", RFC 4122, DOI 10.17487/RFC4122, July 2005, <<https://www.rfc-editor.org/info/rfc4122>>.

<sup>ii</sup> Carolyn Pucket, "The Story of the Social Security Number, Social Security Bulletin, Vol. 69, No 2, 2009, <https://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html>.

<sup>iii</sup> John Leyden, "Gummi bears defeat finger print sensors," The Register, 16 May 2002, [https://www.theregister.co.uk/2002/05/16/gummi\\_bears\\_defeat\\_fingerprint\\_sensor\\_s/](https://www.theregister.co.uk/2002/05/16/gummi_bears_defeat_fingerprint_sensor_s/).

<sup>iv</sup> Pucket, see *Expanding Uses of the SSN*, <https://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html>.