

# PanelWhiz Plugins: Automatic Vector-Oriented Data Cleaning for Large Scale Panel Datasets

Markus Hahn

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen)

5th German Stata Users Group meeting  
(April 2, 2007)



## "Employed Again Last Year" (Source: GSOEP)

Year 2004	Year 2005
[1] First Job	[1] First Job
[2] Job After Break	[2] Job After Break
[3] Job With New Employer	[3] Job With New Employer
[4] New Job-Self Employed	[4] Company Taken Over
[5] Company Taken Over	[5] Changed Job, Same Firm
[6] Changed Job, Same Firm	[6] New Job-Self Employed

## "Number Employees in Firm" (Source: GSOEP)

Year 2004	Year 2005
[1] LT 5	[1] LT 5
[2] GE 5 and LT 20	[2] GE 5 and LT 20
[3] GE 20 and LT 100	[3] GE 11 and LT 20
[4] GE 100 and LT 200	[4] GE 20 and LT 100
[5] GE 200 and LT 2000	[5] GE 100 and LT 200
[6] GE 2000	[6] GE 200 and LT 2000
[7] Self-Employed	[7] GE 2000
	[8] Self-Employed

## What a plugin does

- Ensures time consistency of variables
- Data cleaning / Conversion (DM $\Rightarrow$ EURO, ...)
- Creating new variables (real values, ...)

## How a plugin works

- Executable ado-files for each item
- Plugin only describes what actions and when these actions are to be performed
- Plugin system executes the necessary actions
- Auto-documentation of what is done

## Example Plugin 1/2

```
program define bp_lglohn

*! -----
*! Plugin      : [lglohn] Brutto Lohn- und Gehaltssumme
*! Author      : Markus Hahn
*! Email       : hahn@rwi-essen.de
*! Web         : http://www.panelwhiz.eu
*! Copyright   : Copyright 2006. All rights reserved. Unauthorized copying prohibited.
*! Version     : 1.0
*! Years       : a b c d e f g h i j k l
*! Possible    : a b c d e f g h i j k l
*! Changes     : a b c d e f g h i j k l
*! -----

pwplugin begin, item(lglohn) year(2004)
    .action.dm2euro

    .plugin.newitem
    .action.dm2euro
    .action.real, thisyear
pwplugin end

end
```

## Example Plugin 2/2

```
program define bp_aawest

*! -----
*! Plugin   : [aawest] AA: West-/Ostdeutschland
*! Author   : Markus Hahn
*! Email    : hahn@rwi-essen.de
*! Web      : http://www.panelwhiz.eu
*! Copyright : Copyright 2006. All rights reserved. Unauthorized copying prohibited.
*! Version  : 1.0
*! Years    : a b c d e f g h i j k l
*! Possible  : _ _ _ d e f g h i j k l
*! Changes  : _ _ _ _ _ g h i j k l
*! -----

pwplugin begin, item(aawest) year(2004)
    .action.recode (2=0), g h i j k l
pwplugin end

end
```

# Automatically generated plugin-do-file (Auto-doc)

```
Viewer (#1) [view "d:/panelwhiz_workingdir/iabpp/proj/test/test-plugins.do"]
Back Refresh Search Help Contents What's New News
Command: view "d:/panelwhiz_workingdir/iabpp/proj/test/test-plugins.do"

///
///
///
/// PanelWhiz
///
/// Homepage: http://www.panelwhiz.eu
/// Help: http://forum.panelwhiz.eu
/// Contact: Dr. John P. Haisken-DeNew

///
///
///
/// +-----+
/// [aawest] AA: west-/Ostdeutschland
/// +-----+
///
/// ---[ PREPARATION OF ITEMS ]-----
/// label variable aawest "west-/Ostdeutschland"
///
/// ---[ BEFORE ]-----
///
/// ---[ A - 1993 ]-----
///
/// ---[ B - 1994 ]-----
///
/// ---[ C - 1995 ]-----
///
/// ---[ D - 1996 ]-----
///
/// ---[ E - 1997 ]-----
///
/// ---[ F - 1998 ]-----
///
/// ---[ G - 1999 ]-----
/// recode aawest (2=0) if year == 1999
///
/// ---[ H - 2000 ]-----
/// recode aawest (2=0) if year == 2000
///
/// ---[ I - 2001 ]-----
/// recode aawest (2=0) if year == 2001
///
/// ---[ J - 2002 ]-----
/// recode aawest (2=0) if year == 2002
///
/// ---[ K - 2003 ]-----
/// recode aawest (2=0) if year == 2003
///
/// ---[ L - 2004 ]-----
/// recode aawest (2=0) if year == 2004
///
/// ---[ AFTER ]-----
```

