

# PxEdit 4.0

## Introduction

# About PxEdit

- Developed at Statistics Finland to efficiently **edit and create** statistical tables
- Based on the **px-file** format, but not restricted to it
- Main parts:
  - main window (+ system tray)
  - table spreadsheets
  - metadata editor
  - other special purpose windows
- Multilingual, easy to translate

# Design philosophy

- px tables
  - must be able to handle all **normal** px files
  - should be able to handle incomplete or even damaged tables
  - is able to repair numerous errors
  - is able to point out numerous other errors
  - changes are reflected immediately in the spreadsheet
  - no artificial table size limitations
- Other source file formats
  - imports data or metadata from different sources including text files, Excel tables and clipboard data
- **Feasible** new features are being added regularly

# Terminology

- **Variable**

- datacube edge (~ classification)
- e.g. Country, Gender, Year

- **Value**

- a single variable item (~ class)
- either value texts (compulsory) or value codes (may be missing)
- e.g. Finland, Sweden; man, woman; 1995, 1996, 1997, ...

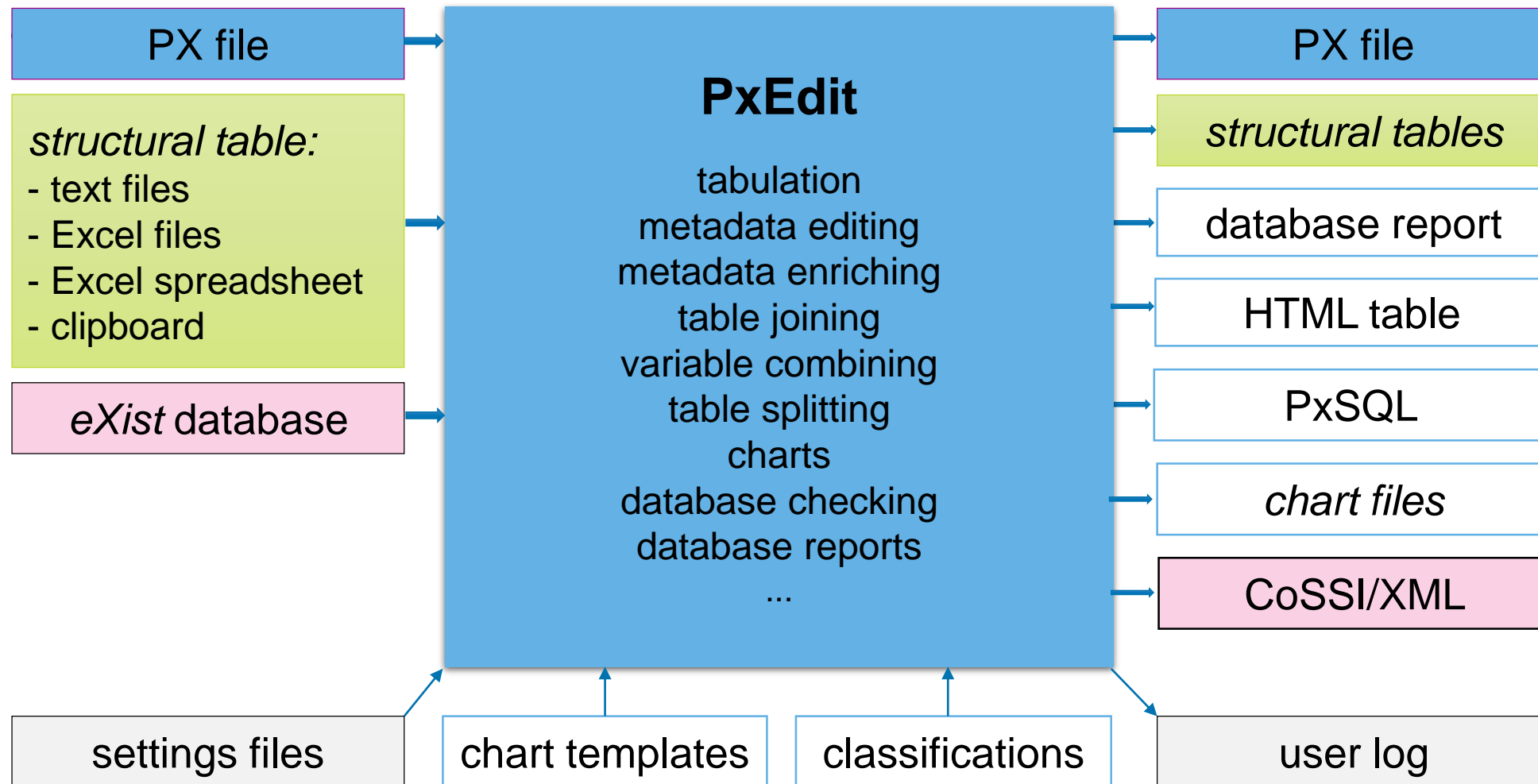
- **Data**

- numeric data
- dot codes for missing, concealed, etc. data: ". " – " . . . . . "

# Settings

- Working environment: workstation or LAN
- Organisational settings: `pxedit_main_40.ini`, `xdf_40.ini`
  - mandatory keywords
  - environmental settings
  - user and error logs
  - own keywords
  - ...
- User settings: `pxedit_40.ini`
  - personal settings

# PxEdit in a nutshell



# Main window

- Open files
  - *File/Open* [**Ctrl+o**], most recently used list
  - drag and drop support (files or directory folders)
  - read *structural* tables from text files, Excel, xls, csv, ...
- Save in different file formats
- Change user settings
  - language, fonts, *Fill item*, user level ...
- Special operations
  - table joining, table cloning, database checking, reporting and standardising, ...
  - help always accessible [**F1**]

# Table window

- Displays the data **visually**
  - range colouring
  - table, variable, value or cell notes
  - data range, variable names, ...
- *File, Edit, Language* and *Window* menus
- Clipboard support
- Multiple tables may be open simultaneously
  - only one table is **active** at a time
- All data is **not** loaded into the spreadsheet!



# Free tabulation

- Change the variable order
  - aggregation is possible, too
- Select variable values
- Sort variable values or codes
