APPLYING SELF-SOVEREIGN IDENTITY PRINCIPLES TO INTEROPERABLE LEARNING RECORDS

PRINCIPLES, CHALLENGES, AND COMMUNITY GUIDANCE

June 2020

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The T3 Innovation Network™ (T3 Network) comprises more than 500 organizations working together to build an open, decentralized, public-private infrastructure for a more equitable talent marketplace where (1) all learning counts, (2) competencies and skills are currency, and (3) learners are empowered with their own data. A project within the T3 Network—Management and Use of Individual-Level Data Records—developed a charter to explore open, self-sovereign protocols and data management guidance for interoperable learning records (ILRs), which is the focus of this paper.

The term "self-sovereign" arises from the term "self-sovereign identity" (SSI), which is associated with both a set of technical standards and a set of community-promulgated principles seeking to enable a shift toward more individual control over digital identities and personal data. The design of SSI-type systems provides a lens to examine how we might restructure such systems to be more equitable, giving learners better access to, and control over, the management of their learning records while maintaining the verifiability of this data. SSI-based approaches could more readily recognize and empower learners while simultaneously improving educators' abilities to teach and employers' and recruiters' abilities to find workforce candidates who suit their needs.

Interest in portable, interoperable, verifiable digital records has expanded in response to COVID-19. At the same time, proposed solutions such as immunity credentials have brought increased awareness of the need to ensure individual rights and privacy in the process. SSI is not a fully formed solution to these concerns. However, individual rights and privacy has been a primary focus of SSI, building on decades of expertise of individuals in the identity space. As such, SSI technologies and concepts can provide valuable insights to jumpstart our efforts and provide opportunities to improve the talent marketplace for all learners and stakeholders by examining ILR systems' potential risks, such as discrimination, manipulation, over-disclosure, tracking, and lock-in/lock-out. Stakeholders should be encouraged to work toward and implement the following principles and tools, when possible, to mitigate risks that are further explained throughout the paper and highlighted below.

- Verifiable Credentials—support flexible proof mechanisms to ensure the credentials are cryptographically reliable and that the issuing institution stands by the statements contained therein and suspends or revokes issued credentials if necessary.
- Decentralized Identifiers—provide a means for both institutions and learners to establish identity without reliance on a centralized party.
- Decentralized Verifiable Data Registries—publish the status of suspended or revoked credentials without requiring verifiers to contact the original issuing authority.
- Privacy-Promoting Credential Status Checks—inspect the current status of a credential without revealing any additional personally correlatable data about the individual.
- Personal Data Stores—provide standards and protocols to support individual control over sharing and access to their data.
- Selective Disclosure—allow an individual the option to share parts of a larger data set.
- Elective Computation—ensure that any processing of an individual's information is explicitly requested.
- Progressive Disclosure—share the minimal amount of information initially and gradually share more information as the value proposition becomes clearer, rapport is built, and trust is developed.

- Embedded Identity Proofing Attributes—include personally identifiable information directly in a credential, only when necessary and appropriate, to ensure usefulness of these standards with digital identity frameworks.
- Minimizing Collected Data—request the absolute minimum information for any particular transaction.
- Information Fiduciaries—support parties with a legally binding obligation to act in the interest of individuals with regard to the acquisition, processing, and distribution of personal information.
- Governance Frameworks—create structures, roles, and policies of an organization or government to adapt SSI approaches to different domains, resolving questions of trust within different stakeholder groups.

This paper further provides technical details of SSI standards and technologies to describe how implementers (such as ILR pilots supported by the T3 Network) can begin applying technical solutions (as described above) to promote self-sovereign management of individual-level data records. Additionally, talent marketplace ecosystems, like the T3 Network, play an important role in developing governance frameworks and promoting sustainable growth of networks committed to self-sovereign management of learner records and learner privacy, while ensuring ethical, equitable outcomes for learners. ILR pilots and other stakeholders should consider utilizing and further testing the principles, technologies, and community guidance outlined in this paper in low-risk, isolated environments and share their findings and best practices with the T3 Network and broader SSI community.

INTRODUCTION

T3 Innovation Network Background and Overview

The T3 Innovation Network™ (T3 Network) comprises more than 450 organizations representing business, government, education, and technology stakeholders working together to build an open, decentralized, public-private infrastructure for a more equitable talent marketplace. The T3 Network is managed by the U.S. Chamber of Commerce Foundation with support from the Lumina Foundation, Walmart, Google, Microsoft, Educational Testing Service (ETS), and the Bill & Melinda Gates Foundation. Currently, the T3 Network is in its second phase of work, standing up eight signature projects depicted in the diagram below. The project titled "Management and Use of Individual-Level Data Records", will explore open, self-sovereign protocols and data management guidance for interoperable learning records (ILRs,) which is the focus of this paper.

The T3 Network guiding principles (specifically Principle 6)[1] state that individuals should have "access and control of their identity attributes and other public and private information about them." However, as stated in the "T3 Innovation Network Phase 1 Report," the downfalls of the current data ecosystem are that an individual's records are locked away in fragmented organizational silos. These records are often not easily accessible, not portable, not based on interoperable standards, and not under the control of the individual to which they relate—resulting in reduced access and utility. This situation "disempowers the individual, decreases choice and accountability, limits access, and reduces equity."[2]

1   "T3 Innovation Network Phase 1 Report," 5.

2   Ibid, 22.
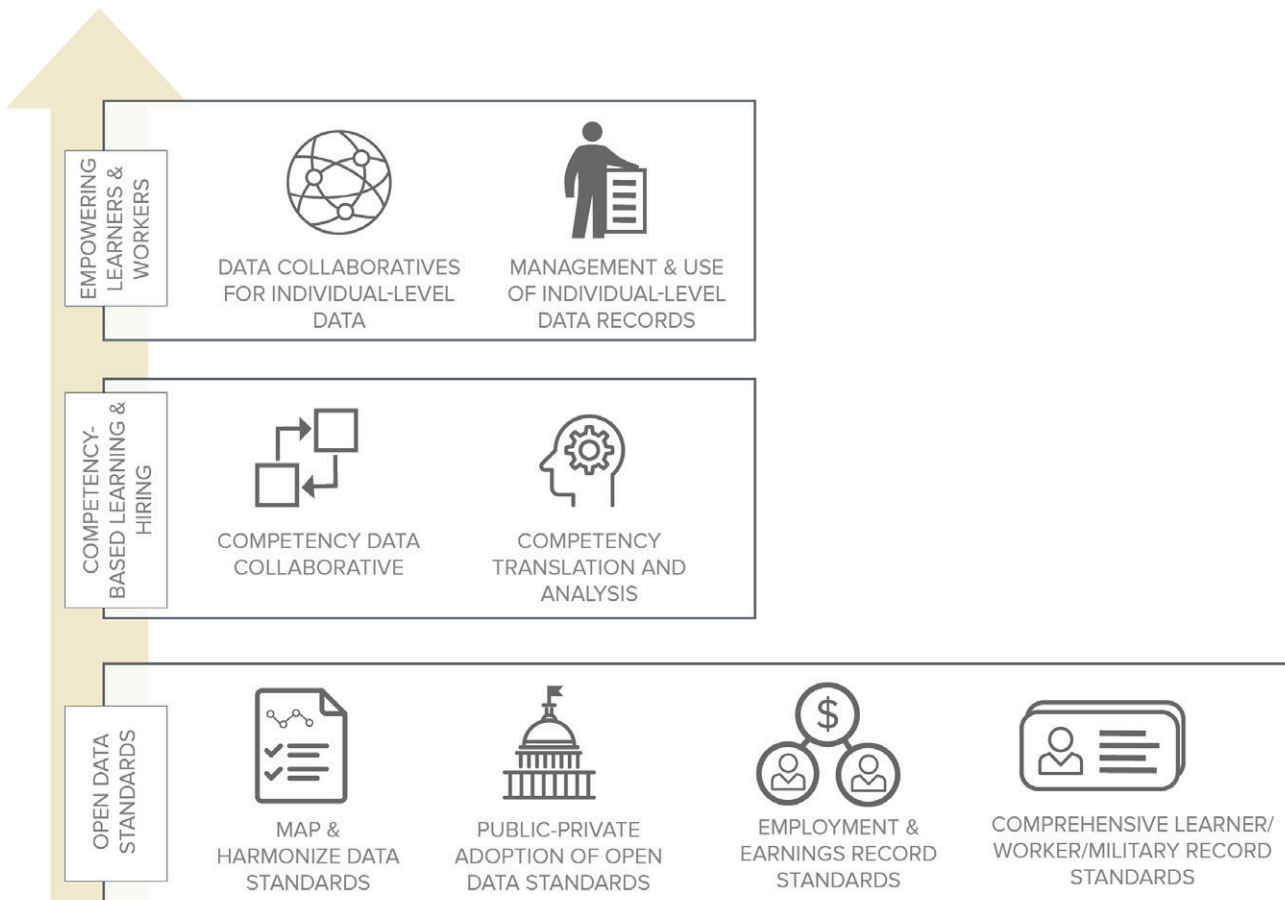
This paper will discuss the risks, opportunities, and provide guidance toward achieving individually controlled, portable learner, worker, and military records. It is important to note that the use of the terms "individual" and/or "learner" in the context of this paper includes students, workers, military personnel, and others as today's talent marketplace is dependent on lifelong learning—whereas an individual takes on different roles at different organizations (sometimes simultaneously) over the course of his or her life.

The protocols and data management guidance offered in this draft paper are the outcomes of ongoing discussions and webinars with SSI experts, government agencies, and participants of the T3 Network Management & Use of Individual-Level Data Records Project. A final version of this report will be published in the second quarter of 2020 to the T3 Network ILR Resource Hub and made available as a resource for use by ILR pilots, standards development organizations, businesses, governments, education organizations, and other stakeholders. The report is published under Creative Commons Attribution 4.0 International License.

Brief Introduction to SSI and How It Can Apply to an ILR

Self-sovereign identity (SSI) refers to an approach to digital identity wherein the goal is to shift control to individuals based on a set of principles and technologies. "At one level, SSI is a set of principles about how identity and personal data control should work across digital networks. At another level,

SSI is a set of technologies that build upon core concepts in identity management, distributed computing, blockchain or Distributed Ledger Technology, and cryptography."[3]

A commonly referenced source of these principles is "The Path to Self-Sovereign Identity," which establishes 10 key principles outlining a new relationship between individuals and their digital identity.[4] The principles outline aspirational goals regarding individuals' rights concerning data collected about them, mechanisms for consent and control of those identities, and the desire for data portability and interoperability to avoid vendor lock-in.

The design of SSI-type systems provides a lens to examine how we might restructure such systems to be more equitable giving learners better access to, and control over, the management of their learning records while maintaining the verifiability of this data. SSI-based approaches could more readily recognize and empower learners while simultaneously improving educators' abilities to teach and employers' and recruiters' abilities to find workforce candidates who suit their needs.

The T3 Network's ILR Prioritized Use Cases describes a set of use cases demonstrating the role of interoperable, individual-level records in improving outcomes for learners.[5] For example, Use Case 4 demonstrates how verifiable, interoperable military records (and subsequent training) could allow a better match of skills in subsequent public or private sector employment. For service members, this leads to faster integration into the workforce and improved job satisfaction. For employers, this leads to improved ability to discover skilled candidates. This paper uses that example to demonstrate how SSI approaches could ensure ongoing verifiability of learner records while giving learners the ability to manage them.