## Why we do **not** use normal ado-files

- Relative high complexity of code (loops, ...)  
⇒ Not that easy to understand
- The code should be self-explanatory (transparency)
- Simpler code could is easier to process  
⇒ for generation purposes ("plugin wizard")  
⇒ for auto-documentation purposes beyond log-files



## Implementation of the plugin system

- Stata and Mata code
- Class system (\*.class)
- Dialog boxes (\*.dlg)
- On-the-fly generation of smcl-files (Output for Plugin Wizard)

## Recoding a variable

```
.action.recode (2=0), a b c d e
```

- Similar syntax to stata's recode
- Item variable gets recoded: 2 becomes 0
- Action is done for the waves a-e

## Changing a variable

```
.action.replace wage/12, a b c d e
```

- Similar syntax to stata's replace
- Item variable gets divided by 12
- Action is done for the waves a-e

## Handling money variables

```
.action.dm2euro
```

- Item variable will be converted to EURO (years < 2002)

```
.action.real, thisyear
```

- Item variable will be converted to real values
- The deflator (price index) of the current year is used

## Label definition

```
.action.label var(Variable Label)
```

- Item variable gets new variable label

```
.action.label val(1 "Yes" 0 "No")
```

- Item variable gets new value label

## Creating new variables

```
.plugin.newitem
```

- A new variable will be created
- Variable name will be automatically assigned
- Following actions/commands will refer to new variable

## Plugin Wizard

- Tool for creating plugins by "clicking"  
⇒ No manual typing necessary
- User interface: generated smcl-files + dialog boxes
- Supports the complete syntax of plugin code
- Once created plugins can be re-read

Viewer (#1) [view "d:/panelwhiz/iabbp/views/wavel\_AA\_DE.smcl"]

Back Refresh Search Help Contents What's New News

Command: view "d:/panelwhiz/iabbp/views/wavel\_AA\_DE.smcl"

Start Page > Wave (1) > AA <Project Page>

IABBP Data File INDEX Browse Page for [ L ]  
Panelwhiz IABBP (1.0) Aug 2006 <john@panelwhiz.eu>

---

AA: Allgemeine Angaben

test2004:

[bu1a2004]	( S T K N )	<b>bundesland 2004 (codierung ab 2000)</b>
	X w p _	<b>aaбуla:</b> Bundesland
[wo2004]	( S T K N )	<b>west/ost</b>
	X w p _	<b>aawest:</b> West-/Ostdeutschland
[188a]	( S T K N )	<b>betriebsrat/personalrat</b>
	X w p _	<b>aaбetr:</b> Betriebs-/Personalrat vorhanden
[188b]	( S T K N )	<b>and. betriebsspez. form der mitarb.vertretung</b>
	X w p _	<b>aaбetra:</b> Andere Form der Mitbestimmung vorhanden
[189]	( S T K N )	<b>betrieb/dienststelle ist ...</b>
	X w p _	<b>aaeinz:</b> Stellung des Unternehmens / Abhaengigkeitsverhaeltnis
[190]	( S T K N )	<b>rechtsform</b>
	X w p _	<b>aaform:</b> Rechtsform des Unternehmens
[191]	( S T K N )	<b>mehrheitl./ausschließl. eigentum</b>
	X w p _	<b>aaherk:</b> Herkunft des Eigentueners
[193]	( S T K N )	<b>gründungsjahr ab 1990</b>
	X w p _	<b>aaग्रjhrb:</b> Gruendungsjahr B : seit wann besteht Betrieb in heutiger Form (nur Ost)
[195b]	( S T K N )	<b>branche (neu ab 2000: w293)</b>
	X w p _	<b>aaбrnoo:</b> Branche in WZ93 Gliederung

