T3 INNOVATION NETWORK

Work Group 4 Report

New Architectures and Uses of
Linked Individual-Level Data

JULY 2018

INTRODUCTION

The T3 Innovation Network organized Work Group 4 to explore the most important stakeholder use cases relevant to new architectures and uses of linked individual-level data consistent with relevant T3 guiding principles. The work group was asked to:

- Identify stakeholder use cases that require improvements in the handling of linked individual-level data.
- Identify the potential for bundled Web 3.0 technologies to give learners and workers more control of their own records (self-sovereignty) while improving the discoverability and verifiability of their personal data.
- Identify potential solutions and pilot project opportunities to address these critical use cases (all work groups were given this task).
- Identify implications for the longer-term vision of a shared public-private data infrastructure as well as ethical considerations.

Work Group 4 held two webinars. The first webinar focused on reviewing the relevant stakeholder use cases from Work Group 1 and then identifying additional use cases, challenges and potential solutions. The second webinar focused on potential solutions, pilot project ideas, and implications for a public-private data infrastructure.

This work group report first summarizes the most important stakeholder use cases for the application of Web 3.0 technologies. Next, the report summarizes major issues and technological challenges that cross-cut these use cases; enumerates potential pilot project ideas that have emerged through the group's discussion; and mentions potential ethical considerations.

STAKEHOLDER USE CASES

During the first webinar, Work Group 4 identified the following as the most important stakeholder use cases relevant to linked individual-level data:

Pooled Analytics & Record Linkage

- Institutional stakeholders such as government agencies, large employers, researchers, and educational institutions would like to ethically and responsibly combine large amounts of individual-level data to make better decisions in the talent marketplace.

Verification of Claims

- Employers would like to request and obtain complete and verified information related to learning claims, work history, and other hiring requirements that are necessary for screening and hiring decisions.
- Training providers would like to provide services to learners to document, verify, and distribute comprehensive educational records.
- Governments would like to verify education, employment, and earnings to determine program eligibility.

Self-Sovereign Records

- Potential learners and employees want to maintain, manage, and distribute resumes and comprehensive learner records along with verifications on the open web, through professional profiling tools and systems. These tools and systems also would be used to provide information necessary to complete applications for and participate in employment, education, and credentialing opportunities.

During the second webinar, Work Group 4 further explored and extended the major stakeholder use cases to include the following:

Pooled Analytics & Record Linkage

- Workers and learners would benefit from persistent digital identifiers which stakeholders could use to transfer, join, and verify their learning and career records over many years. This digital identity may also be used for authentication across training provider and employer IT systems[1]. It may also be used to obtain government services in multiple contexts, including but not limited to education and work.
- Analysts at institutions such as government agencies would benefit from a Master Client Index, allowing them to cross-link records from multiple sources. In this context, the cross-linked records would likely be anonymized or pseudo-anonymized, in compliance with regulations such as GDPR, and used largely to generate aggregate statistics and performance metrics.
- Employers and educational institutions would like to link their own internal individual-level data to external sources in order to obtain more comprehensive credential, student, and work records. These records may be used to inform internal decision making and screening applicants.

Verification of Claims

- Workers and learners would benefit from the ability to make verifiable claims about their skills, competencies (e.g. linking to training provider records, assertions by peers, and assessments), and from the ability to be linked to potential employers based upon those verifiable claims.
- Employers and educational institutions would like to verify assertions of credentials, work history, skills, and competencies made by applicants in a simple way that is less expensive than their current business process.
- Governments and financiers would like to verify wage and employment records to confirm eligibility for means-tested programs or scholarships.

Self-Sovereign Records

- Individuals would like to grant or deny consent to access their education and work records, in a way that allows for curation and prevents forward-sharing. In particular, they

---

[1]Nascent efforts by the European Union provide a useful example of this idea: https://www.enisa.europa.eu/publications/mami/at_download/fullReport

may wish to identify and suppress or remove data about themselves that is incorrect (or simply unflattering).

- Employers, governments, and educational institutions would like to ensure regulatory requirements for privacy, erasure, and consent are met. Additionally, they would like to create auditable records of consent for individual-level data access.
- Employers and educational institutions wish to make requests for specific pieces of data from applicants, such as complete academic transcripts, which applicants may choose to explicitly accept or decline.