This paper focuses on the aspects of SSI relevant to ILRs, and is not a complete map of the SSI landscape. The "Comprehensive Guide to Self-Sovereign Identity"[6]offers a comprehensive view of the history, goals, and landscape of SSI.

Benefits to an SSI-Based Approach

Several key benefits for learners, issuers, and the ecosystem that has applied SSI-based technological solutions include:

- Learners have control over who can access their record(s), including aspects of their records, and when they are allowed to access them.
- Authentication is cryptographically secured, most typically on distributed ledgers, making the credentials verifiable, and accessible regardless of the state of the issuing organization at the time of verification.
- Online verification of learners and issuers can be secured and streamlined.
- Verifiable credentials can support verification of non-traditional achievements providing evidence of learning in various contexts.

---

3   Preukschat and Reed, Self-Sovereign Identity.

4   Allen, Christopher, "The Path to Self-Sovereign Identity." Note: While many original SSI writings focused on identity or identities, the community has since evolved to focus on the more precise "personal data" provided by the GDPR.

5   "Interoperable Learner Record (ILR) Prioritized Use Cases."

6   Vescent and Young, Comprehensive Guide to Self-Sovereign Identity.

Beyond SSI

This paper uses the broad term "SSI-based approaches" as a shorthand, where necessary, to encompass relevant SSI terms and concepts. This is because, as described in "Brief Introduction to SSI", the term "SSI" can refer to a set of technologies and principles, neither of which are fully formed or up to date. Terminology problems are further compounded by ambiguous scope: the goals of SSI-based approaches extend into areas for which SSI cannot provide complete answers.[7] This paper draws attention to these broader concerns affecting real-world ILR deployments, such as regulations affecting learner privacy and access. Gathering and sharing such information is an area where the T3 Network can provide support to ILR pilot partners and implementers.

POTENTIAL RISKS OF ILR SYSTEMS AND RESPONSES

To illustrate how SSI might alleviate certain problems with ILR systems, we look at ILR Prioritized Use Case 4: Apply for and Accept Employment Opportunities[8], starting with an analysis of potential risks to a learner's agency and privacy, followed by specific areas where these risks may elevate to regulatory considerations. We'll conclude this section with possible responses for these risks, including technical, policy, and architectural approaches.

Use Case 4 describes the experience of Amanda, a college student transitioning from the military, augmented by a university certification program, who is now looking for work. She interacts with multiple systems, using her interoperable learner record to keep track of her accomplishments and share them selectively with various institutions. These systems include:

- Career navigation tool(s)
- Recruiting/Applicant tracking system(s)
- Assessment system(s)
- Background check system(s)
- Human resource information system(s)
- Data collaborative system(s)

We refer to this use case as UC4 and to particular elements within that use case with appropriate sub-indices; for example, UC4.4.a is step "a" in the 4th section of the use case (the "Flow of Events").

Potential Risks

We look at potential risks that face Amanda in her use of these systems: discrimination, manipulation, over-disclosure, tracking, and lock-in/lock-out. In each section of the risks, we reference possible responses based on both SSI and privacy-focused frameworks, which are subsequently described.

---

7   Fry and Renieris, "SSI? What We Really Need Is Full Data Portability."

8   "Interoperable Learner Record (ILR) Prioritized Use Cases."

In the traditional learner context, the general approach to address these risks is laws and regulations, such as the Family Educational Rights and Privacy Act (FERPA). In the SSI world, a more decentralized approach is favored. The SSI approaches aren't substitutes for regulations and the rule of law, but additional mechanisms that can provide relief even in those cases where the rules may not be applied evenly.

DISCRIMINATION

Service providers in Use Case 4 have access to sensitive information and are, unfortunately, in a position to (potentially unwittingly) discriminate based on gender, race, age, and other legally protected categories.

For example, machine learning algorithms re-affirm the biases present in training data, often without the realization of the well-intentioned providers using such optimization algorithms. Offering "custom recommendations" to learners also means restricting those same recommendations to those who don't fit the criteria for the custom offer, sometimes propagating unintended bias and causing illegal discrimination. For example, offering higher management roles to Ivy League graduates re-affirms the privileges of those who made it through Ivy League schools, discriminating against those groups who are underrepresented in the population of Ivy League graduates. Machine learning algorithms can "learn" to apply such bias even if "Ivy League graduate" is never formally labelled in the training data.

Discrimination can also result from systems that are not designed for inclusion from the start. The World Wide Web Consortium (W3C) calls out accessibility, usability, and inclusion in its accessibility fundamentals[9] as critical considerations when considering how users will interact with websites and tools.

Responses

- Selective disclosure
- Embedded identity proofing attributes
- Minimizing collected data
- Governance frameworks

MANIPULATION

Providers of the career navigation tool may place their own interests above the learner's. If their business model is tied to the interests of other parties, they will, inevitably, fine tune their system to better meet the needs of paying customers. If those customers are primarily advertisers and recruiters, there is a risk that the system becomes designed to manipulate more users toward favorable ends in the eyes of these customers. Unchecked algorithmic decision making may lead to features and designs that favor those who pay the tool provider rather than serve the interest of learners.

Responses

- Elective computation (instead of automatic analytics)
- Information fiduciaries

---

9   Henry, Abou-Zahra, and White, "Accessibility, Usability, and Inclusion."

OVER DISCLOSURE

Any time information is requested or shared, there is a risk of sharing more information than is strictly required. For example, this happens when a driver's license, that includes a person's address, is shown to purchase an age-restricted item(s). It also happens when learners share an unedited transcript of their learning history to demonstrate mastery of a particular skill, typically exposing the entire set of coursework undertaken at that institution.

Responses

- Selective disclosure
- Progressive disclosure
- Embedded identity proofing attributes
- Minimizing collected data

TRACKING

A common verification pattern is to check with the initial authority to verify the status. For example, when pulled over at a traffic stop, the police officer runs a real-time check against the department of motor vehicles (DMV) records to verify the license is valid. Similarly, companies may contact the issuing organization (or trusted third party) to verify the graduation status of applicants.

In isolation, contacting the initial issuer is not an unreasonable solution to verify the latest status, and in fact may be required by regulations or norms. However, as records become more accessible through digital means, overdependence on architectures that check with a centralized source creates its own privacy risks. Educational institutions and service providers generally have no legal need to know where a learner is sharing his or her accomplishments. Solutions that require checking with the school to verify claims of accomplishments expose the learner to risks of inappropriate parties learning about his or her activities.

Responses

- Verifiable credentials
- Decentralized identifiers
- Decentralized verifiable data registries
- Privacy-promoting credential status checks

LOCK-IN OR LOCK-OUT

In addition to the risk of tracking through centralized verification solutions, another risk is that a learner who relies on their credentials is unable to use them. For example, if verification of the credentials relies on services, and those services become unavailable, the credentials are rendered unusable to the learner and relying parties.

Other risks presented by reliance on digital credentials is that the learner loses his or her credential (or ability to effectively use it due to name change, etc).

Responses

- Verifiable credentials
- Decentralized identifiers
- Decentralized verifiable data registries
- Privacy-promoting credential status checks
- Personal data stores
- Governance frameworks

Responses to Risks

VERIFIABLE CREDENTIALS

A key enabler of interoperability is the Verifiable Credential (VC) data model, which provides a standard, lightweight data model for presenting statements of verifiable authenticity. It allows representation of different data standards appropriate to the use case.

Verifiable Credentials support flexible proof (verification) mechanisms to ensure the credentials are cryptographically reliable and that the issuing institution stands by the statements contained therein. They can also use privacy-respecting and decentralized credential status mechanisms so that institutions can selectively suspend or revoke issued credentials if necessary.

Lastly, they support advanced proof formats such as selective disclosure and zero-knowledge proofs, enabling use within systems of minimal and progressive disclosure.

In UC4.4, the verifiable credentials data model would benefit all steps by enabling credentials to be expressed in a single format, enabling portability and interoperability across systems, and helping avoid lock-in.

DECENTRALIZED IDENTIFIERS

A design goal for decentralized identifiers (DIDs) is to provide a means for both institutions and learners to establish identity without reliance on a central, trusted third party. This can help learner records survive the failure or aggregation of learning institutions, as some forms of decentralized identifiers retain their usability even after the institution ceases to exist. A proposed benefit is enabling individuals to maintain lifelong learner records by ensuring updatable cryptographic material used for proving control over an identifier. Whether for institutions or individuals, decentralized identifiers also offer the ability to verify the cryptography without relying on the issuing authority.

In UC4.6.a, DIDs obviate the need for a single unique identity across platforms, which increases privacy and flexibility for the learner.

DECENTRALIZED VERIFIABLE DATA REGISTRIES

Decentralized verifiable data registries, such as revocation lists, are a pattern for publishing the status of suspended or revoked credentials without requiring verifiers to contact the original issuing authority. Decentralization in SSI systems is commonly enabled by blockchains—or more generally, distributed ledgers. This helps to avoid the risk of any single point of failure; for instance, the value of the information may persist even if the institution issuing credentials goes out of business or the issuing system is taken offline.

In UC4.4.g, decentralized verifiable data registries help ensure credentials are verifiable by the hiring manager, increasing confidence in moving forward toward the face-to-face interview.

PRIVACY-PROMOTING CREDENTIAL STATUS CHECKS

Privacy-promoting credential status checks are a pattern for checking the current status of a credential without revealing any personally correlatable data to the original issuer of that credential. A naïve, but common, approach to status checks is to simply ask the issuer if a given credential ID represents a valid credential. Unfortunately, that leaks the identity of the subject back to the issuer, often in a way that can be trivially back-traced to the entity performing the check.

In educational contexts, that would mean the educational institution can definitely know that a given learner is triggering verification requests and may disclose the schools or employers the learner is talking with. Privacy-promoting checks avoid this by enabling verification without explicitly disclosing any identifier or personally identifiable information (PII). Possible implementations are discussed in a subsequent section.

In UC4.4.g, privacy-promoting credential status checks the hiring manager to verify Amanda's credentials without revealing additional data Amanda chooses not to share at the time.

PERSONAL DATA STORES

A personal data store is a collection of data repositories in which an individual has the ability to control sharing and access. These repositories might be owned by the individual or simply exposed by a service provider in a way that ensures the individual is in charge—or any combination of the two. A common form of limited data store is a personal blog. It is a place where individuals can store articles they've written that can be shared—either publicly or privately. Data storage providers (such as Dropbox and Google Drive) are generalized data stores while social media sites (such as Facebook and Twitter) are highly structured.

For learners, it is imperative that they are able to manage and use credentials regardless of where they are stored. This may mean an integrated service that pulls in credentials from various sources and gives individuals a user-friendly interface for curating and presenting various representations. It may also be different services (even those of employers) exposing standard interfaces so employees can dynamically export specific accomplishments via a URL on a digital resume.

In UC4.4.c, personal data stores enable Amanda to submit verifiable application data, minimizing tedious duplicate effort.

SELECTIVE DISCLOSURE

Selective disclosure is a technical approach to minimizing disclosure, where either cryptographic or structural mechanisms are used to allow an individual to share just parts of a logically larger data set. For example, an employment record with selective disclosure capabilities would allow a learner to share his or her employment position, start date, and end date without including information about his or her job responsibilities or pay. There are a variety of technical approaches to accomplish selective disclosure, ranging from more advanced techniques—such as zero-knowledge proofs and redaction signatures— to the brute-force method of issuing a range of oversampled, fine-grained statements from which an individual can choose which details they want to expose.