Data Browse Page: Start Plugin Wizard

Viewer (#1) [view "c:/panelwhiz/temp/plugin\_wizard.smc"]

Back Refresh Search Help Contents What's New News

Command: view "c:/panelwhiz/temp/plugin\_wizard.smc"

Start Page > wave (1) > AA > Plugin Wizard <Project Page>

---

Panelwhiz IABBP - Plugin Wizard  
Panelwhiz IABBP (1.0) Aug 2006 <john@panelwhiz.eu>

---

**bp\_aabu1a.ado - (p)** Based On: NEW Plugin

[Create] [View] [Erase] [New] [Open (Original)] [Open (Personal)]

---

**Plugin Information** [Default] [Edit]

Plugin : [aabula] Bundesland  
 Author : Markus Hahn  
 Email : hahn@rwi-essen.de  
 web : http://www.panelwhiz.eu  
 Copyright : copyright 2006. All rights reserved. unauthorized copying prohibited.  
 version : 1.0  
 Years : a b c d e f g h i j k l  
 Possible : a b c d e f g h i j k l  
 Changes : -----

---

**Plugin Initialization** [Default] [Edit]

Update year : 2004  
 Depends on : <nothing>

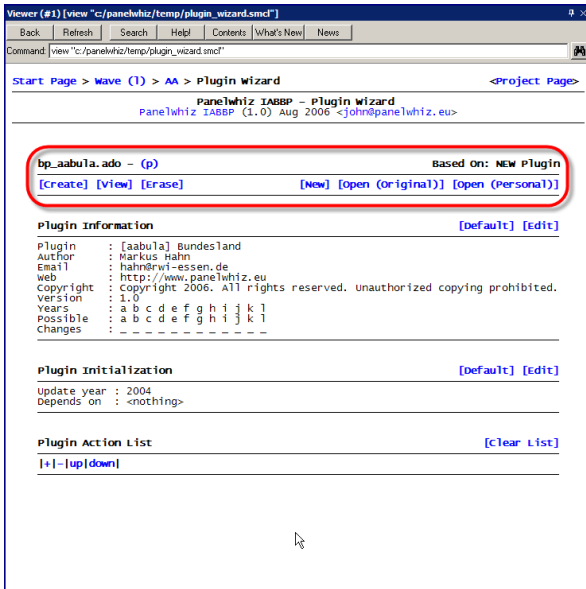
---

**Plugin Action List** [Clear List]

[+]-[up|down]

## Plugin Wizard





Create, view, erase, open, ...

Viewer (#1) [view "c:/panelwhiz/temp/plugin\_wizard.smc"]

Back Refresh Search Help Contents What's New News

Command: view "c:/panelwhiz/temp/plugin\_wizard.smc"

Start Page > wave (1) > AA > Plugin Wizard <Project Page>

Panelwhiz IABBP - Plugin Wizard  
Panelwhiz IABBP (1.0) Aug 2006 <john@panelwhiz.eu>

---

bp\_aabula.ado - (p) Based On: NEW Plugin

[Create] [View] [Erase] [New] [Open (Original)] [Open (Personal)]

---

**Plugin Information** [Default] [Edit]

Plugin : [aabula] Bundesland  
Author : Markus Hahn  
Email : hahn@rwi-essen.de  
web : http://www.panelwhiz.eu  
Copyright : copyright 2006. All rights reserved. unauthorized copying prohibited.  
Version : 1.0  
Years : a b c d e f g h i j k l  
Possible : a b c d e f g h i j k l  
Changes : - - - - -

---

**Plugin Initialization** [Default] [Edit]

Update year : 2004  
Depends on : <nothing>

---

**Plugin Action List** [Clear List]

[+]-[up|down]

Information about plugin

Viewer (#1) [view "c:/panelwhiz/temp/plugin\_wizard.smc"]

Back Refresh Search Help Contents What's New News

Command: view "c:/panelwhiz/temp/plugin\_wizard.smc"

Start Page > wave (1) > AA > Plugin Wizard <Project Page>

Panelwhiz IABBP - Plugin Wizard  
Panelwhiz IABBP (1.0) Aug 2006 <john@panelwhiz.eu>

