


# 344. Import §21 Files

## Background

The §21 dataset is a German benchmarking dataset mandated by the law on hospital reimbursement ([Paragraph 21: "Übermittlung und Nutzung von Daten - des Gesetzes über die Entgelte für voll- und teilstationäre Krankenhausleistungen" \(Krankenhausentgeltgesetz - KHEntgG\)](#)). It is primarily used to derive the German inpatient billing system based on Diagnosis Related Groups (DRGs) and has to be provided by every German hospital on a yearly basis. It is thus supported by all hospital patient administration and billing systems established in Germany. Due to its standardization and high availability, the §21 dataset has been established as a source dataset for benchmarking projects and other regional and national collaborations. It is thus also an ideal candidate for a "quick win" i2b2 implementation, as its format is well-structured and documented and it can be produced easily by any German hospital, rapidly providing a functional prototype with local data.

 It should be noted, however, that the §21 dataset is primarily reimbursement-oriented, which may limit its usefulness in the research area.

The §21 dataset is based on a standard delimited ("CSV" = comma separated values) format with the following files and contents of interest, among others ([full specification in German language](#))

- fall.csv: encounter-related data and patient demographics
- fab.csv: transfer history within encounter
- icd.csv: diagnosis codes based on ICD10-GM catalog (GM: German Modifications)
- ops.csv: procedure codes based on German OPS medical procedures catalog
- drg.csv: diagnosis related group codes

The drg.csv file is not part of the standard §21 dataset, but can be generated from the other files using a certified DRG grouping software.

The §21 dataset itself does not contain the relevant terminologies - all coded items are represented just as their codes (e.g. an ICD diagnosis code). Thus, the terminologies have to be imported and joined to achieve an interpretable dataset.

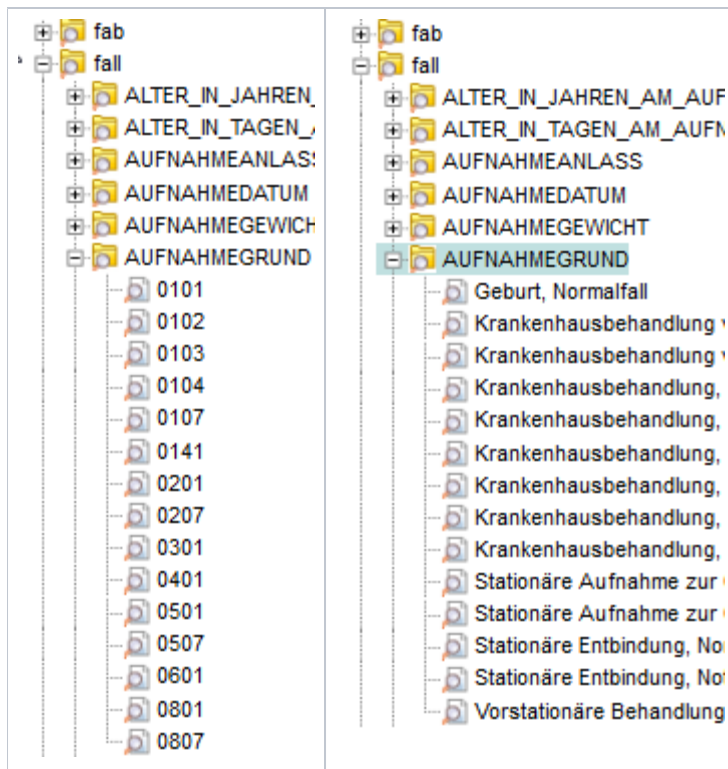
## Concept

As the §21 is provided in CSV format, the standard CSV extraction capabilities of the IDRT platform can be used, either through the [Import and Mapping Tool \(IMT\)](#) or through direct use of the underlying Talend ETL jobs. The IDRT §21 provides ready-to-use configuration files for the full set of §21 source data files, with versions adapted to the several releases of the §21 specification over the past years.

Since the §21 files are linked together only by an encounter number, the patient number has to be added in order to achieve full compatibility with the IDRT CSV import module. This addition of patient IDs is carried out in an integrated pre-processing step before the actual import.

After the initial import, the §21 dataset would be represented in i2b2 in a very low-level format still close to the source CSV files: all coded items would just be displayed with their codes (and not the related names), and items that relate to large, hierarchical terminologies (e.g. ICD diagnosis classification) would be presented as very long lists. In a final post-processing step, the relevant standard terminologies (in this case ICD10, OPS and G-DRG) as well as §21-specific codelists (e.g. for admission type, clinical service) are imported and merged with the dataset to provide a well-structured and human-readable hierarchical representation (see figure below).

Raw import	Import after terminology integration
------------	--------------------------------------



## Importing §21 files with the IDRT Import & Mapping Tool

The §21 import functions are fully integrated into the [IDRT Import and Mapping Tool \(IMT\)](#), which provides the following dialogue to configure an import job:

**P21 Import Settings**

P21 Import Settings

Truncate i2b2 Project? ☐

Truncate Previous Queries? ☐

Set patient\_count after import ☐

P21 Folder  ...

P21 Version

Date Pattern

Save Settings? ☐

Name	Description
Truncate i2b2 Project?	If you check this, the project will be truncated before the data is uploaded.
Truncate Previous Queries?	This will truncate the previous queries tables.
Set patient_count after import	This will fill the c_totalnum column.
P21 Folder	Here you can select the folder that contains all P21 files you want to upload.
P21 Version	Here you can select the corresponding version for your P21 dataset.
Date Pattern	Select the date pattern used in the P21 Files.
Save Settings?	This will save the settings for the next import. <b>Note:</b> The truncate checkboxes are never saved.

When you click **finish**, the upload starts. You can observe the progress in the progress bar of the [Import Browser](#).

## Using TOS ETL Jobs for §21 import

The underlying §21 import jobs are provided as part of the IDRT Talend Open Studio distribution. They consist of 3 separate master jobs for the pre-processing of the raw data files, the actual loading and the post-processing of the imported data. As there exist several versions of the §21 dataset (depending on the year of patient discharge), the correct version of the import configuration files need to be copied into the directory containing the raw data files.

Using the TOS jobs instead of the IMT frontend enables the creation of automated ETL scripts (e.g. for nightly updates) or combining several import steps into complex ETL pathways.