StreamServe™ Persuasion™

StreamServe Persuasion SP5 PageIN

User Guide

Rev A

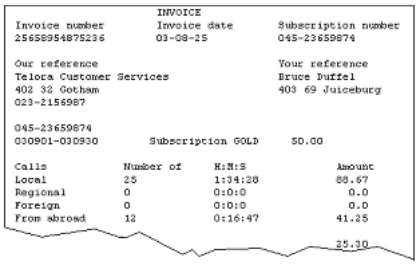
StreamServe Persuasion SP5 PageIN User Guide
Rev A

StreamServe Web Site
http://www.streamserve.com

Contents

About PageIN

The StreamServer can receive and process page formatted input in plain text as shown in the example below.



```
                    INVOICE
Invoice number      Invoice date        Subscription number
25658954875236      03-08-25            045-23659874

Our reference                           Your reference
Telora Customer Services                Bruce Duffel
402 32 Gotham                           403 69 Juiceburg
023-2156987

045-23659874
030901-030930       Subscription GOLD    50.00

Calls           Number of     H:M:S             Amount
Local           25            1:34:28           88.67
Regional        0             0:0:0              0.0
Foreign         0             0:0:0              0.0
From abroad     12            0:16:47           41.25

                                                25.30
```

The instructions to the StreamServer about how to handle page formatted input are created using the PageIN tool.

Samples

You can take a snap-shot of input data, and use this snap-shot when you create the PageIN configuration. See *Input data samples* on page 5.

Configuration

When you create a PageIN configuration, you specify which information to pick from the input, and how to label and structure this information as blocks and fields. See *Creating a PageIN configuration* on page 9.

4 About PageIN

Input data samples

The easiest way to map data is to take a snap-shot of the actual input, create a sample of the input data, and load the sample into the PageIN tool. The mapping is done by drawing rectangles around the displayed text segments.

UTF-8 encoding in the sample file

Use UTF-8 encoding in the input sample file. If you do not, you may run into problems with special characters etc.

1. Open the file in Notepad.
2. Enter `//!codepage UTF-8!` as the first line.
3. Use **Save As** and encoding **UTF-8** to save the file.

Creating and loading input samples

You create a sample by recording the input, and writing the output to a file. To make the file available to the PageIN tool, you must import it to a resource set connected to the corresponding Message. To encode the sample file according to a specific code page, you can use the *-grbcodepage* startup argument, see the *Startup argument* reference.

To record the input

With this method, you create a simple Project with a Platform and a Runtime configuration – and no Message.

1. Create a new Platform
2. Add the appropriate input connector and a Null output connector to the Platform.
3. If you intend to use filter chains in the real Project, you must add the same filter chain to this Project's input connector.
4. Specify the startup arguments *-reconly* and *-grb* . See the *Startup argument* reference.
5. Create a Runtime configuration to which you add the Platform. You do not have to configure the Runtime configuration - just add the Platform to it.
6. Export the Project.
7. Run the Project and record the input.

The result file is shown in the path defined by the `-grb` argument. The file is called `allpages.<input connector name>`. Import this file to the appropriate resource set.

To load the sample

1 Select **File** > **Open Sample**. The Select Resource dialog box opens.

2 Browse to and select the resource.

Adjusting the loaded sample

After you have loaded the sample, you must compare the result displayed in the PageIN tool with the original input in the sample resource.

Delimiters

The PageIN tool uses the standard hex-codes for line feed (<0d,0a> and <0a>), page break (<0d,0c> and <0c>), and carriage return (<0d>). If the input contains other hex-codes, you must add these codes to the PageIN tool's list of delimiters.

1. Select **File** > **Page Setup**. The Page Setup dialog box opens.
2. On the Delimiters tab, select the delimiter type.
3. Add the new hex-code at the end of the list and click **OK**.

Page size

The page size defined in PageIN tool must conform to the page size of the actual input. If the sample does not fit the page size defined in the PageIN tool, you must change the page size settings.

1. Select **File** > **Page Setup**. The Page Setup dialog box opens.
2. On the Size tab, specify the page size.

Creating a PageIN configuration

In the PageIN tool, you create instructions describing which parts of an input page the StreamServer should extract and process, how it should organize the extracted data as labelled fields, and how it should structure the fields. This structure of fields will be used as input in the next stage in the processing chain.

The first thing you do when you create a PageIN configuration is to load the sample file, i.e. snap-shot of the input, into the sheet area in the PageIN tool. Then you create fields and trigger patterns by drawing areas on the PageIN sheet. Each drawn area on the PageIN sheet will in turn generate a node in the PageIN tool Message view. These nodes constitute the actual PageIN configuration.