---

**bp\_aabula.ado - (p)** Based On: NEW Plugin  
[\[Create\]](#) [\[View\]](#) [\[Erase\]](#) [\[New\]](#) [\[Open \(Original\)\]](#) [\[Open \(Personal\)\]](#)

---

**Plugin Information** [\[Default\]](#) [\[Edit\]](#)

Plugin : [aabula] Bundesland  
Author : Markus Hahn  
Email : hahn@rwi-essen.de  
web : http://www.panelwhiz.eu  
Copyright : copyright 2006. All rights reserved. unauthorized copying prohibited.  
Version : 1.0  
Years : a b c d e f g h i j k l  
Possible : a b c d e f g h i j k l  
Changes : - - - - -

---

**Plugin Initialization** [\[Default\]](#) [\[Edit\]](#)

Update year : 2004  
Depends on : <nothing>

---

**Plugin Action List** [\[Clear List\]](#)

[|+|](#) [|up|](#) [|down|](#)

## Action List

Viewer (#1) [view "c:/panelwhiz/temp/plugin\_wizard.smc"]

Back Refresh Search Help Contents What's New News

Command: view "c:/panelwhiz/temp/plugin\_wizard.smc"

Start Page > wave (1) > AA > Plugin Wizard <Project Page>

Panelwhiz IABBP - Plugin Wizard  
Panelwhiz IABBP (1.0) Aug 2006 <john@panelwhiz.eu>

---

**bp\_aabula.ado - (p)** Based On: NEW Plugin  
[\[Create\]](#) [\[View\]](#) [\[Erase\]](#) [\[New\]](#) [\[Open \(Original\)\]](#) [\[Open \(Personal\)\]](#)

---

**Plugin Information** [\[Default\]](#) [\[Edit\]](#)

Plugin : [aabula] Bundesland  
Author : Markus Hahn  
Email : hahn@rwi-essen.de  
web : http://www.panelwhiz.eu  
Copyright : copyright 2006. All rights reserved. unauthorized copying prohibited.  
Version : 1.0  
Years : a b c d e f g h i j k l  
Possible : a b c d e f g h i j k l  
Changes : - - - - -

---

**Plugin Initialization** [\[Default\]](#) [\[Edit\]](#)

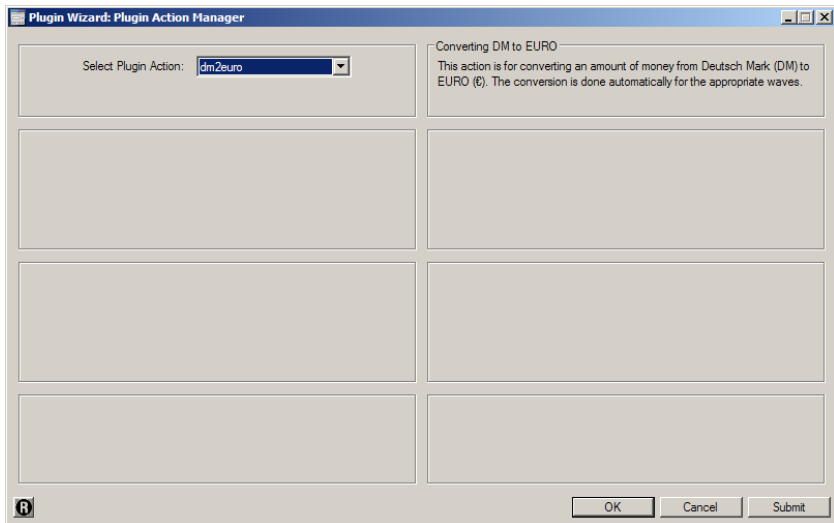
Update year : 2004  
Depends on : <nothing>

