

# File Source

- [Modes](#)
- [Installation](#)
- [Configuration](#)
  - [Archive Directory](#)
  - [Change Threshold](#)
- [File Format](#)
  - [CSV v2](#)
  - [CSV v1 \(Legacy\)](#)
- [See Also](#)

The File Organizational Identity Source Plugin is designed to integrate using CSV files.

## Modes

Org Identity Source Mode	Support
Manual Search and Linking	Supported
Enrollment, Authenticated	Not supported
Enrollment, Claim	Not supported
Enrollment, Search	Supported
Enrollment, Select	Supported

Org Identity Sync Mode	Support
Full	Supported
Query	Supported
Update	Supported
Manual	Supported

## Installation


This is a non-core plugin, see [Installing and Enabling Registry Plugins](#) for more information.

## Configuration

The path specified in *File Path* must be readable by the COnamange Registry web server.

### Archive Directory

As of Registry v3.1.0, FileSource supports an *Archive Directory*. If specified, this directory must be writeable by the COnamange Registry web server, and FileSource will maintain two previous copies of the source file (using the basename of the file, appended with .1 or .2). This allows FileSource to perform a diff in order to determine changed records, which performs significantly better for larger files.

 Use of this configuration option is especially recommended for source files containing more than a few hundred sync'd records.

### Change Threshold

As of Registry v4.0.0, FileSource supports a *Warning Threshold*, which can prevent processing if an unexpectedly large number of changes (including new records) are detected. The threshold is expressed as a percentage (integer).

If a large number of changes are expected, the *Warning Threshold Override* can be enabled to skip the threshold check. The override will be automatically unset after processing. Note the first time a source is processed, the threshold check is automatically skipped. (The [Core Job Plugin](#) Sync job force parameter will also skip threshold checking, but the override is the recommended mechanism.)

## File Format

As of Registry v4.0.0, multiple file formats are supported:

- CSV v1: Legacy CSV format
- CSV v2: Flexible CSV format

For earlier versions, the only supported version is CSV v1.



CSV, or comma-separated value format, is not consistently implemented everywhere. While there is an [RFC](#) defining the format, a given implementation may or may not support it fully. For the avoidance of doubt, FileSource uses PHP's [CSV parsing functions](#), and so for any particular inconsistency PHP's behavior should be considered normative.

## CSV v2

The first line of the file is a header documenting which columns are present in the file:

- The first column in this file *must* always be `SORID`, a unique string identifying the record.
- Remaining columns are specified as follows:
  - `Model.field.[type]`, for example `EmailAddress.mail.personal` or `OrgIdentity.affiliation`
    - Values for `OrgIdentity.valid_from` and `OrgIdentity.valid_through` may be in any format supported by `strtotime()`, but are expected in UTC
    - Values for `OrgIdentity.date_of_birth` are expected in `YYYY-MM-DD` format
    - `Name.primary_name.type` may be used to specify a primary name, otherwise the first name found will be considered primary
  - `Identifier.identifier.type+login`, to flag that this column represents a login identifier, for example `Identifier.identifier.network+login`
  - `AdHocAttribute.tag`, for example `AdHocAttribute.color`

Remaining lines in the file each represent one inbound record, with columns as specified by the header.

Example files are [attached to this page](#).

CSV v2 format can be used with [Pipeline Relationship Syncing](#) by using the column names `OrgIdentity.manager_identifier` and `OrgIdentity.sponsor_identifier`.

## CSV v1 (Legacy)

The file format is fixed, with the following columns:

Column	Description	Required?	Notes
1	Unique Record Key	Yes	
2	<i>blank</i>		
3	<i>blank</i>		
4	Given Name	Yes	
5	<i>blank</i>		
6	Family Name		
7	Street Address		
8	Locality		
9	<i>blank</i>		
10	State		
11	Postal Code		
12	Country		
13	Email Address		
14	Telephone Number		
15	Telephone Country Code		
16	National Identifier		
17	Title		
18	Organization		
19	Match Reference Identifier		

20	Valid From		Any format supported by <a href="#">strptime()</a> , expected in UTC
21	Valid Through		Any format supported by <a href="#">strptime()</a> , expected in UTC
22	Personal URL		Added in v3.1.0
23	Affiliation		Affiliation corresponding to values of AffiliationEnum (Extended Types not supported)  Added in v3.2.0
24	Identifiers		Semi-colon (:) separated list of identifiers of the form <code>type[+login]:value</code> , where <i>type</i> corresponds to the values of IdentifierEnum (Extended Types not supported), the literal <i>+login</i> denotes that the identifier may be used for logging in, and <i>value</i> is the actual identifier  Added in v3.2.0
25	Date of Birth		YYYY-MM-DD format. Added in v3.3.0.
26	Ad Hoc Attributes		A JSON encoded list of ad hoc attributes, as tag/value pairs. The JSON document cannot contain linefeeds, and must be appropriately escaped for CSV. For example:  <code>"{"bonusPoints":585,"cashBalance":"\$1725.00"}"</code>  Note the comma is handled by the outer quotes.  Added in v3.3.0

## See Also

- [cm\\_file\\_sources](#)