Task

For understanding:

Various files are downloaded daily from a third party system and synchronised with a database. The files are delivered in a **TAB-separated text format**.

The file and field specification can be found in the specification document under **/doc/FileSpec_1.0.pdf**.

The relevant file for this task is: /sample/Substances.dat (specification: chap. 3.3.2 on page 4 f).

The file contains a list of pure chemical substances (substance). Every pure chemical substance contains a series of internationalised pure substance names (synonym).
z.B.

R	1061	7439-92-1		231-100-4	1	0	1	0	0	0
RN	1061	0	DE	Blei						
RN	1061	0	EN	Lead						

Representation in UML:

Task:

Please develop a class library that reads the file **/sample/Substances.dat** and makes it available as Java objects for further processing.

The following requirements must be observed:

- For each line, the number of attributes contained and the length of the individual attributes must be checked according to the specification document.
- Further file types (e.g. the enclosed /sample/Companies.dat) as well as resulting objects should be able to be added with as little intervention as possible in the existing code while maximising its reuse.
- Since the files can be of almost any length, it is important that the Reading is done in blocks - i.e. in the case of Substances.dat per pure substance - so that the entire file or the resulting objects do not have to be loaded into the main memory before further processing. The sorting of the file can be used here.

These classes should be stored separately from the test under **/src/prod/.**
The package structure can be freely determined.

Please develop a unit test that reads the file using the **/sample/Substances.dat** file with your class library and writes it in the format

```
CAS-NR, first synonym of the current locale
```

on the console.

Example:

```
118725-24-9, (1,3-dioxo-2H-benz(de)isoquinolin-2-ylpropyl)hexadecy...
```

Check in your test whether the last substance is called "TestD" or "TestE" according to the locale set:

```
R       90008145822222-22-2     1       -1      1       0       0       1
RN      9000814580      DE      TestD
RN      9000814580      EN      TestE
```

Source folder for unit test is **/src/test/.**

Note:

The two classes Dummy.java and DummyTest.java are only present so that the two source folders are not empty and may be renamed or removed as desired.
The language or the locale used in the JVM can be specified with the command line option -Duser.language
command line option.