---

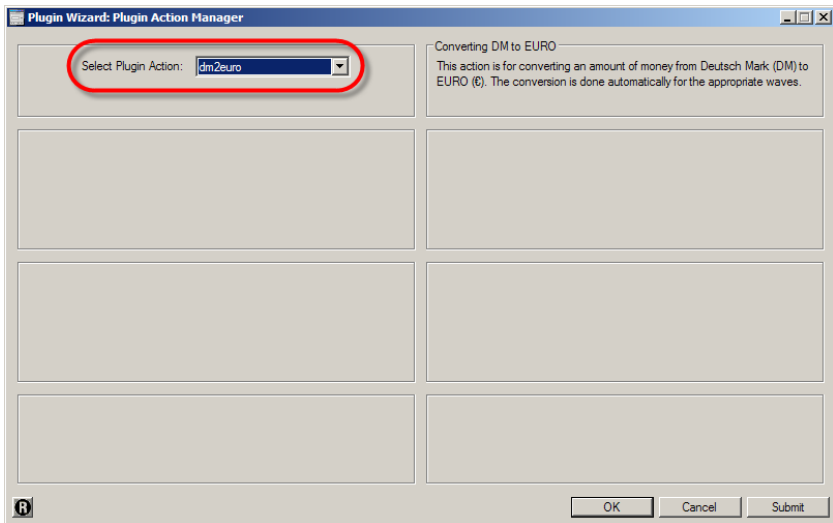
**Plugin Action List** [\[Clear List\]](#)

