NIST Interagency Report
NIST IR 8480 ipd

Attribute Validation Services for Identity Management

Architecture, Security, Privacy, and Operational Considerations

Initial Public Draft

Ryan Galluzzo
Connie LaSalle
Maria Vachino
Richard Newbold

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8480.ipd

NIST Interagency Report
NIST IR 8480 ipd

Attribute Validation Services for Identity Management

Architecture, Security, Privacy, and Operational Considerations

Initial Public Draft

Ryan Galluzzo
Connie LaSalle
Applied Cybersecurity Division
Information Technology Lab

Maria Vachino
Richard Newbold
Calvert Consulting, LLC

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8480.ipd

October 2024

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

How to Cite this NIST Technical Series Publication
Galluzzo R, LaSalle C, Newbold R, Vachino M (2024) Attribute Validation Services for Identity Management: Architecture, Security, Privacy, and Operational Considerations. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) NIST IR 8480 ipd. https://doi.org/10.6028/NIST.IR.8480.ipd

Author ORCID iDs
Ryan Galluzzo: 0000-0003-0304-4239
Connie LaSalle: 0000-0001-6031-7550
Maria Vachino: 0000-0002-3494-5307
Richard Newbold: 0009-0008-5033-6684

Public Comment Period
October 7, 2024 –November 8, 2024

Submit Comments
Digital_Identity@nist.gov

National Institute of Standards and Technology
Attn: Applied Cybersecurity Division, Information Technology Laboratory
100 Bureau Drive (Mail Stop 2000) Gaithersburg, MD 20899-2000

Additional Information
Additional information about this publication is available at https://csrc.nist.gov/pubs/ir/8480/ipd including related content, potential updates, and document history.

Abstract

Attributes provide information about an individual that can be used to confirm the individual's identity or ability to access information or services. Attributes and the processes for validating and asserting them are essential for securely identifying individuals and can also be utilized for authorization and other purposes. This report provides a foundation upon which federal, state, and local government agencies can design and develop attribute validation services. Agencies with authoritative data are well-positioned to provide attribute validation services to other organizations that need to confirm the accuracy of self-asserted identity and authorization attributes. Ultimately, the intent is to facilitate greater use of government data in a manner that preserves user privacy while also enabling increased equity by decreasing reliance on incomplete commercial data.

Keywords

attribute validation; attributes; digital identity; identity management; identity proofing; identity verification.

Reports on Computer Systems Technology

The Information Technology Laboratory (ITL) at the National Institute of Standards and Technology (NIST) promotes the U.S. economy and public welfare by providing technical leadership for the Nation's measurement and standards infrastructure. ITL develops tests, test methods, reference data, proof of concept implementations, and technical analyses to advance the development and productive use of information technology. ITL's responsibilities include the development of management, administrative, technical, and physical standards and guidelines for the cost-effective security and privacy of other than national security-related information in federal information systems.

Audience

The primary audience for this report is program and project managers who are interested in standing up attribute validation services for federal and other government agencies. Others may also find the contents of the report to be beneficial. Previous knowledge of attribute validation and attribute validation services is not a prerequisite for reading this report.

Call for Patent Claims

This public review includes a call for information on essential patent claims (claims whose use would be required for compliance with the guidance or requirements in this Information Technology Laboratory (ITL) draft publication). Such guidance and/or requirements may be directly stated in this ITL Publication or by reference to another publication. This call also includes disclosure, where known, of the existence of pending U.S. or foreign patent applications relating to this ITL draft publication and of any relevant unexpired U.S. or foreign patents.

ITL may require from the patent holder, or a party authorized to make assurances on its behalf, in written or electronic form, either:

a) assurance in the form of a general disclaimer to the effect that such party does not hold and does not currently intend holding any essential patent claim(s); or

b) assurance that a license to such essential patent claim(s) will be made available to applicants desiring to utilize the license for the purpose of complying with the guidance or requirements in this ITL draft publication either:

    i. under reasonable terms and conditions that are demonstrably free of any unfair discrimination; or

    ii. without compensation and under reasonable terms and conditions that are demonstrably free of any unfair discrimination.

Such assurance shall indicate that the patent holder (or third party authorized to make assurances on its behalf) will include in any documents transferring ownership of patents subject to the assurance, provisions sufficient to ensure that the commitments in the assurance are binding on the transferee, and that the transferee will similarly include appropriate provisions in the event of future transfers with the goal of binding each successor-in-interest.

The assurance shall also indicate that it is intended to be binding on successors-in-interest regardless of whether such provisions are included in the relevant transfer documents.

Such statements should be addressed to: Digital_Identity@nist.gov

Table of Contents

List of Tables

List of Figures

1. Introduction

An *attribute* is a "quality or characteristic ascribed to someone or something." [1] Attributes provide information about an individual that can be used to confirm the individual's identity or ability to access information or services. Attributes and the processes for validating and asserting them are essential for securely identifying individuals. They can also be utilized for online transactions — for example, determining eligibility based on state of residence, enabling granular and more reliable access control decisions, and supporting timely authorization decisions. The uses are nearly endless — from supporting security architectures such as zero trust to enabling more accessible and secure online benefit services. As a result, the processes by which attributes are used, validated, stored, transferred, and managed are increasingly important for scaled digital identity models.

1.1. Purpose and Scope

In support of the CHIPS and Science Act [2], this report provides a foundation upon which federal, state, and local government agencies can design and develop attribute validation services. Agencies with authoritative data are well-positioned to provide attribute validation services to other organizations that need to confirm the accuracy of self-asserted identity and authorization attributes. Ultimately, the intent is to facilitate greater use of government data in a manner that preserves user privacy while also enabling increased equity by providing access to a broader array of authoritative data sets.

The decision to build and enable attribute validation services is the responsibility of the agencies with data custodianship. While this report is intended to be helpful to agencies, it is not a comprehensive or normative document defining what must or must not be done. Instead, it provides a high-level overview of the space and its technologies and acts as a starting point for agency-specific implementation discussions, development, and business activities. Similarly, this report does not address all challenges that an agency may face. Legislative, regulatory, and other policy constraints may prevent an agency from providing the services as described, regardless of technical feasibility. Such challenges are organizational in nature, and they need to be addressed through non-technical means that are outside the purview of this report.

This report focuses on applying attribute validation services and architectures to support identity use cases, specifically identity proofing (data validation) and support for authorization decisions. However, the principles and considerations contained herein can support use cases beyond those explicitly addressed and may be adapted by readers to support their own needs.

1.2. Approach

This report provides an overview of the current and emerging environment, explores operational considerations for deciding how to build and manage a service, discusses data management strategies, and details three archetypes for attribute validation services: query/API-based models, brokered attribute hubs, and verified attribute models. For each of these, this document presents a generalized architecture and set of components as well as a set of considerations for how to secure the service and preserve user privacy in a standards-based manner.

The information for this report was developed through a structured market research and technical evaluation process. This began by canvassing current technologies and standards, researching real-world implementations, and interviewing providers and consumers of attribute validation services both within and outside of government. These engagements with ecosystem participants focused on both the state of the present — covering successes, limitations, and challenges — as well as the art of the possible, including emerging models, technologies, and standards. To preserve the privacy and intellectual property of those who participated in the market research interviews, their input has been anonymized and aggregated into the considerations reflected in the report.

2. Attribute Validation Service (AVS) Overview

Attribute validation services (AVSs) are not new and, in many cases, represent core government services that have existed for decades. In practice, however, these government systems have focused tightly on specific uses of the data related to core business operations, from validating Social Security numbers (SSNs) for payroll purposes to validating taxpayer identification numbers (TINs) to enable tax filing. Similarly, in the commercial sector, online services from different sectors have long leveraged AVSs provided by organizations with access to high-fidelity data, such as credit files, and with proprietary means to evaluate, process, and score vast amounts of data collected from open and closed sources.

This report does not attempt to determine whose services and data are more valuable or accurate. Instead, it focuses on lessons learned to provide organizations with a set of considerations for navigating a complicated ecosystem and providing high-value services to individuals and entities seeking reliable information to establish digital identities and support trusted, identity-based transactions. Furthermore, it attempts to set the stage for an ecosystem-wide set of capabilities that can provide the flexibility needed to promote user choice, consent, and interoperability of reliable identity and authorization attributes beyond today's constrained systems.

2.1. AVS Uses

An AVS is valuable because it reduces errors, inconsistencies, and fraudulent data by verifying that attributes conform to predefined rules, standards, and constraints and compares them against reliable data sets to confirm accuracy. This process is especially vital in identity proofing, where attributes such as names, dates of birth, and identification numbers must be accurate. As such, attributes and AVSs play a crucial role in many fields, with wide-ranging applications that promote data integrity, user experience, and security.

These services are regularly encountered across a wide array of high-risk interactions. Within the financial sector, banks and other institutions leverage such services to confirm the accuracy of critical data (e.g., SSNs) prior to account opening to deter and prevent fraud. In the healthcare sector, AVSs are leveraged to confirm critical identifiers such as e-prescribing numbers and to increase fidelity in patient identification and matching. In federal zero-trust architectures and access control systems, granular user attributes such as clearance level, time of access, and location can be compared against authoritative sources and policies in order to make access control decisions.

2.1.1. Identity Proofing

Identity proofing is the process of confirming, to a stated level of certainty, that individuals are who they claim to be for the purposes of establishing a digital identity. Essentially, it is the process of a user going from unknown to known through *identity evidence and attribute collection* (e.g., a driver's license or passport), *identity resolution* (whether we are talking about the correct person), *evidence validation* (whether it is genuine and not tampered with), *attribute validation* (whether attributes related to the person are accurate), and *user verification* (whether the person presenting this information is the true owner of the evidence and information). As indicated by the "attribute validation" step, services that can validate data about an individual, or that can validate the information presented on identity evidence, are essential to the overall confidence in the identity-proofing process. Increasing confidence in the attributes of an individual enrolling for a digital identity, and in the attributes contained in presented identity evidence, improves security by detecting potentially fraudulent data and increases the accuracy of collected data to ensure the right services are delivered to the right people at the right time. Table 1 provides examples of identity proofing attributes.

Table 1. Identity Proofing Attribute Examples

Attributes	Description
Name	Given name, family name, and often middle name (based on the needs for resolution or service provisioning) for the individual seeking to establish the digital identity.
Mailing or Physical Address	A physical location at which an individual can receive identity-related communications and is often used to verify identity out-of-band — for example, through delivery of a one-time enrollment code. Also helpful in verifying the user when a code is sent to an address strongly associated with the individual.
Government or Other Unique Identifier	A unique government identifier, such as a driver's license number or SSN, used to resolve the user to existing records and often to link associated records across systems.
Phone Number	A digital location to which communications can be delivered. Often used to verify identity, for example, through the delivery of a one-time-enrollment code to a number strongly associated with that individual. Also helpful in resolving the user.
Date of Birth (DOB)/Age	The date of the enrolling user's birth; used primarily for resolution of the user.

NIST's Digital Identity Guidelines, specifically NIST SP 800-63A: *Enrollment and Identity Proofing* [3], provide detailed requirements for collecting and validating attributes during the identity proofing process. They also provide characterizations of the evidence validation sources and their appropriateness for identity proofing. For more discussion of attribute usage in identity-proofing scenarios, refer to NIST SP 800-63A-4.

2.1.2. Authorization and Access Control

Authorization and access control encompass a system's ability to evaluate and determine whether a person or entity should have access to data, applications, or services. As pointed out in NIST SP 800-205, *Attribute Considerations for Access Control Systems* [4], "[v]irtually all authorization systems are dependent on attributes for rendering access control decisions and ultimately enforcing policy over subject access requests to system objects." Whether this attribute is a role issued to a user within an organization to support role-based access control (RBAC) or a fine-grained attribute associated with a specific access policy in an attribute-based access control (ABAC) model, it is critical to have accurate attributes validated with sources that can confirm their veracity in order to enable access control decisions that support intended security outcomes. Attributes commonly used in making access control decisions are listed in Table 2.

NIST SP 800-205 provides detailed considerations for the handling of attributes within access control contexts, while this report focuses on the ability to establish services that can support that document's intended outcomes. In particular, this report discusses external services that can augment enterprise systems — such as HR systems, entitlement stores, and access governance products — with additional attribute data to support or enrich access decisions.

Table 2. Authorization and Access Control Attribute Examples

Example	Description
Certification and Credentialing	An individual's specific claim of professional or organizational training status. This may be a technical certification (e.g., CISSP) or, more likely, certification of having completed training required for access (e.g., Security Training, Privacy Training, Rules of Behavior).
Clearance	An individual's clearance level within an organization or government context (e.g., Secret, Top Secret), which is compared against object classifications to determine access.
Employer or Entity Affiliation	The organization with which the user is associated. May be compared against object or system policies to enforce access to proprietary or company-sensitive data sets.
Location	Associated with a transaction; may be compared against access policies to determine access capabilities for remote users or to detect anomalous access attempts.
Role or Group	Assigned to an individual or group of individuals to define their role within an organization and, subsequently, the entitlements associated with holding that position. These can be general or more specific based on the complexity of the implementing organization.

2.1.3. Fraud Prevention

An outcome of the identity-proofing or authorization process is identifying and preventing fraudulent attempts to gain access to a system or service. This may include impersonation of a real person through the misuse or theft of identity evidence and information or use of a synthetic identity, which typically combines real information with newly created data to establish an identity that appears legitimate. While identity proofing and, in particular, attribute and evidence validation steps, go a long way to detecting when a synthetic or compromised identity is being used, basic attribute validation is often insufficient to address the full threat environment. To enhance fraud prevention, attributes not explicitly related to the natural person may be collected to aid in decision making. Attributes used in fraud prevention, such as the examples in Table 3, are often related to devices, historic transactions, online behavior, or a corpus of compromised data and can be used to identify possible anomalies that may indicate a potential bad actor. Having valid and accurate data improves user experience by preventing legitimate transactions from being delayed and improves security by preventing fraudulent transactions from being executed.

Table 3. Fraud Prevention Attribute Examples

Attribute	Description
Account Tenure	Typically associated with a digital or physical address; can indicate an attribute that may warrant further inspection, such as a phone number that is less than a week old.
Date of Death/Deceased Status	Indicates that users are no longer alive.
Device ID or Fingerprint	Generated by a service or organization to uniquely identify a single device on return interactions with a protected website or property. This is often compared against historical fraud records to determine if a single device is being used to commit fraud through multiple accounts.
Fraud, High Risk, or Blocklist Status	Such lists may be established by a diverse set of entities and indicate individuals or devices that have been associated with some indication of or actual bad behavior. Appearance on these lists may then be used to triage or block a transaction.
Location	The location from which a transaction originated. Not necessarily bound to the user; typically determined relative to IP addresses for the device initiating the transaction.
Risk Score	Generated relative to the user or the device; typically based on proprietary algorithms intended to evaluate transactional indicators of risk.

Since other identity-proofing attributes may be inputs to these services (e.g., submitting names and SSNs for a Date of Death check), the importance of accurate attributes is compounded. This makes it critical that only validated attributes, where available, be leveraged in seeking further signals and indicators of compromise. For more discussion regarding the use of fraud prevention attributes, refer to NIST SP 800-63A-4.

2.2. Current AVS Technologies and Standards

AVSs are indispensable tools that promote accuracy, security, and efficiency across a wide range of applications and industries. Table 4 provides examples of operational AVSs that solve discrete real-world problems today. Each of these services represents a spectrum of capabilities ranging from heavily manual legacy programs to more modern systems with automated processes and built-in onboarding services. Each has its own set of pros and cons, many of which are synthesized in this report.

Table 4. Operational Attribute Validation Services

Service	Provider	Description
Consent-Based SSN Verification (CBSV) Service	Social Security Administration (SSA)	With the consent of the SSN holder, CBSV can verify if the SSN holder's name, DOB, and SSN match SSA's records. Typically used by companies that provide banking and mortgage services, process credit checks, provide background checks, satisfy licensing requirements, etc.
Electronic Consent-Based SSN Verification (eCBSV) Service	SSA	Electronic service that offers registered members, such as banks, the ability to confirm the SSN, name, and DOB of their customers, with the customer's consent.
Social Security Number Verification System (SSNVS)	SSA	Application that allows employers and third-party representatives to verify employees' names, DOBs, and SSNs against SSA records.
Driver's License Data Verification (DLDV) Service	American Association of Motor Vehicle Administrators (AAMVA)	Provides commercial and government entities with the real-time capability to verify DL/ID information against data from the issuing agency.
E-Verify	SSA and U.S. Citizenship and Immigration Services (USCIS)	A web-based system through which employers electronically confirm the employment eligibility of their employees.
Income Verification Express Service (IVES)	Internal Revenue Service (IRS)	Allows designated entities within the mortgage ecosystem to retrieve tax transcripts and data to support mortgage decision-making.

Existing AVSs typically take the form of query-based systems that make use of APIs or custom integrations to request and exchange information between RPs, AVSs, and the end user.

The following is a typical workflow for such a service:

1. User navigates to the RP's application (e.g., a registration page)
2. User inputs attributes (e.g., name, DOB, address)
3. RP application packages these attributes into a payload
4. RP conveys the attribute fields and values to the AVS via an API or custom integration
5. AVS compares the data to its records
6. AVS conveys a response to the RP (e.g., yes/no or specific attribute values)

The authentication and authorization of API calls are often — but not always — protected using protocols such as OpenID Connect, OAuth, and SAML.

