CSV-Spezifikationen für  
ICRED-CSV-Inputplugin

Inhalt

[2. Einleitung 1](#_Toc430703497)

[3. Dateien 2](#_Toc430703498)

[1. Kopfzeilen der CSV-Dateien 2](#_Toc430703499)

[2. Datei: 1\_HEADER.csv 2](#_Toc430703500)

[3. Datei: 2\_PERIODS.csv 3](#_Toc430703501)

[4. Entitäten-CSV-Dateien 3](#_Toc430703502)

[Datensatzkennung 3](#_Toc430703503)

[Datenfelder 3](#_Toc430703504)

[Gruppendatenfelder 3](#_Toc430703505)

[5. Format der CSV-Dateien 4](#_Toc430703506)

# Einleitung

Für den International Converter for Real Estate Data (kurz „ICRED“) steht als Inputplugin ein CSV-Plugin zur Verfügung. Dieses Plugin erwartet im Folgenden näher definierte CSV-Dateien, deren Daten dann vom ICRED für die Konvertierung verwendet werdenReferenzen

|  |  |
| --- | --- |
| **[ISO8601]** | ISO 8601:2004 Data elements and interchange formats -- Information interchange -- Representation of dates and times, International Organization for Standardization and International Electrotechnical Commission, 2004. |
| **[RFC2045]** | N. Freed and N. Borenstein, Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies, http://www.ietf.org/rfc/rfc2045.txt, IETF, 1996. |
| **[MIMETYPES]** | MIME Media Types, http://www.iana.org/assignments/media-types/, IANA. |
| **[RFC5646]** | A. Phillips, M. Davis, Tags for Identifying Languages, http://www.ietf.org/rfc/rfc5646.txt, IETF, 2009. |
| **[UNICODE]** | The Unicode Consortium. The Unicode Standard, Version 5.2.0, defined by: The Unicode Standard, Version 5.2 (Boston, MA, Addison-Wesley, 2009. ISBN 978-1-936213-00-9). |

# Dateien

Die folgenden Dateibezeichnungen sind als Empfehlung zu verstehen und dürfen abweichen. Jede Datei ist als UTF-8 ohne Byte Order Mark zu encodieren.

Welche Dateien für das CSV-Import-Plugin zwingend notwendig sind, hängt vom gewählten zgif-Subset hab.

## Kopfzeilen der CSV-Dateien

Sämtliche Spaltenüberschriften der CSV-Dateien sind nach dem Uppercase-Underscore-Muster zu schreiben – d.h. in Großbuchstaben und einzelne Wortbestandteile ggf. durch einen Unterstrich getrennt.

Besteht eine Spaltenüberschrift aus zwei Teilen (z.B. Entitätsbezeichnung + Feldbezeichnung) wird beides durch ein Punkt getrennt. Beispiel:  
 UNIT.OBJECT\_ID\_SENDER

Entitätsbezeichner als Präfix vor dem Feldbezeichner sind nur notwendig, wenn in einer Entitäten-CSV-Datei Datenfelder anderer Entitäten auftauchen müssen oder wenn Gruppendatenfelder enthalten sind.

## Datei: 1\_HEADER.csv

Diese Datei enthält Metadaten zur Datenlieferung wie Angaben zum Ersteller, Erstelldatum, Sprache, etc.

Diese CSV-Datei ist immer 2 zeilig – erste Zeile Kopfzeile einer CSV-Datei und zweite Zeile die Daten.  
Nachfolgende Zeilen sind fehlerhaft und zu ignorieren.

Vergleiche auch mit Aufbau der meta.xml – beschrieben in der instance\_meta.xsd

Liste der Datenfelder:

|  |  |  |
| --- | --- | --- |
| Bezeichner | Beschreibung | Type |
| FORMAT | MIME-Type | Text (immer ”application/vnd.gif-ev.zgif”) |
| VERSION | Verwendete zgif-Version | Text nach Schema „A-X.Y.Z“ |
| PROCESS | Nummer des gewählten Subset | Text nach Schmea „X.Y“ |
| LANGUAGE | Sprache der Inhalte | nach RFC 5646 |
| CREATED | Erstelldatum | Datum & Uhrzeit nach ISO8601 |
| VALIDATED | Datum der Validierung, wird automatisch von einem Validator gesetzt (optionales Feld) | Datum & Uhrzeit nach ISO8601 |
| PUBLISHER | Erstveröffentlicher / Ersteller der Datei | Text |
| CREATOR | Software die zur Erstellung verwendet wurde, möglichst inklusive Versionnummer. Hilfreich zur möglichen Fehleraufklärung | Text |
| VALIDATOR | Software die zur Validierung verwendet wurde, möglichst inklusive Versionnummer. Hilfreich zur möglichen Fehleraufklärung (optionales Feld) | Datum & Uhrzeit nach ISO8601 |
| DESCRIPTION | Beschreibungstext zur Datei (ggf. Vermerke zur Verwendung; optionales Feld) | Text |

## Datei: 2\_PERIODS.csv

Diese Datei enthält die Auflistung sämtlicher verwendeter Perioden in der gesamten Datenlieferung, sofern das gewählte Subset Perioden unterstützt.