[|+|-|up|down|](#)  
\_dm

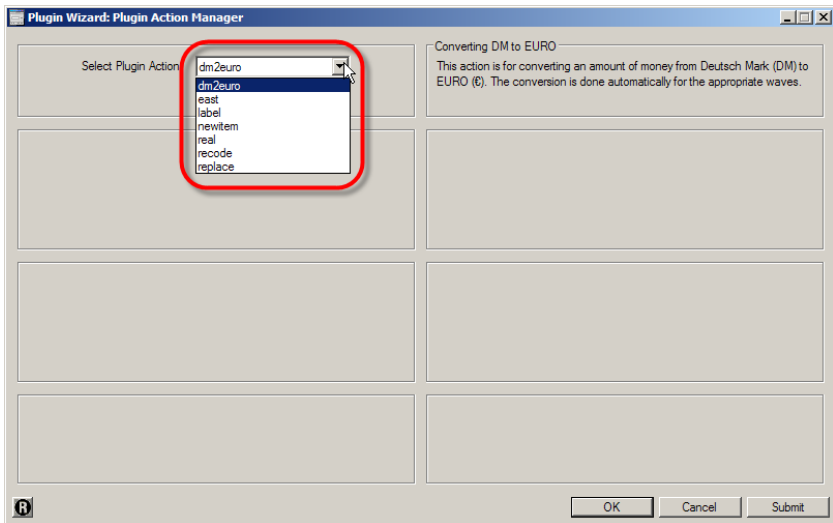
Add an action to the list



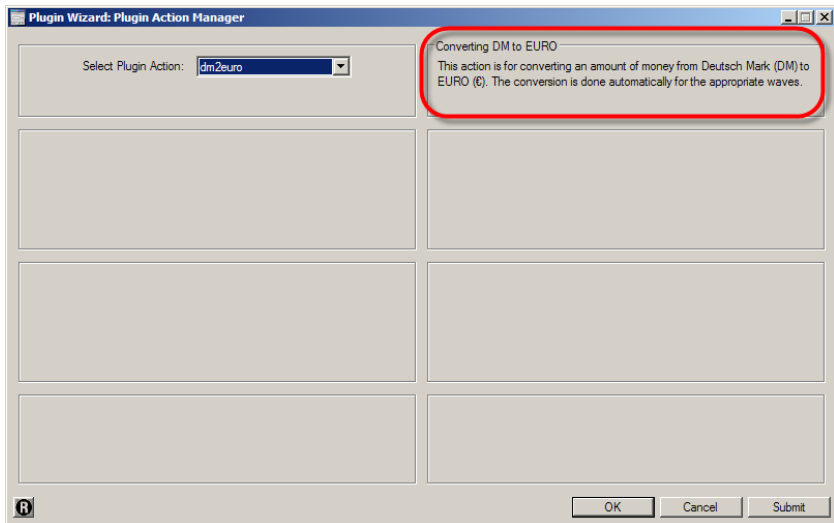
Plugin Action Manager



Select an action

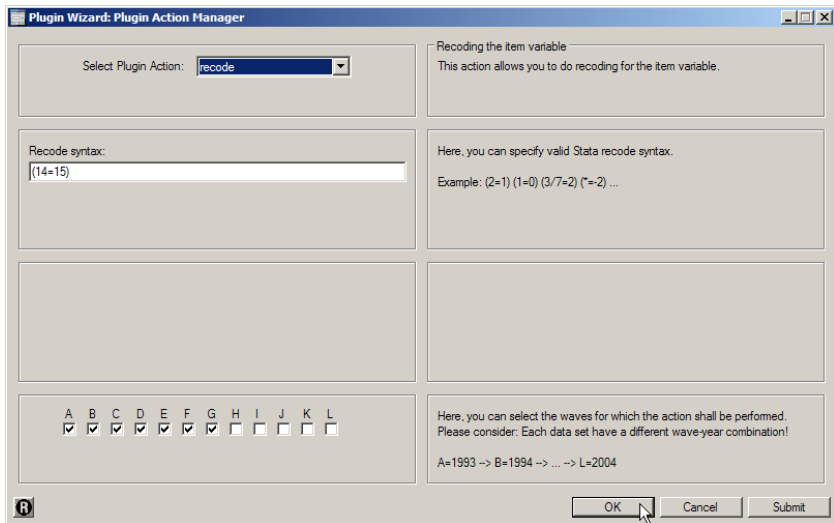


Select an action



Description of action





Action "recode"

Plugin Wizard: Plugin Action Manager

Select Plugin Action:

Replacing the item variable  
This action allows you to do replacing (in the sense of Stata) for the item variable.

Replace syntax:

Here, you can specify a Stata expression.  
Example: itemname/10

A  B  C  D  E  F  G  H  I  J  K  L

Here, you can select the waves for which the action shall be performed.  
Please consider: Each data set have a different wave-year combination!  
A=1993 -> B=1994 -> ... -> L=2004

Action "replace"

Viewer (#1) [view "c:/panelwhiz/temp/plugin\_wizard.smc"]

Back Refresh Search Help Contents What's New News

Command: view "c:/panelwhiz/temp/plugin\_wizard.smc"

Start Page > wave (1) > AA > Plugin Wizard <Project Page>

Panelwhiz IABBP - Plugin Wizard  
Panelwhiz IABBP (1.0) Aug 2006 <john@panelwhiz.eu>

---

**bp\_aabula.ado - (p)** Based On: NEW Plugin  
[\[Create\]](#) [\[View\]](#) [\[Erase\]](#) [\[New\]](#) [\[Open \(Original\)\]](#) [\[Open \(Personal\)\]](#)

---

**Plugin Information** [\[Default\]](#) [\[Edit\]](#)

Plugin : [aabula] Bundesland  
Author : Markus Hahn  
Email : hahn@rwi-essen.de  
web : http://www.panelwhiz.eu  
Copyright : copyright 2006. All rights reserved. unauthorized copying prohibited.  
Version : 1.0  
Years : a b c d e f g h i j k l  
Possible : a b c d e f g h i j k l  
Changes : a b c d e f g - - - - -

---

**Plugin Initialization** [\[Default\]](#) [\[Edit\]](#)

Update year : 2004  
Depends on : <nothing>

---

**Plugin Action List** [\[Clear List\]](#)

[\[+\]](#)[\[-\]](#)[|up|down|](#) **recode** [Recording the item variable]  
-> .action.recode (14=15), a b c d e f g

[\[+\]](#)[\[-\]](#)[|up|down|](#) **replace** [Replacing the item variable]  
-> .action.replace (-8=), a b c d e f g

---

⤴

Two actions were added

Viewer (#1) [view "c:/panelwhiz/temp/plugin\_wizard.smc"]

Back Refresh Search Help Contents What's New News

Command: view "c:/panelwhiz/temp/plugin\_wizard.smc"

Start Page > wave (1) > AA > Plugin Wizard <Project Page>

Panelwhiz IABBP - Plugin Wizard  
Panelwhiz IABBP (1.0) Aug 2006 <john@panelwhiz.eu>

---

bp\_aabula.ado - (p) Based On: NEW Plugin

**[Create]** [View] [Erase] [New] [Open (Original)] [Open (Personal)]

---

Plugin Information [Default] [Edit]

