CONTENT

1. Introduction

CURL Site is a software application which normalizes and hashes clear-text data used in privacy-preserving record linkage. CURL Site supports two deployment options:

- **Option 1:** Docker image from Docker hub
- **Option 2:** Java application from Github

2. Linkage data preparation

- **Linkage data preparation:** Sites extract, transfer, and load (ETL) clear-text identifiers used in the linkage process from local source data into the structure and format of the *CURL Data Model* into a flat file on a local file server housed entirely within the data partner's secure environment. Clear-text identifiers will not leave sites' server at any point in the CURL record-linkage process. The clear-text flat file will be hashed locally using one-way hashing functions (i.e., SHA-512) by the CURL Site application to produce hash data, also in flat file format. Only hash data will be securely transported to the honest broker. CURL team will assist sites with the ETL process if needed. Encrypted hash data can be transferred to the CURL Honest Broker using Secure FTP, encrypted thumb drive, or PopMedNet.
- **Salt file for hashing:** Sites will receive and store a salt file from the CURL KeyMaster. The salt file is a clear-text JSON file with very large random salts used in the secure hashing process. Each salt file will be compatible with ONLY one linkage job. **Important:** Salt file must NOT be shared with the CURL Honest Broker. A salt file can be transported from the CURL KeyMaster to CURL Sites using Secure FTP, encrypted thumb drive, encrypted email, or PopMedNet.

3. Software and hardware requirements

- Accounts
  - Docker hub account (Option 1) or Github account (Option 2)
- Hardware requirements
  - Processor:
    - Minimum: no specific requirement
    - Recommended: Dual or Quad core processors

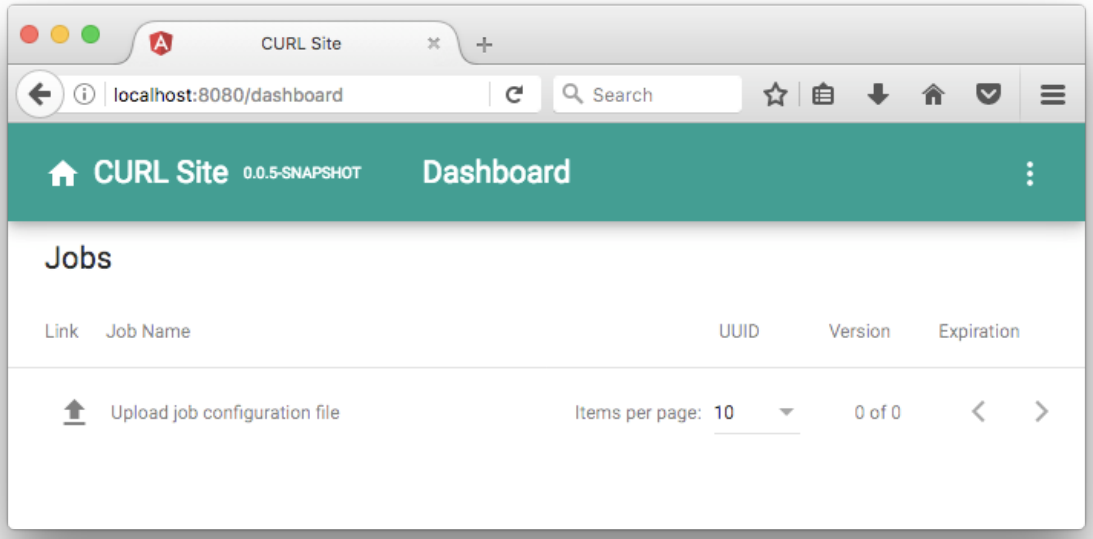
    - Memory:
        - Minimum: 4GB
        - Recommended: 8GB
    - Disk space: at least 5GB (depends on size of the dataset)
- Software requirements:
    - Option 1 - Docker image:
        - Docker version 17.06.0-ce or later
        - Operating system (Linux, Windows or MacOS) with Graphical user interface
    - Option 2 - Java application
        - Java 1.8 or later
        - Operating system (Linux, Windows or MacOS) with Graphical user interface
    - Both options:
        - Web browser:
            - Microsoft Edge (Version 40)
            - FireFox (Version 55) - Used in screenshots
            - Chrome (Version 60)
            - Safari (Version 10.1)