There are numerous benefits to this approach. First, it requires minimal infrastructure changes for AVS providers since existing components or services can be used, with only the need to develop and maintain external APIs or connections. Second, it uses existing, common deployment patterns for online services such as APIs and common access and authorization standards. However, there are also vast disparities in the way these services are deployed, resulting in a lack of standardization in the matching algorithms and APIs, and substantial inconsistencies in how they are protected.

Two other models related to AVSs are brokered models and Public Key Directories (PKDs). A *brokered AVS* allows a single broker to integrate with multiple AVSs through a "hub and spoke" model where the RP application sends its attribute queries to the broker, who then parses and distributes them to the appropriate AVS. Such services ease integration for AVSs by limiting the number of endpoints they need to interact with. Like traditional query-based systems, hubs typically rely on common patterns (APIs) and standards such as OpenID Connect and OAuth to manage access to the APIs and data.

In some instances, AVSs do not validate the attributes themselves. Instead, they provide cryptographic means by which an RP can confirm the accuracy and integrity of attribute data. The RP receives a payload signed by an AVS using public key cryptography. The AVS then makes its public key available to RPs through a PKD. RPs, in turn, download the key to verify signatures on signed attribute bundles from the AVS, confirming their accuracy and integrity before leveraging them in business processes. The PKD also often provides trust services on top of a key distribution role by ensuring that participants follow common standards, protocols, and business processes. AVSs could also provide their public keys to support validation directly to RPs without a third party playing this role. For the purposes of this discussion, they would also be considered PKD AVSs.

PKD services are less common today than query-based models, although excellent examples exist such as the International Civil Aviation Organization (ICAO) PKD, which provides public key services for over 200 national e-Passports. That said, they have much more in common with emerging approaches to attribute validation than more traditional models.

2.3. Emerging AVS Technologies and Standards

Enter the digital wallet.

Emerging digital identity models are rapidly converging on the ability to prove identity and other attributes through cryptographically protected attributes in an individual's digital wallet. The two most popular forms of this are Mobile Driver's Licenses (an ISO-standardized digital representation of the physical card and its associated data, which can be used for any type of credential) and Verifiable Credentials (a W3C-defined data model). For the purposes of this paper, we will refer to them collectively as *User-Controlled Verified Attributes* (UCVAs). Essentially, these are attributes that are signed by the issuing source using public-key cryptography to ensure the integrity and accuracy of the data when asserted to an RP and are issued to the user described by those attributes. This is similar to the signed data elements on e-Passports that can be validated using the ICAO PKD. In fact, most architectures that support UCVAs will have a PKD (or similar service) to help manage and distribute keys at scale. The difference is that these signed attribute bundles reside on a device *and* in an application controlled by the user.

The benefits of these emerging systems are twofold. First, users are given greater control over their personal data, allowing them to present and assert their information when and where they want. The second benefit is that the data is signed by the issuer at the time of issuance, preserving the integrity and, in many cases, the accuracy of the attributes. However, these models place a substantial burden on the issuing source to provide the technical infrastructure for signing, distributing, and protecting keys — a role they do not often play today — and on the business processes to securely manage the enrollment of users and the issuance of the verified attributes to user-controlled devices. There are additional post-issuance concerns that will need to be addressed, such as how to manage reports of compromised UCVAs and how to prevent unauthorized RPs from accessing them.

3. Validation Logic

To increase confidence in identity proofing results or authorization decisions, it is crucial to validate self-asserted attributes by comparing them against authoritative data sets. This process involves several key roles: the relying party (RP) that requests attribute validations, the end user whose attributes need to be verified, and the AVS that performs the validation. The algorithms the AVS uses for attribute matching and the responses they generate must be carefully designed to meet the needs of all parties involved while complying with statutory and regulatory requirements. Here, we explore the complexities and challenges of attribute matching and provide options for balancing accuracy, usability, and privacy.

The simplest form of validation logic determines whether the authoritative attributes exactly match the string provided and then returns only a yes/no response. However, this simplistic approach can result in unacceptably high false negative rates and rarely meets the needs of RPs or users. The addition of simple fuzzy matching, such as algorithms that use Levenshtein distance [5], accounting for common typos, or matching only on the first few letters of a name or street address, can reduce some false negatives but can also introduce risk if not done carefully and transparently. Simple matching algorithms can also have adverse equity impacts, particularly for members of cultures who do not follow the typical U.S. first-middle-last name pattern. A significant percentage of name mismatches are not due to fraud but rather are the result of input typos, unreported name changes, use of a nickname, and other inconsistencies that, though harmless, could lead to low double-digit mismatch rates [6].

The AVS will have to understand the requirements of the anticipated RPs as well as their end users to design matching algorithms that meet their needs. The matching requirements of RPs will vary depending on their use cases and risk tolerance. For example, one RP may require a precise match on both the unique identifier and DOB, while another may find fuzzy matching on DOB acceptable. Ideally, the AVS will provide RPs with the option to pass flags at the attribute level to indicate whether a precise match is required, or whether fuzzy matching is acceptable.

The AVS can provide further value to its customers by closely monitoring the impact of their fuzzy matching logic on both false negatives and false positives by providing feedback mechanisms for RPs and end users and analyzing the responses over time. This will allow an AVS to understand the approximate percentage of false positive and false negative results that their matching algorithm generates for a given population, which can be used as feedback to improve their algorithms and can allow RPs a greater understanding of the risks associated with the matching service.

3.1. Names

In general, name matching is problematic, so carefully designing fuzzy matching algorithms and user input fields for names can be particularly useful. Common challenges with name matching include:

- **Nicknames:** Some use cases may require strict matching on given names, while others may allow the use of nicknames. RPs should be able to set a flag indicating whether nicknames are allowed. If nicknames are allowed, it is best to use a flexible datastore for nicknames that can be updated. If a nickname was matched, consider returning an indicator to the RP that the match was on a nickname, even if the nickname was flagged as allowable.

- **Name Changes:** Name changes are especially common with changes in marital status, but individuals may continue to use both their marital and birth names, depending on the context. Whether a match should be allowed on a previous name depends on the use case, risk tolerance, and whether previous names are maintained in the data source.

- **Long Names:** Some names are so long that they become truncated in databases and on official documents [7]. Since official documents have different character length restrictions, the surname can vary among authoritative sources[1]. Individuals may provide their full name or a truncated version from a document.

- **Compound Names:** Knowing which name to provide for a particular validation service can be challenging for individuals with compound names. For example, the famous artist Salvador Dalí's full name was Salvador Domingo Felipe Jacinto Dalí i Domènech[2], which could be stored in a variety of ways.

  Compound surnames are common and can follow several patterns that make matching challenging. In Spanish-speaking countries, it is often traditional to have two surnames, one from each parent, and these names can include a coordinating conjunction. Some databases may store both names together in the surname field, some may store the first surname in the middle name field, and some may drop one or the other surname altogether. The compounding conjunction may be present or could have been dropped. Dutch surnames traditionally have prefixes. Those prefixes can end up partially or entirely affixed to the name, can be distributed across the middle and surname fields, or can be dropped altogether. For example, the surname Van Der Hof could be stored as Van Der Hof, Vanderhof, Der Hof, or Hof. Hyphenated surnames are increasingly common but may be stored in a database without the hyphen, with a space instead of a hyphen, or with only the first or second part of the surname.

- **Diacritical Marks:** Diacritical marks can be allowed in the user interface but can be removed for matching purposes. Examples include the caron (ˇ), tilde (~), umlaut (¨), and cedilla (¸).

- **Romanized Names:** The Latin, or Roman, alphabet used in English is only one of over a hundred scripts currently in use [8], and thirty-two scripts have over a million users each

[1] U.S. passports limit given names to 24 characters but do not limit surnames. SSA limits given names and surnames to 26 characters each. https://mh.usembassy.gov/wp-content/uploads/sites/83/ds11.pdf, https://secure.ssa.gov/poms.nsf/lnx/0110205120
[2] https://www.rem.routledge.com/articles/dali-i-domenech-salvador-domingo-felipe-jacinto-1904-1989

[9]. The romanization of names from other scripts is an inexact science that leads to inconsistent translations and spellings[3].

- **Surname First or Absent:** It is common in Asian cultures [10] for the surname to be placed first and, in some cases, it is altogether absent [11]. When absent, the given mononym may be stored in the surname field. A suggested user interface that accommodates a variety of naming conventions is to allow entries in two fields: [Given Name(s)] [Surname(s)]. The length of each field should be long enough to capture multiple given names and surnames as well as the long names common in some cultures.

3.2. Dates

Most of the world uses the Day/Month/Year format, so the Month/Day/Year format common in the U.S. can lead to attribute validation challenges, particularly for birthdates. User input fields for dates should be easily usable by both U.S. and international populations. When the AVS does not control the user interface, RPs may benefit from being given the option to tolerate the transposition of the month and day.

Leap years can present additional issues. Some individuals have a recorded DOB of February 29 during a non-leap year. Since modern databases will prohibit entering a date of February 29 on a non-leap year, an individual may provide one of three days for their DOB: February 28, February 29, or March 1. Therefore, attribute validation sources should consider allowing fuzzy matching for birthdays in this range.

3.3. Addresses

Addresses can include postal addresses, email, and phone numbers. Address validation is often used during identity proofing but presents several challenges. Individuals can have multiple addresses, addresses are subject to change, and there is no authoritative source for any type of personal or business address. Also, since addresses in the U.S. are associated with names rather than unique identifiers, an individual's records can easily become contaminated with address information for individuals with the same or similar names, either accidentally or purposefully. Bad actors will often use change of address mechanisms to add addresses they control to the records of individuals whose identities they have stolen.

If an AVS is performing address validation, additional vigilance is required due to the potential for malicious address injections into its records. An AVS can assist RPs by including metadata in the response such as the date the address was last modified and the original source of the address (e.g., driver's license verification, commercial data broker query, or self-asserted data).

[3] For example, the Arabic writing system often omits vowels, and it contains sounds that cannot be represented using the Latin alphabet used in English. https://www.academia.edu/82526032/Transliteration_of_Arabic_Names. Consistently translating Arabic names to the Latin alphabet is an area of ongoing research. https://thescipub.com/pdf/jcssp.2021.776.788.pdf

3.4. Transparency, Risk, and Trust

RPs can only manage risks of which they are aware. So, by providing RPs with complete and accurate information about the risks associated with a validation service, the AVS provider improves trust and provides RPs the ability to better control risk and improve the experience of their users.

Many authoritative attribute sources will contain errors. This is especially true if manual data entry has been involved, translation from a non-Latin alphabet has been performed, or identifiers intended to be unique were issued in a decentralized manner.

The use of fuzzy matching algorithms can hide such errors, so to improve both data quality and trust, AVS providers should consider informing RPs when fuzzy matching was required to obtain a match for a particular attribute, otherwise errors in the data source may go undetected. When errors are discovered, the AVS provider should consider providing redress options so the data source can be corrected. Redress should be carefully designed to reduce the risk that an imposter does not subvert the redress process. Estimated error rates in the data source should be tracked, and if sufficient transparency and opportunities for secure redress are provided, data quality should improve over time.

To further reduce risk, the service should consider implementing controls that ensure that its matching logic cannot be used to reconstitute partial attributes, as well as controls that can detect patterns indicative of an attempt to verify stolen data.

3.5. Responses

A global "no match" response rarely meets the needs of RPs, so when legally permissible, AVSs should consider providing matches at the attribute or field level in those cases where the AVS has confidence that the RP is using the service appropriately. Granular responses can improve usability, reduce risk for RPs, and improve data quality over time when combined with secure redress methods that allow errors in data to be corrected. At the same time, granular responses can increase certain risks for the AVS, particularly if either an RP or one of its end users is attempting to abuse the service to validate stolen PII. To mitigate that risk, RPs should be carefully vetted, and the user agreement between the RP and the AVS should prohibit the RP from passing along field-level responses and matching indicators to its end users. The additional risk associated with providing granular responses can be further mitigated by increased access controls, adding controls that analyze request and response patterns, and prohibiting repeated attempts to submit information for the same person with slight variations.

For "Yes" responses, it is useful to provide an indication of the degree of the match — whether the match was exact or near exact (single character error), or if fuzzy matching was required to match an attribute. This information is needed to reduce risk for the RP and can result in improved data quality even when an RP explicitly indicated that fuzzy matching was acceptable for a particular attribute.

Finally, when an AVS is asked to validate information pertaining to an individual in their records who is deceased, they should strongly consider returning a death indicator. Data for recently deceased individuals can be highly vulnerable to identity theft [12][13].

3.6. Derived Attribute Values

Derived attributes avoid the transfer of PII, improving privacy and security. Support should, therefore, be provided for derived attributes whenever the full attribute is not required. For example, if an RP does not require a full DOB and only needs to know whether a user is over a particular age, the SP could support derived attributes such as “IsOver18:yes” or “IsOver18:no.” Similarly, an RP may need to know whether someone is married (or not) but may not need to know the spouse’s name. In that case, support for an “Is Married” attribute could be provided.

4. Data Management

Data management involves the "development, execution, and supervision of plans, policies, programs, and practices that deliver, control, project, and enhance the value of data and information assets throughout their life cycles." [14] This process is essential for maintaining data quality and integrity, especially in the large and complex systems found throughout the public sector. To help guide project managers and developers, this report discusses some key aspects related to data management in the context of attribute validation within federal systems. By addressing these key issues, an AVS can make significant strides toward maintaining high-quality data in an efficient manner, which is essential for informed decision-making, regulatory compliance, and overall system reliability. For more details on data management, see the DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK).[4]

4.1. Origination and Sources

Attributes derive from a variety of sources, both direct and indirect. Some attributes are inherent while others are randomly assigned or assigned according to proprietary formulas. A federal AVS may generate, maintain, and/or process the official source of truth for consumers, data brokers, and third parties. The federal government sometimes relies on credit agencies and third parties to validate or augment its own data holdings. One notable source, for example, is credit bureaus that collect and maintain a wide variety of data on hundreds of millions of individuals and assign consumer credit scores.

Agencies manage dozens and sometimes hundreds of systems and applications. A system functioning as an AVS may be considered a *system of records*[5] for a particular kind of federal record, or a system could support an agency mission and become the de facto system of record even though it has not been officially designated as such. The same information may reside in multiple federal systems and may be shared among agencies, with law enforcement as part of an investigation, or with other third parties in accordance with the system of records notice (SORN) associated with the system. The point to note here is that the federal government may be the overall source of much data, but the same PII elements (e.g., address, telephone number) may reside in multiple databases at multiple agencies (e.g., VA, FHA). At one or more locations, the data may be stale or inaccurate depending on when an individual last used agency benefits or accessed services (e.g., VA health care or a home loan). This highlights the need for metadata that accurately reflects information such as the date the information was captured and the source of the information.

To better understand how and where to locate data within an enterprise and how data is collected, stored, accessed, and used, organizations should consider conducting a data inventory to systematically catalog their data assets. Agencies should determine whether there are specific datasets within their inventory that are more accurate, better managed, or more

[4] Earley, S, et al. (2017) The DAMA Guide to the Data Management Body of Knowledge. Bradley Beach, NJ: Technics Publications.
[5] A system of records is a "group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual." Source: Privacy Act of 1974, as amended (5 USC 552a(5))

easily accessible. A data inventory brings situational awareness and clarity to organizations that would otherwise struggle to navigate data residing in a variety of data management systems spread across multiple offices and regions (and likely in different formats). Perhaps data was originally collected for a different purpose but now the agency would like to use it for an AVS. This may require obtaining additional consent from affected individuals as well as updates to the public notices. Agencies should also budget for any costs associated with the repurposing of existing data.

Every AVS data source contributes unique pieces of information, which are cross-referenced and validated and contribute to the goal of creating a comprehensive verified profile of an individual's identity, thereby minimizing the risk of fraud and enhancing trust in digital transactions.

4.2. Quality

High-quality data and trust go hand in hand, and data's availability and proper use instill confidence in providers and consumers alike. Some of the characteristics exhibited by high-quality data include accuracy, completeness, consistency, and currency. *Accuracy* is the correctness of the data content as compared to an agreed-upon and accessible authoritative reference source. *Completeness* measures values in the fields (fill rate). *Consistency* is achieved when data is uniform and coherent across various databases, systems, and applications. Data and information should also be *current* and ready for use as specified and within an anticipated timeframe.

Unlike most resources, today's digital data is easily replicated yet persists — even after multiple uses. This highlights the importance of quality data, so that "bad" data is not perpetuated, which results in higher costs as well as higher frustration levels [15]. There are two basic approaches to improved data quality: error prevention, or error detection followed by correction. Error prevention is closely associated with the processes of data acquisition and data entry. While many organizations have undergone process improvements, errors in large data sets are still common [16] and should be anticipated.

A case study in process improvement, SSA has been issuing SSNs for decades, as technology has shifted from typewriters to punch cards to databases. The further back in time one goes on the technology implementation scale, the greater the likelihood that errors exist. Like other agencies and departments, SSA operates on the scale of hundreds of millions of identities, so the potential for error is high although miniscule in relative terms. Potential remedies include an individual contacting the agency, where processes are in place to correct errant records. If federal agencies rely on data from states' bureaus of vital records or other non-federal entities, errors should first be corrected at the source and then updated at the federal level.

Traditionally, data quality has been managed as close as possible to the source, but this is becoming increasingly difficult. This can shift the burden of data quality management to data consumers since relying on data producers to supply data of adequate quality may not be practical [17]. A public sector model would ensure high data quality without the associated cost pressures that could otherwise result in data management burdens for the public.

