**Data Quality Assessment – Credential Engine’s Credential Registry**

**Rubric for Assessing Dataset Quality**

1. **Relevance**
   - **What is the total number of items relevant to credentials?**
     There are 43 variables in the database’s Credential section, including a credential’s name, description, type, subject, status, webpage, Credential Transparency Identifier (CTID), language, organization, jurisdiction, audience level, estimated cost, estimated duration, industry type, occupation type, requirements, external quality assurance, online or physical location, financial assistance availability, related credentials, holder number and characteristics, employment and earnings outcome, renewal requirements and frequency, revocation criteria and process, and maintenance process.
   - **What are the measures of credential attainment like?**
     Degree, diploma, license, certificate, certification, courses, and training programs.
   - **Are there any indicators related to education attainment that are unique to this dataset?**
     Cost, duration, financial assistance, and labor market outcome data of a credential.
   - **What is the purpose of the dataset, and how closely does that purpose align with the following use cases?** (Evaluate as relevant or not relevant.)
     The Credential Registry is created by Credential Engine to house up-to-date information about all credentials with a common description language. It is intended to promote credential transparency, enable credential comparability, and support customized search about credentials.
     a. **Measuring the rate of attainment of credentials within the U.S. skilled technical workforce**
        Relevant. For each credential, it is optional for the credentialing organization to upload the number and characteristics of credentialed individuals and their geographic location. This information, when available, can be used to calculate the rate of attainment of a credential.
     b. **Measuring aggregate returns to credentials by credential type**
        Relevant. Organizations can choose to upload employment and earnings data of a credential.
     c. **Identifying disparities by race and gender in the attainment of credentials**
        Relevant. The holders profile includes demographic information.
     d. **Identifying which credentials are associated with the strongest labor market returns for individuals in the skilled technical workforce**
        Relevant. Employment and earnings data are available for a credential if the credentialing organization chooses to upload them.
     e. **Evaluating the effectiveness of public policies that support the attainment of credentials?**
        Relevant. Researchers can use the database to assess the labor market outcome and financial assistance landscape of credentials and help policymakers make relevant supporting policies.

2. **Coverage**
   - **What is the frame of reference for the dataset, what population does the dataset attempt to cover?**
     The Credential Registry attempts to cover all credentials in the United States.
What is the number of cases, and how does that number compare to known estimates of the relevant population?

As of September 2022, the Credential Registry includes 40,890 credentials. Credential Engine identifies nearly one million credentials in the United States, so the database seems to cover about four percent of them.

How does the publisher of the data ensure that data is collected for cases that should be in the dataset?

Credential Engine depends on credentialing organizations to upload information about their credentials. It also cooperates with 28 states and regions for bulk upload of credential information in their jurisdiction.

What percent of cases lack data for key variables of interest? (Direct assessment if not in metadata.)

For each credential, its name, description, type, status, webpage, CTID, language, and organization are required. Other information is optional. Among all 40,890 credentials, 9,759 have a cost profile, 16,814 have a duration profile, 3,408 have industry information, 14,086 have occupation information, and 1,034 have financial aid information. Coverage for a selection of optional information can be found here.

3. Granularity

How granular (i.e., how many different categories exist, if not continuous) is data for key variables of interest (attainment, field of study, income)? What about for different levels of aggregation researchers might consider, such as geography, age, and race?

- Type: Apprenticeship Certificate/Associate Degree/Bachelor’s Degree/Badge/Certificate/Certification/Degree/Diploma/Doctoral Degree/General Education Development (GED)/Journeyman Certificate/License/Master Certificate/Master’s Degree/Micro-Credential/Open Badge/Professional Doctorate/Quality Assurance Credential/Research Doctorate/Secondary School Diploma
- Status: Active/Deprecated/Probationary/Suspended/Teach Out
- Audience level: Beginner/Intermediate/Advanced/Lower Division/ Upper Division/ Secondary School or Equivalent/Post-Secondary (Associates Degree/Bachelor’s Degree/Master’s Degree/Doctoral Degree)/Undergraduate/Graduate/Professional/Remedial
- Industry type: recommend NAICS code recommended
- Occupation type: O*NET code recommended
- Financial assistance: General Financial Assistance (Grant/Loan/Scholarship/Work-Based)/Military

Is this data granular enough (yes or no) to perform analyses for each of the use cases identified under “relevance”?

Yes.

4. Timeliness

How often is the dataset updated?
According to the technical documentation, the update frequency varies with the credential. For each credential published, the entity issuing that credential will be asked for an estimate of how frequently the information about that credential is likely to change. If there have been no updates provided by the owner about the credential in that timeframe, Credential Engine will automatically follow-up with the owner to confirm that the existing information remains accurate. If changes need to be made, the owner, working with Credential Engine, will determine if a new version is warranted or not. If no changes are made, the currency clock on that credential is reset.

- Is data collected continuously or in waves? If in waves, what is the duration of those waves? Are some variables/records more frequently updated than others?
  
  Continuously.

- What is the time lag between when an event occurs, when it is recorded, and when that data is available to researchers?
  
  As the registry is continuously updated, there should be little or no lag between reporting and data availability. However, it is ultimately up to the credential issuers and collaborating state higher ed systems to report data in a timely manner.

5. **Integrity**

- What are the risks to the integrity of this dataset?
  
  Information is self-reported by credentialing organizations, which theoretically can have an incentive to misreport the cost, duration, holder, and employment profiles of the credential to attract more students/users. Credential Engine claims to authenticate the information uploaded by organizations.

- Were there changes to the dataset that may have resulted from political influence? If so, do those changes threaten the overall quality of the data?
  
  None that we are able to identify.

6. **Accessibility**

- How do researchers access this dataset?
  
  Organizations with approved Credential Engine accounts can download the Registry data. There is no published application form or process for independent researchers to apply to access the registry microdata beyond looking up individual credentials via [Credential Finder](#). However, researchers and other users are invited to contact Credential Engine to learn about microdata access options.

- Are any variables or cases withheld from researchers? If so, does that withholding or censoring affect researchers’ ability to use the data?
  
  No.

- Are there direct costs or indirect costs (e.g., training, resources) associated with accessing and using the data?
  
  Yes. It is possible that Credential Engine may charge fees for registry data access depending on the intended/licensed use of the data. There may also be opportunity costs associated with becoming
proficient in the use of Credential Transparency Description Language (CTDL) that is used to classify all credentials in the registry.

7. **Interoperability**

- Is there a unique identifier for individual cases? If so, is it one that can be found in other datasets?
  Yes, the Credential Transparency Identifier (CTID), assigned by Credential Engine, is unique in the database but does not apply in other databases.

- Is it common? (e.g., a SSN might be higher value than an address, though even name/address might be sufficient to match in some cases).
  No.

- Do occupation and industry coding schemes correspond to commonly used frameworks such as O*Net and NAICS? If not, are they well documented in metadata?
  Credential Engines recommends that organizations submit NAICS and O*Net codes, but organizations may choose to use other classification systems.

8. **Suitability for Longitudinal Research?**

- Is the metadata consistent over time, at least for key variables?
  Metadata has been modified over time, but the reporting standards for key variables seem to remain consistent. Metadata release history can be found [here](#).

- What is the construction of the dataset like? Is the microdata organized in “waves”? Can multiple observations for the same unit of analysis (i.e., person) over time be easily linked together?
  It seems that each credential has a unique entry, and whether historic versions of the credential can be accessed is unclear.

- How far back do administrative records from this dataset go?
  2017.