Plugin : [aabula] Bundesland  
Author : Markus Hahn  
Email : hahn@rwi-essen.de  
web : http://www.panelwhiz.eu  
Copyright : copyright 2006. All rights reserved. unauthorized copying prohibited.  
Version : 1.0  
Years : a b c d e f g h i j k l  
Possible : a b c d e f g h i j k l  
Changes : a b c d e f g - - - - -

---

Plugin Initialization [Default] [Edit]

Update year : 2004  
Depends on : <nothing>

---

Plugin Action List [Clear List]

[+]-[up|down] recode [Recording the item variable]  
-> .action.recode (14=15), a b c d e f g

[+]-[up|down] replace [Replacing the item variable]  
-> .action.replace (-8=.), a b c d e f g

Create plugin

Viewer (#1) [view "c:/panelwhiz/temp/plugin\_wizard.smc"]

Back Refresh Search Help Contents What's New News

Command: view "c:/panelwhiz/temp/plugin\_wizard.smc"

Start Page > wave (1) > AA > Plugin Wizard <Project Page>

Panelwhiz IABBP - Plugin Wizard  
Panelwhiz IABBP (1.0) Aug 2006 <john@panelwhiz.eu>

---

bp\_aabula.ado - (p) Based on: PERSONAL Plugin

[Create] **[View]** [Erase] [New] [Open (Original)] [Open (Personal)]

---

Plugin Information [Default] [Edit]

Plugin : [aabula] Bundesland  
Author : Markus Hahn  
Email : hahn@rwi-essen.de  
web : http://www.panelwhiz.eu  
Copyright : copyright 2006. All rights reserved. unauthorized copying prohibited.  
Version : 1.0  
Years : a b c d e f g h i j k l  
Possible : a b c d e f g h i j k l  
Changes : a b c d e f g - - - - -

---

Plugin Initialization [Default] [Edit]

Update year : 2004  
Depends on : <nothing>

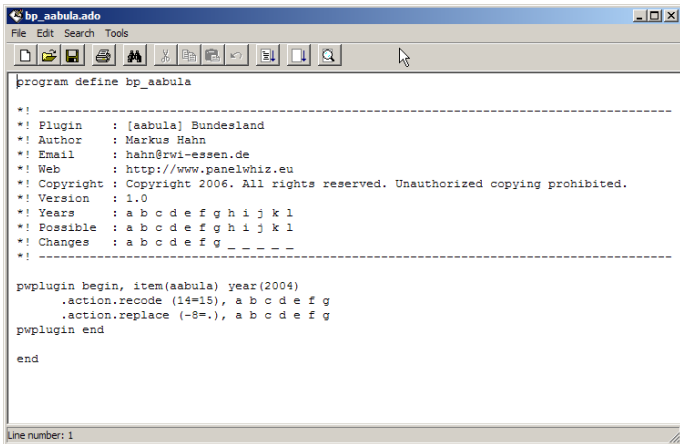
---

Plugin Action List [Clear List]

[+|-|up|down] recode [Recording the item variable]  
-> .action.recode (14=15), a b c d e f g

[+|-|up|down] replace [Replacing the item variable]  
-> .action.replace (-8=.), a b c d e f g

View plugin



```
program define bp_aabula

*! -----
*! Plugin      : [aabula] Bundesland
*! Author     : Markus Hahn
*! Email      : hahn@rwi-essen.de
*! Web        : http://www.panelwhiz.eu
*! Copyright   : Copyright 2006. All rights reserved. Unauthorized copying prohibited.
*! Version    : 1.0
*! Years      : a b c d e f g h i j k l
*! Possible   : a b c d e f g h i j k l
*! Changes    : a b c d e f g _ _ _ _ _
*! -----

pwplugin begin, item(aabula) year(2004)
    .action.recode (14=15), a b c d e f g
    .action.replace (-8=.), a b c d e f g
pwplugin end

end
```

Line number: 1

Created plugin code

## When a plugin is finally created...

- Can be used directly with PanelWhiz retrievals
- Can be shared with the PanelWhiz community
  - Embedded knowledge, which is respected by contract  
⇒ PanelWhiz Plugins must be cited
  - Modularized piece of code ⇒ Shareable

## Plans for future releases

- Add more predefined actions
- More flexibility through "free coding"
- Extend the auto-documentation feature