Data quality is best viewed from the perspective of RPs (those using the data), because they will judge whether a product is fit for use. Errors may occur due to delays in processing times, lengthy correction times, or insufficiently stringent data edits [18]. For federal Privacy Act systems, individuals may seek redress in several ways, to include contacting the system owner listed in the system SORN, submitting a FOIA or Privacy Act request, requesting assistance from their member of Congress, or filing a civil rights lawsuit. While redress needs to be accessible enough that legitimate users can correct errors in their records, organizations should keep in mind the potential for abuse by impersonators seeking to contaminate records through redress mechanisms.

Agencies should consider defining and establishing clear data quality standards for each attribute within a federal system. These standards should support accuracy, completeness, consistency, and currency. Having well-defined standards helps in setting expectations and guidelines for attribute validation. Comprehensive validation rules and checks help enforce existing data quality standards. Such rules can include format checks, range validations, referential integrity checks, and other business-specific rules. An AV service may be designed to apply these rules systematically across all relevant data attributes.

4.3. Refresh and Maintenance

Refreshing data means importing data from the original data source based on a refresh schedule or on demand. Following a data refresh, previously cached query results may no longer be valid. After a refresh schedule has been established, notifications should be set up that go out to multiple individuals at each RP (perhaps via a distribution list) to avoid a single point of failure. It is best to schedule a refresh during less busy time periods, to keep refresh limits in mind, and to verify that refresh time does not exceed maximum refresh duration [19]. The refresh process should be documented, communicated, and well understood by both the AVS and RPs for reasons of both accuracy and service availability.

Data maintenance is the ongoing process of collecting and organizing data in a way that is accessible and useful to an organization. The process ensures that organizations retain high-quality data and can make better decisions as a result. Maintaining high-quality data requires motivation, knowledgeable personnel, a willingness to make difficult decisions, and sustained funding. The same data may reside in multiple locations, but often no one has the clear authority or the willingness to delete duplicates, so they persist and proliferate. If multiple copies of data exist within an agency — especially if some have been modified — it is critical to know where to go to find the “original.” This issue may be exacerbated by a lack of associated metadata.

Unique identifiers used for AVSs can exhibit various levels of persistence, and attributes may require different rates of refresh. For example, setting aside hospital record-keeping errors, date of birth in the U.S. is extremely persistent. Portability has also allowed cell phone numbers to become increasingly persistent, thereby increasing their value as unique identifiers. In contrast, postal addresses can change relatively frequently, especially for renters and young adults.

Regularly refreshing data helps maintain data integrity by identifying and addressing any inconsistencies or data errors. Refreshed data allows for accurate and meaningful data-driven insights, resulting in a more informed business strategy. Data maintenance improves overall data quality and reliability, enhances the accessibility and usability of data, reduces redundancy and inconsistency, improves data privacy and security, and helps optimize storage.

4.4. Storage and Security

Data is retained for various lengths of time depending on the reason(s) for its collection, the agency mission, and the applicable federal compliance requirements. In most cases, federal system proponents work in conjunction with the National Archives and Records Administration (NARA) to develop a records retention schedule for federal records contained in each system. There are several issues to consider, such as whether data and information in each system qualifies as a federal record. A second consideration is how records will be tagged or identified for disposal after the end of the approved retention period.

It is the responsibility of system and business owners to ensure that sensitive data is protected and that access to the validation processes is appropriately controlled. Compliance with federal statutes such as the Federal Information Security Modernization Act (FISMA) [20] and regulations related to data security and privacy should be a top priority. The NIST Risk Management Framework (RMF) provides a flexible, holistic, and repeatable multistep process to manage security and privacy risk, and it links to a suite of NIST standards and guidelines to support the implementation of risk management programs to meet FISMA requirements [21].

Federal information security programs are responsible for protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, and destruction and to ensure the confidentiality, integrity, and availability of federal data. Federal systems will establish or inherit many of the controls presented in NIST SP 800-53, *Security and Privacy Controls for Information Systems and Organizations* [22]. For example, it is important to implement comprehensive audit trails and logging mechanisms to track changes and activities related to attribute validation. This helps in monitoring the effectiveness of the validation processes, identifying patterns of data quality issues, and facilitating compliance with audit requirements. FedRAMP also uses NIST guidelines and procedures to provide standardized security requirements for cloud service offerings [23]. Taken together, the referenced security frameworks and measures create a multi-layered defense strategy, fortifying the digital infrastructure against a spectrum of cyber threats and bolstering the overall security posture of organizations.

4.5. Metadata

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, and manage an information resource. It is often referred to as data about information (or information about information) [24], and it describes the content, quality, condition, and other characteristics of data while facilitating many functions associated with data such as organization and management, long-term preservation, indexing and discovery, and retention [25].

Metadata covers data elements that pertain to information carriers as well as those that pertain to the information (content) itself. It can, among other things, help confirm the existence of information and support effective access to information resources. Metadata records follow a standard format that enables operability [26], and producing effective metadata involves using appropriate values to record correct and carefully considered elements [27]. For additional guidance and consideration in this area, NIST has released a report [24] containing a metadata schema for attributes asserted about an individual during online transactions.

Especially relevant for this report, metadata provides information pertaining to the freshness, sourcing, and confidence level for third-party attributes. Indicators allow an RP to determine if the underlying data is trustworthy and whether the verification should be refreshed. In other words, data quality metadata answers the question: “Is this data of sufficient quality for me to use it for a specific purpose?” [28]

Attribute service providers should develop and implement a metadata schema to support their RP with associated decision making. Metadata requirements will vary depending on the AVS architecture used and the attributes verified.

5. Deciding Whether to Establish an AVS

There are many factors to consider when establishing an AVS, such as the time, knowledge, and resources required to effectively scale the offering as well as the legal authorities required. The attribute validation function sits at the center of several key processes and technologies essential for enabling trust online, most notably identity proofing, which considers resolution, validation, and verification. The validation of identity attributes relies upon the cooperation of an issuing authoritative or credible source that acts as a steward of identifying information. Generally, an authoritative source has the most complete dataset for a given attribute, such as a driver's license number, date of birth, or SSN.

5.1. Attribute Sources

Before planning to establish an attribute service, it is critical that an agency determine its relationship with the data it intends to offer. This relationship defines the type of “source” that may be considered for each attribute it offers. This document considers three types of sources, each with impacts on the degree to which an external entity or RP may wish to trust the provided attributes:

- **Issuing Source**: The organization is the original source of the attribute’s value. For example, a DMV is the issuer within its jurisdiction of the driver’s license number issued to a uniquely identified individual.
- **Authoritative Source**: The organization has a regulated business process for collecting, validating, and maintaining attributes for which it is not the issuer/originator. For example, the DMV that issues a driver’s license number may be considered authoritative for a mailing address, given its need to maintain an accurate location to communicate with an individual effectively. While it does not generate this attribute, it has established regulated processes to validate the information for its business needs. An authoritative source may also have direct access to issuer records.
- **Credible Source**: The organization has a defined business process for collecting, validating, and maintaining attributes, has directly received data from an authoritative or issuing source, or has established processes to gather and correlate data from multiple sources. For example, a data aggregator may leverage public records and purchased data sets to correlate an individual’s name and physical address to provide an RP with an “address” attribute that meets a defined confidence level for that attribute.

As the examples indicate, a single source will have different relationships to different attributes, acting as the issuer for some and an authoritative or credible source for others. While the decision regarding what types of sources are acceptable for specific use cases ultimately resides with the RP, it is important for the AVS provider to determine its relationship to each of its offered data elements and effectively convey that information to the consuming entity. This is often done through trust agreements but may be supported in greater detail and at runtime through attribute metadata, which is discussed in more detail in Section 4.5.

It is also important to note that, while these source types describe the relationship between the source and the data, they do not presuppose or indicate accuracy. While it could be reasonably assumed that the closer to the issuer one is, the more accurate the data will be, this is not always the case. Data management, testing, rationalization, fraud prevention techniques, and matching algorithms can all contribute to the accuracy of services, resulting in variances in the performance of AVSs regardless of their relationship to specific data elements. For example, a credible source pulling from commercial retail data sets — which would not likely be considered authoritative — may have more accurate address information for a user than a Department of Motor Vehicles (a source that would likely be considered authoritative for this data). Therefore, while attribute source types are helpful for characterizing services at a high level, they should be accompanied by due diligence and testing to inform the most viable path to accuracy for specific attributes.

While government agencies are often an issuing or authoritative source of identity data, they seldom provide AVSs and, when they do, are often severely constrained in terms of to whom they can provide such services. For example, SSA validates the name/DOB/SSN combination for financial institutions through the eCBSV program, but — as of the time of this writing — these validation services are not available for identity proofing in citizen-facing applications.

5.2. Mission, Authorities, and Legal Environment

A proper understanding of the legal and regulatory environment is necessary to establish and effectively operate an AVS. For example, sharing prohibitions, constraints, and mandates will dictate or at least influence what can and cannot be shared and how that sharing may or may not occur. The agency mission and authorities must align and may need to support a broader agency or national framework, strategy, or plan. Does the service have access restrictions on some individuals or entities? Does the information fall into a special category (e.g., tax, immigration) requiring additional protections? What information could be accessed and under what conditions? Are there other sector- or jurisdiction-specific legal and regulatory requirements affecting either the offering or its customers? An agency will not know if they can offer a particular service until these critical questions are answered.

5.3. Governance, Buy-In, and Service Demand

Securing buy-in from organizational leadership, appropriators, consumers, communities affected by the offering, and other relevant parties is critical. Given the sensitive nature of identity proofing and, therefore, attribute validation, a well-defined multidisciplinary governance model is needed for the health and success of the service offering. Absent a formal structured governance process, organizational leaders must handle governance decisions on an ad hoc basis, but such decisions may be at odds with broader organizational goals [29]. Governance requires commitment at a strategic level, involves personnel at multiple levels of an enterprise, and encapsulates governing structure, leadership, processes, and relational mechanisms to address performance while providing assurances that information is sufficiently protected from threats [30]. In a digital identity risk management context, risk factors include, but are not limited to, information security, privacy, equity, usability, and legal and regulatory requirements. It is important for risk management efforts, including those whose scope includes AVSs, to weigh these factors as they relate not only to enterprise assets and operations but also to individuals, other organizations, and society more broadly.

One of the Cybersecurity Framework Core Functions is *Govern (GV)*, which includes organizational context, roles, responsibilities, authorities, policy, and the establishment of cyber strategy and supply chain risk mitigation [31]. Minimally, the governance model should account for the functions of the service offering, define who is responsible for which functions, and document these items through policies, plans, and procedures that are communicated clearly across the organization. The model should also consider the role of the service's customers and others with equities in the decision-making process, and it should specifically note how their feedback will be requested and collected. In some cases, an existing governance body, or a combination of bodies, might already include in its scope matters pertinent to an AVS offering — for example, an agency identity and access management council, a data governance working group, or an external advisory committee. In other cases, it may be necessary to establish a new multidisciplinary governance body or some other mechanism for consultation and feedback. As a first step or interim governance model, consider establishing a multidisciplinary steering committee, perhaps positioned under the CIO, or modeling the effective governance structure of a partner agency.

A high degree of buy-in can be achieved with the assistance of leadership at all levels, a strategic communications plan, and consistent interagency messaging. If champions within the organization come forward or are otherwise identified, they can be trained and leveraged throughout the organization and perhaps be integrated into the emerging governance structure. For example, many governance bodies have non-voting members or observers who possess a particular expertise or are just excited about the project and eager to assist. Several NIST publications, including the Cybersecurity Framework, the Privacy Framework, and the Digital Identity Guidelines, call out the need for strong governance processes and offer further guidance on this topic.

In addition to governance and buy-in, organizations should consider the demand for an AVS. Service demand can be organic, expressed as a groundswell of support from a large number of constituents. However, demand drivers usually derive from new legal or policy mandates or from a shift in leadership priorities that are supported by existing authorities. Organizations should consider the specific gap or opportunity that an attribute validation service will address within the identity ecosystem and whether there are already services offered that could address those same needs. Understanding the existing market, including the landscape of complementary or substitute services already offered, could not only inform the organization's decision whether to develop an offering or not, but could also support the identification of specific features of the offering that would differentiate it from existing alternatives.

An essential part of understanding demand is understanding who is asking for the service and their motivations for doing so. Different customer segments have different needs and, therefore, might advocate for a wide variety of business requirements. Determining what these minimum requirements are can then inform whether and how a service offering should be pursued. To accommodate initial service demands, the U.S. Digital Services Playbook [32] recommends building a service using agile and iterative practices, structuring budgets and contracts to support flexible delivery, deploying in a flexible hosting environment, and relying on data to drive decision making.

5.4. Anticipated Impact

In many cases, the user population of an AVS consists of RPs or intermediary service providers operating on behalf of RPs rather than the individuals to whom the attributes relate, such as when a financial institution contracts with a third-party service provider to verify the identities of individuals applying for checking accounts. Therefore, when estimating the impact of an AVS and evaluating its actual impact, several audiences should be acknowledged, and the anticipated impact on them should be considered separately. By separately evaluating the potentially affected populations, and by considering the impact to individuals whose personally identifiable information is being processed, a richer, more comprehensive understanding of the service's potential reach, role, benefits, and risks can be brought to light to inform the decision whether to instantiate a service.

Several dimensions of impact can be considered across potentially affected populations, as well as to the broader identity ecosystem and to the government organization that is considering providing the service. For example, the service's impact could be estimated and assessed based on the following:

- Identity proofing process outcomes and performance (e.g., accuracy, timeliness, cost-effectiveness).
- Improved accuracy of authorization decisions.
- Type(s) and amount of fraud that the service is expected to address.
- Extent to which the service model promotes an approach to identity verification that improves protections for individuals' privacy and civil liberties.
- Secondary risks of offering the service, such as creating a single point of failure in the market, in the case of a service that outcompetes commercial alternatives.
- Secondary benefits of offering the service, for example, those associated with positive identity proofing process outcomes (e.g., improved, faster, or broader access to other essential services); or
- Potential for the AVS to expand digital services to end-users.

5.5. Privacy, Notice, and Consent for End Users

With personal data constantly being collected, analyzed, and shared, it can be important to end users to understand how their data is being used by an AVS. In an ideal world, consumers would be given the choice to provide consent or denial for particular uses of their personal data (or instances of personal data use). Ensuring privacy, obtaining informed consent, and providing clear notice not only respects the rights of individuals but also fosters trust between the government and citizens.

The notion of privacy protection has expanded from mere control over data flows to encompassing issues of autonomy, protection from bias, and the view of data holders as data fiduciaries with a legal obligation to act in the best interest of others. In the context of data privacy, consent is intended to allow certain data practices that would otherwise be off-limits. However, the way consent is currently obtained is often weak or unclear, starting from the moment data is collected. End users may be only vaguely aware of the extent of data about them that is regularly collected. Obtained consent follows the data as it moves among various parties, for example from a mobile app developer to a data broker to an advertiser, so it is important that consent is clear to users. The existing consent paradigm does not work in favor of users.

To address this AVSs should provide clear privacy notices and obtain proper consent. Providing clear privacy notices enables users to knowingly agree to an organization's intended purposes. Without clear communication and consent, users may unknowingly allow their personal data to be used in ways they do not intend. It is important to be transparent about how data is to be used, and to ensure that it is not used in ways that the user did not provide consent for, used in ways that exceed the user's expectations, or shared with additional parties without the user's informed consent [33]. Properly informing users and gaining explicit consent ensures that data is handled responsibly, especially as it moves through different services and organizations for validation.

5.6. Key Questions for Agencies

Agencies considering whether they should attempt to design and offer an AVS can ask a few key questions to determine what role, if any, they might play, for instance:

Table 5. - Key Questions for Agencies

Factor	Questions
Attribute Sources	1. Does my agency serve as an authoritative or issuing source? a. If so, for which attributes? b. Is my agency the only authoritative source for a particular attribute type? i. If not, does another authoritative source already provide an AVS?
Mission, Authorities, and Legal Environment	1. Has my agency been granted the requisite authority to offer an AVS? a. If not, why not? b. Is trying to obtain the requisite authority appropriate, given my agency's core mission and anticipated ability to deliver? 2. What is the state of the international, national, and sub-national policy environment on relevant topics such as privacy, cybersecurity incident reporting, data sharing, and protection? How would those policies impact an AVS?

Factor	**Questions**
Governance, Buy-In, and Service Demand	1. Is there a demand for an AVS? If so, what gap, challenge, or opportunity would the service address? Why has demand not been fully addressed? 2. How saturated is the market? a. Who is already competing to meet the demand? b. Are other organizations already offering validation services for the same attributes? Are complementary or substitute services currently offered? c. What factors are contributing to the current market saturation status? 3. Does my agency have buy-in from leadership, appropriators, and other relevant parties to pursue an AVS? 4. Who is asking for the service? What customer segments would be served? What requirements and limitations do potential customers have? Can my agency effectively address them?
Anticipated Impact	1. What is the service's intended impact on the portion of the population that it would serve? a. What percentage of people struggling with identity proofing are expected to benefit from the service? Are there anticipated secondary benefits (e.g., improved access to other services)? b. What kinds of fraud might the service address, and how does the service perform compared to other approaches? 2. What factors will affect whether this intended impact is delivered? What are some unintended (positive or negative) consequences to anticipate should the service be launched?
Privacy, Notice, and Consent for End Users	1. Is my system a Privacy Act system? If so, is it covered under an existing SORN(s), or do staff attorneys in conjunction with the system owner need to prepare one for publication in the Federal Register? 2. Is my federal system subject to any other jurisdictional legal requirements (e.g., international, state, or local)? If so, does the system meet them? Requirements could include mandatory periodic reporting, additional notice, enhanced consent, deletion upon request, or a partial to total ban. 3. Is someone at my agency monitoring privacy trends and working with developers to ensure compliance with existing mandates and best practices related to, for example, meaningful notice and active consent? This is especially relevant for upstream and downstream systems where consent may be nonexistent or implied.