Liste der Datenfelder:

|  |  |  |
| --- | --- | --- |
| Bezeichner | Beschreibung | Type |
| IDENTIFIER | Eindeutige techn. Bezeichnung/ID der Periode | Text |
| FROM | Beginn-Datum der Periode | Datum nach ISO8601 |
| TO | End-Datum der Periode | Datum nach ISO8601 |
| PERIOD\_TYPE | Dauer einer einzelnen Periode | Zeitspanne nach ISO8601 |
| VALUE\_TYPE | Art der Periode | Ausprägung („ACTUAL“, „PLAN“, „FORECAST“, „BUDGET“, „OTHER“) |
| VALUE\_TYPE\_NAME | Bezeichnung der Periode (optional) |  |

## Entitäten-CSV-Dateien

Je Entitätsebene (Mandaten, Immobilien, Gebäude, Mieteinheiten, …) gibt es eine CSV-Datei. Die Dateibezeichnung sollte den Entitätbezeichner enthalten. Empfehlenswert ist eine logische/hierarchische Sortierung der Dateien durch einen numerischen Präfix – beispielsweise: 7\_UNIT.csv

Die Spalten dieser CSV-Dateien teilen sich in 3 Bereiche auf (in dieser Reihenfolge):

1. Datensatzkennung
2. Datenfelder
3. Gruppendatenfelder
4. Datenerweiterungsfelder

### Datensatzkennung

Zur eindeutigen Einsortierbarkeit einer Entität in den kompletten hiearchischen Entitätenbaum sind das Datenfeld OBJECT\_ID\_SENDER der in der Hierarchiestruktur höherliegenden Entitäten aufzulisten, ggf. inklusive dem IDENTIFIER der Periode, sofern das gewählte Subsets Perioden unterstützt.  
Diese Spaltenbezeichner erhalten immer das Präfix der zugehörigen Entität bzw. PERIOD.IDENTIFIER bei einem Periodenbezug.

Anschließend folgen die Spalten OBJECT\_ID\_SENDER, OBJECT\_ID\_RECEIVER, LABEL, VALID\_FROM, VALID\_TO zu jeder Entität.  
Hier entfällt das Präfix.

### Datenfelder

Die Liste der Datenfelder sind den jeweilgen Subsetspezifikationen zur jeweiligen Entität zu entnehmen. Die Bezeichnung der Spalten richtet sich dabei nach der Datensatzkennung, des entsprechenden Datenfeldes.

### Gruppendatenfelder

Manche Datenfelder einer Entität sind zwecks Übersichtlichkeit, Vermeidung von Verwechselungsgefahr und möglichen Namenskonflikten in Gruppen unterteilt. So gehören sämtliche geografischen Angaben wie Adressedaten (Straße, PLZ, Stadt, …), sowie die geogr. Koordinaten zur Gruppe ADDRESS. Angaben zum Energieausweis zur Gruppe ENERGY\_RATING. Entsprechenden Datenfeldern ist der Gruppenname als Spaltenpräfix voranzustellen.

### Datenerweiterungsfelder

Zu den vordefinierten Feldern je Subset gibt es die Möglichkeit weitere Daten je Objekt/Entität zu ergänzen, welche nicht zum Standard gehören. Diese zusätzlichen Datenfelder erfordern somit eine Absprache zwischen Sender und Empfänger.

Die Bezeichner der Datenfelder können frei gewählt werden, sollten jedoch nur aus Buchstaben, Ziffern und Trennzeichen wie - und \_ bestehen. Die Daten selbst werden immer zunächst als reiner Text interpretiert und ggf. vom Empfänger zu konvertieren/decodieren.

Aufbau der Bezeichner:  
Es gibt 2 Arten von Erweiterungsfeldern. Die erste Art ist ein einfaches Datenfeld mit einem Datensatz. Dem eigentlichen Bezeichner ist der Präfix EXTENSION.ITEM. voranzustellen, beispielsweise: EXTENSION.ITEM.Pruefsumme

Die Zweite Art entspricht einer einfachen Liste je Bezeichner und wird mit dem Präfix EXTENSION.LIST. gekennzeichnet, beispielsweise: EXTENSION.LIST.PARENT\_IDS

## Format der CSV-Dateien

Als Zeilentrennzeichen werden die Zeichen Carriage Return (0x0D) plus Line Feed (0x0A) erwartet. Das Zellentrennzeichen wird das Semikolon (0x3B).

Bei Textwerten wird als Textbegrenzungszeichen das doppelte Anführungszeichen (0x22) als Prä- und Suffix erwartet. Bei Vorkommen dieses Zeichen innerhalb des Textes wird dieses durch dasselbe Zeichen escapet.

Null-Werte sind ohne Angabe eines Zeichens zwischen den Semikolon zu übertragen.

Die Formatierung der einzelnen Datentypen entspricht der offiziellen zgif-Spezifikation.