4. CURL Site Installation

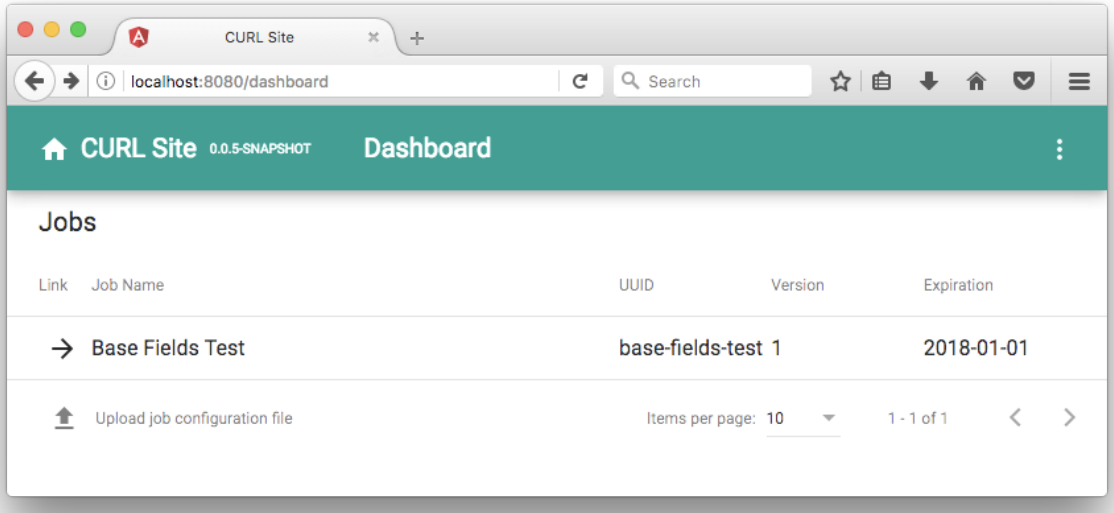
- Send an email to CURL team at toan.ong@ucdenver.edu. This email should include the following information:
    a. Full Name
    b. Organization
    c. Work Title
    d. Work email address. Note: This email address will be used to receive hash key from CURL KeyMaster.
    e. Work phone number
    f. DockerHub username (Option 1) or Github username (Option 2)
- Receive an email from CURL team. This email includes:
    a. Instruction to download and install CURL Site
    b. Hashing job configuration file
- Follow the instruction in the email from CURL team to download and install CURL Site.

5. CURL Site hashing workflow

- Launch CURL Site
    - Open a new browser. In the address bar, type in: http://localhost:8080
    - The homepage of CURL Site is a Dashboard