6. Considerations for Designing and Deploying an AVS

Once a decision has been made to move forward with planning to deploy an attribute validation service, several critical decisions and operational factors require consideration.

6.1. Existing Capabilities

Agencies that are authoritative sources for attributes useful for identity proofing or eligibility determinations often have existing data exchanges or AVSs. Many of these are bespoke services that address a single use case and can include a variety of attribute verification and sharing models, including mainframe-to-mainframe data exchanges, individual and batch queries, web interface queries, CSV file uploads, or more modern protected APIs. Different parts of the organization may own these services, which may have proliferated over time in response to specific needs or statutory or regulatory requirements.

For agencies that already offer such services, it may be worth investigating whether a generalized service for core common capabilities could be created to support the new attribute validation use case as well as some existing ones. Consolidating services can have several long-term benefits, including reduced expenses, more efficient utilization of agency resources, and improved security and fraud detection capabilities. One common service to explore could be a core attribute validation API or microservice that allows the application of tunable fuzzy matching algorithms and inexact matching rules, and which can provide error and non-match responses at different levels of granularity, depending on the use case and RP.

While organization structure and appropriations can create barriers to service reuse and consolidation, the benefits over time can be substantial, including reducing the technical debt created when distinct services must be maintained for capabilities that could be consolidated.

6.2. Direct or Brokered Service

Your organization can provide the AVS either directly to end users or through a third-party attribute validation broker. Third-party brokers integrate with multiple external attribute providers or validation services to create a shared service for RPs who require AVSs. This model can simplify validation for RPs by reducing the number of authoritative sources or other AVSs they must integrate with.

Providing services through a broker can significantly simplify service development, deployment, and maintenance and dramatically reduce customer support needs. It can also substantially decrease the initial and ongoing costs for the service. Access control and customer support needs are greatly simplified when an organization only has the broker as its customer. With a brokered model, the authoritative source can provide a copy of the attributes to the broker, who will perform the validation requested by the RP. Alternatively, the authoritative source can retain full control of the data and perform the validation themselves, sending the results to the broker for further transmission to the RP.

AAMVA[6] and Naphsis[7] are two nonprofits that function as brokers to provide attribute validation services. Naphsis services include brokering state vital record death information [34] through its Electronic Verification of Vital Events - Fact of Death (EVVE FOD) service. AAMVA currently provides six attribute validation and verification services, including the Social Security Number Online Verification (SSOLV) service[8], where AAMVA acts as a broker for SSA so states can perform SSN verifications when issuing driver's licenses, and the U.S. Passport Verification Service (USPVS)[9] where they act as a broker for passport data held by the Department of Homeland Security. The General Services Administration has also considered providing an external interface for federal customers to its Identity Verification API (IDVA) [35].

While utilizing a brokered model simplifies deployments, there may not be a single centralized broker that represents an RP community, or there may be other reasons for an AVS provider to offer services directly to RPs. For example, SSA offers the eCBSV (electronic Consent Based SSN Verification)[10] service directly to financial institutions and to brokers who provide services to eligible financial institutions.

6.3. Requirements

Developing a successful attribute validation service requires careful planning and a detailed understanding of end-user and RP needs as well as the requirements of other stakeholders. Internally, stakeholders include representatives from security, fraud analytics, operations, IT, customer service, program management, privacy, and legal departments. The design and architecture of the service will also be driven by any statutory or regulatory requirements, so early and comprehensive requirements discovery can be critical to the success of a project.

When gathering performance requirements, it is essential to understand anticipated demand over time at a granular level. What is the maximum number of validations anticipated each day, each hour, each minute? What is the maximum number of concurrent validations expected during peak hours? What are the availability requirements for the service's users? Is a 99.9% availability rate sufficient, or is five 9s required? Do these availability requirements vary over time? For example, federal customers may have minimum availability requirements on holidays and during non-core hours, whereas some private sector customers may require consistent 24x7 availability. What is the acceptable number of outages each year? What are the least-impactful times for outages? If demand increases, can the current infrastructure scale, or may an infrastructure upgrade be required? It may be helpful to draft specific and comprehensive service level agreements (SLAs) with potential customers during discovery. SLAs that reflect a detailed understanding of customer needs can be helpful when making design decisions.

An understanding of performance and availability requirements will help drive foundational technical decisions, including how much of the organization's current infrastructure can be leveraged, whether it is necessary to replicate the data that will be used in verifications,

[6] Verification Systems - American Association of Motor Vehicle Administrators - AAMVA
[7] On Demand (naphsis.org)
[8] Social Security Online Verification (SSOLV) Service - American Association of Motor Vehicle Administrators - AAMVA
[9] U.S. Passport Verification Service (USPVS) - American Association of Motor Vehicle Administrators - AAMVA
[10] https://www.ssa.gov/dataexchange/eCBSV/

whether a full cloud solution is necessary, or whether a hybrid cloud and on-prem solution may be sufficient.

Other questions to consider are whether existing open-source, commercial, or cloud service solutions can be leveraged or whether extensive custom development is required. Commercial solutions have many advantages over custom-developed capabilities. They allow agencies to focus on their core missions and capabilities, and contracts can include requirements to stay current with evolving standards and guidelines. This is especially useful as the number of cybersecurity threats and requirements to mitigate those threats continues to rapidly evolve. However, integrating new commercial software or services with existing systems and the impact on current workflows must be considered.

Requirements cannot be driven exclusively by customer desire for functionality; data privacy and security requirements must be primary drivers of system development. Attribute validation is not only beneficial to legitimate users but has also become increasingly profitable for criminals and other bad actors. It is critical to understand their potential incentives for exploiting the service and to put in place protections to guard against misuse of the service as well as means of detecting abuse. Security, digital identity, fraud, and privacy risk assessments should be integrated into product development early enough to influence and enhance requirements and implementation decisions. Requirements must also include compliance with all relevant laws and regulations, including privacy and data protection laws.

6.4. Access Control

Access control is required to restrict access to the AVS to those users who meet all authorization requirements. It is relatively simple to implement if the agency offers the service through a third-party attribute validation broker, which requires the agency to establish legal agreements and secure connections to only a single organization. Access control increases in complexity as the number of direct connections with RPs increases.

6.4.1. RP Registration and Enrollment

Since AVSs are not typically offered directly to the public but rather to authorized organizations (RPs) and the individuals supporting those organizations, entity proofing, registration, and enrollment are often necessary. It may also be necessary to determine whether a particular individual is authorized to act on behalf of a specific organization in a capacity governed by the agreement between the agency providing the validation service and the organization consuming those services. Self-enrollment for organizations at scale is particularly challenging since there is no authoritative source in the U.S. for the information required to validate and authorize organizations. Registration, enrollment, and entering into data sharing agreements or other legal contracts should be done by an individual within an organization who is legally authorized to enter into contractual agreements on behalf of that organization; however, for most organizations,[11] there is no comprehensive source of authoritative information that agencies can query to determine who within any given organization possesses those roles.

Also, while most communication with organizations is now typically done through email, many authoritative sources for entity attributes only provide physical mailing addresses and phone numbers as contact information. The lack of authoritative email domain information can make it challenging to ascertain whether communication is occurring with someone from the correct organization. For larger organizations, third-party data brokers can provide some level of confidence in the association between an organization and an email domain, but they typically only have that information for larger organizations, and some organizations may only use free email providers. There is also no authoritative source of information for which organizations control which API client endpoints or Identity Provider endpoints, which creates challenges when attribute validation is done through APIs and connections with RPs must be done at scale. Legal agreements and extended validation[12] (EV) TLS certificates can be used to mitigate these risks. EV certificates can help address the gap in binding domains to organizations[13], but they impose an additional cost on the RPs.

When an agency expects to support a significant number of RPs, a registration portal may be required that allows legal agreements to be completed and authorization evidence to be uploaded, if necessary. A Digital Identity Risk Assessment (DIRA) should be conducted per NIST's Digital Identity Guidelines to understand which digital identity controls are needed to access a particular portal or API.

6.4.2. Federated Authentication and Authorization

When providing services to RP organizations, there are two options for authenticating individuals — directly connecting to the RP's Identity Provider[14] (IdP), which allows the affiliates of an organization to use their organizational credential to authenticate to their IdP, which then passes an authentication assertion to the Service Provider (SP) hosting the AVS, or by using a third-party federated credential[15].

A direct connection between the IdP and SP is preferable whenever individuals' authorization to access a service or application is associated with their affiliation with an organization. When individuals are no longer associated with an organization, they will lose their ability to authenticate to the RP's IdP. They will, therefore, automatically lose their access to the service. When federated credentials issued by a third-party Credential Service Provider (CSP) are used[16], the relationship between the RP organization and that individual must be maintained by the SP.

[11] The exception is publicly traded corporations, for which the U.S. Securities and Exchange Commission (SEC) provides a searchable database: https://www.sec.gov/edgar/search-and-access

[12] https://cabforum.org/info-for-consumers/ To further enhance security, it may be useful to restrict EV certificates to those issued by members of the CAB forum who are headquartered in the U.S. The company name and state listed on the certificate should match the name associated with the EIN in IRS or other financial records.

[13] The European Union uses Qualified Certificates for Website Authentication (QWACs), which have features similar to EV certificates. Qualified certificates for website authentication (europa.eu)

[14] This is typically done using either OpenID Connect or SAML.

[15] OMB M-19-17 strongly encourages the use of federation and federated credentials: "Agencies shall leverage existing credentials and identity federations that meet the agency's determined acceptable risk level rather than standing up processes or capabilities to issue new credentials to users." https://www.whitehouse.gov/wp-content/uploads/2019/05/M-19-17.pdf

[16] Kantara provides a list of credential service providers that have met the NIST Digital Identity Guideline requirements.

This creates an additional burden on the SP and may result in individuals retaining access to the SP service even after their association with an RP has ended and they are no longer authorized to do so.

Both options require a DIRA to determine the identity assurance, authenticator assurance, and federation assurance needed to access the service. When users are utilizing their organizational credentials rather than an agency or third-party credential, a legally binding user agreement is needed that requires that the organization's credentials meet the assurance levels identified in the DIRA. User agreements should be specific and outline terms, conditions, and penalties for non-compliance.

6.5. Budget Considerations

Budget estimation and planning for both initial deployment and long-term sustainability is another important consideration. It is critical to accurately estimate total costs for initial service development, annual maintenance, periodic assessments, and updates to secure sufficient funding to implement the service well.

Cost considerations include staffing needs, infrastructure upgrades or additions (if required), software, development, testing, and integration. Custom development tends to cost substantially more over time than leveraging commercial solutions, is much harder to maintain as standards and security requirements evolve and can lead to unexpected and significant cost overruns. Staffing needs beyond the development team include program and project management, analysts, risk assessments, legal, communications, and customer support. Periodic outreach to potential RPs should also be planned.

If an agency has data exchanges or similar services with substantial technical debt that provide similar functionality, budgeting to modernize and consolidate those services will reduce the total cost over time. Consolidation and modernization also improve fraud detection, privacy, and security controls. Sustainability is improved with service consolidation and reuse.

If cost recovery is a requirement, it is essential to understand what an acceptable cost per transaction is for RPs and to estimate the total usage per year. The service's total cost should be constrained to match the expected reimbursement rate times the anticipated volume, also taking ancillary costs into consideration. If this is not possible yet the service is mission critical, additional funding should be sought that is not tethered to a cost recovery requirement. Letters of commitment from RPs can help plan the budget and ensure that investments in the service will be effectively leveraged.

6.6. Development and Testing

Identity attribute validation services can be valuable targets for criminal organizations, identity thieves, and other bad actors. Additional care must therefore be taken to ensure that such services are protected from misuse and are resilient to hackers and cybersecurity attacks. This requires the project development team to have expertise in secure software and service design, expertise with the security and other standards and protocols that will be used, and be supported by competent software, security, and test engineers. Software engineers and architects should have experience with all COTS, open source, and SaaS products that will be used and an understanding of the infrastructure that will support the deployment. The security engineer should have an in-depth understanding of cryptographic requirements, other relevant NIST security standards, and any specialized security knowledge required for the deployment, such as cloud security.

The test engineer should be highly skilled in developing comprehensive usability tests and automated unit, integration, and security tests. Tests and testing infrastructure should be designed to evaluate compliance with all requirements, including functionality, performance, security, access control, and privacy requirements. In addition, having an experienced red team assess the system's resilience against various attacks and attempts at misuse, including social engineering attacks, is extremely useful.

It is helpful to have test engineers and red teams involved early in the development process to ensure that the solution is developed in a way that maximizes resiliency and can detect attempts at misuse. Program and business leadership should also be involved early so the solution can be designed to automatically provide the management information and metrics needed to understand the system's health and use. If the service will be offered directly to RPs rather than through a broker, customer service should also be involved early in the process. Their involvement will help the development team understands what tools and information are required to support end users, and how privacy, consent, and notices for end users will be handled.

During the project's planning stages, it is critical to understand all roles, expertise, and skills required for the service's success. An evaluation must then be conducted to determine whether the expertise and skills are already available within the organization. Once the individuals with the appropriate expertise are identified, the impact on other agency efforts must be evaluated to determine when they will be available to support the development of the new service. If there are gaps in the team's skills and expertise, staff may require additional training, or additional contractors may be required. Depending on the skill sets needed, the agency's existing contracting vehicles may not provide ready access to the necessary expertise, so contract amendments or new contracts may be required. Assembly of a team with the skills needed to ensure the success and security of the service can require significant lead time, so should begin as early as possible.

6.7. Planning for Deployment and Post-Deployment

Starting a deployment with a pilot is beneficial even for organizations with rigorous testing programs and highly involved usability experts. Pilots allow service providers to refine internal and external documentation, customer service tools, and training. They may also uncover usability or performance issues that should be addressed before a full-scale deployment. Pilots with a core set of committed RPs are especially critical when a broker is not utilized. The AVS can then be released to a broader audience once the lessons learned from the pilot have been incorporated into the service.

If the agency will be onboarding and supporting multiple RPs, preparation should be made for significant customer support. Potential RPs will need a point of contact to whom they can make inquiries regarding eligibility requirements, technical requirements, and cost. Enrollment support may need to be provided by multiple components within the agency to provide contractual and legal support in addition to technical support. A technical support team will need to work with each RP to conduct end-to-end testing, ensure all technical and security requirements are met, and troubleshoot any issues the RP may encounter. Thorough documentation and user support tools, such as test endpoints, validation tools, open-source example client code, and potentially a sample IdP configured to meet the agency's requirements, should be provided early to any RP. However, an agency should be prepared to provide technical support to each direct RP regardless of how thorough the documentation is or how simple the tools may be. Dedicated Tier 1 support staff should be available during onboarding, with reach-back to support that requires greater expertise. Providing multiple channels for support, including chat, email, and phone, can be useful.

Inevitably, some of the individuals whose data is validated will discover errors or outdated information in the data used by the AVS, so the AVS provider will need to establish clear redress guidance and mechanisms. It is also likely inevitable that impersonators will attempt to leverage the redress procedures to contaminate or alter legitimate data. A fraud risk assessment can be conducted to better understand that potential threat and implement appropriate controls. This can include defining acceptable forms of evidence for correcting data and processes for confirming user identity in the absence of the AVS having accurate data on record for the user. The AVS should also consider establishing an appeals process if a user disagrees with a decision to deny a request.

Plans must also be made for other types of ongoing post-deployment support. Service monitoring should be continuous, with logging and analytics performed to understand usage and performance. The effectiveness and impact of the validation logic should also be monitored and assessed (see Section 3, Validation Logic).

Periodic customer engagement should be planned to understand the needs of the RPs over time, including any new functional requirements or concerns. Requested improvements or changes should be tracked. There should also be ongoing risk assessments and testing, which should reflect the evolving security threat landscape, as well as changes to federal guidelines and requirements.

Validation services will require periodic maintenance to ensure they remain compliant with evolving security standards and requirements, meet changing customer needs, and address findings from service monitoring, testing, or periodic risk assessments. Finally, an incident response plan should be developed for emergent service or security issues.

7. AVS Architectures and Deployment Models

There are three primary deployment models for AVSs: API query-based services, the shared service broker model, and user-controlled verified attributes (UCVA). Each model has unique benefits and limitations, and different implications for security, privacy, and user experience.

API query-based services represent the most traditional approach to attribute validation. In this model, RPs interact directly with attribute validation services through APIs to verify user-provided attributes. This architecture typically involves an RP sending user-provided data to a validation service, which checks the accuracy of the data against its records and returns a validation response. The primary advantage of this model is its scalability and real-time validation capabilities. However, the AVS must onboard and monitor multiple RPs, and users can only have their information validated by those RPs that have a relationship with the AVS.

The shared service broker model introduces an intermediary, or broker, that facilitates the interaction between RPs and AVS providers. The broker acts as a central hub, streamlining integrations for RPs and managing RP onboarding and management for the AVS. This model can make it easier for organizations to implement and maintain AVSs. However, the broker must be trusted by all parties and is a viable option only when a broker exists that provides services to all the potential RPs for a service. AVSs will often integrate with both brokers and individual RPs when brokers exist that cover only a part of the population of RPs that the AVS must support.