It is important to note that selective disclosure does not preclude the requesting party from requiring more information than the learner is comfortable disclosing. It merely gives the subject of the information the freedom to choose which details are shared and which are not, rather than requiring an all-or-none interaction such as showing all of the information on an individual's drivers license when he or she only need to provide information related to age. It is the reciprocating minimal disclosure policy of the requesting party that would ask for just the minimal detail; in response, a learner uses selective disclosure to share just what is needed.

In UC4.4.c and d, selective disclosure enables Amanda to reveal only the necessary information required by the recruiting/applicant tracking system at the time.

ELECTIVE COMPUTATION

Elective computation is the practice of ensuring that any processing of an individual's information is explicitly requested. Background surveillance and "automated" optimizations often ignore this principle, taking a patronizing role with regard to "what is best" for the customer. Google is a prime example of elective computation in which search results on Google are based on explicit user queries. Users are never surprised that the search results page they are looking at actually has results related to their search. It's understood. In fact, that is the point. Compare this to the feeling an individual gets when a product visited on one site is advertised on subsequent sites due to ubiquitous ad technology "retargeting" users automatically.

In UC4.4.a, the career navigation tool should require explicit direction from Amanda to perform this search for jobs that best match Amanda's background. However, the matching of career goals and background requires processing of information with potential for privacy harms, including discrimination.

PROGRESSIVE DISCLOSURE

Progressive disclosure is the practice of sharing the minimal amount of information initially and gradually sharing more and more as the value proposition becomes clearer, rapport is built, and trust is developed. One way in which we see this every day on the World Wide Web is that visiting most online retail stores does not require a membership, credit card, or address. Individuals also normally don't start by telling the store what they need, but often clicking on links and entering search terms discloses an interest in particular types of products. Once an individual is ready to make a purchase, sharing payment

and shipping details is necessary to complete the transaction. In this manner, most online stores practice progressive disclosure as a matter of course.

Applying progressive disclosure to UC4.4.b, the elements selected for sharing with applicant tracking systems would initially be the bare minimum required—e.g., the attestation of a completed degree in a specific field or certifications earned—without regard to which school, what grades, or when they were earned. As a candidate progresses through an application process, further disclosures may be requested, giving an opportunity for the learner to divulge more details as he or she feels more comfortable in the process. The requirement to submit sensitive personal data may be deferred to later stages of the recruiting/hiring process, potentially even until after a job offer is made "pending additional documentation." For example, personal data related to name, age, and even the school attended could be deferred to protect an individual's rights against discrimination.

EMBEDDED IDENTITY PROOFING ATTRIBUTES

Even with the availability of sophisticated selective disclosure approaches, it is sometimes necessary and appropriate to include PII directly in a credential. This is how driver's licenses and passports enable a human observer to evaluate whether someone presenting an ID is the person who was issued that ID. In addition to checking the physicality and various traits of the ID card like holograms and microprinting, an individual checks the photo, age, eyes, hair, and other "soft" biometrics.

Combined with minimal and selective disclosure, embedding identity proofing attributes in a given credential can enable employers to more robustly correlate a given credential with their new hire. By the time learners are filling out the paperwork for employment, they are required by law to reveal their full legal name with appropriate documentation. Comparing the name revealed in the process to one embedded in a credential is entirely appropriate, provided that such embedded attributes may be selectively disclosed and are only asked when absolutely required (minimal disclosure,) such as in an information exchange using progressive disclosure.

In UC4.4.g, embedded identity proofing attributes allow efficient background checks within the same credential exchange processes.

MINIMIZING COLLECTED DATA

Data minimization is the policy and practice of requesting the absolute minimum information for any particular transaction. When ordering a pizza by phone, the clerk doesn't need to know the person's age, but the clerk does need the person's address. Often service providers ask for more information than needed, placing individuals in a bind: refuse and they can't access the service or accept and they expose themselves to unnecessary risk. The practice of data minimization goes hand in hand with progressive disclosure, favoring the minimum at each stage in a relationship rather than immediately requiring everything up front.

The National Institute of Standards and Technology's (NIST's) "Digital Identity Guidelines: Enrollment and Identity Proofing" highlights the need to minimize collected data as a privacy consideration, and furthermore highlights the data security risks related to retention of PII, which can "become

vulnerable to unauthorized access or use." Minimizing the amount of data collected reduces this threat, and "encourages trust in the identity proofing process."[10]

In UC4.4, minimizing collected data increases Amanda's privacy and her trust in the process.

INFORMATION FIDUCIARIES

Information fiduciaries (also called data fiduciaries) are an emerging concept of parties with a legally binding obligation to act in the interest of individuals with regard to the acquisition, processing, and distribution of personal information.[11] Similar to fiduciary obligations that bolster societal willingness to trust doctors, lawyers, and accountants, this is a mechanism to help individuals navigate information decisions that have become too complicated for the typical person to understand.

In UC4.4.a, the provider of the career navigation tool should be a fiduciary, with an explicit commitment to putting the interest of the learner over that of the service. This would ameliorate concerns over exploitation by the service, including unintended scope creep that sometimes favors advertisers and paying customers (like recruiters) over the rights and interests of individual learners.

Fiduciaries are not a panacea for bad actors—there are malpractice suits for a reason. However, fiduciaries provide a well-established solution for resolving conflicts of interest when individuals, such as learners, must rely on the good faith of professional service providers.

GOVERNANCE FRAMEWORKS

Governance frameworks, referring to the structures, roles, and policies of an organization or government, are also commonly used to adapt SSI approaches to different domains, resolving questions of trust within different communities of education, government, and banks.[12]

In UC4.6, governance frameworks help mitigate the points of failure, helping ensure records use interoperable formats, are available to the learner, and use privacy and other best practices to minimize bias in the application process.

SSI for ILRs

Of course, there are other lessons to be learned from SSI that can improve ILR systems. This summary focused on a single use case to highlight five risks that could be addressed in part by applying SSI approaches. Self-sovereign identity itself is an emerging understanding of how modern identity systems can respect the human dignity of individuals. It is an ongoing conversation that continues to challenge its own ideals and ideas as developers, regulators, and administrators work together to bring SSI into practice. As employers, educators, and institutions explore how best to recognize the principles of SSI for learners

---

10   Grassi and Fenton, "Digital Identity Guidelines: Enrollment & Identity Proofing," 26.

11   Balkin and Zittrain, "A Grand Bargain to Make Tech Companies Trustworthy."

12   Windley, "Four Pillars of an SSI Network."

and workers there will undoubtedly be additional insights and approaches developed. We anticipate that these approaches can be gradually integrated into existing systems. Together we can explore and develop specific standards and practices for the talent marketplace that make the most of SSI.

SSI ECOSYSTEM ROLES AND HUMAN EXPERIENCE

An essential SSI concept is the *credential*, which is a set of claims (or assertions about a subject) made by an issuer. This definition of credential includes degrees, certificates, certifications, licenses, and digital badges, among others. Credentials are customarily verified by humans, and while humans are central to the SSI approach, the technology and architecture run on the congruence of standards, cryptography, distributed ledgers, and front-facing applications that enable machines to perform the verification. The next section describes how credentials are verified in an SSI ecosystem by providing an explanation of the roles and human experience.

Roles and Relationships

There are four key roles in an SSI ecosystem[13]:

- The *subject/holder*[14] is the individual (learner) in a central role in the exchange of their verifiable credentials.
- The *issuer* is the agent (person or organization) that created the verifiable credential.
- The *verifier/relying party*[15] is the agent that receives the verifiable credential and typically wants to verify its authenticity.
- A *verifiable data registry* is a human-governed system that maintains information needed to verify a credential.

In a verifiable credentials ecosystem, an issuer issues credentials about a subject. The subject presents credentials (in the form of a verifiable presentation) to a relying party. The relying party will typically want to verify credentials, which it does through a standards-based verification protocol (implemented by software/services). This relies on the credential itself and instances of a verifiable data registry, which maintains information such as credential status and the issuer's cryptographic keys. Verifiable data registries enable the issuer to control the lifecycle of a credential and indicate authentically issued credentials, without requiring relying parties to directly contact the issuer.

---

13   Sporny et al., "Verifiable Credentials Data Model 1.0."

14   The architecture allows for separate subject and holder roles (e.g., if the credential is about a child, but the parent is the holder), but these will be the same for the purposes of this document (and we'll refer to them interchangeably).

15   The architecture allows for separate relying party and verifier roles, but these will be the same for the purposes of this document (and we'll refer to them interchangeably).
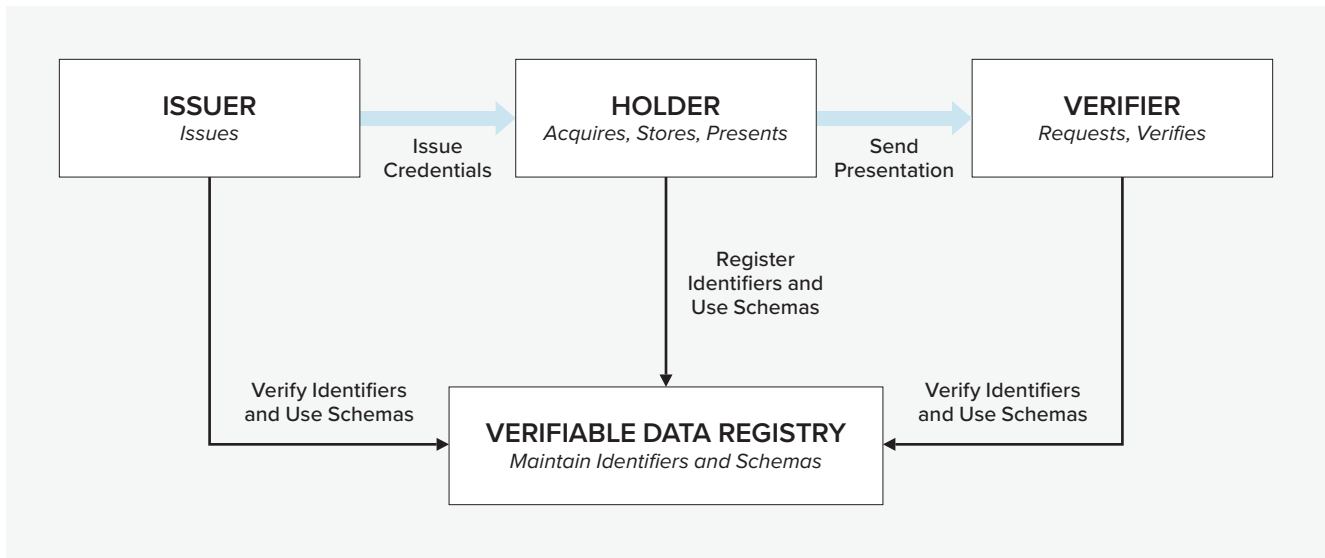
Placing the individual at the center of credential exchange necessitates a shift in the way credentials can be verified. This happens as follows:

- The issuer creates a verifiable credential to and about the subject/holder.
- The subject/holder saves the credential to his or her mobile or web-based ILR wallet/app.
- Using his or her wallet or app, the subject/holder presents a credential or the relying party requests the credential.
- The relying party then initiates the verification process (through a code library or service provider).
  - » Confirms the credential has not been tampered with (typically via a cryptographic signature).
  - » Checks the authenticity of the credential (i.e., confirms that the issuer is a known party and credible for the assertions made in the claims). This is typically done by checking a cryptographic signature in the credential against the issuer's "identifiers" available from an instance of a verifiable data registry.
  - » Checks that the credential is in good standing (e.g., has no revocations) against an instance of a verifiable data registry.