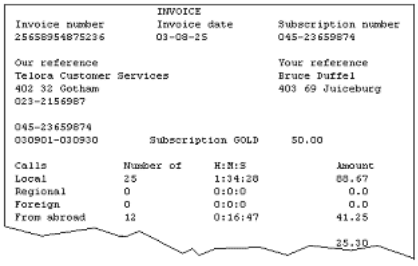
Static and floating positions

When you analyze the input data, you will identify portions of data that are in static positions, and other portions that are floating. For example, an article list contains recurring data – the number of lines depends on the number of articles. All text segments below line 1 in this list are floating.

When you map static portions, you define a field for each text segment. When you map floating data, you must first define a frame for the data, and one or more blocks where you define the fields.

Example: creating a PageIN configuration

This example describes how to configure a PageIN Event for the following type of input data:



```
                    INVOICE
Invoice number      Invoice date        Subscription number
25658954875236      03-08-25            045-23659874

Our reference                           Your reference
Telora Customer Services                Bruce Duffel
402 32 Gotham                           403 69 Juiceburg
023-2156987

045-23659874
030901-030930       Subscription GOLD    50.00

Calls           Number of    H:M:S            Amount
Local           25           1:34:28          88.67
Regional        0            0:0:0             0.0
Foreign         0            0:0:0             0.0
From abroad     12           0:16:47          41.25

                                              25.30
```

Load the sample

The sample is available as the resource `PhoneInvoice.txt` in the resource set.

1. Select **File** > **Open Sample**. The Select Resource dialog box opens.
2. Browse to, and select, the **PhoneInvoice.txt** resource. The corresponding sample is loaded.

Define static fields

Select **Insert** > **New Field** and draw rectangles around each text segment that you want to define as a field. For each field, select the field and configure the properties in the Properties view.

Define areas for floating text

Select **Insert** > **New Frame** and draw a rectangle around the area that contains text in floating positions. A new block is automatically added to the top of the frame.

Define blocks

1. Move and stretch the first block so that it covers all text you want to include in the block.
2. Rename the block.
3. Select **Insert** > **New Block**. A new block is added to the frame.
4. Move and stretch the new block so that it covers all text you want to include in the block.
5. Rename the block.
6. Repeat steps *2* - *5* for all blocks.
7. For each block, select the block and configure the properties in the Properties view.

Define fields within blocks

For each block:

1. In the Message tree, select the block. The block is activated on the PageIN sheet.
2. Select **Insert** > **New Field** and click the first text segment in the block.
3. Configure the field properties in the Properties view.
4. Click the next field, configure the field properties, and so on.

Specify a Message pattern

Select **Insert** > **New Pattern** and draw a rectangle around the text segment "`INVOICE`". The selected text segment "`INVOICE`" is automatically specified as the match criterion.

Specify block patterns

For each block:

1. Select the block.
2. Select **Insert** > **New Pattern** and draw a rectangle around the text segment you want to use as pattern.
3. Rename the pattern, and specify the match criterion.

Managing patterns, frames, blocks, and fields

Patterns

Patterns are match criteria for input data. You must specify at least one pattern at Message level. This pattern will be used as the Event trigger. If the StreamServer finds matching input in the specified area, the Event will be triggered.

Block patterns

If the PageIN configuration contains blocks, you must specify at least one pattern per block in order to uniquely identify the block.

To create a pattern

1. Select **Insert** > **New Pattern**.
2. On the PageIN sheet, draw a rectangle around the text segment you want to use as pattern.
3. In the Properties view, specify the properties. See *Pattern properties* on page 24.

Multiple patterns

You can use multiple patterns. By default, the match criteria defined for all patterns must be fulfilled in order to trigger the Event, or to identify a block. This corresponds to specifying the following rule for the patterns:

`pattern1 AND pattern2 AND ... pattern`*`N`*

where `pattern1` etc. is the pattern name, i.e. the **Label** property for the pattern.

The **Rule** property for the Message and blocks enables you to specify custom rules for when to trigger an Event or identify a block. You can use the logical operators `AND`, `OR` and `NOT` in a rule.

Valid characters for the pattern names in a rule
A-Z
a-z
0-9
_ (underscore)
. (dot)

Example 1 *A rule*

```
Pattern1 AND (Pattern2 OR Pattern3)
```

Example 2 *Another rule*

```
Pattern1 NOT $countrycode=LAT
```

Frames

You cannot add blocks directly to the PageIN sheet. First you must draw a frame, and then add the blocks to the frame. The purpose of the frame is to limit the area in which the StreamServer will search for blocks.