The UCVA model is an emerging model that offers a more decentralized approach. Here, users have control over their verified attributes, usually stored in digital wallets. Users can share these pre-verified attributes directly with RPs as needed, giving them greater control over their data. This model has the potential to reduce challenges with data quality and minimizes the need for repeated verifications by an AVS, but it presents several additional management and technical challenges for the AVS, the RPs, and the users, which can make for complicated implementations.

7.1. API Query-Based Validation Services

As noted in Section 2.2, most existing attribute validation services typically take the form of query-based systems that make use of APIs or custom integrations to request and exchange information among RPs, AVSs, and the end user. They are most commonly seen in identity proofing schemes where users submit their data to an RP who packages the attribute values or claims and conveys them to a validation service via an API. The attribute validation services convey a response indicating the accuracy of the attributes that have been sent based on a comparison against their data. In some instances, these API queries can be designed to minimize the passing of PII by limiting queries to derived attribute values, with “yes/no” or “true/false” responses to structured input from the RP service. For example, if an RP service has an obligation to provide services only to individuals over the age of 18, rather than asking the AVS to confirm the user’s date of birth, it may instead request confirmation that the user is over 18, limiting the transmission to a claim of “over 18” with a derived attribute value of “Yes or No.”

7.1.1. Architectural Overview

A typical API query-based architecture contains the following participants, as shown in Fig. 1:

- **User**: Interacts with the RPs (and optionally a Credential Service Provider [CSP] operating on their behalf) through an agent — typically a browser or mobile application — to gain access to a service, benefit, or data. The user may submit their personal attributes as part of a proofing process.
- **RP**: The entity relying on the AVS to confirm the accuracy of any submitted attributes needed for identity proofing or approving access to protected services, benefits, or data. The RP may have a CSP operating on its behalf; however, for simplicity, we will consider CSPs as RPs of the AVS.
- **AVS**: The organization that receives queries from the RP and compares data against their records to help determine the accuracy of the submitted attribute(s).

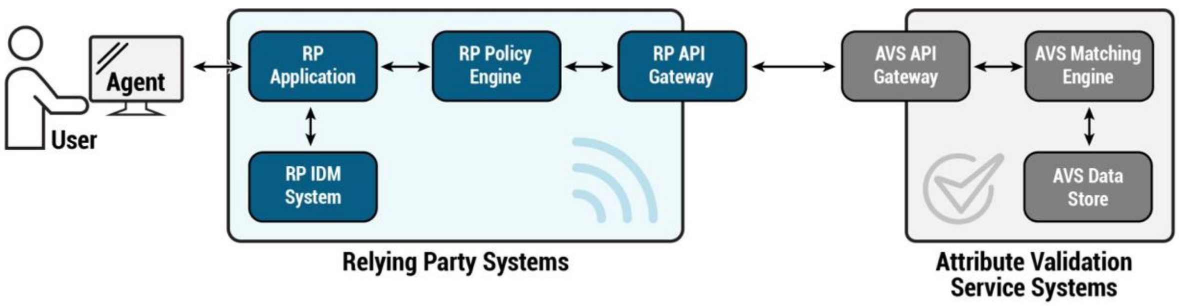

Fig. 1. Typical API query-based architecture

A typical API query-based architecture consists of the following components:

- **API**: A system access point or library function that has a well-defined syntax and is accessible from application programs or user code to provide well-defined functionality.[17] Provides a standardized method for interacting with and requesting information from the AVS. This is typically provided by the AVS but may be dictated by the RP in certain situations. More than one API may be used through the process of a complete workflow; for the purposes of this discussion, we will focus on the interaction between the RP and the AVS.
- **UI/RP Application**: A web or mobile application maintained by the RP (or a CSP acting on behalf of the RP) with which end users interact to provide their data for validation. This may be part of an onboarding, access, or service request workflow, depending on the use case being applied.
- **RP Policy Engine**: A policy enforcement point that serves multiple purposes including basic data validation (such as whether everything is present and correctly formatted), packaging of the data into the API-defined data structures, communicating it to the AVS, and validating the results of the AVS call to make a policy decision.

[17] NIST Glossary: https://csrc.nist.gov/glossary/term/application_programming_interface - :~:text=Definitions%3A,to provide well-defined functionality.

- **API Gateways (AVS and RP)**: A security and network traffic appliance protecting both RP and AVS APIs that enforces authentication and access for the API requests that are transitioning between the AVS and RP during a transaction. These can also be used for translation and a degree of orchestration when needed to support calls and responses.
- **AVS Matching Engine**: A policy or algorithm that compares the API-received data with the stored data to generate the appropriate response for RP consumption. In an AVS system where there are multiple integrated sources, this engine (or another) may parse the data received in the API calls and internally query the most appropriate data stores for validation purposes.
- **AVS Data Store(s)**: The repository (or repositories) within the AVS where data is stored to which data in the API calls are compared. How these are queried, and how the data is handled, will depend on the type, structure, and technology of the data store. For example, cloud-based databases can typically be queried using internal APIs or microservices.

7.1.2. Standards Consideration

API query models should consider the following standards:

Data Query and Interchange Standards: At their core, AVSs are services that exchange data between the RP, the AVS, and connected systems. They rely, in large part, on well-established data standards and information exchange/interchange protocols. This document is not intended to explore the value of JSON versus XML. The best data format for a particular service will depend on the participants, the technologies involved, and the limitations and capabilities of each. What is important is that as services are developed by an AVS, it is critical to have established common data interchange standards for the users of the service that will function with their intended consumer population.

Standard APIs: At the heart of query-based models are the APIs that support interacting with the services that the AVS offers. This report is not intended to serve as a guide to the development of APIs as these will be highly dependent upon the systems, architectures, and technologies comprising the AVS service. OWASP's Secure API Project [36] provides extensive recommendations on the secure development of APIs that can be considered in the context of AVS development.

What is important in the context of this paper is for AVS services to deploy well designed and standardized APIs that provide secure RP access and ensure appropriate protections for PII and other data. Two key elements in this process are defining within the API what kind of data can be requested and enforcing this through appropriately structured parameters. It is particularly useful for AVSs to consider the level of granularity they offer in their parameters, since offering field level parameters can reduce RP and end-user risk while improving AVS data quality. For example, the API could allow RPs to indicate in their request whether they require an exact match or if fuzzy matching is acceptable and can provide parameters in the response that indicate whether fuzzy matching was required to make a match.

Regardless of the form they take, AVSs must have well-structured and clearly defined APIs defined that effectively support integration with RPs and should make all information required to integrate with the service available through developer and integration guides. Providing a sample open-source API client should also be considered, particularly when providing the API to multiple RPs.

Enrollment and Proofing Standards: As discussed in Section 6.4.1, managing access to API services starts with an enrollment and registration process for consumers of the service. While it is certainly possible to establish APIs as open and publicly available, the sensitivity associated with the data used for attribute validation for identity and access scenarios mandates that APIs are appropriately protected to ensure they are only successfully called by consumers who should have access to them. This dictates the need for an enrollment process where consumers of the API service register for access, have their identity verified by the AVS, and are issued the necessary credentials to be able to access the services on a defined ongoing basis consistent with an established user agreement. The enrollment and proofing process should be done consistent with an established risk assessment and management process, for example the DIRA defined in the NIST Digital Identity Guidelines [1].

NIST SP 800-63A: *Enrollment and Identity Proofing* [3] provides a basis for resolving, validating, and verifying the identity of individuals seeking to establish digital identities. This can, and should, be used as a starting point for organizations seeking to establish a consistent program and process for vetting users prior to granting access to any API service. However, it should also be clearly noted that the processes defined in NIST SP 800-63A are intended to be applied to individuals and do not cover entity verification (e.g., whether this a legitimate business with legitimate needs to access an individual's data), nor do they cover the process of binding an individual to a business (e.g., Person 1 works for Company A). Processes from the guidelines can be used — for example, leveraging authoritative sources for validating information about an entity and its affiliates — but would need to be augmented with organizationally standardized processes for confirming and binding entities to individuals. At a minimum, the following steps must be taken prior to granting access to an AVS:

1. **An Enrollment Risk Assessment**: The process used to determine the level of risk or rigor related to accessing the API and providing a standardized set of processes and controls that can be applied to the enrollment and registration process. All the supporting processes should be consistent with defined legal and policy requirements.
2. **Vetting of Consuming Entities:** The process of vetting the entities that are registering to consume the information provided by the AVS. This provides a process to confirm whether the entity is a real entity and whether it is a legitimate consumer of the services with a legitimate reason to request the information. The AVS should not allow access to any APIs before these issues are addressed.
3. **Identity Proofing of Individuals:** The processes used to resolve, validate, and verify the identity of specific individuals who may be requesting data through the AVS. While there is no expectation that all individuals who work at a consuming entity will need to be identity proofed, this step may be required for administrators, users with elevated

privileges, or users who are granted authority by the AVS and the RP to manage accounts that interact with the AVS. Identity proofing should follow NIST SP 800-63A when the AVS is provided by a federal agency.

4. **Binding Entities and Individuals**: The process used to validate that an individual represents a specific entity and ensure that relationship is captured and represented in the AVS identity and access management systems whenever direct federation with that entity is not available or does not provide all required information. This requires the AVS to support processes — whether manual or automated — to confirm with entities the role an individual plays and maintain that role over the duration of the relationship. The specific business processes, policy environment, and technology stacks will dictate how and how often this binding is confirmed. In some scenarios, where direct federation is viable, this step can be outsourced by leveraging credentials and roles issued and managed by the responsible entity.

Authentication and Federation Standards: Authentication and federation standards provide the means and mechanisms for verifying that a returning user is the same individual that registered, and for conveying authentication information between the RP and AVS. The type of authentication and federation standards that are used will depend heavily on how the APIs are accessed, the scale of support required, and the underlying technology stack. Human user access to APIs or UIs related to APIs needs to be protected with phishing-resistant MFA, for example a FIDO2 Web Authentication credential or similar PKI-based cryptographic authenticator. NIST's Digital Identity Guidelines provide guidance on the selection and implementation of authenticators and management of authentication processes. Human user authentication for any AVS run by or operated for federal agencies should be consistent with NIST SP 800-63B, *Authentication and Lifecycle Management* [37], and, where applicable, NIST SP 800-63-C, *Federations and Assertions* [38].

Additionally, authentication of human users may be done through direct authentication to an API dashboard or UI or through a federation set up between the RP and AVS.

There are two core options that are used by most services today:

1. **OpenID Connect.** OpenID Connect (OIDC) is an interoperable authentication protocol based on the OAuth 2.0 framework of specifications [39].[18] Essentially, it provides a consistent way for expressing authentication, consent, and authorization information through identity tokens between RPs and the AVS when user authentication is required for access to an API or application. The OIDC specifications offer extensive flexibility, making them suitable for a wide range of needs. Profiles tailor the specifications to meet the requirements of specific use cases or user groups, which also improves interoperability.
    a. **iGov Profile**. The International Government Assurance Profile (iGov) profile of OIDC [40] is designed to meet the needs of government agencies that provide online services to the public.

[18] https://openid.net/developers/how-connect-works/

2. **Security Assertion Markup Language (SAML).** SAML 2.0 **Error! Reference source not found.** is an XML-based standard that defines a framework for exchanging security information between online business partners.[19] It is an older standard than OIDC but can achieve similar outcomes.

Selection between SAML and OIDC will be determined based on the capabilities and capacities of the RPs and AVS providers. At their core, both standards support the ability to convey information between parties in a secure manner. Service accounts and client-to-client calls should make use of valid authentication and authorization tokens bound to a set of organizational credentials and maintained by clients on the RP and AVS infrastructure. For additional guidance on Federation, AVSs that are operated by federal agencies or in the federal space should leverage NIST SP 800-63C, *Federation and Assertions* [38].

Access and Authorization Standards: Access to AVS APIs must be managed effectively to prevent unauthorized exposure of information. Unprotected APIs can be queried over the internet by attackers, potentially ingesting highly sensitive data and the PII of unsuspecting users. Additionally, given the high-value nature of the data an AVS can either provide or validate, organizations that choose to offer such services must anticipate being the target of such attacks. To help counter these threats, AVS providers can turn to several standards:

1. **OAuth 2.0.** OAuth 2.0 is an authorization standard that may be used to support access control objectives by API services. The standard defines a set of technical specifications for the generation, protection, and delivery of authorization tokens (JSON Web Tokens or JWT) to different connected endpoints (e.g., servers). The authorization (or access token) is used to define what actions an endpoint may take relative to a specific service. For API protection, these tokens are typically issued to consumers of the service, allowing them to make requests to the API service and allowing the API service to confirm that such requests are coming from a valid and approved source. To be effective, they are combined with authentication standards and protocols such as OIDC or SAML to provide confidence that the requesting endpoint is the same one that participated in the enrollment or registration process.
2. **Financial-grade API Security Profile 2.0 (FAPI).** FAPI is an Open Identity Foundation (OIDF) profile of the OAuth and OIDC specifications intended to provide a high-security model for API access and protection and the secure authentication of endpoints. While built to address financial APIs, it can be applied to support any high-risk use of API-based services, including those that may be offered by an AV service.

Regardless of the approach taken and the standards applied, API consumers must be both authenticated and authorized to ensure that only approved services are making calls and receiving data from the AVS API services.

[19] https://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0.html

7.1.3. Security Considerations

NIST SP 800-95 [42] provides comprehensive guidance on securing web services, including guidance on securing APIs for both internal microservice architectures and external facing interactions. For the purposes of this document, API-based query model AVSs should focus on addressing the following threats to confidentiality, integrity, and availability:

1. **Threat**: Data exchanged between the AVS and RP is intercepted.

   Mitigation Strategies: Exchange all data between the AVS and RP over an encrypted channel. When highly sensitive data is exchanged, the AVS and RP should encrypt the data at the message level when in transit. Use only approved cryptography.

2. **Threat**: Data at rest is subject to unauthorized access.

   Mitigation Strategies: Implement AVS internal identity and access controls consistent with FISMA moderate baselines. Restrict authentication to AVS data sets to phishing-resistant MFA mechanisms. Encrypt data at rest with approved cryptography.

3. **Threat**: Access tokens from the RP are stolen by an attacker and used to create new requests.

   Mitigation Strategies: Employ capabilities to time-bound and restrict calls to a single event. Within the context of OAuth, this is achieved by using mutual TLS and by limiting the lifetime of access tokens. Receipt of a previously used access code or token results in the denial of access.

4. **Threat**: Data exchanged between the AVS and RP is modified in transit.

   Mitigation Strategies: As part of each exchange, the AVS and RP use message authentication codes or digital signatures consistent with the agreed data standards employed by the AVS. For example, when using JSON Web Tokens, JSON Signing and Encryption (JOSE) can be used to protect the integrity of tokens and JSON responses passed between the RP and AVS. However, it is critical that the RP cryptographically verifies the signature to ensure that no changes have been made.

5. **Threat:** Data is exposed by an attacker setting up an illegitimate RP endpoint.

   Mitigation Strategies: Authenticate and constrain senders and audience endpoints using an approved and agreed-to standard for authentication and authorization. This can be achieved using access and authorization standards such as OAuth coupled with authentication standards such as OIDC. In high-risk scenarios, mutual TLS (mTLS) should be used to support sender and client authentication. Other mitigation techniques can include allowlists at the AVS to ensure that only registered entities and endpoints are eligible to make calls to the service.

7.1.4. Privacy Considerations

NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management [43] provides a comprehensive model for evaluating privacy risks associated with technology implementations within an enterprise. It focuses on providing outcomes for systems and processes intended to preserve the predictability, manageability, and disassociability[20] of systems. AVS and RPs should leverage this resource to evaluate and understand the potential problematic data actions that can result from the design of an AVS and integration with an AVS. Additionally, AVS and RPs consuming their services should address the following:

1. **Problematic Data Action**: Unnecessary data is exchanged between the RP and AVS.

   Mitigation Strategies: A minimum step for all API query-based models is to minimize the amount of information passed in each call and response. Even where data is passed over encrypted channels or where end-to-end encryption may be used, data minimization prevents unnecessary aggregation by both the RP and the AVS. A key aspect of this minimization is to leverage claims and derived attribute values where possible. In this model, the RP submits minimal attributes to support AVS resolution (e.g., an identifier such as an SSN) and requests the evaluation of a series of claims rather than attribute values (e.g., 21 or older). The AVS leverages this information to correlate the claim to a user in their system and computes a response to the claim rather than providing the attribute value itself (e.g., 23 years old rather than a birthdate of 12/12/2001).

2. **Problematic Data Action:** The AVS creates or aggregates user information and behavior across RPs (i.e., user surveillance).

   Mitigation Strategies: The AVS will be exposed to a wide range of transactions and data regardless of how the system is designed and implemented. As a result, they are also able to actively aggregate and leverage data used across transactions and RPs. To an extent, this is expected and often used to improve the accuracy of data and services offered and to detect potentially fraudulent activity. However, this could easily transition from well-intentioned efforts to improve accuracy and prevent fraud into surveillance of users. This is particularly true within query-based API systems where technical controls such as privacy-enhancing technologies are limited. Regardless, it is essential that data is not used for any purpose other than that which has been defined by the RP and consented to by individuals interacting with the RPs. It is therefore critical for RPs and AVS providers to have well-defined terms of service and use for the data they exchange and convey. Additionally, while an AVS may use data from multiple sources to gain fidelity and accuracy, they should not be tracking calls relative to individuals across their RPs. Data related to specific calls should be retained for well-defined allowable purposes (e.g., audit, fraud prevention, investigation) with user notice and never for tracking and profiling users. The NIST Privacy Framework emphasizes the importance of building customer trust through ethical decision-making and the need to facilitate communications about privacy practices with individuals, partners, assessors, and regulators[21].