TECHNICAL CHALLENGES
& POTENTIAL SOLUTIONS

The fundamental challenges addressed by this work group are summarized in the figure below. Individual-level data is generated by learners, training providers, and employers. This data is connected in a way that all parties can trust and where each party maintains a sufficient level of control, representing both a conceptual and a technical challenge.
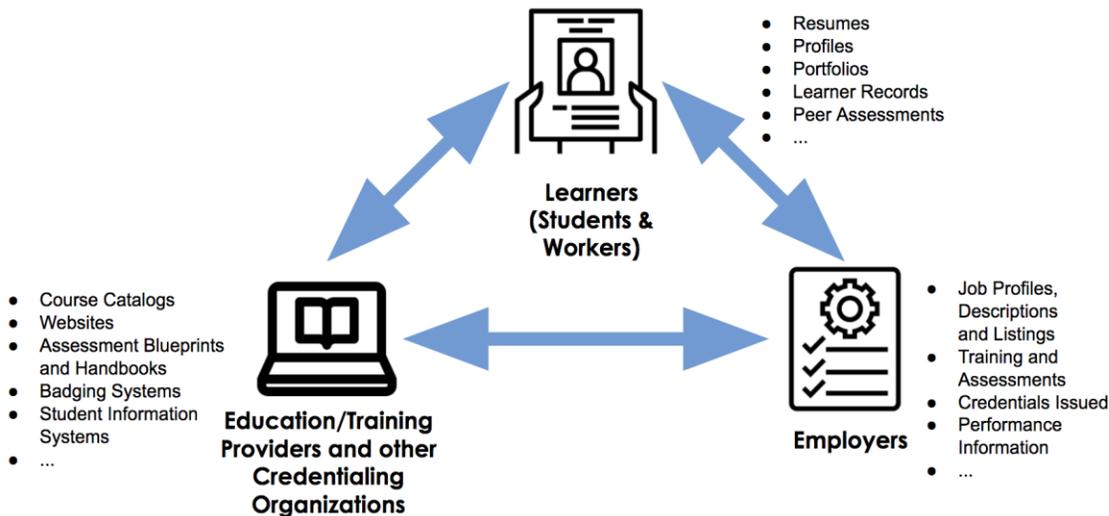
It is useful to analyze challenges from multiple perspectives. Data specifications and standards, including those for the semantic web, ease the sharing and linking of data. Some examples of data specifications that enable linked open data include: schema.org, Credential Engine's Credential Transparency Description Language (CTDL), CTDL Achievement Standards Network (ASN), and the OpenBadge 2.0 Specification. Leveraging data specifications, artificial intelligence (AI, including machine learning) can be used to structure, merge, and generate data sets. Distributed ledger technology (DLT) along with smart contracts may offer useful features for data verification and self-sovereign management of life-long learning and employment. DLT is being actively explored for use in verifiable educational records[2] and electronic consent[3], similar to how it is being used in the medical community[4].

---

[2]See, for example, https://www.learningmachine.com/
[3]See, for example, https://id2020.org/
[4]See, for example, https://medrec.media.mit.edu/

Work Group 4 views immutability and decentralization as key advantages, while the rapid state of flux in the technical landscape and uncertainties surrounding sustainable business models were cited as reasons to proceed in an exploratory fashion. Additionally, participants in the work group highlighted the possible synergies between DLTs and other efforts (e.g. how a more mature digital badge ecosystem might be used for off-chain details and validation).

Implications for a Public-Private Infrastructure

Given the use cases and challenges discussed above, the work group proposed a set of implications for the long-term development of an open and distributed data and technology infrastructure for the talent marketplace. High-level goals were identified by the group:

- Public-private digital identity infrastructure
- Mechanism for individuals to consent to, and manage, the sharing of their personal data
- Standards and protocols for verifying claims about education and employment

Potential Pilot Projects

In line with these high-level implications, and in light of the use cases and technical challenges discussed above, the work group identified potential pilot projects:

Employer HRIS Projects
- Explore how HRIS vendors could use distributed ledgers to provide more complete and trusted verification of job histories provided by applicants, including job-specific information about job titles and job responsibilities/tasks.

Education and Training Provider Projects
- Explore how multiple credential/transcript vendors could publish to the same network or interoperable networks that allow for complete information about an individual to be accessed and verified from HRIS systems.
- Develop a prototype employment and/or educational data consent and distribution service using a distributed ledger.

Learner (Student/Worker) Projects
- Explore ways in which skills and competencies can be asserted by applicants and verified by educational institutions or potential employers.
- Create a prototype record linkage service for associating a student from one education provider with a student from another provider.

Large-Scale Data Analytics Projects
- Develop an application prototype that derives labor market indicators from pseudo anonymized linked data reported from multiple employers/vendors.
- Develop a secure multi-party computation application prototype which draws on data from multiple employers and/or educational institutions.

Employer and Government Projects
- Explore how HRIS vendors could provide more current and verified employment and earnings information for means-tested programs.

Ethical Considerations

Recent events surrounding social media platforms and financial and consumer credit data systems illustrate the importance of keeping security and ethical practices at the forefront of the discussion when personally identifiable information is involved. Key considerations highlighted by the work group include:

- **Data security.** Every organization which serves as a source or a custodian of individual-level data should be following industry best-practices.
- **Privacy.** Individuals should have the opportunity to opt out of data collection, and to exert a measure of sovereignty over their own identity and records. Increasingly, these considerations are likely to be supported by regulatory requirements.
- **Equity.** Care should be taken to ensure that new technological infrastructures do not have adverse impact on disadvantaged groups and individuals who may not have the same access to, awareness of, or comfort with the relevant technologies. Moreover, attention should be given to the educational providers and employers participating in pilot projects, to ensure that the benefits and challenges uncovered are relevant to the broadest possible audience.