As shown in the SSI ecosystem figure above,[16] the holder of the credential(s) has control over the storage and sharing except for those credentials managed by other parties, including the original issuers and relying parties. Credential hosting (public or permissioned) is provided by an issuer or trusted party.

The end result is a process that enables verification without a technical need to directly consult the issuer. At the same time, the issuer can control the state of credentials it issues (e.g., valid, revoked) through the use of the verifiable data registry, which is also consulted during verification.

---

16   Sporny et al., "Verifiable Credentials Data Model 1.0."

How Does SSI Work with ILR?

ILR Prioritized UC4 from above provides a frame to describe each role and the experience of a job applicant named Amanda when verifiable credentials in an SSI ecosystem are in use:

> Amanda is a first-generation college student who recently transitioned from the military and enrolled at a community college near her home in a mid sized Midwest metropolitan area. She enrolled in an IT program, to build on the military training and courses that she already completed, to earn an industry certification in network administration. She then applied for and enrolled in a university certificate program that provided another internship and industry certification. She recently completed her university program, including an internship, and received a certification. She is now interested in applying for jobs in network administration and cybersecurity.
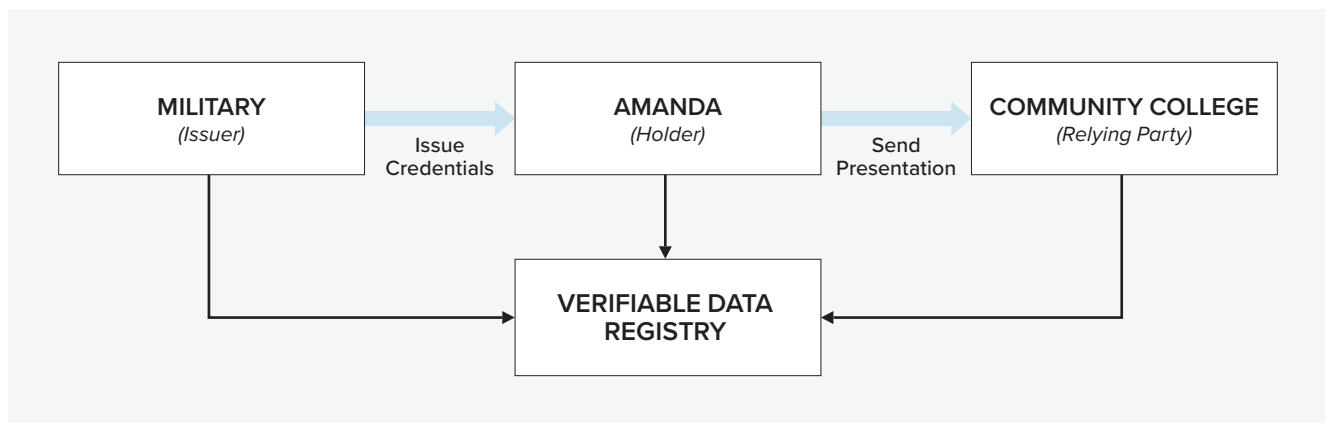
Beginning with Amanda's military credentials, in an SSI-based system, the military (issuer) issues credentials to Amanda (subject/holder). Amanda can upload them to her mobile or web-based ILR wallet/app. When she applies to the community college (verifier/relying party), she sends her military credentials from her wallet to the community college, which should be able to verify through the verifiable data registry that the credentials have not been tampered with, are authentic, and still have a valid status.

The community college is able to verify Amanda's credentials, and her application is successful. On completion of events (such as completing a class, earning a certification), the community college would issue verifiable credentials—one of these being the industry certification in network administration. Amanda adds these credentials to her ILR wallet, organizes them, and then uses this certification and a few chosen military credentials to apply for the university program, which would determine the validity of Amanda's credentials through the same process. The community college may belong to a different verifiable data registry than the military, but the registries would serve the same purpose.
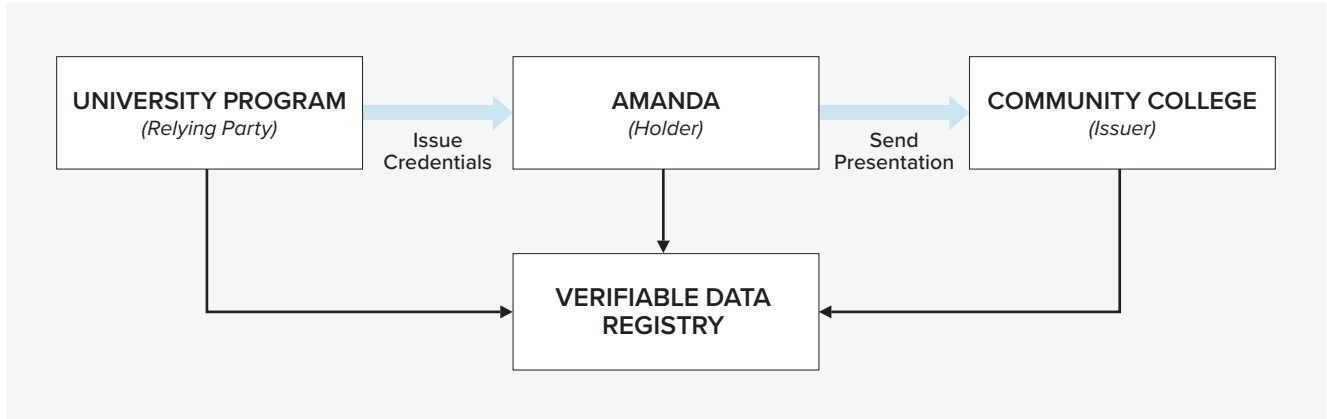
When Amanda wants to apply for jobs leveraging her skills, she has a variety of verifiable credentials in her ILR wallet, including her military records, industry certification in network administration (and other industry certifications), internship records, and university program certificate. She could apply to two different employers with tailored presentations of her credentials from her ILR wallet. The two potential employers (the verifier/relying parties) would follow the same verification process as the community college and university program.
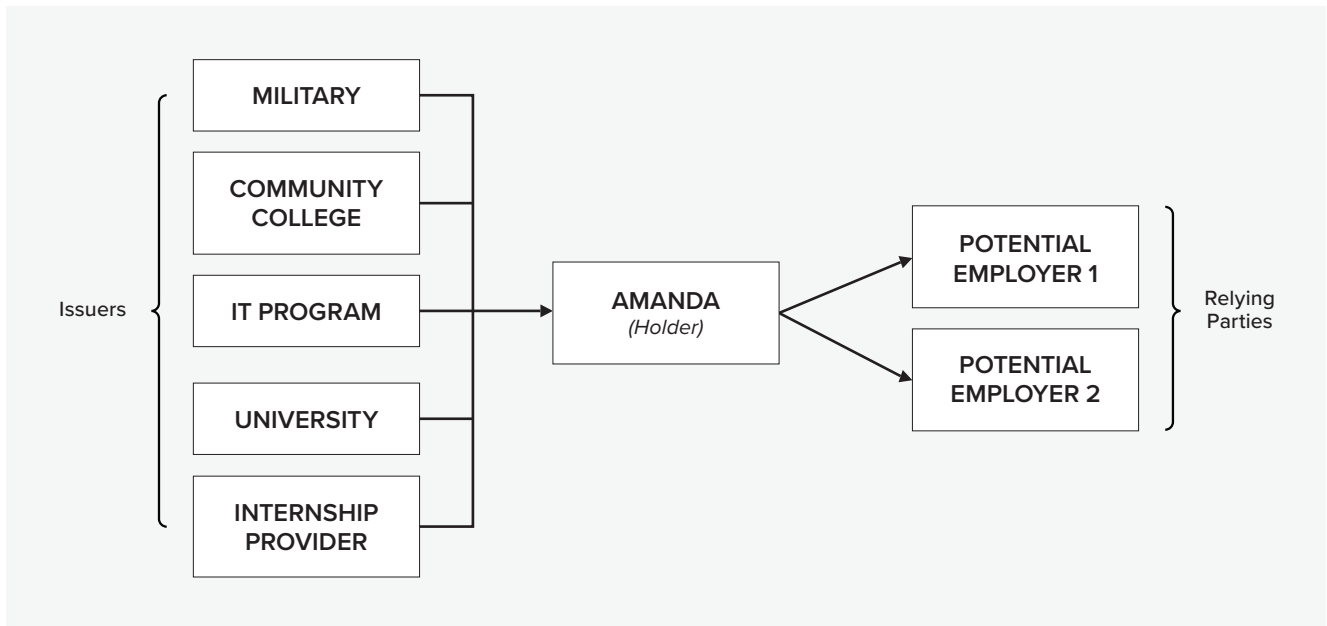
THE PATH TOWARD SELF-SOVEREIGN MANAGEMENT OF INDIVIDUAL-LEVEL DATA RECORDS
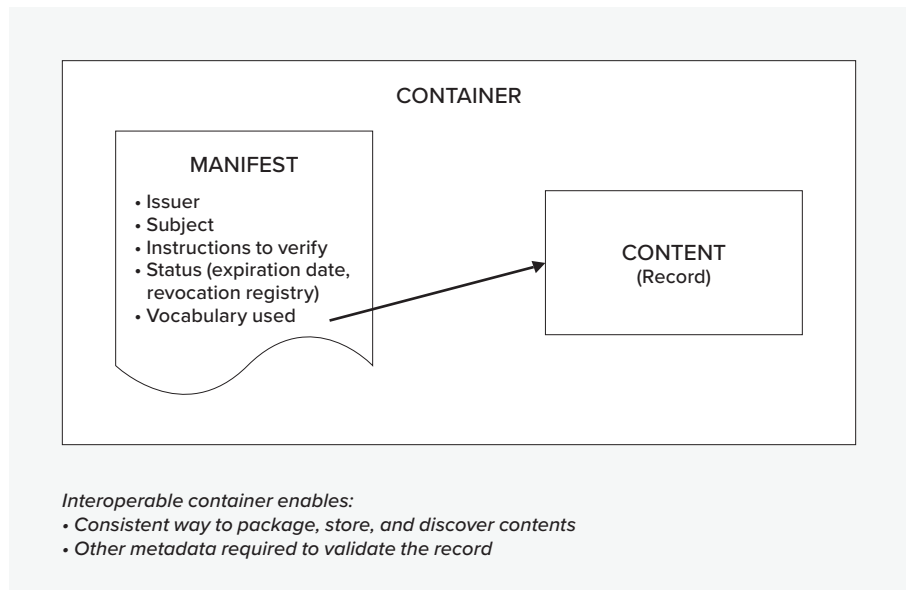
PART 1: STANDARDS AND TECHNOLOGIES

The goal of this section is to describe the technical details of SSI standards and technologies to describe how implementers (such as ILR pilots supported by the T3 Network) can begin implementing the technical solutions described above. This section also describes opportunities for research and development-focused projects to contribute to forward-looking technologies that promote learner privacy and control over individual-level data records.