  - alphabetically
  - according to the data values
  - free sorting
- All settings can be **reset** to their original state

# Input data

- px files (px) [**Ctrl+o**]
- Metadata files (pxk)
- **Structural** text files (txt, dat, csv, ...)
  - tabulator, semicolon, comma or space delimited; arbitrarily large
- Structural Excel files (xls, xlsx, csv)
- Structural tables from Excel worksheet [**Ctrl+e**, **Shift+Ctrl+e**]
- Archives (zip)
- Structural tables from clipboard
- *eXist* database (Statistics Finland's)
- New empty tables [**Ctrl+n**]

# Opening a px file: Unicode or *CODEPAGE*

- If the file contains the *Byte Order Mark* (BOM) header → **Unicode**
- Selecting px *Unicode files* in the *File Open* window
  - the Unicode check is passed → **Unicode**
- *Character conversion/Ignore coding*
  - is not set, and *CODEPAGE* is
    - `utf-8` and the Unicode check is passed → **Unicode**
    - a valid iso-8859 coding → **ISO-8859**
    - a valid WinANSI coding → **WinANSI**
    - a valid DOS codepage → **DOS**
  - is set, or *CODEPAGE* is not recognised
    - *Settings/Check for Unicode* is set, and the Unicode check is passed → **Unicode**

# Opening a px file: *LANGUAGE*

- The language dependent codes are based on the main table language
  - the system language will be used if *Character conversion/System language coding* is set or there is no *LANGUAGE* setting in the px file
  - if the code cannot be deducted, the Western code will be used instead
- *Character conversion/ISO-8859* is set
  - a valid iso-8859 coding (e.g. `iso-8859-15`) → **ISO-8859**
- *Character conversion/DOS coding* is set and *CHARSET* is not ANSI
  - a valid DOS codepage (e.g. `codepage-858`) → **DOS**
- Otherwise a valid WinANSI coding (e.g. `Windows-1252`) → **WinANSI**

# Structural tables

- A structural table contains the **necessary** data
  - table heading
  - language code (optional)
  - variable names
  - variable values and/or codes
  - data part
  - hierarchy level (optional)
  - keyword block (optional)
  - structural files may be multilingual
- Two different file types
  - general table format
  - record mask for fixed-length sequential files

# Syntax checking

- Severe errors
  - mandatory keywords: table will not be opened
  - others: errors will be bypassed
- All found errors will be shown
  - message contains the table and file names
- Some errors may be **repaired**
  - e.g. missing *UNITS*, timestamps and *DECIMALS* problems
  - multiple table opening includes the **Repair all** option