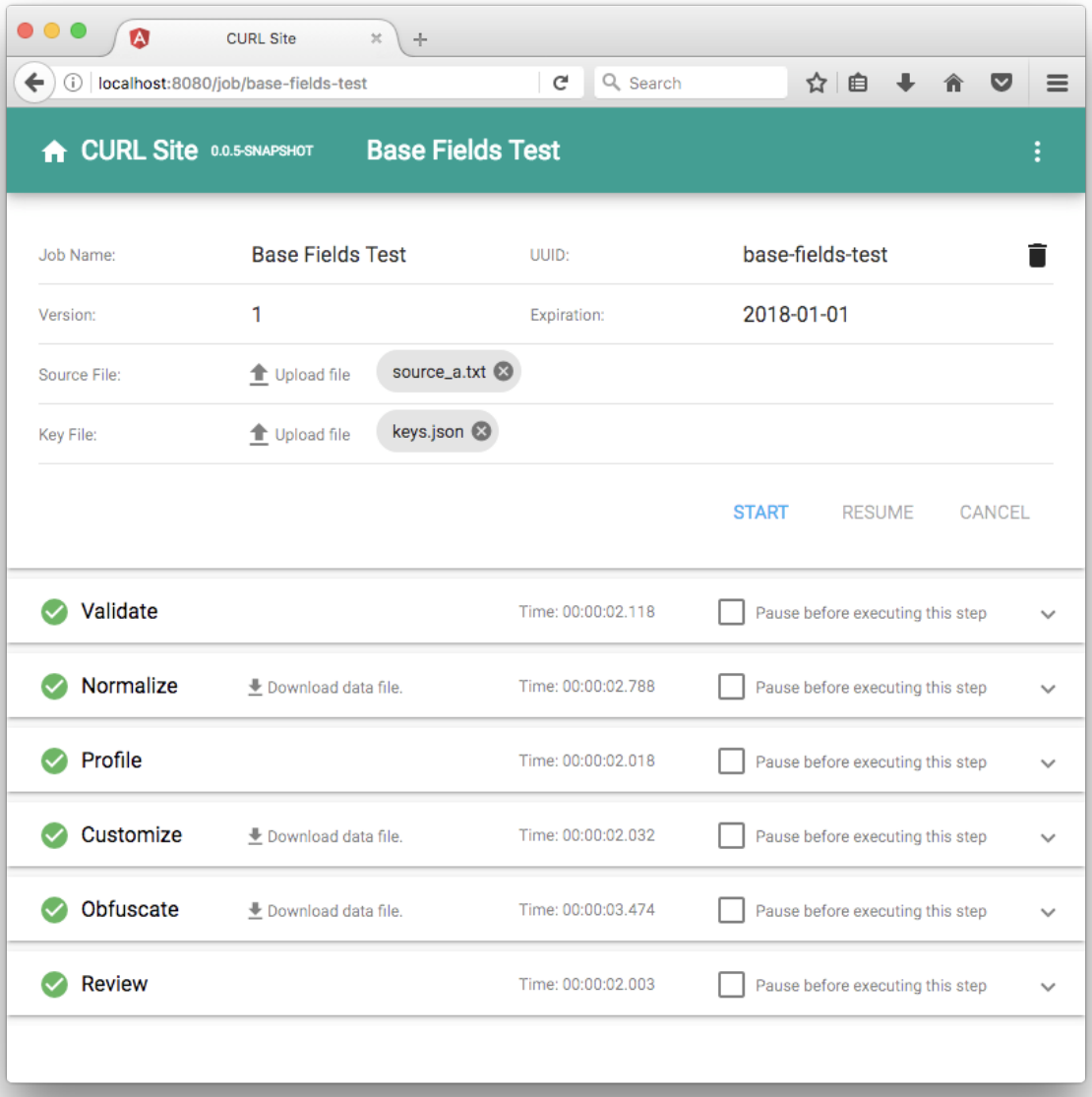


- Load a hashing job
    - Load CURL site job configurations:
        - Click “Upload job configurations file” link on the CURL Dashboard
        - Browse to the CURL Site configuration file from CURL team. CURL Site configuration filename ends with “.site-config.json”.
        - Click Open



- Start a hashing job

- Click on the arrow in front of the job name to open the job
- Click on “Upload file” link next to “Source File” to upload source data file. Source data file must be in delimited (tab or comma delimited) file format with headers as column names. Column names must match the column named defined in the configurations.
- Click on “Upload file” link next to “Key File” to upload hashing key from CURL KeyMaster. Filename of key file ends with “.keys.json”
- Click on “START” to start the hashing process.
- After the process is Done, a green checkmark will appear in front of all steps.



- Review hash data

    - Click on “Review” to review hash data. **Important**: make sure no clear-text data are included in the result.
- Download hash results
    - Click on “Download data file” next to Obfuscate to download. The downloaded hash file will be located in your “Download” folder.
- Data transfer to CURL Honest Broker
    - Transfer file with hash data to the CURL Honest Broker using one of the following methods
        - Secure file transfer protocol (sFTP)
        - Encrypted flash drive
        - PopMedNet

6. Other operations

- Multiple jobs
    - CURL Site supports loading multiple jobs. Simply go back to the Dashboard and load new job. Note that if the new job has the same job job universal unique identifier (UUID) as one of the existing ones, the existing job with the same UUID will be overwritten.
- Linkage job versioning:
    - CURL Site supports different versions of the same linkage job. Different versions of the same job will be identified by the same job UUID and the version number.
- Expiration date
    - An expired linkage job will not be able to execute. You will receive a error message if you try to start an expired job.
- Job settings
    - Click on the  icon on the title bar, then “Properties” to:
        - Set the number of threads to be used for the job. With more processing cores (e.g., CPUs) available, more threads will increase the processing speed.
        - Set the delimiter of the data file. Use “\t” for tab delimiter.
        - Set max warnings displayed. Default value = 10 warnings.
        - Set thread timeout. Default value = 60 minutes.
- Download log file of backend Java engine.
    - Click on the  icon on the title bar, then “Download log file”
- Warning and error messages:
    - Click on the title of a step to review any warning or error messages generated during the execution of that particular step.

- Pause a linkage job
    - Click on “Pause before executing this step” next to the step you would like to pause **before** the job is started.
    - Click on “RESUME” to resume a paused process.
- Download data file
    - Click on “Download data file” to download the output of a step in the workflow.
- Delete a linkage job
    - Click on the recycle bin on icon  to delete a linkage job. **Important:** Once a linkage is deleted all related data and hash files will also be deleted. However, the original data file, the hash key and the configuration file of the job will not be deleted and must be deleted separately.

If you have any additional questions or comments, please send an email to:
toan.ong@ucdenver.edu

THIS IS THE END OF THIS USER GUIDE