Verifiable Credentials

A verifiable credential (VC) is a tamper-evident credential, as described by the W3C Verifiable Credentials Data Model specification.[17] A VC serves as a lightweight, interoperable wrapper around its content, which is entirely up to the issuer.[18] A VC supports use cases beyond the credentialing scenarios described above, in which it can flexibly express a driver's license, an educational degree, a certification of completion of an online course, employment records, and other records. VCs support linked data—allowing the content to be anchored to competency definitions and other information relevant to the credential—enabling digital and semantic interoperability among different types of credentials. These features could support the ILR ecosystem by enabling learner-focused tools (such as ILR wallets) to store credentials in an interoperable way.

It's helpful to think of a VC as a container enabling a range of functionality. First it offers an interoperable means of packaging the content and inspecting key metadata about the contents. These allow tools and services consistent ways of storing and inspecting the contents. For example, regardless of the type of record, an ILR wallet would be able to store and expose metadata about the kind of record for grouping and searching.

CONTAINER

MANIFEST

• Issuer
• Subject
• Instructions to verify
• Status (expiration date, revocation registry)
• Vocabulary used

CONTENT
(Record)

Interoperable container enables:
• Consistent way to package, store, and discover contents
• Other metadata required to validate the record
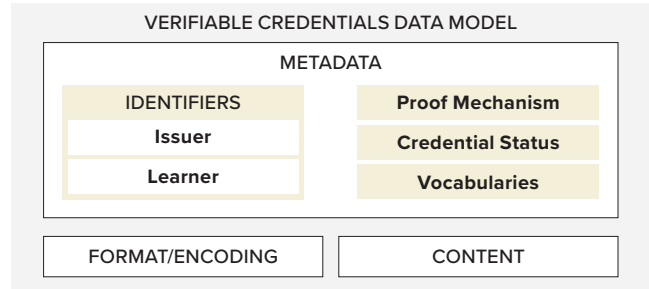
---

17   Sporny et al., "Verifiable Credentials Data Model 1.0."

18   No additional conclusions about the truth of the statement can be made; simply that the issuer has asserted this claim.

This lightweight structure also allows the content of the container to use vocabulary that makes sense in its domain, while using vocabularies and linked data to allow service providers to use these credentials without needing to understand a different format for every domain.

More precisely, a verifiable credential can be viewed as a template for expressing:

- A common set of metadata about the credential
- A container for the credential content
- Flexible format/encoding (such as JSON-LD, JSON, and XML)

VERIFIABLE CREDENTIALS DATA MODEL

METADATA		
IDENTIFIERS		Proof Mechanism
Issuer		Credential Status
Learner		Vocabularies

FORMAT/ENCODING	CONTENT

This standard set of metadata provides a consistent means to discover information about:

- Whether the issuer is authentic ("issuer identifiers")
- How to authenticate the learner ("learner identifiers")
- How to verify the credential ("proof mechanism")
- Credential status information (such as expiration date and how to determine revocation status)
- Vocabulary/taxonomies used in the credential

Recommendations:

- The Verifiable Credential Data Model can inform key metadata for the interoperable ILR wrapper while allowing flexible content.
- Verifiable Credentials Implementation Guidelines 1.0[19] contain valuable recommendations for implementers authoring (or designing systems to author) credentials, such as enabling content integrity checks and providing evidence.

Verifiable Credentials Proof Mechanisms

The Verifiable Credentials Data Model is designed for compatibility with a variety of existing and emerging proof mechanisms. These mechanisms, as well as many other useful topics for implementers, are covered in the "Verifiable Credentials Implementation Guidelines" (Guidelines).[20] This section is not intended to override the Guidelines, but to provide commentary and recommendations based on the current state of these approaches.

---

19   Chadwick et al., "Verifiable Credentials Implementation Guidelines 1.0."
20   Chadwick et al., "Verifiable Credentials Implementation Guidelines 1.0."

COMMONLY USED PROOF FORMATS: LINKED DATA PROOFS AND JSON WEB TOKENS

Linked Data Proofs (LD Proofs)[21] and JSON Web Tokens (JWTs)[22] are the most commonly used proof formats in current verifiable credential deployments. It can be difficult for implementers to understand the relative merits of each, so below is a unified side-by-side comparison table of the proof formats to summarize the Guidelines' tables.

FEATURE	JWTs	LD Proofs
Mature libraries in a wide range of languages	✓	
Open-world data modeling support		✓
JSON-native file storage		✓
Part of native platform toolchain (doesn't require special libraries)	✓	
Implementation complexity in the context of verifiable credentials	**No clear winner**	

Mature libraries in a wide range of languages: JWTs have the advantage that they are familiar and widely used, with mature libraries available in a variety of languages. LD Proofs, on the other hand, are a newer standard with less widespread library support at the moment.

Open-world data modeling support: LD Proofs go hand in hand with the open-world approach of linked data (as in the "linked data" of JSON-LD), which enables semantically rich, unambiguous statements by specifying a "context" in which statements are made. This enables anchoring learner credentials to competency frameworks, taxonomies, and ontologies, which is precisely why linked data approaches are appealing for enabling portable, interoperable ILRs.

JSON-native file storage: LD Proofs use a "canonicalization" step that guarantees a consistent ordering before signing and verifying,[23] and therefore the signed credential can be stored as a JSON-native document. The ability to use standard JSON tools on the signed document after issuance has previously

---

21 Longley and Sporny, "Linked Data Proofs."

22 Jones, Bradley, and Sakinura, "JSON Web Token (JWT)."

23 Recall that JSON objects are an unordered set of name/value pairs. Accordingly, JSON libraries do not guarantee a consistent ordering of an object's properties. This introduces the following question: How does one guarantee the consistency of the JSON payload when signing and verifying JSON? LD Proofs handle this through a canonicalization step, whereas JWTs sign and verify a base64 encoding of the JSON payload. The result of the latter is a string—not a JSON document—that is needed for subsequent verification.

been considered an advantage for educational verifiable credentials,[24] but this may change over time as verifiable credential tool sets improve. Note that this canonicalization step is considered a downside by JWT proponents, who claim it introduces additional computation and complexity.

Part of native platform toolchain: Unlike JWTs, JSON-LD libraries are relatively new and generally not part of native platform toolchains.

Implementation complexity in the context of verifiable credentials: In the end, there is no (current) clear winner in implementation complexity in the context of verifiable credentials. As a new standard, there is not an abundance of clear guidance for working with JSON-LD, and many implementers find it difficult to get up to speed. On the other hand, JWTs have challenges as well—the VC data model describes three different JWT-encoding implementation choices, which introduces complexity for implementers.

We believe a promising approach for near-term implementations is to use JSON-LD canonicalization with JSON Web Signatures (JWS), as demonstrated in the lds-jws signature suite.[25] This approach uses LD canonicalization while using well-known and well-audited JWS libraries.

EMERGING PROOF MECHANISMS

The proof formats described above are compatible with more advanced (emerging) proof mechanisms. The potential of increased privacy offered by these approaches warrants additional study through pilots specifically focusing on usage scenarios and fitness.

"Zero-knowledge" proofs allow proving a statement without revealing additional information. While not commonly adopted in current systems, there is a growing interest due to the increased privacy offered by this approach. The BBS+ Signatures 2020 Draft Specification[26] was recently proposed as an easily implementable approach to zero-knowledge proofs.

Another interesting category of proof mechanisms warranting further development is those that enable "progressive disclosure." The progressive disclosure proof mechanisms primary needs include tooling for users, such as how to choose what to disclose, and Resource Data Graph (RDF) patterns for selective disclosure, as discussed in the "Part 2: Beyond Technologies" section.

Recommendations:

- Consider using lds-jws signature suites for near-term implementations of VC proof mechanisms.
- Explore zero-knowledge and progressive disclosure proof usage scenarios and tooling support.
- Relevant standards groups and communities should continue to improve implementation guidance for VC proof mechanisms.

---

24 Duffy, "How Blockcerts Uses JSON-LD Normalization (Canonicalization) during Verification."

25 Steele, Linked Data Signatures for JWS.

26 Looker and Steele, "BBS+ Signatures 2020."

Decentralized Identifiers

A verifiable credential includes Uniform Resource Identifiers (URIs) to refer to the issuer and learner (subject/holder). This URI can take a variety of forms, such as a web address (i.e., an issuer profile hosted at a URL), as is commonly done in current deployments of Open Badges v2.[27] Another type of URI that can be used in VCs to identify both issuers and learners is called DID.[28] A DID is a new type of identifier intended to support permanent, verifiable, decentralized digital identities. DIDs were developed alongside VCs as a privacy-enabling way to prove control over an identifier, associated with authentication methods, signing keys, and other secure means of interacting with the subject.
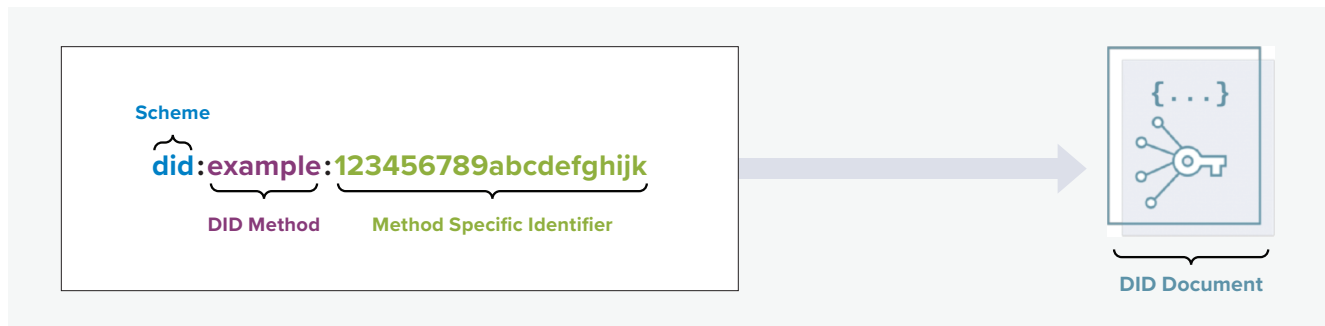
DIDs offer theoretical improvements to longevity over traditional public key infrastructure (PKI) management, for example, by allowing for planned key rotation, or enabling for recovery in case of device loss or other catastrophic failures. However, more work needs to be done to realize these theoretical improvements, both technically and more critically from a usability perspective. Usability studies of learner-focused applications suggest we are still in the early phases of understanding how a user will successfully interact with this ecosystem. Many DID implementations are based on blockchains, and while these may offer improved longevity and control, they introduce additional risk associated with new technical dependencies. We recommend further research and development into these promising areas.

One approach to DIDs that may offer near-term advantages is the ability to use DIDs as an interoperability "bridge" between centralized, federated, and decentralized digital identity systems. For example, the did:web method proposes using web-based DIDs to "bootstrap trust using a web domain's existing reputation" as a way to incentivize adoption.[29] This provides benefits without bringing in the additional uncertainty of a blockchain-based DID method. While these DIDs may not be fully learner controlled, they can provide the learner increased control and usefulness, for example, if implemented to bridge into existing authentication mechanisms such as OpenId Connect.[30]

27   "Open Badges v2.0."