# Saving a px file: Unicode or *CODEPAGE*

- *Settings/Save in Unicode* is set → **Unicode**
- Selecting px *Unicode files* in the *File Save* window → **Unicode**
- If *CODEPAGE* exists and no *Character conversion* is set
  - `utf-8` → **Unicode**
  - a valid iso-8859 coding → **ISO-8859**
  - a valid WinANSI coding → **WinANSI**

# Saving a px file: *LANGUAGE*

- The language dependent codes are based on the main language
  - the system language will be used if *Character conversion/System language coding* is set or there is no *LANGUAGE* setting in the px file
  - if the code cannot be deducted, the Western code will be used instead
  - *CODEPAGE* will be set as the used conversion code
  - *CHARSET* will be set as ANSI
  - DOS conversion is **not** supported in saving
- *Character conversion* is set to *iso-8859*
  - a valid iso-8859 coding (or *iso-8859-15*) → **ISO-8859**
- Otherwise a valid WinANSI coding (or *Windows-1252*) → **WinANSI**
- Unicode files will be saved with the Byte Order Mark (BOM) header



# Output formats

- px file (px) [**Ctrl+s**]
- Pure metadata file (pxk)
- Excel files (xls, csv)
- Structural files (csv)
- Text files (txt, ...) [**Ctrl+t**]
  - freely adjustable
- Export to Excel
- Copy to clipboard [**Ctrl+c**]
- Other types: HTML files and **tables** (htm, html), CoSSI/XML table formats (xml), PxSQL files (sql) and *eXplorer* tables (txt)

# Values and codes editing

- Manual editing
- Search and replace
- *Value* Tab menus
  - character case changes
  - remove leading characters or dots
  - handle leading codes
  - clipboard support [**Ctrl+c/v**]
  - ...
- Adjust classifications with external classification file (pxc)
- Code creation

# Metadata editing

- Multilingual keywords
  - single line keywords
  - multiline keywords
- Menu selections
- Date fields
- Special input fields (e.g. *TIMEVAL*)
- Fill empty fields with default keywords [**Ctrl+k**]
- **Import** keywords from another file or open table [**Ctrl+i**]
- *Meta part* Tab for checking the file format

# Data editing

- Cell by cell editing (may also be prohibited)
- Paste from clipboard [**Ctrl+v**]
- Search and replace [**Ctrl+f**]
- Aggregate variables
- Mask or delete values
- Data anonymising
- *Calculator*
  - make new values
  - change existing values

# Special functions

- Merge two or more tables into a new table [**Ctrl+j**]
- Add a new variable or new values
- Combine or **split** variables
- Remove values
- Check, standardise or convert database
- Database report
- Compare variables from two tables
- Clone (duplicate) tables [**Ctrl+d**, **Shift+Ctrl+d**]
- Split the table into partial tables along one variable

# Multilingualism

- The multilingual tables will be manipulated with the keywords *LANGUAGE* and *LANGUAGES*
- *LANGUAGE* is the main language, the others are extra languages
  - almost all keywords have language-based variants in the px file
  - the language code of the extra language must be given after the keyword in brackets (*KEYWORD[fi]*)
- The current language can be changed with the table or metadata editor language menus
- PxEdit will fill in the **missing** keyword contents for each extra language from the corresponding main language keyword

# There are many ways to add languages

- Add a new language to a monolingual table with the **Add** function in the *LANGUAGES* input field
  - other metadata input manually
- Join two or more tables with different languages
  - the tables should be uniform
- Import metadata from another table with different languages
  - the import window will then have a separate language menu
- Import structural tables with a multilingual keyword block