A page can contain any number of frames, and a frame can contain any number of blocks. A frame cannot contain other frames.

To create a frame

1. Select **Insert** > **New Frame**.
2. On the PageIN sheet, draw a rectangle around the area you want to include in the frame. The first block is automatically added to the frame.

Blocks

You must define blocks for floating data. For example, an article list contains floating data – the number of lines depends on the number of articles. All text segments below line 1 in this list are floating.

A block can contain any number of fields. A block cannot contain sub-blocks.

Example 3 *Block defined for an article list*

The shaded area within the frame indicates the outline of the block.

010	Ball Leather, black/white	2.00	15.00	30.00
020	Bat Wood, pine	2.00	150.00	300.00
030	Cap Leather, black/white	1.00	2.00	20.00

Block patterns

You must specify at least one pattern for a block. See *Patterns* on page 11.

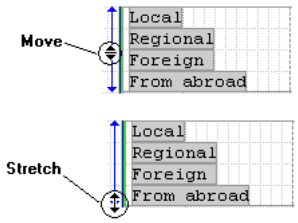
Block order

The StreamServer checks blocks in the order (top-down) they are displayed in the Message tree. It takes one line of input data, and checks whether or not there is a matching pattern for the first block. If there is no match, it checks the next block and so on until it finds a match. You must arrange the blocks (drag-and-drop in the Message tree) as follows:

- Begin with blocks where the patterns are easy to identify, for example a unique text string such as "`Due`".
- End with blocks that contains free text, i.e. text in any position within the block, where the match criterion is "`?`".

To insert a block

1. Select the frame where you want to have the block, and select **Insert** > **New Block**. A new block is added to the top of the frame.
2. Move the block (drag the double-arrow) to the appropriate position within the frame.
3. Stretch the block (drag the bottom arrow-head) so that it spans over all lines you want to include in the block.



To configure a block

1. Select the block you want to configure. The block properties are displayed in the Properties view.
2. Edit the properties. See *Block properties* on page 26.

Fields

When you configure a PageIN Event, you map text segments in the input to areas in the PageIN tool. A field in the PageIN tool corresponds to an area with one or more text segments.

To insert a static field

Select **Insert** > **New Field** and draw a rectangle around the text segment that you want to define as a field.

To insert a fields in a block

1. In the Message tree, select the block.
2. Select **Insert** > **New Field** and draw a rectangle around the text segment that you want to define as a field.

To configure a field

1. Select the field you want to configure. The field properties are displayed in the Properties view.

2 Edit the properties. See *Field properties* on page 26.

Field variables

You can define a field variable, and later on refer to the variable instead of a static value. The use of variables does affect performance, so only use variables when necessary.

To define a variable

1 Select the field you want to configure. The field properties are displayed in the Properties view.

2 Enter the name of the **Variable** and press `Enter`.

Before Message script

You can also define variables in a Before Message script using the following syntax:

```
$<variable> = <column> <row> <length>;
```

For example:

```
$text = 8 14 9;
```

where `$text` is assigned the string on row 14 between column 8 and 17.

Numeric and date formats

There are three input format categories for the fields:

- **General**. Data will be treated as a regular string of characters. This is the default format.
- **Numeric**. Enables the StreamServer to handle input data as numeric data.
- **Date**. Enables the StreamServer to handle input data as date formatted data.

Format tables

Numeric and date formats are made available through format tables. Before you specify numeric or date formats for the fields, you must add a format table to a resource set connected to the Message. You can import `Formats.txt` from

`<StreamServe installation>\Applications\StreamServer\<version>\Tools\Samples`

to the resource set.

The first time you specify a numeric or date format for a field, a resource selection dialog box opens. In this dialog box you must browse to and select the format table you want to use. This table will be selected by default the next time you specify a format for any of the fields in the Event configuration.

To select a numeric | date format for a field

1 Select the field. The field properties are displayed in the Properties view.

2 Select the **Input format** property.

3 At the input field, click **Select**. The Formats dialog box opens.

4 Select the **Numeric** | **Date** category.

5 Double-click the **Format** that corresponds to the input format.

Example 4 *Numeric formats*

Input `1000000,25` corresponds to `k= d=,`

Input `1000,000.25` corresponds to `k=,d=.`

Example 5 *Date formats*

Input `31/10/03` corresponds to `dd/mm/yy`

Input `2003-10-03` corresponds to `yyyy-mm-dd`

To add a new format

Enter the new format in the **Format** field, and click **Add**.

You can also add new formats directly to the format table resource.