28   Reed, Sporny, and Sabadello, "Decentralized Identifiers (DIDs) v1.0."

29   Terbu, Zagidulin, and Guy, "Did:Web Decentralized Identifier Method Specification."

30   Terbu et al., "Self-Issued OpenID Connect Provider DID Profile."

Recommendations:

- Explore URI- or DID-based implementations that demonstrate increased usability to learners, such as integrating with secure authentication mechanisms.
- Explore pilots that test fitness, usability, and tooling around more advanced DID features and recovery scenarios.

Decentralized Verifiable Data Registries

Decentralized verifiable data registries are used to manage credential status, such as whether a credential has been revoked, and for management of identifiers to enable DID discovery. These are commonly implemented on blockchains, but can also be based on distributed ledgers such as the InterPlanetary File System (IPFS).[31] These approaches allow the issuer to keep information needed during credential verification up to date on the latest status, without the need for verifiers to directly contact the issuer.

For a comprehensive study of different approaches, refer to NIST's "Taxonomic Approach to Understanding Emerging Blockchain Identity Management Systems."[32] A brief summary of design considerations from that report, with commentary, are below:

- Design decisions behind any given blockchain-based IDMS result in different implications for security, privacy, governance, and user guarantees. Similarly, characteristics of any given SSI system may not apply broadly.
- To understand the design decisions, it's helpful to consider separate architecture decisions around identifier management and credential management.
- For both factors (id and credential) there is variation in what is stored on-chain (on the blockchain) and privacy implications thereof.
  » Storing a credential on a blockchain is strongly discouraged for privacy reasons. Even in non-public, permissioned chains, there are concerns with doing this due to: the inability to delete a blockchain entry; the indefinite storage of records; and the distributed storage of records in indeterminate jurisdictions, etc. This risks compliance with right-to-be-forgotten and erasure requirements, storage minimization and limitation principles, and data localization requirements, among others.
  » Some approaches anchor a hash or merkle root on-chain, but still there can be privacy concerns as the resulting data is often pseudonymous rather than anonymous. This can introduce the same privacy risks as the raw data itself.[33]
  » Still others leave credentials entirely off-chain, only anchoring DIDs on-chain (i.e., signing key material). These identifiers may still be linkable to personal data or personal information, and therefore even this approach does not eliminate privacy, data protection-related concerns, or sidestep legal requirements.
- There is significant variation in system governance. This includes factors such as ownership and funding, and whether the network is open or permissioned, along with other internal rules for participation and software management. These factors are key to whether its participants consider the system trustworthy.

---

31   More precisely, IPFS and InterPlanetary Name System (IPNS).

32   Lesavre et al., "A Taxonomic Approach to Understanding Emerging Blockchain Identity Management Systems." This report describes both top-down and bottom-up organizational structures, the latter of which they associate with SSI.

33   "Introduction to the Hash Function as a Personal Data Pseudonymization Technique."

Recommendations:

- For blockchain (or distributed ledger-) based credential solutions, it's helpful to separate out identity management and credential management considerations. NIST's "Taxonomic Approach to Understanding Emerging Blockchain Identity Management Systems" provides a useful framework for architectural approaches.
- Decentralized verifiable data registries may offer advantages from an availability perspective, but it's also critical to consider privacy implications of anything stored on a blockchain.
- System governance of blockchains themselves is also essential to consider. Blockchains are not a silver bullet; responsible system design is still necessary.

Privacy-Promoting Credential Status Checks

Issuers need the ability to update the status of credentials, such as indicating a credential has been revoked, but the Verifiable Credentials Data Model indicates, as a desirable ecosystem characteristic, that this should be done in a privacy-promoting way. For example, revocation checks should not reveal any additional information about the learner. The Verifiable Credential Data Model also recommends avoiding implementations of revocation status checks that allow the issuer (or similar parties) to know that a verification check has been performed on a given credential (or learner).

There are a range of ways to accomplish this. Some approaches publish a list of hash of the credentials that have been revoked. Anyone checking a given credential simply performs the (well-known) hash, downloads the current list, and checks to see if the hash of the credential in question is on that list. Combining with the previous recommendation, this list should further be stored in a decentralized fashion.

The recently-proposed Revocation List 2020[34] draft specification offers improved privacy characteristics by publishing a bitstring containing revocation status of a batch of credentials. Cryptographic accumulators are another privacy-promoting approach in which verifiers perform a simple inclusion check without learning any additional information beyond whether the credential is in the set[35].

Recommendations:

- Implement credential status registries in a privacy-promoting manner. A simple option for current implementations is to publish a pseudonymous list, such as credential hashes.
- Investigate pilots for more advanced options, such as cryptographic accumulators.

Personal Data Stores

While more standards work is needed to support storage and exchange of credentials, implementation of ILR wallets can begin based on a common ILR wrapper data model, such as that provided by VCs. Usability and fitness of learner-facing tools are essential because they introduce entirely new ways of interacting with one's data. Existing attempts to implement learner-focused credential management tools have seen challenges. For example, a Georgia Tech study on the usability of the Blockcerts

---

34   Sporny and Longley, "Revocation List 2020."

35   Tobin, "What Goes on the Ledger."

application concluded that the experience, workflow, and even concepts were confusing to learners. This led to concern that this new method was less secure/proven. Further, learners worried that risk would be shifted to them, whereas, at least with the current system, they could enlist the help of the school rather than have the blame shifted to them.[36]

Beyond usability, additional work needs to be done on standards. For example, standard access patterns, indexing, and encryption will be critical for building tools to manage credentials as learners start receiving more and more digital credentials. These additional standards/protocols, which are critical to prevent lock-in and achieve interoperability, are in the early incubation stage at the W3C and the Decentralized Identity Foundation, including:

- The protocols and data standards by which credentials are exchanged between the subject and a relying party. Draft standards include the Credential Handler API,[37] Verifiable Presentation Request Specification,[38] and Presentation Exchange.[39]
- The standards by which credentials can be securely stored, accessed, searched, and exported/imported (enabling portability). Draft standards include encrypted data vaults[40] and identity hubs,[41] which are inputs to a combined secure data storage effort, recently launched by the W3C Credentials Community Group and Decentralized Identity Foundation.[42]

Recommendation:

- Design tools for the learner experience, with the understanding that exchange protocols are still under development.

Cautions and New Challenges

SECURITY

Finally, with any new technical stack comes the need to perform security audits, threat assessments, privacy and data protection impact assessments, and more. Existing threats may be replaced by new ones. For example, in an ecosystem in which the issuer doesn't need to be consulted may allow for issuing-side threats. Solutions based on fit-for-purpose blockchains may introduce entirely new security issues that are not yet understood. Pilots should carefully consider new threat models and remediations.

---

36   Kelly, "Blockcerts Wallet Usability Testing Results."

37   Longley and Sporny, "Credential Handler API 1.0."

38   Longley, Varley, and Zagidulin, "Verifiable Presentation Request Specification."

39   Buchner, Zundel, and Riedel, "Presentation Exchange."

40   "Encrypted Data Vaults 0.1."

41   Buchner, *Identity Hub Github Repository.*

42   Sporny, "Secure Data Storage Call."

From NIST's "Taxonomic Approach to Understanding Emerging Blockchain Identity Management Systems":

There are still scalability, security, and privacy considerations that must be carefully scrutinized to build viable digital identity solutions using blockchains, zero-knowledge proofs, second layer protocols, and related technologies. That said, if properly addressed, blockchain-based identity could become a fundamental architectural component of tomorrow's web.[43]

SUSTAINABILITY AND RESILIENCE

A full exploration of sustainability concerns is outside the scope of this paper, but should continue to receive attention by appropriate experts. This section includes specific technical recommendations on developing sustainable ILR pilots.

The excitement for blockchain technology can confuse what advantages it actually offers. Principles of responsible system design still apply, and in general blockchains should be used sparingly and intentionally. Permissioned blockchains especially can encourage a hasty "everything on-chain" approach, which can risk vendor lock-in.

At the same time, concerns aren't limited to permissioned blockchains; proof-of-work consensus mechanisms commonly used on non-permissioned blockchains have environmental impacts that cause resistance of some to recommend this as an environmentally sound approach.[44]

Finally, pilots should be mindful of the technical barriers they may introduce. Assuming that the learner has access to certain devices or even regular internet access may lead to bias and unequal ability to access the system. Pilots should design for accessibility for the broad range of learners.

PART 2: BEYOND TECHNOLOGIES

T3 Network Phase 2 projects are starting to produce reports, guidance, protocols, and tools for the broader public to use, and ILR pilots will soon be underway. Nonetheless, the work needed to build an open, decentralized public-private data and technology infrastructure is ongoing. This section explores how ecosystems such as the T3 Network can promote sustainable growth of networks committed to self-sovereign management of learner records and learner privacy, while ensuring it's achieving goals of ethical, equitable outcomes for learners.

Aligning with Privacy/Security Frameworks

The technologies described in this paper seek to improve individual privacy and control over personal data. However, while technical solutions can offer some assistance in accomplishing this goal, the heavy lift occurs through privacy and security regulatory frameworks. These may be specific within and across countries (e.g., European Union (EU) General Data Protection Regulation (GDPR)), state and local governments (e.g., California Consumer Privacy Act (CCPA)), and/or specific sectors (e.g., FERPA).

43   Lesavre et al., "A Taxonomic Approach to Understanding Emerging Blockchain Identity Management Systems."

44   Duffy, Pongratz, and Schmidt, "Building the Digital Credential Infrastructure for the Future."

This means that the privacy/security concerns and impact to implementers will vary according to jurisdiction, sector, and use case. At the same time, even privacy frameworks that are locale-specific might have extraterritorial reach. This is the case for the European Union GDPR, which grants privacy rights to individuals located in the EU when their personal data is processed by non-EU companies that offer goods or services to them or monitor their behaviors.

Systems should be built taking into account the privacy-by-design principle under GDPR. Some systems might then have to comply with different legal frameworks depending on the type of information they store and the type of processing they carry out.

For example, systems may need to provide notices to individuals on the use of the data and any sort of automated decision making or profiling activities may be subject to specific requirements under GDPR and other regulations and laws. Furthermore, some of this data might fall under the definition of special categories of data (and require higher protection, or specific consent).

Specific examples, in relation to GDPR and other regulations and laws, include:

- Applicant tracking system: Specific attention should be paid to whether this involves profiling or automated decision making.
- Background check system: The information involved in this activity might be considered special categories of data and require, as a matter of example, specific consent and higher standards for protection.
- Data collaborative system.[45] Specific attention should be paid to whether this involves profiling or automated decision making. Consider requirements on sharing of data (e.g., limitation on transfers, contractual arrangements among parties).

This doesn't mean that privacy guarantees should be the lowest common denominator required for compliance. As stated in the Digital Credentials Consortium (DCC) white paper, "placing learner privacy at the core of our design is not just for compliance with legal frameworks, but also an ethical position we choose to take."[46] Citing the protection of privacy as a human right (per the Universal Declaration of Human Rights, UN, art. 12), the DCC adopted the definition of personal data use for GDPR by the European Commission, which is that "personal data is any information that relates to an identified or identifiable living individual." Different pieces of information that are collected together that may lead to the identification of a particular person also constitutes personal data. Further, because it is developing a global solution, the DCC chose to align, at least initially, with the EU GDPR for the reason that it "appears to prioritize individual rights and establish globally respected, broadly applicable, and widely used standards." For ILR use cases, it will be essential to explore these impacts beginning in the earliest stages of ILR pilots.