[20] [43], pg. 34.
[21] [43], pg. i.

7.2. Shared Service Attribute Broker Model

A shared service broker model provides a multi-party platform into which agencies can integrate to: 1) provide attribute validation capabilities; 2) consume attribute validation services; or 3) both provide and consume. Such services are intended to simplify integration by consolidating technical connections while maximizing value by providing access to an array of integrated attribute services. The architecture is similar in structure and standards to the attribute query model but with the addition of a broker who operates as a "hub" for both AVSs and RPs seeking to streamline integrations. This may be done for several reasons, including where RPs are seeking to integrate with many AVSs from a specific community (e.g., federal government) or where many RPs are seeking to integrate with an AVS that has administrative, policy, or implementation constraints that make a broker model more attractive to consumers of the service. While the latter is often seen as a "workaround," it may also be a legitimate model for accessing services based on the conditions that exist at the AVS.

7.2.1. Architectural Overview

A shared service attribute broker model typically consists of the following participants, as shown in Fig. 2:

- **User**: Interacts with the RPs (and optionally an SP operating on their behalf) through an agent — typically a browser or mobile application — to gain access to a service, benefit, or data. May submit their personal attributes as part of a proofing process.
- **Broker:** A service provider that sits between RPs and the AVS providers to serve as a common integration point and to direct API calls and queries to the correct services and consumers. The broker may play a role in intermediating requests to support interoperability, for example translating between protocols.
- **RP(s)**: The entity relying on the AVS to confirm the accuracy of any submitted attributes needed for identity proofing or approving access to protected services, benefits, or data. In a brokered model there are often many RPs from a community with similar needs and requirements.
- **AVS(s)**: The organization that receives queries from the RP and compares data against their records to help determine accuracy of the submitted attribute.

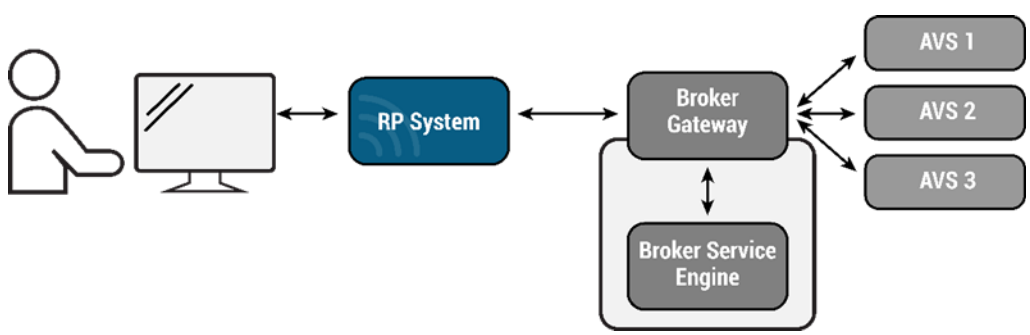

Fig. 2. Typical shared service attribute broker model architecture

A shared service attribute broker model typically includes the following components. Due to the overlap with attribute query models, this description focuses on components critical to the broker model only:

- **API(s)** - A system access point or library function that has a well-defined syntax and is accessible from application programs or user code to provide well-defined functionality. APIs in a brokered model are often defined by the broker and the AVS, though this is subject to the specific conditions of the integration and community. For example, the broker may provide a common API for RP integration but then integrate with established AVS APIs on the back end.
- **Broker Service Engine** - A mechanism or mechanisms used to route API requests to the correct integrated endpoints and, where necessary, translate between protocols to allow for consumption of responses between an RP and AVS, for example by translating from SOAP to REST or OIDC to SAML. In some instances, it may also function as a policy evaluation point to generate binary Y/N responses that may not be directly provided by the AVS or data sources. In other instances, the broker service engine can also provide privacy enhancing qualities by stripping unnecessary data, blinding RPs and AVSs from the sources of specific requests and preventing the tracking of users across different participants. The degree and capacity of these entities to enforce privacy enhancing technology will be highly dependent on the integrated partners and underlying technologies.
- **API Gateways (AVS, Broker, RP)** - Security and network traffic appliances that protect APIs. They enforce authentication and access for the API requests and secure the responses back to the RP. These can also be used for load balancing, translation, and a degree of orchestration when needed to support calls and responses.

7.2.2. Standards Considerations

The standards considerations related to implementation of a broker-based AVS model are similar in nature to those introduced by an API query-based validation model. Essentially, they revolve around the protection of the APIs coming into and out of the broker service. This includes standards such as OAuth 2.0 for authorization — and profiles such as FAPI — and Open ID Connect for authenticated calls in some instances. The unique characteristics of this model lie not with the standards but instead with the security and privacy implications introduced by the broker and its role in the process of orchestrating and directing calls.

7.2.3. Security Considerations

This section covers only new risks introduced by the inclusion of a third party (i.e., broker) into the architecture of an AVS. Other risks are similar to those discussed in Section 7.1.3 relative to a query-based attribute service.

1. **Threat**: Broker Compromise

Mitigation Strategies: In a broker-based model, the broker is placed in a position of elevated privilege. All API calls coming in and going out may be visible to their systems. A compromise of the broker system could result in the exposure of sensitive information coming from the RPs and the AVS providers. In most cases, architectures should be designed to prevent the broker from viewing or accessing any PII. This should be achieved by encrypting all PII that may need to be sent to the broker with a key only available to the RP and the AVS. In these instances, the AVS acts as nothing more than a pass-through, directing calls and requests to the appropriate endpoints and back again. However, in many instances, the broker has a more robust role to play in managing calls and directing attributes to different endpoints. In these instances, the broker must manage PII and other sensitive information to appropriately broker calls to connected AVSs. In such cases, brokers must not retain data for any longer than is necessary to complete calls between the RP and connected AVS provider. At a minimum, data retention policies need to be defined in trust agreements with RPs and AVS providers and, ideally, destruction of stored data should be automated to enable greater confidence in compliance to data retention rules. Additionally, the broker must not create individual profiles for users within their system, all data at rest must be encrypted using approved cryptography, and all exchanges of data must take place over a protected channel.

7.2.4. Privacy Considerations

This section covers only new problematic data actions introduced by the inclusion of a third party (i.e., broker) into the architecture of an AVS. Other problematic data actions are like those discussed in Section 6.1.4 relative to a query-based attribute service.

1. **Problematic Data Action**: User Surveillance and Data Aggregation by the Broker

   Mitigation Strategies: Where feasible, encrypt all PII passed through the broker to prevent the broker from gaining visibility into the specific attributes and data elements being passed. Where this is not possible, controls should be put into place at the broker that prevent the correlation of data across different requests. This can be policy based but should also include automated technical controls such as enforced deletion after a certain timeframe, and granular access controls for humans and system accounts.

7.3. User-Controlled Verified Attributes (UCVAs)

API-based verification services are not currently available for all attributes required for identity resolution, identity proofing, or authorization decisions. Some attributes are only available in physical documents, which cannot be easily or securely utilized for online transactions. Physical documents can be outdated and are easily forged. Privacy concerns also arise when a document contains more information than is required for a transaction, resulting in overcollection.

When an AVS is available, the person described by the attributes has little control over which RPs are allowed access to those verifications and has no input into how they are performed. If there are inaccuracies with the data, the user may not learn of them until after being denied access to a service and, if fuzzy matching is used, any data quality issues may remain hidden.

A UCVA architecture has the potential to overcome these limitations, giving individuals greater access to and control over their information in a way that enables both secure online and in-person data sharing and increases trust and privacy while reducing fraud. With a UCVA architecture, the authoritative source provides users a digitally signed copy of their verified attributes or claims. Users can then share those verified claims directly with RPs to prove their identity, access services, or obtain benefits.

7.3.1. Architectural Overview

A typical UCVA architecture contains the following participants, as Fig. 1 depicts:

- **User or Subject:** Obtains a UCVA from an issuer, stores or "holds" it in an app such as a wallet, and then shares the UCVA, or some of the attributes it contains, with RPs (and optionally an SP operating on their behalf) to gain access to a service, benefit, or data.
- **Relying Party (RP):** The entity relying on the UCVA to obtain and confirm the accuracy of attributes needed for identity proofing or approving access to protected services, benefits, or data.
- **Issuer**: An authoritative source that creates and digitally signs a UCVA, then issues it to the individual it has identified as the legitimate owner of those attributes and claims.

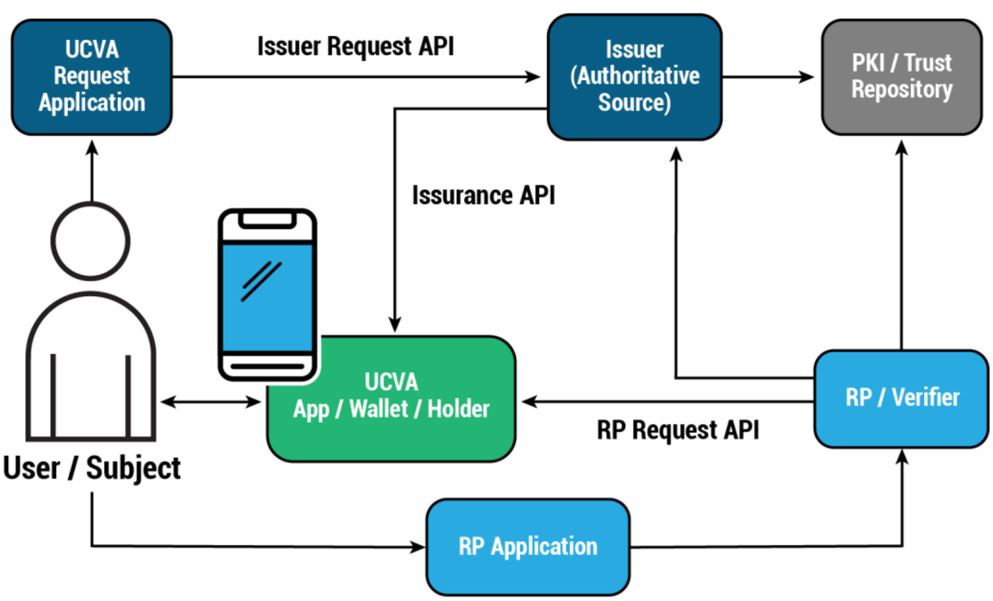

Fig. 3. Typical UCVA (User-Controlled Verified Attributes) architecture

A typical UCVA architecture includes the following components, also depicted in Fig. 1:

- **UCVA:** A credential or set of attributes and claims that have been verified by an authoritative source, packaged into a standardized data model, digitally signed, and then securely issued to the individual identified as the legitimate owner of those attributes and claims.

- **Wallet or UCVA Holder**: The digital wallet is an application that acts as a secure interface to the UCVA. It provides a UI that allows the user to manage their UCVAs and may provide APIs to issuers and RPs, as well as NFC interfaces for physical readers.
- **Verifier or Reader:** The verifier reads and evaluates the UCVA to determine its authenticity and validity.
- **Public Key Infrastructure (PKI)**: RPs must be able to obtain and verify the public digital signature (ds) certificate of the issuer. The method used to verify the public key will depend on the trust infrastructure used by the issuer and RP. Issuers may, for instance, be issued certificates from a CA trusted by the RPs, or their keys or certificates may be on a trust list shared from a central authority. Once the issuer ds key is verified, the RP can use it to cryptographically prove that the UCVA and its data elements were signed by the issuer and have not been altered.
- **Trust Repository:** A trusted service provider or broker can provide centralized access to the issuer ds certificates an RP may require for UCVA verifications. The repository may provide additional services such as revocation checks on the issuer certificates or UCVA integrity and issuer checks.
- **API**: This is a system access point or library function that has a well-defined syntax and is accessible from application programs or user code to provide well-defined functionality. It provides standardized methods for Issuers to provision UCVAs into a wallet, and for RPs to request or query the UCVA.

Once the UCVA architecture is fully realized, information currently trapped in physical documents, which are vulnerable to manipulation, theft, and forgery, will be available to share more easily and in a more trustworthy manner. UCVAs may include any attributes or information that requires verification today, such as identity resolution data, names, name history, DOB, age verification, proof of address, address history, proof of income, licenses, student or employee IDs, marital status, degrees and certifications, proof of employment, income history, proof of relationships (such as parent of a minor child), and proof of benefit entitlement.

In recognition of this potential, several commercial and governmental efforts are underway to understand and implement the components required for this architecture to succeed and to create the standards and protocols necessary for interoperable solutions. In the United States, agencies such as NIST and DHS have joined leading industry and international efforts to actively support the development of critical enabling standards and protocols. In the European Union, there are efforts underway to create “a trusted, user-controlled identity, allowing each citizen to control their online interactions and presence” in a privacy-preserving manner.

7.3.1.1. Mobile Wallets

UCVAs are issued to the individual and stored in a “holder” or “container,” most often taking the form of a mobile wallet. UCVA issuers and verifiers need to determine which wallets they will support based on whether the wallet sufficiently protects UCVAs from theft and misuse. Which standards a wallet supports is another consideration. Most wallets today support standards for the UCVA itself, including the ISO mdoc/mdl standard and W3C VC data model; however, issuance and presentation APIs and protocols are often proprietary, something that may change as those standards mature.

While trust in the information contained within UCVAs is obtained using PKI and digital signatures, trust that the individual presenting the UCVA is the same person the UCVA was issued to is anchored in the security behaviors and characteristics of wallets, including the presentation protocols. The security of both the wallet and the APIs it utilizes is therefore critical to understand. Risk assessments can help determine the complete set of requirements that end-user software and hardware must meet to store and share UCVAs in a way that sufficiently protects the confidentiality of the data. However, there are currently limited standards and certification programs for wallets. Current efforts are being led by the Digital Identification and Authentication Council of Canada (DIACC), which has created recommendations for Digital Wallets, and the European Union, whose EUDI Wallet will be supported by an existing regime to establish “qualified” wallets and service providers. A comparable framework has yet to emerge in the U.S.

Considerations for implementing an appropriate wallet don’t end with security. Once wallets that meet all usability, risk, and legal requirements are identified, a decision needs to be made as to which wallets the attribute provider will issue to. APIs may need to be developed for each wallet that will be utilized. A decision then needs to be made regarding which data model to use for the attributes, which digital signature algorithm and key lengths to use to digitally sign the attributes or credentials, and how to make the agency’s digital signature certificate available and easily discoverable for RPs or verifiers. These decisions will be constrained by which data models, standards, and protocols are supported by the wallets.

While mobile wallets are likely to be the primary mechanism for storing and presenting UCVAs, they are not the only possibility. Laptops with hardware-based Trusted Execution Environments (TEEs) such as Trusted Platform Modules (TPMs) have similar security capabilities that wallets could leverage, although proving ownership and control of a laptop is even more challenging than doing so for a mobile phone. However, if these challenges can be overcome, allowing TEE-based wallets to hold credentials that need to be secured could reduce the need for cross-device workflows. Other verified attributes and claims may be deployable using a cloud-based solution if the convenience and easier deployment outweigh the risks of the UCVA being stolen or copied, which increases when hardware-based security isn’t available.

Issuers should conduct risk and usability assessments before choosing a deployment model and deciding which wallets to support. Rigorous assessment can ensure that the choices made meet security and privacy requirements for government-issued verified attributes and claims and that the wallets will be usable by a sufficiently high percentage of the target user population.

7.3.1.2. Issuance Considerations

An authoritative source that wishes to issue UCVAs must also create a secure user interface that allows an individual to request a verified copy of their claim. The list of supported wallets or applications must be provided to users in advance so they can download and install the software required to receive, store, and use their UCVA.

Since the verified claims must be issued to the correct individual, those requesting a verified copy of their attributes or credentials must be identity-proofed at an assurance level that is proportionate to the potential negative impacts that could arise should a bad actor gain access to and control over that information. A DIRA should be conducted following the current NIST *Digital Identity Guidelines* to make this determination. Other security controls can help increase confidence that individuals requesting the UCVA are who they claim to be, such as by ensuring they are not using a high-risk VPN or connecting from a high-risk location or device, and by using third parties that provide risk scores for phone numbers, postal addresses, and emails. Procedures also need to be implemented that allow issued claims to be revoked if it is discovered that someone obtained them by impersonating a legitimate user. User-initiated revocation procedures must be established so that the user may request a revocation for any reason, including a concern that their UCVA has been compromised. These revocations must be easily discoverable so that an RP, when presented with a UCVA, can quickly ascertain whether it is still valid. Several standards are under development for how to manage revocations in a privacy-preserving manner.

Once the individual has been identity-proofed at the appropriate assurance level and has requested their verified attributes or credentials, the Issuer must encrypt the data and send it to the user’s wallet or a suitable alternative application to which the user has access.

7.3.1.3. RP Considerations

UCVAs are only useful if an ecosystem of RPs is available to utilize them. There are compelling use cases for RPs to utilize UCVAs once they become more widely available, such as a way to more reliably identity-proof individuals, verify their claims, or ascertain their entitlements. UCVAs may also reduce the need for RPs to store user PII and documentation [44]. If users can assert the information they need for each transaction, the RP’s need to retain and maintain that data diminishes. Decentralizing identity data and sensitive PII also reduces the amount of information a bad actor can obtain with a single breach.