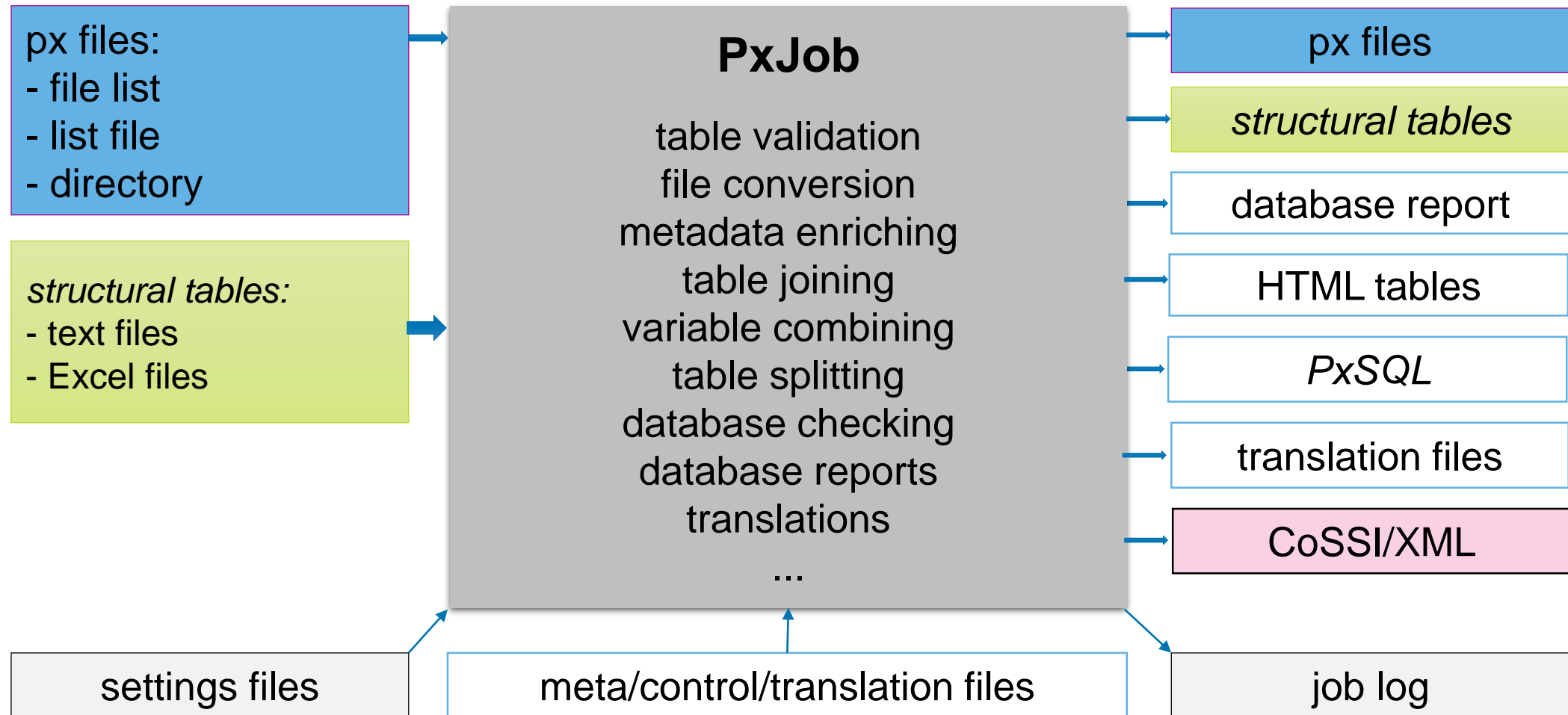
# PxJob: command line interface (batch)

**PxJob** {job} [**in**] {out} {err} {copy} {meta} {set} {path} {log} {-} {!}

job	job type (default: px)
<b>in</b>	source directory, file, file list or list file ( <b>mandatory</b> )
out	output directory or file (default: source files)
err	directory for erroneous files
copy	directory for source files
meta	metadata file or directory
set	settings file
path	common directory path
log	log file
-	options
!	switches



# PxJob in a nutshell



# PxEdit history

- 0.9 1999 Basic structure (Kimmo *Linna*)
- 1.0 2003 GUI, memory handling, variable compare
- 1.1 2003 Data anonymising, structural tables, sparse matrices
- 2.0 2003 Batch, StatXDF, keyword fetching, cell notes
- 2.1 2004 Huge tables, *TIMEVAL*, New table, font charsets
- 2.2 2004 Batch templates, classification adjusting, PX-Tool
- 2.3 2005 System Tray, aggregations, formulae, table mask
- 2.4 2006 CoSSI/XML, *DATASYMBOL*\*, Excel comments
- 3.0 2011 Multilingual tables, Unicode, database standardisation
- 3.1 2013 PX-Job, PxSQL, database conversions
- 3.2 2015 User levels, metadata injection, reporting, variable combining
- 3.3 2017 Data report, Join/merge, PxJob enhancements
- 4.0 2019 **Keyword blocks, variable splitting, multilingual structural tables**