A concrete recommendation is to begin privacy planning with the NIST Privacy Framework,[47] which is a voluntary tool organizations may use to identify and manage the privacy risks related to their activities. In particular, it helps organizations take privacy into account when they design and develop systems, services, and products that affect the privacy of individuals. In this context, this framework enables

---

45 Data collaborative guidance will be addressed in the T3 Network project, "Data Collaboratives for Individual-Level Data."

46 Duffy, Pongratz, and Schmidt, "Building the Digital Credential Infrastructure for the Future."

47 "NIST Privacy Framework: A Tool For Improving Privacy Through Enterprise Risk Management."

organizations to identify and prioritize the privacy risks related to their operations and develop effective solutions that help minimize the adverse consequences for individuals. Furthermore, this framework facilitates the way an organization communicates about its privacy practices and encourages collaboration among an organization's different teams (e.g., legal and information technology).

This framework is a flexible tool that can be easily adapted to each organization's different needs and can be used by the organization as a way to demonstrate compliance with the different privacy frameworks it is subject to. Implementing the framework at an early stage of the development and designing of the tools would certainly be of help in managing privacy risks and compliance with the different legal frameworks.

Recommendations:

- Systems should be built taking into account the privacy-by-design principle under GDPR to comply with different legal frameworks depending on the type of the information they store and the type of processing they carry out.
- Use the NIST Privacy Framework to take privacy into account when designing and developing systems, services, and products that affect the privacy of individuals.

Aligning with Digital Identity Frameworks

VCs and DIDs support embedded identity proofing attributes, but further work is needed to align with digital identity standards frameworks such as NIST's "Digital Identity Guidelines." Work is already underway to align with Electronic Identification, Authentication and Trust Services (EIDAS) levels of assurance.[48]

Recommendation:

- Support efforts to align digital credentials with identity proofing guidelines.

Anti-Bias

When accelerating access to digital information, it is vital to ensure that the most vulnerable do not become unwitting victims and compromised by inadvertent investments in the next, great fad. Lifelong learning is vital to keeping pace with the ever-changing talent marketplace, especially for empowering vulnerable and under-represented populations. Therefore, it is important that ILRs directly and explicitly tackle the challenge of diversity. This is especially true when digital processes enable analysis and communication to occur automatically and without transparency.

Recommendations:

There is no single solution to removing bias. The T3 Network and its participants must seek out ways to:

- Maximize diversity in its process from individual learners to institutions, from contractors and consultants to executive leadership.

---

48   Alamillo, "SSI EIDAS Legal Report."

- Incorporate diverse opinions from both traditional and non-traditional sectors. This should include underrepresented minorities based on race, age, gender, or physical disability, as well as learners with diverse backgrounds and education goals and needs, such as immigrants, career changers, and single parents.
- Embed bias protections at the architectural level to prevent unchecked algorithmic decision-making that can lead to unintended and unexpected discrimination.
- Establish processes that incorporate diversity strategies at every stage of the technology life cycle that may include requirements, analysis, design, implementation, deployment, operations, maintenance, and corrections.
- Develop standard policies that can be adopted by institutions and regulators to help guide the creation of services that respect the inherent diversity of learner populations.
- Develop practices that can be employed by institutions and regulators to both realize and ensure diverse populations are welcomed and supported.
- Educate all participants in the system, from learners to educators about the risks and possible responses for reducing bias.

Query Capability

To ensure true, privacy-enabled interoperability across digital records, it is vital to standardize data representations that can be queried across large populations with diverse experiences while minimizing unnecessary and unconsented disclosure of data.

The practices of minimal, selective, and progressive disclosure need to be supported by common data structures that can be used with advanced techniques such as zero-knowledge proofs, homomorphic encryption, and differential privacy (technologies with specific limitations on types of information and analyses that can be used). Unless the underlying data architecture enables these kinds of techniques, no amount of subsequent innovation can reverse the over-disclosure of information.

For example, unless a specific credential explicitly enables the separate disclosure of accomplishments and accomplishments tied to a legal name, it will be impossible to use that credential in an anonymous yet verifiable fashion. Additionally, without a standard query mechanism for recipients of learner record compositions to ask for specific, selected information, learners will be forced to disclose the entirety of various credentials. Finally, unless the cryptographic means of verification explicitly support decomposition and recomposition, it is generally impossible to do without losing verifiability.

For these approaches to work, there must be a marriage between the types of queries that need to be answered and the structure of the underlying records that enable verifiable responses.

Recommendations:

- Work with both recipients and educators to understand the most valuable queries for the most common use cases.
- Design a data architecture that supports minimal, selective, and progressive disclosure of records with just enough granularity and detail to satisfy those queries without unnecessarily leaking additional private information.

Technology-Independent Requirements

Before diving too deep into detailed technical specifications of specific formats, protocols, and performance requirements, we recommend first establishing a foundational understanding of the human needs for ILRs from learners, educators, and regulators. There are a number of tools and techniques that can be used to capture technology-independent requirements. Essential use cases and engagement models are two techniques highlighted below.

Essential use cases, also called abstract use cases, describe a single value creating interaction in a short paragraph focused on the human motivation and experience. Essential use cases were developed by Larry Constantine and Lucy Lockwood as part of Usage-Centered Design to "capture the essence of problems through technology-free, idealized, and abstract description"[49] thereby avoiding pitfalls commonly occurring in conventional use cases, such as inadvertently built-in assumptions that unnecessarily constrain design decisions. The 30 Use Cases in the Verifiable Credentials Use Cases[50] demonstrate examples of essential use cases.

Engagement models illustrate the human experience across the entire information lifecycle. They take a single individual as a protagonist (e.g., learner) and follow their experience from before contact through exit and re-engagement, documenting at least one interaction with the system, including the experiences of the people supporting their process (e.g., teachers, administrators, recruiters, employers). These interactions are described in brief paragraphs explaining what the individuals actually do when they interact with the system, resulting in a coherent narrative. Engagement models focus on the human experience, leaving the underlying implementation details to a later stage of development. With a focus on a single protagonist and their path through a complex system, readers can step into that individual's role and evaluate the feasibility of the imagined system, both technically and experientially. The result is a concise technology-independent description of the system that reads like a simple story, accessible to laypeople and technologists without loss of rigor. Published examples of engagement models include Joram 1.0.0[51] and Amira 1.0.0.[52]

Recommendation:

- Review and expand the ILR Prioritized Use Cases in order to deeply understand the human requirements prior to designing and implementing solutions.

Open Data Standards

To achieve scalable interoperability, we must continue to strengthen public and private collaboration in the development and use of voluntary consensus standards. Guiding principles, practices, and implementation methods for supporting and participating in standards development processes are examined in the T3 Network paper, "Public-Private Standards Development and Use by Government."

---

49   Biddle, Noble, and Tempero, "Essential Use Cases and Responsibility in Object-Oriented Development."

50   Ibid.

51   Andrieu and Clint, "Joram 1.0.0."

52   Andrieu et al, "Amira 1.0.0."

Additionally, it is vital to not only explore and implement existing and emerging standards, such as VCs and DIDs, but also to actively participate and contribute to these ongoing standards discussions. As discussed in greater detail in the "Model Roles and Processes for Standards Development" paper, all stakeholders in the standards development process—employers; government agencies; SDOs; and education, training and credentialing organizations—have an important role to play to ensure that standards for education and workforce partners are appropriately developed, implemented, and utilized as a foundational part of an open, public-private data and technology infrastructure that meets the needs of an ever-changing talent marketplace.

Recommendation:

- Encourage public and private collaboration in the development, implementation, and use of open, voluntary consensus standards, whenever possible.

Governance Frameworks

Finally, thinking about governance frameworks for ILR pilots can help guide the above consideration along with other learner-focused concerns. For example:

- Provide resources and guidance promoting longevity of linked data, such as links to competencies, accomplishments, and registries of credential status (e.g., revocation lists).
- Promote resilience, addressing the impact to users if any aspect of the system becomes unavailable.
- Encourage stakeholders to promote patterns for privacy-protecting technologies and ethical outcomes.
- Establish guidelines and protocols for re re-obtaining lost credentials.
- Define an appeals/correction process.

Recommendations:

The participants of the Management and Use of Individual-Level Data Records Project developed draft recommendations as a starting point for ILR governance frameworks. This is not a final set of recommendations; rather, it's intended to be a starting point for continued refinement.

1. SSI-Aligned Access Control (all are assumed to be within the bounds of regulation or law)
   a. Learner Rights
      1.a.1. Access: Learner (or his or her legal guardian) has access to, and ability to control access to, their credentials.[53]
      1.a.2. Usage: Learner has the ability to set terms of use of his or her credentials (e.g., the ability to share with particular entities for a particular purpose and period of time, and to revoke access).
      1.a.3. Transparency: Learner has the option to be notified about, and have access to, the details regarding the flow of his or her data.

---

53 "SSI-Aligned Access Control" begins with "All are assumed to be within the bounds of regulation or law." So these statements should be read as "Within the bounds of regulation of law, the learner (or their legal guardian) has access to, and ability to control access to, their credentials," and so on, for each statement.

    b.   Issuer Rights and Requirements
- 1.b.1. Non-override: The rights granted to the learner do not override data retention/usage guidelines or regulations applying to the issuer (i.e., the issuer should comply with usual requirements to retain and restrict access per regulation or law).
- 1.b.2. Consent: Issuer obtains explicit consent from the learner or his or her legal guardian to share the personal record outside the organization except where specifically authorized by regulation or law.
- 1.b.3. Longevity: Issuer makes a best effort to ensure credentials are available to the learner.
- 1.b.4. Audit: Issuer maintains auditable logs of access, including chain of custody, except where regulation or law dictates otherwise.

    c.   Third parties
- 1.c.1. Audit: Third parties audit logs of access, including chain of custody, except where regulation or law dictates otherwise. The sharing terms may require destruction of the data after a certain period of time, but the third party maintains records of how the data has been handled (i.e., collected in, controlled, analyzed, stored, shared, and disposed of).

2. **Authoring and Issuing Credentials—The following guidelines for authoring and issuing credentials help ensure credentials are maximally useful to the learner.**
   a.   Credentials with well-understood, formal vocabularies—Linked data mechanisms provide a means for disambiguating potentially confusing terms through explicit, rigorous labeling of the meaning of each particular value in a credential.
   b.   Interoperable data standards and formats—Use of open data standards appropriate for the given domain and expected use enable interoperability with broader credentialing ecosystems.
   c.   Learner privacy and identity linkability—Credentials should allow learners flexibility to link credentials to other identity data in a privacy-respecting manner.
   d.   Verifiability—Credentials should be verifiable as being from the issuer in a manner that maximizes availability. This includes minimizing centralization points and other sources leading to potential downtime and unusability (e.g., a centralized credential repository, an institutional verification service, or an institutionally centralized revocation registry)

3. **Credential Storage and Exchange Ecosystem—Credential storage and exchange ecosystems are based on open standards that enable portability and interoperability for the learner, as well as security and privacy by design.**

4. **Governance Frameworks—Credential ecosystems are based in a set of governance frameworks responsible for adapting this or similar recommendations—in addition to other relevant principles— to appropriate domain- and locale- (and other) decisions to ensure compliance, usability, privacy, security, fitness-for-purpose, and any other social or legal concerns related to the specific use case(s).**

CONCLUSION

This paper is the beginning of an ongoing conversation on management and use of individual-level data records in relation to the T3 Network and ILR pilots. The application of SSI-based approaches and technologies to ILRs enables educational institutions, learners, and employers to streamline effective use of learner records, while ensuring access and privacy of data for all individuals. Advancements in the use of individual-level data records are accompanied by potential risks to the learner's data, which were outlined in this paper, along with potential solutions to mitigate the risks. Finally, technical details of SSI standards and technologies are described above to begin implementing technical solutions.