However, the UCVA ecosystem may not be a good fit for all RPs or all attributes that need verification. For RPs whose business cases rely on real-time data or who have existing data exchanges with attribute validation or data exchange services, there may not be a compelling reason to accept many user-controlled verified attributes. The most compelling initial use cases for RPs may be replacing physical document inspections such as passports and driver’s licenses with secure UCVAs to improve accuracy and privacy. While an API is available through AAMVA[22] to verify some of the data found on physical driver’s licenses, the allowed use cases are limited,

[22] https://www.aamva.org/it-systems-participation-map?id=594

and not all states participate. The service also does not provide biometric match capabilities, which severely limits its utility in reducing fraud.

RPs that do decide to accept UCVAs will need to decide whether they need to perform full revocation checks of the attributes or credentials in addition to digital signature (ds) verifications. Depending on the risk, a revocation check may also need to be performed for the ds certificate[23] and the certificates in its chain of trust. A risk assessment can determine whether that is a requirement for a particular attribute and should consider the degree of confidence in the identity of the individual making an assertion, the likelihood that the attribute values may have changed since it was issued, and the negative impacts that could arise if a no-longer valid attribute or revoked credential was accepted. Revocation checks also have privacy implications. Depending on the implementation, the attribute or credential issuer may be able to gain knowledge of its use by a particular RP, but that is also the case with API verification.

For use cases beyond mobile driver's licenses, it is not yet known which standards and protocols will be most widely adopted by issuers or RPs. Also, government AVSs will need to continue to provide alternative paths for users who are unable or unwilling to take advantage of UCVAs.

7.3.2. Usability Considerations

A 2021 Executive Order [45] on transforming federal customer experience and service delivery states: "The Federal Government must design and deliver services in a manner that people of all abilities can navigate … and implement services that are simple to use, accessible, equitable, protective, transparent, and responsive for all people of the United States."

Since UCVAs that contain sensitive data require users to own up-to-date technology and have the technical literacy required to obtain and use it properly, it may be challenging to implement UCVAs in a way that meets the EO's requirements. At a minimum, alternatives will need to be provided for those individuals who will not be able to utilize UCVAs or who chose not to do so due to security or privacy concerns. For those users who wish to obtain a UCVA, extensive user support may be required. Tutorials and other support will need to be specific to the mobile operating system and should include support for iOS, Android, and Windows devices. Users will need assistance on installing and using the UCVA, will need instructions on how to protect their information, and will need to understand how to handle suspected theft or misuse of their data. Tutorials will need to be kept up to date as changes are made to mobile operating systems or to wallets that impact the installation or use of the UCVA. Finally, it is important to know your audience and deploy technology that makes sense under a given set of circumstances.

Any user-facing interfaces should prioritize a human-centered approach by applying well-established design principles and best practices [46][47][48]. This is important to ensure good usability and satisfactory user experiences. In addition, interactive systems should prioritize accessibility from the outset to achieve the highest possible level of accessibility [49][50].

[23] If an issuer's private DS key is compromised, it could be used to sign false UCVAs before the theft is discovered. There have also been cases where RAs have been compromised and have issued certificates to bad actors.

Designing user interactions with usability and accessibility in mind promotes greater effectiveness, efficiency, and satisfaction for individuals with diverse capabilities and preferences.

7.3.3. Standards Considerations

The standards space surrounding UCVA is nascent, and many of the core standards — particularly for online presentation — are still in development. That said, for the architecture to be successful, an ecosystem must emerge that consists of issuers who create and provide UCVAs that can be transmitted securely to users; end-user software or wallets that can receive, secure, and provide access to those UCVAs; and RPs that can request information or verifications from wallets in a trusted, privacy-preserving, and consent-respecting manner. For this ecosystem to work, the issuer must use a data model and issuance protocol compatible with the end-user software, typically a digital wallet. The RPs or verifiers must be able to interact with the wallet that acts as the interface to the data and should be able to access and verify the public digital signature certificate used by the issuer. To improve privacy, the wallets and supporting standards must allow for selective disclosure, derived attributes, and meaningful consent. There are competing standards under development for the data model. Issuance, user consent, request/response, and revocation protocols are under development as well. This report is not intended to provide direction on which emerging standards agencies or organizations must implement — that will be highly dependent on the use cases being implemented, the applications being deployed, and the supporting ecosystem or community expected to issue and accept the UCVA.

7.3.3.1. UCVA Data Model Standards

Data model standards provide a consistent means of expressing a credential or claim so that systems can be designed to properly handle the data and manage issuance, presentation, and verification. There are two primary data models being explored today:

1. **ISO/IEC 18013 Mdoc Standard.** Mdoc is the data model used today for representing mobile driver's licenses, but the standard can be used to represent other credentials or sets of attributes. Mobile driver's licenses (mDLs) are an example of a UCVA and use the ISO/IEC 18013-5 specification [51] for issuing, storing, verifying, and displaying mDLs. The data model in the ISO specification is the mdoc, which has cryptographic features that are not present on physical driver's licenses, which make mDLs less susceptible to forgery if appropriately implemented and allow revocation to be checked more easily. mDLs issued by several states are now accepted by TSA at a limited number of airports, and AAMVA has produced guidance for states that wish to issue mDLs that comply with TSA requirements. Inspection of the mDL can be done visually or by using a scanner. Although it was designed for mDLs, the ISO mdocs can be used for any type of license or set of attributes.

2. **W3C Verifiable Credentials (VC) Data Model.** The W3C Verifiable Credentials Data Model [52] is one of several alternatives to the mdoc format, and there are several

standards and protocols under development for the required issuance and revocation capabilities, digital signatures, user consent protocols, and support for RP requests for information and verifications. The current version of the W3C model is designed for online use cases and has greater flexibility than the ISO model, but that greater flexibility could make interoperability more challenging. Some training certifications are already being issued using the Open Badges Specification [53], which leverages the W3C Verifiable Credentials (VC) Data Model. Unlike the ISO standard, the W3C data model does not specify all protocols required for full interoperability. As a result, there are multiple competing proposals for implementing the VC data model [54].

7.3.3.2. Encoding and Credential Representation Format

Encoding format and credential representation standards describe the structure of the data and objects that will be transmitted to the verifier during presentation. There are several standard models that have been advanced based on the mdoc and Verifiable Credentials data models:

1. **ISO/IEC 18013-5 - Mobile Security Object (MSO) [51]:** A structured data element that allows the verifier to confirm the accuracy and validity of the data elements in the mdoc data model when transmitted. The MSO is a concise binary object representation (CBOR). It does not contain the mdoc data itself but rides along as part of the payload to support encryption and validation. The rest of the mdoc format is also encoded as CBOR and exchanged during the presentation.
2. **Internet Engineering Task Force (IETF) Selective Disclosure JSON Web Token (SD-JWT) Verifiable Credentials [55]**: The SD-JWT-based Verifiable Credentials provide both a data model and encoding format for the deployment of Verifiable Credentials as JWTs. Though it is referenced as a Verifiable Credential, it does not strictly follow the data model defined by W3C and instead leverages the existing structure of JWT claims. It can be used — as represented by the European Commission's Architectural Reference Framework — to convert a credential stored as a W3C Verifiable Credential or mdoc into a JWT for presentation in online scenarios.
3. **W3C JSON [56] for Linking Data [57] (JSON-LD):** Similar to SD-JWT, this encoding standard supports the representation of Verifiable Credentials in a JSON format. However, unlike SD-JWT, JSON-LD representations of Verifiable Credentials follow the W3C Data Model, incorporate the ability to use linked data signing, and support extensibility by allowing verifiable credentials to have additional context added by members of a supporting community.

As noted previously, there are substantial departures within the overall identity community as to the "best" model for representing UCVAs in online models. It is unlikely there will be a single model to "rule them all," and AVS providers are encouraged to explore each standard relative to their own technology capabilities, the inclinations of their serviced communities, and technologies available to their end users to select a data model and representation that works for their ecosystem.

7.3.3.3. Identity Proofing and Credential Issuance

An authoritative source that wishes to issue UCVAs must also create a secure user interface that allows individuals to request a verified copy of their claim. The list of supported wallets or applications must be provided to users in advance so they can download and install the software required to receive, store, and use their UCVA.

Since the verified claims must be issued to the correct individual, those requesting a verified copy of their attributes or credentials must be identity-proofed at an assurance level that is proportionate to the potential negative impacts that could arise should a bad actor gain access to (and control over) that information. A DIRA should be conducted following the current NIST *Digital Identity Guidelines* to make this determination. Where appropriate, NIST SP 800-63A guidance (as previously discussed) can be applied to help provide confidence in the identity of the individual requesting a UCVA.

Once individuals have been identity-proofed at the appropriate assurance level and have requested their verified attributes or credentials, they must be issued into the designated user's wallet. There are two core standards focused on this to date; both are drafts. However, proprietary processes for mDL issuance have been in place since the development of ISO/IEC 18013-5, though it does not define issuance protocols.

1. **OpenID for Verifiable Credential Issuance (OpenID4VCI):** A draft specification [58] that defines an API for issuing any UCVA, including mdocs and VCs. Support for OpenID4VCI issuance is required for the EU Digital Wallet. The specification uses OpenID Connect (OIDC), which is a widely supported federation standard.
2. **23220-3 Cards and security devices for personal identification — Building blocks for identity management via mobile devices — Part 3: Protocols and services for issuing phase**: Provides general requirements for issuance protocols, processes, and services [59]. Once completed, this will likely include reference to protocols such as OpenID4VCI. This is currently in development with ISO/IEC Joint Technical Committee 1, Subcommittee 17.
3. **Verifiable Credentials API v0.3**: A draft specification [60] for managing the lifecycle of VCs within or across security domains. Endpoints are specified for issuing, retrieving/reading, updating, verifying, and presenting VCs. Additional privacy-enhancing capabilities include functions for deriving credentials and for creating and retrieving presentations.

7.3.3.4. Online Presentation

Online presentation standards define the protocols and processes that enable the user, wallet, and verifier to exchange information to support online (often called unattended) uses of digital wallets, for example the presentation of a UCVA for access to a protected website.

1. **ISO/IEC TS 18013-7 - ISO-compliant driving license Part 7: Mobile driving license (mDL) add-on functions:** Identifies acceptable mechanisms and protocols for the online presentation of ISO-compliant mDLs [61]. There are two methods that are defined; the

first is a basic REST API used to request data directly from the wallet, and the second uses Open ID for Verifiable Presentations.

2. **OpenID for Verifiable Presentations:** A draft specification [62] to allow OIDC to be used for the presentation of VCs to RPs or Verifiers. This has been selected as the online presentation protocol for the EUDI Wallet and is referenced as an acceptable protocol in ISO/IEC TS 18103-7 [61].

3. **Verifiable Presentation Request v2024**: A specification [63] for requesting or querying VCs from wallets or agents that use DIDs (Decentralized Identifiers).

7.3.4. Security Considerations

The verified information and credentials contained within digital wallets will be a target for bad actors and criminal organizations who will attempt to gain access to the information by exploiting weaknesses in the implementations or through social engineering, including attempting to impersonate trusted RPs.

1. **Threat:** Illegitimate RPs will attempt to access the UCVA.

   Mitigation Strategies: One of the benefits of UCVAs is that users have greater control over which RPs they can share their verified information with. Unfortunately, that ability is also a security vulnerability. Wallet providers must therefore establish sufficient vetting procedures to minimize the ability of bad actors to obtain sensitive information directly from wallets, including mechanisms that strongly identify the RP or verifier to the user and that obtain meaningful and granular user consent before releasing information to a verifier. For sensitive information, cryptographic security may not be sufficient; strong governance and access controls that restrict which RPs/verifiers are allowed access to certain attributes and verified information may also be required. One option is for wallets or other UCVA holders to use allowlists of trusted RPs.

   For example, the Apple Wallet restricts verifiers to specific categories and requires that they apply to access the API, justifying their request [64]. Approved RPs are then added to an allowlist. However, Apple's criteria for evaluating potential RPs are not publicly available. Federal requirements for wallet security and RP/verifier vetting, along with third-party certifications to ensure compliance, would increase trust in the ability of this architecture to protect user information. An alternative would be for federal issuers to work with leading wallet providers to restrict RPs to those explicitly approved by the issuer.

2. **Threat:** UCVA can be exfiltrated/stolen from a wallet/device.

   Mitigation Strategies: UCVAs should only be installed into wallets that meet the security requirements necessary to protect the information contained within the UCVA. A risk assessment should be conducted by issuers to understand the minimum software security requirements for each UCVA, including the requirements for both the mobile operating system and the wallet. Best Bring Your Own Device (BYOD) security practices should be considered such as requiring that the mobile operating system be up-to-date, and restricting issuance to those wallets that meet all identified security requirements. Users should also be provided information explaining how to safeguard their UCVA after installation and should be provided a way to request that their UCVA be revoked if they suspect that it has been stolen. The process of revoking UCVA should be easy for end-users to understand and follow. However, it should also prevent accidental revocation, especially if the option to revoke is located close to other frequently used features on the user interface. Additionally, these revocations must be easily discoverable so that an RP that is presented with a UCVA can quickly ascertain whether it is still valid. Several standards are under development for how to manage revocations in a privacy-preserving manner.

For a UCVA that contains especially sensitive or valuable information, the data may need to be cryptographically bound to the device and stored encrypted in a secure element, accessible only by trusted software that the user has accessed using multi-factor authentication.

3. **Threat:** User impersonation.

   Mitigation Strategies: The individual requesting a UCVA must be identity-proofed at an assurance level proportionate to the potential negative impacts that could arise should a bad actor gain access to and control over a legitimate user's UCVA. A DIRA should be conducted to make that determination, and risk assessments should be conducted on the available credentials to understand their ability to withstand impersonation and post-issuance compromise. For guidance on identity proofing controls – inclusive of resolution, validation, and verification processes – UCVA providers should consult NIST SP 800-63A and apply processes to prevent impersonation of applicants. An additional mitigation to consider is restricting the number of UCVAs issued to different individuals that can be stored on a single device. Procedures will also need to be established for handling suspected cases of imposters being issued UCVAs.

4. **Threat:** Compromised digital signature key or digital signature from an illegitimate issuer.

   Mitigation Strategies: Digital signatures are widely used to validate message integrity and to verify that the message was signed by the expected organization or individual, which requires confidence that the key was issued to the correct entity and that the entity who was issued the private signing key has maintained complete control of the key so that it could not have been used fraudulently. RPs can increase their confidence in the digital signature by verifying it using a public key obtained directly from the issuer or a trusted broker and running revocation checks on the digital signature certificate as well as the certificates in its issuance path. An SCVP (Server-based Certificate Validation Protocol) can be used to ensure that no certificate in the chain of trust has been revoked, which could invalidate the end certificate even if the end certificate has not yet been revoked.

For threats and mitigations related to the API calls required to request and transmit the UCVA, please see Section 7.1.3.

7.3.5. Privacy Considerations

1. **Problematic Data Action**: More information is disclosed than is required.

   Mitigation Strategies: A common characteristic of a UCVA is its ability to support technical selective disclosure. This process allows a user to present a subset of attributes signed by the issuer based on either a defined use case or through the allowance for optional versus mandatory attributes. This can also take the form of common derived attributes that can be generated, signed, and included in the UCVA. If DOB is an available attribute, the issuer can digitally sign commonly requested assertions such as “Is 18 or older” or “Is 21 or older.” An alternative approach allows the wallet to derive requested attributes or sign subsets of verified attributes, but that requires trust in the wallet and in the wallet’s verification of the UCVA, and it pushes the RP relationship from a direct one with the issuer to a relationship with the wallet. In the future, additional cryptographic techniques may become practical that could allow UCVA holders to prove, to a high degree of probability, that they possess information, such as proof that they are over 21, without revealing their age and without requiring pre-signed derived attributes. Zero-knowledge proofs are being explored for this purpose. However, such techniques are still nascent.

2. **Problematic Data Action**: Inadequate consent.

   Mitigation Strategies: The approved Routine Uses described in system SORNs are often broad and not well understood by information collection system and downstream counterpart users. When a service is mandated (or strongly incentivized) while alternative systems are deprecated, users may feel they have no choice but to be a part of a system they, along with most of the public, do not fully understand. This is especially true as the uses of collected information relate to third-party information sharing arrangements, for example with law enforcement and intelligence agencies. Meaningful active consent at the point of collection is preferred to implied consent derived from, for example, scrolling through a lengthy privacy notice. Consent in the form of a signature (or other physical act) at the point of collection — and preferably at additional processing points along the information pathway — is an example of active consent (as opposed to passive implied consent). The goal of adequate or enhanced consent is to ensure the public is aware of what could potentially happen to (or *is happening to*) their information *before* that information is collected. If this is not possible, post-collection consent is better than no consent at all. Meaningful consent mechanisms for the public result in greater transparency, public discourse, and buy-in while also protecting the agency should its AVS come under scrutiny at any point.

3. **Problematic Data Action**: Usage tracking by the wallet.

   Mitigation Strategies: Tracking is an additional concern with proprietary wallet platforms that often track usage and monetize information collected about users. Issuers may need to enter into agreements with wallet platforms to prohibit the platform from tracking usage of the UCVAs they have issued.

4. **Problematic Data Action**: Usage tracking by the issuer.

Mitigation Strategies: If RPs query an issuer OCSP responder to check for revocation, that provides issuers the opportunity to track usage of the UCVAs they have issued. Issuers can create policies to only use usage data for the purpose of monitoring for (and prosecuting) fraudulent use or acquisition of a UCVA. Issuers could also provide CRLs to trusted RPs who can then check for revocation without providing the Issuer information regarding the UCVAs verified by the RP. However, providing CRLs publicly would make it more difficult to detect fraudulent use.