The design of SSI-based approaches provides a lens to examine how the talent marketplace might restructure systems to give learners better access to, and control over, the management of their learning records. While SSI-based approaches are relatively new, this paper outlines how they can serve as a starting point and guide for T3 Network participants to articulate their plans to empower individuals with their own data through interoperable learner records. As ILR pilots begin to form in the second and third quarter of 2020, pilot teams should consider utilizing and further testing these principles, technologies, and community guidance in low-risk, isolated environments and report their findings to the larger T3 Network and broader SSI community to learn from.

ACKNOWLEDGMENTS

This paper was authored by Kim Hamilton Duffy of MIT's Digital Credential Consortium with significant contributions from Joe Andrieu of Legendary Requirements and Kerri Lemoie of OpenWorks Group, LLC.

We would also like to thank the participants of the Management and Use of Individual-Level Data Project and the project's technical consultants, Matt Gee of BrightHive, Jim Goodell of Quality Information Partners, Taylor Kendal of Learning Economy Foundation, Greg Nadeau of Public Consulting Group, Elizabeth Renieris of Hackylawyer, Bob Sheets of George Washington University, and Stuart Sutton of Sutton and Associates.

BIBLIOGRAPHY

Alamillo, Ignacio. "SSI EIDAS Legal Report." April 2020. https://joinup.ec.europa.eu/sites/default/files/document/2020-04/SSI_eIDAS_legal_report_final_0.pdf.

Allen, Christopher. "The Path to Self-Sovereign Identity." *Life with Alacrity* (blog), April 25, 2016. http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html.

Andrieu, Joe, and Christopher Allen. "Amira 1.0.0." Rebooting Web of Trust, n.d. https://github.com/WebOfTrustInfo/rwot5-boston/blob/master/final-documents/amira.md.

Andrieu, Joe, and Phil Archer. "Use Cases and Requirements for Decentralized Identifiers." W3C Editor's Draft. W3C, April 25, 2020. https://w3c.github.io/did-use-cases.

Andrieu, Joe, and Bob Clint. "Joram 1.0.0." April 14, 2017. https://github.com/WebOfTrustInfo/rwot3-sf/blob/master/final-documents/joram-engagement-model.pdf.

Balkin, Jack, and Jonathan Zittrain. "A Grand Bargain to Make Tech Companies Trustworthy." *The Atlantic*, October 3, 2016. https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/.

Biddle, Robert, James Noble, and Ewan Tempero. "Essential Use Cases and Responsibility in Object-Oriented Development." Victoria University of Wellington, May 20, 2001. http://www.mcs.vuw.ac.nz/research/object/Papers/euc-html/paper.html.

Buchner, Daniel. *Identity Hub Github Repository*. Decentralized Identity Foundation, n.d. https://github.com/decentralized-identity/identity-hub.

Buchner, Daniel, Brent Zundel, and Martin Riedel. "Presentation Exchange." Decentralized Identity Foundation, n.d. https://identity.foundation/presentation-exchange/.

Chadwick, David, Dave Longley, Manu Sporny, Oliver Terbu, Dmitri Zagidulin, and Brent Zundel. "Verifiable Credentials Implementation Guidelines 1.0." W3C Working Group Note. W3C, September 24, 2019. https://www.w3.org/TR/vc-imp-guide.

Cimpanu, Catalin. "IOTA Cryptocurrency Shuts Down Entire Network after Wallet Hack." *ZDNet*, February 16, 2020. https://www.zdnet.com/article/iota-cryptocurrency-shuts-down-entire-network-after-wallet-hack.

Duffy, Kim Hamilton. "How Blockcerts Uses JSON-LD Normalization (Canonicalization) during Verification." May 2017. https://community.blockcerts.org/t/how-blockcerts-uses-json-ld-normalization-canonicalization-during-verification/102.

Duffy, Kim Hamilton, Hans Pongratz, and Philipp Schmidt. "Building the Digital Credential Infrastructure for the Future." Digital Credentials Consortium, n.d. https://digitalcredentials.mit.edu/wp-content/uploads/2020/02/white-paper-building-digital-credential-infrastructure-future.pdf.

"Encrypted Data Vaults 0.1." Draft Community Group Report. W3C, January 26, 2020. https://digitalbazaar.github.io/encrypted-data-vaults/.

"Family Educational Rights and Privacy Act (FERPA) and the Disclosure of Student Information Related to Emergencies and Disasters." June 2010. https://www2.ed.gov/policy/gen/guid/fpco/pdf/ferpa-disaster-guidance.pdf.

"FERPA General Guidance for Students." n.d. https://www2.ed.gov/policy/gen/guid/fpco/ferpa/students.html.

FIDO Alliance. "FIDO2: WebAuthn & CTAP." April 1, 2020. https://fidoalliance.org/fido2/.

Fry, Emily, and Elizabeth Renieris. "SSI? What We Really Need Is Full Data Portability." *Women in Identity Blog* (blog), n.d. https://womeninidentity.org/2020/03/31/data-portability/.

Grassi, Paul, and James Fenton. "Digital Identity Guidelines: Enrollment & Identity Proofing." National Institute of Standards and Technology, June 2017. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63a.pdf.

Henry, Shawn Lawton, Shadi Abou-Zahra, and Kevin White. "Accessibility, Usability, and Inclusion." May 6, 2016. https://www.w3.org/WAI/fundamentals/accessibility-usability-inclusion/.

"IEEE P7004: Standard for Child and Student Data Governance." Accessed April 8, 2020. https://site.ieee.org/sagroups-7004/.

"Interoperable Learner Record (ILR) Prioritized Use Cases." T3 Innovation Network, January 29, 2020. https://www.uschamberfoundation.org/sites/default/files/T3%20Network_ILRUseCase_March2020.pdf.

"Introduction to the Hash Function as a Personal Data Pseudonymization Technique." Agencia Española de Protección de Datos, October 2019. https://edps.europa.eu/sites/edp/files/publication/19-10-30_aepd-edps_paper_hash_final_en.pdf.

Jones, M, J Bradley, and N Sakinura. "JSON Web Token (JWT)." Proposed Standard. Internet Engineering Task Force (IETF), May 2015. https://tools.ietf.org/html/rfc7519.

Kelly, Lindsay. "Blockcerts Wallet Usability Testing Results." January 2019. https://www.dropbox.com/s/tmwlfx7wwyc7exm/Usability%20Testing%20Take-Aways_Updated_1.28.19_De-Identified.pdf?dl=0.

Lesavre, Loïc, Priam Varin, Peter Mell, Michael Davidson, and James Shook. "A Taxonomic Approach to Understanding Emerging Blockchain Identity Management Systems." White Paper. National Institute of Standards and Technology, January 14, 2020. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01142020.pdf.

"Linked Data Proofs." Draft Community Group Report. W3C, n.d. https://w3c-ccg.github.io/ld-proofs/.

Longley, Dave, and Manu Sporny. "Credential Handler API 1.0." Draft Community Group Report. W3C, February 19, 2020. https://w3c-ccg.github.io/credential-handler-api/.

Longley, Dave, Mike Varley, and Dmitri Zagidulin. "Verifiable Presentation Request Specification." Unofficial Draft. W3C, May 24, 2020. https://w3c-ccg.github.io/vp-request-spec/.

Looker, Tobias, and Orie Steele. "BBS+ Signatures 2020." Draft Community Group Report. W3C, May 19, 2020. https://w3c-ccg.github.io/ldp-bbs2020/.

"NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management." National Institute of Standards and Technology, January 16, 2020. https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework_V1.0.pdf.

"Open Badges v2.0." Final Release. IMS Global, April 12, 2018. https://www.imsglobal.org/sites/default/files/Badges/OBv2p0Final/index.html#Profile.

Otto, Nate, and Kim Hamilton Duffy. "Open Badges Are Verifiable Credentials." July 9, 2018. https://github.com/WebOfTrustInfo/rwot6-santabarbara/blob/master/final-documents/open-badges-are-verifiable-credentials.pdf.

Preukschat, Alex, and Drummond Reed. *Self-Sovereign Identity*. Manning, 2020.

Reed, Drummond, Manu Sporny, and Markus Sabadello. "Decentralized Identifiers (DIDs) v1.0." W3C Working Draft. W3C, March 23, 2020. https://w3c.github.io/did-core/.

Sporny, Manu. "Secure Data Storage Call." March 10, 2020. https://lists.w3.org/Archives/Public/public-credentials/2020Mar/0025.html.

Sporny, Manu, and Dave Longley. "Revocation List 2020." Draft Community Group Report. W3C, May 12, 2020. https://w3c-ccg.github.io/vc-status-rl-2020/.

Sporny, Manu, Grant Noble, Dave Longley, Daniel Burnett, and Brent Zundel. "Verifiable Credentials Data Model 1.0." W3C Recommendation. W3C, November 19, 2019. https://www.w3.org/TR/vc-data-model/.

Steele, Orie. *Linked Data Signatures for JWS*. W3C, n.d. https://w3c-ccg.github.io/lds-jws2020/.

"T3 Innovation Network Phase 1 Report." U.S. Chamber of Commerce Foundation, October 2018. https://www.uschamberfoundation.org/sites/default/files/T3Phase1_Report_FINAL_0.pdf.

Terbu, Oliver, Ivan Basart, Kyle Den Hartog, Christian Lundkvist, David Stark, Dmitri Zagidulin, Danny Strockis, and Orie Steele. "Self-Issued OpenID Connect Provider DID Profile." Decentralized Identity Foundation, n.d. https://identity.foundation/did-siop/.

Terbu, Oliver, Dmitri Zagidulin, and Amy Guy. "Did:Web Decentralized Identifier Method Specification." Editor's Draft, March 19, 2020. https://w3c-ccg.github.io/did-method-web/.

Tobin, Andrew. "What Goes on the Ledger." White Paper. Sovrin, n.d. https://sovrin.org/wp-content/uploads/2018/10/What-Goes-On-The-Ledger.pdf.

U.S. Chamber of Commerce Foundation. "T3 Network Pilot Projects." n.d. https://www.uschamberfoundation.org/t3-innovation/pilot-projects.

"Verifiable Credentials Use Cases." W3C Working Group Note. W3C, n.d. https://www.w3.org/TR/vc-use-cases/.

Vescent, Heather, and Kaliya Young. *Comprehensive Guide to Self-Sovereign Identity*. 2nd ed. The Purple Tornado, Inc., 2019.

Windley, Phil. "Four Pillars of an SSI Network." *Technometria* (blog), January 7, 2020. https://www.windley.com/archives/2020/01/four_pillars_of_an_ssi_network.shtml.