8. Conclusion and Next Steps

Authoritative government data is a powerful tool for identity proofing, improved access control, and fraud reduction. Government attribute validation services have the potential to increase equity by expanding access to services for individuals with thin credit files, protect US citizens and taxpayers by reducing fraud, reduce barriers to service access, and increase data accuracy and privacy.

Perhaps the most significant immediate impact of the increased availability of government attribute validation services will be for individuals with thin credit files. Traditional credit reporting systems often exclude those with limited credit histories, disproportionately affecting marginalized communities. By leveraging authoritative data, government agencies can validate the attributes of these individuals, thereby enabling their inclusion in financial and other essential services. Other individuals face obstacles in accessing services due to discrepancies or inaccuracies with their identity data. Government agencies, with their authoritative data, can provide accurate attribute validation, simplifying the verification process and making it easier for individuals to access necessary services. Attribute validation is therefore expected to play an increasingly important role in delivering public sector digital services.

In the attribute validation landscape, there are two primary architectures: traditional API query-based services and the emerging User-Controlled Verified Attributes (UCVA) model, such as those found in mobile driver's licenses (mDLs). A traditional API query-based AVS involves systems directly querying government databases to validate attributes. It is a mature architecture that is in widespread use today.

UCVAs are an emerging approach that gives individuals more control over their personal data. In this model, individuals can present pre-verified digitally signed attributes directly to relying parties. This method has the potential to expand the use of verified attributes and improve data quality. However, implementing such systems requires careful consideration to ensure security, interoperability, and widespread adoption. The standards for using UCVAs for remote identity proofing and authorization are still under development. NIST will continue participating in and monitoring the development of UCVA standards to ensure their usability for future U.S. government use cases.

Choosing an architecture and implementing the technical solution is only one aspect of standing up an AVS, and this report discusses several non-technical considerations that may prove equally challenging. Operational, policy, security, and privacy considerations are all critical when planning for an AVS deployment. A reliable and high-quality data source is fundamental, as is ensuring that implementers are well-versed in the standards that facilitate interoperability and data sharing.

Designing the system with scalability in mind is critical to accommodate future growth in users, functionality, and data volume or type. Thorough pre-deployment testing and red teaming are essential to uncover any performance or usability issues or potential security flaws. Additionally, any AVS project requires a robust change management system to handle updates and upgrades in a controlled manner. Engaging in early and ongoing discussions with all stakeholders, including potential customers, can improve the project's success.

References

[1] Grassi PA, Garcia ME, Fenton JL (2017) Digital Identity Guidelines. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-63-3, Includes updates as of March 02, 2020. https://doi.org/10.6028/NIST.SP.800-63-3

[2] CHIPS and Science Act of 2022, Pub. L. 117-167, 136 Stat. 1366. https://www.congress.gov/117/plaws/publ167/PLAW-117publ167.pdf

[3] Grassi PA, Fenton JL, Lefkovitz NB, Danker JM, Choong Y-Y, Greene KK, Theofanos MF (2017) Digital Identity Guidelines: Enrollment and Identity Proofing. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-63A, Includes updates as of March 02, 2020. https://doi.org/10.6028/NIST.SP.800-63A

[4] Hu VC, Ferraiolo DF, Kuhn DR (2019) Attribute Considerations for Access Control Systems. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-205. https://doi.org/10.6028/NIST.SP.800-205

[5] D. Medhat, A. Hassan and C. Salama, "A hybrid cross-language name matching technique using novel modified Levenshtein Distance," 2015 Tenth International Conference on Computer Engineering & Systems (ICCES), Cairo, 2015, pp. 204-209, doi: 10.1109/ICCES.2015.7393046. Available at: https://ieeexplore.ieee.org/abstract/document/7393046/authors - authors

[6] SentiLink (2022) The Electronic Consent Based SSN Verification Service. Available at https://insight.sentilink.com/hubfs/Whitepapers/sentilink-ecbsv-ssn-verification-service-final.pdf

[7] Memmott M (2013) Hawaiian Woman Gets IDs That Fit Her 36-Character Last Name. NPR. Available at https://www.npr.org/sections/thetwo-way/2013/12/31/258673819/hawaiian-woman-gets-ids-that-fit-her-36-character-last-name

[8] SIL International (2024) ScriptSource Alphabets, Abugidas, Syllabaries and Others (2024). Available at https://www.scriptsource.org/cms/scripts/page.php?item_id=script_overview&sort_scripts_current=script_family

[9] World Atlas (2024) The World's Most Popular Writing Scripts. Available at https://www.worldatlas.com/articles/the-world-s-most-popular-writing-scripts.html

[10] Asia Media Centre (2022) A basic guide to Chinese names. Available at https://www.asiamediacentre.org.nz/features/explainer-chinese-names/

[11] Aribowo, EK, Herawati, N. (2016) Trends in Naming System on Javanese Society: A Shift from Javanese to Arabic. *Lingua Cultura*, 10(2). 117-122. https://doi.org/10.21512/lc.v10i2.1730

[12] Cybercrime Support Network (2024) Deceased Family Member Identity Theft. Available at https://fightcybercrime.org/scams/identity-theft/deceased/

[13] Internal Revenue Service (2023) Deceased Person Identity Theft. Available at https://www.irs.gov/individuals/deceased-person-identity-theft

[14] DAMA-DMBOK: Data Management Body of Knowledge (Technics Publications, Sedona, AZ), 2nd Ed.

[15] McGilvray D (2021) Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (Academic Press, Cambridge, MA), 2nd Ed.

[16] Dalcin EC (2005) Data Quality Concepts and Techniques Applied to Taxonomic Databases. https://doi.org/10.13140/2.1.4440.2562

[17] Maung I, Maydanchik O, Bardmesser J (2017) Data Governance in Big Data Platforms. *Journal of Digital Banking*, (2(1), 29.

[18] Wang RY, Reddy MP, Kon HB (1995) Toward Quality Data: An attribute-based approach. *Decision Support Systems*, 13(3-4), 349-372.

[19] Microsoft (2023) Data Refresh in Power BI.

[20] Federal Information Security Modernization Act of 2014, Pub. L. 113-283, 128 Stat. 3073. Available at https://www.govinfo.gov/app/details/PLAW-113publ283

[21] NIST (2024) NIST Risk Management Framework. Available at https://csrc.nist.gov/projects/risk-management/fisma-background.

[22] Joint Task Force (2020) Security and Privacy Controls for Information Systems and Organizations. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-53, Rev. 5. Includes updates as of December 10, 2020. https://doi.org/10.6028/NIST.SP.800-53r5

[23] Hamilton J (2020) FedRAMP's NIST Rev5 Transition Plan. Available at https://www.fedramp.gov/FedRAMP-NIST-Rev5-Transition-Plan/

[24] Grassi PA, Lefkovitz NB, Nadeau EM, Galluzzo RJ, Dinh AT (2018) Attribute Metadata: A Proposed Schema for Evaluating Federated Attributes. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) 8112. https://doi.org/10.6028/NIST.IR.8112

[25] Carnegie Mellon University (2023) Metadata Guide. Available at https://guides.library.cmu.edu/Metadata

[26] U.S. Geological Survey (2024) Metadata Creation. Available at https://www.usgs.gov/data-management/metadata-creation

[27] Hider P (2012) Information Resource Description: Creating and Managing Metadata. (Facet Publishing, London, UK).

[28] Forshay N, Taylor A, Mukherjee A (2014) Winning the Hearts and Minds of Business Intelligence Users: The role of metadata. *Information Systems Management*, 32(2), 169.

[29] Dawson GS. et al. (2016) An Examination of Effective IT Governance in the Public Sector Using the Legal View of Agency Theory. *Journal of Management Information Systems* 33, no. 4: 1180-1208.

[30] Mhamed N, Jasber K (2012) A Conceptual Framework for Information Technology Governance Effectiveness in Private Organizations. *Information Management & Computer Security* 20, no. 2: 88-106.

[31] National Institute of Standards and Technology (2024) The NIST Cybersecurity Framework (CSF) 2.0. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) NIST CSWP 29. https://doi.org/10.6028/NIST.CSWP.29

[32] Young S (2024) Office of Management and Budget U.S. Digital Services Playbook. Available at https://playbook.usds.gov/

[33] NIST (2024) Catalog of Problematic Data Actions and Problems. Available at https://github.com/usnistgov/PrivacyEngCollabSpace/blob/master/tools/risk-assessment/NIST-Privacy-Risk-Assessment-Methodology-PRAM/catalog-PDAP.md

[34] Davies PS (2021) The Social Security Administration's Death Data: In Brief. Congressional Research Service. Available at https://crsreports.congress.gov/product/pdf/R/R46640

[35] ACT-IAC and Better Identity Coalition (2022) ACT-IAC and Better Identity Coalition White Paper: Identity, Credential, and Access Management (ICAM). Available at https://www.actiac.org/system/files/2022-03/ACT-IAC ICAM.pdf

[36] OWASP Secure API Project. Available at: https://owasp.org/www-project-api-security/

[37] Grassi PA, Newton EM, Perlner RA, Regenscheid AR, Fenton JL, Burr WE, Richer JP, Lefkovitz NB, Danker JM, Choong Y-Y, Greene KK, Theofanos MF (2017) Digital Identity Guidelines: Authentication and Lifecycle Management. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-63B, Includes updates as of March 02, 2020. Available at https://doi.org/10.6028/NIST.SP.800-63B

[38] Grassi PA, Nadeau EM, Richer JP, Squire SK, Fenton JL, Lefkovitz NB, Danker JM, Choong Y-Y, Greene KK, Theofanos MF (2017) Digital Identity Guidelines: Federation and Assertions. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-63C, Includes updates as of March 02, 2020. https://doi.org/10.6028/NIST.SP.800-63C

[39] OpenID Foundation (2024) What is OpenID Connect. Available at https://openid.net/developers/how-connect-works/

[40] Varley M, Grassi P (2023) International Government Assurance Profile (iGov) for OpenID Connect 1.0. Available at https://openid.net/specs/openid-igov-openid-connect-1_0.html

[41] Campbell B, et al. (2015). Security Assertion Markup Language (SAML) 2.0 Profile for OAuth 2.0 Client Authentication and Authorization Grants. Internet Engineering Task Force. Available at https://datatracker.ietf.org/doc/html/rfc7522 - ref-OASIS.saml-core-2.0-os

[42] Singhal A, Winograd T, Scarfone KA (2007) Guide to Secure Web Services. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-95. https://doi.org/10.6028/NIST.SP.800-95

[43] National Institute of Standards and Technology (2020) NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) NIST CSWP 10. https://doi.org/10.6028/NIST.CSWP.10

[44] Digital ID & Authentication Council of Canada (2023) Perspectives on the Adoption of Verifiable Credentials. Available at https://diacc.ca/wp-

content/uploads/2023/05/Perspectives-on-the-Adoption-of-Verifiable-Credentials-1.pdf

[45] Executive Order 14058 (2013) Executive Order on Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government (The White House, Washington, DC), DCPD-202101050, December 13, 2021. Available at https://www.govinfo.gov/content/pkg/DCPD-202101050/pdf/DCPD-202101050.pdf

[46] International Organization for Standardization (2018) *ISO 9241-11:2018 – Ergonomics of human-system interaction – Part 11: Usability: Definitions and Concepts* (ISO, Geneva, Switzerland). Available at https://www.iso.org/standard/63500.html

[47] International Organization for Standardization (2020) *ISO 9241-110:2020 – Ergonomics of human-system interaction – Part 110: Interaction principles* (ISO, Geneva, Switzerland). Available at https://www.iso.org/standard/75258.html

[48] International Organization for Standardization (2017) *ISO 9241-112:2017 – Ergonomics of human-system interaction – Part 112: Principles for the presentation of information* (ISO, Geneva, Switzerland). Available at https://www.iso.org/standard/64840.html

[49] International Organization for Standardization (2008) *ISO 9241-171:2008 – Ergonomics of human-system interaction – Part 171: Guidance on software accessibility* (ISO, Geneva, Switzerland). Available at https://www.iso.org/standard/39080.html

[50] W3C Web Accessibility Initiative (WAI) W3C Accessibility Standards Overview. Available at https://www.w3.org/WAI/standards-guidelines/

[51] International Organization for Standardization/International Electrotechnical Commission (2021) *ISO/IEC 18013-5:2021 – Personal identification – ISO-compliant driving license – Part 5: Mobile driving license (mDL) application* (ISO, Geneva, Switzerland). Available at https://www.iso.org/standard/69084.html

[52] World Wide Web Consortium (W3C) (2022) Verifiable Credentials Data Model v1.1. Available at https://www.w3.org/TR/vc-data-model/

[53] 1EdTech Consortium (2024) Open Badges Specification 3.0. Available at https://www.imsglobal.org/spec/ob/v3p0

[54] Flanagan H (2024) More on the Options and Diversity of Verifiable Credentials. Available at https://sphericalcowconsulting.com/2024/01/15/more-on-the-options-and-diversity-of-verifiable-credentials/

[55] Terbu O, Fett D, Campbell B (2024) SD-JWT-based Verifiable Credentials (SD-JWT VC). (Internet Engineering Task Force (IETF)), IETF Internet-Draft draft-ietf-oauth-sd-jwt-vc-05. Available at https://datatracker.ietf.org/doc/draft-ietf-oauth-sd-jwt-vc/

[56] World Wide Web Consortium (2024) Verifiable Credentials JSON Schema Specification: JSON Schemas for Verifiable Credentials. Available at https://www.w3.org/TR/vc-json-schema/

[57] World Wide Web Consortium (2023) JSON-LD 1.1: A JSON-based Serialization for Linked Data. Available at https://www.w3.org/TR/json-ld11/

[58] Lodderstedt T, Yasuda K, Looker T (2024) OpenID for Verifiable Credential Issuance. Available at https://openid.net/specs/openid-4-verifiable-credential-issuance-1_0.html

[59] International Organization for Standardization/International Electrotechnical Commission (2024) *ISO/IEC CD TS 23220-3 – Cards and security devices for personal identification – Building blocks for identity management via mobile devices – Part 3: Protocols and services for issuing phase* (ISO, Geneva, Switzerland). Available at https://www.iso.org/standard/86783.html

[60] World Wide Web Consortium (2024) Verifiable Credentials API v0.3: An HTTP API for Verifiable Credentials lifecycle management. Available at https://w3c-ccg.github.io/vc-api/

[61] International Organization for Standardization/International Electrotechnical Commission (2024) *ISO/IEC TS 18013-7 – Personal Identification – ISO-compliant driving license – Part 7: Mobile driving license (mDL) add-on functions* (ISO, Geneva, Switzerland). Available at https://www.iso.org/standard/82772.html

[62] Terbu O, Lodderstedt T, Yasuda K, Looker T (2024) OpenID for Verifiable Presentations. Available at https://openid.net/specs/openid-4-verifiable-presentations-1_0.html

[63] World Wide Web Consortium (2024) Verifiable Presentation Request v2024: A data model for requesting presentations of verifiable credentials. Available at https://w3c-ccg.github.io/vp-request-spec/

[64] Apple (2024) Get started with the Verify with Wallet API. Available at https://developer.apple.com/wallet/get-started-with-verify-with-wallet/

Appendix A. List of Symbols, Abbreviations, and Acronyms

AAMVA
American Association of Motor Vehicle Administrators

ABAC
attribute-based access control

API
application programming interface

AV
attribute validation

AVS
Attribute validation service

CA
Certificate Authority

CBOR
concise binary object representation

CBSV
Consent-Based SSN Verification

CRL
Certificate Revocation List

CSP
Credential Service Provider

DIRA
Digital Identity Risk Assessment

DL
driver’s license

DLDV
Driver’s License Data Verification

DMV
Department of Motor Vehicles

DOB
date of birth

DS
digital signature

eCBSV
Electronic Consent-Based SSN Verification

EU
European Union

EUDI
European Union Digital Identity

EV
extended validation

EVVE FOD
Electronic Verification of Vital Events – Fact of Death

FIDO
Fast IDentity Online

ICAO
International Civil Aviation Organization

IDP
Identity Provider

IDVA
Identity Verification API

IEC
International Electrotechnical Commission

IRS
Internal Revenue Service

ISO
International Organization for Standardization

IVES
Income Verification Express Service

JOSE
JSON Signing and Encryption

JSON
JavaScript Object Notation

JSON-LD
JSON for Linking Data

JWT
JSON Web Token

mDL
mobile driving license

mDoc
mobile document

MFA
Multi-Factor Authentication

MSO
Mobile Security Object

NFC
Near Field Communication

OAuth
Open Authorization

OCSP
Online Certificate Status Protocol

OIDC
OpenID Connect

OMB
Office of Management and Budget

PKD
Public Key Directory

PKI
Public Key Infrastructure

RBAC
role-based access control

REST
REpresentational State Transfer

RP
relying party

SAML
Security Assertion Markup Language

SD-JWT
Selective Disclosure JSON Web Token

SLA
service level agreement

SOAP
Simple Object Access Protocol

SORN
system of records notice

SP
Service Provider

SSA
Social Security Administration

SSN
Social Security number

SSNVS
Social Security Number Verification System

SSOLV
Social Security Number Online Verification

SVCP
Server-based Certificate Validation Protocol

TIN
Taxpayer Identification Number

TLS
Transport Layer Security

TSA
Transportation Security Administration

UCVA
User-Controlled Verified Attributes

USCIS
United States Citizenship and Immigration Services

USPVS
United States Passport Verification Service

VC
Verifiable Credential

W3C
World Wide Web Consortium