Sorting

You can use sort criteria to specify the order in which data will be delivered to the subsequent Processes. If no sort criteria is used, data will be delivered in the same order as it arrives. Performance will be less affected if data is sorted at Event level compared to sorting at Process level.

You specify the sort criteria by assigning priorities to blocks. Data associated with blocks with the highest priority will be delivered first, and so on.

To specify sort criteria

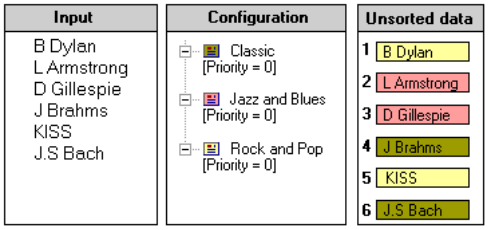
1. Select the **Message** node in the Message tree. The Message properties are displayed in the Properties view.
2. Set **Use block sort priority** to **Yes**.
3. Select the first block. The block properties are displayed in the Properties view.
4. Set the **Block sort priority** level. The lower the number, the higher the priority.
5. Repeat steps 3 and 4 for all blocks.

Sorting examples

The following examples illustrate how the order of the output from an Event changes when sort criteria is specified for the blocks in the Event.

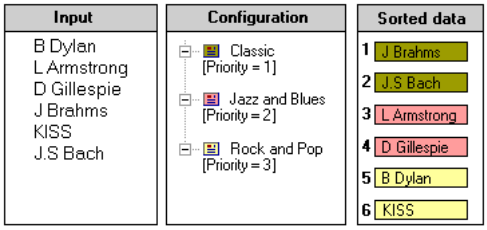
Example 6 *Event output without sorting*

All blocks have sort priority set to 0. Output is delivered in the same order as it arrived.



Example 7 *Event output with sorting*

The blocks have sort priority set to 1, 2, and 3. Output is delivered in this order.



PageIN tool GUI reference

Main window

The Main window contains three views:

- Message view
  This is where you navigate in, and configure, the PageIN structure. Each block, field, and pattern in the Message view corresponds to a block, field, or pattern on the PageIN sheet.
- PageIN sheet
  This is where you draw the frames, blocks, fields, and patterns.
- Properties view
  This is where you configure the properties for the field or block currently selected in the Message view.

File menu

New	Create a new PageIN configuration.
Open	Open an existing (stand-alone) PageIN configuration file.
Save	Save the PageIN configuration as data embedded in the corresponding Message file in the Design Center Project.
Save As	Save the PageIN configuration as a separate file.
Open Sample	Open a sample on the PageIN sheet.
Close Sample	Close the active sample.
Reload Sample	Reload the active sample.
Set Sample Font	Set the font for the sample on the PageIN sheet.
Page Setup	Open the *Page Setup dialog box* where you can specify the page setup options.
Exit	Exit the PageIN tool.

Edit menu

Go To If there is more than one page, use this option to navigate between the pages.

View menu

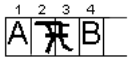
Patterns High-light the patterns on the PageIN sheet.

Tools menu

Change String Order If arabic text – or other type of right-to-left ordered text – is not displayed correctly, you can use this option.

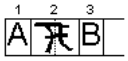
Map Data By This option is used when handling double-byte characters, such as chinese text.

Grid Position:



All cells have the same width, and a double-byte character can occupy more than one cell.

Character Position:



The cell has a variable width, and always contains one character.

Page Setup dialog box

Size options

Used for: Specifying page setup options.

Settings	
Format	Select a predefined format, or select **Custom** and specify width and height.
Size	Specify a custom width and height. **Minimum width** – If all lines in the input data are shorter than this width, the input data will be ignored.

Delimiter options

Used for: Specifying hex-codes for line break, page break, and carriage return.

Settings	
Type	**Line –** Line break. **Page –** Page break. **CR –** Carriage return
Delimiters	Lists the hex-codes for the selected type. You can add new hex-codes to the list to define new delimiters.

Filter options

Used for: Specifying a character substitution filter table. The filter can, for example, be used to translate special characters.

Settings	
Disabled	Select whether or not to enable the filter.
Filter file	Select an external filter that contains the substitution table.
Internal	Specify the filter in this dialog box.
Output file	Save the filtered sample as a new file. The file must exist in the specified path.

Syntax *input string*

output string

Description Substitute *input string* with *output string*. Use plain text, or hexadecimal notation within angle brackets (*<hex>*).

Case sensitive – Text strings are case sensitive.

Multiple hex values – You can separate multiple hex values with a comma, for example:

```
<0D,0A>
```

Comments – Use * to comment rows.

Example

```
* Substitute / with \
<2F>
<5C>
```

Message view

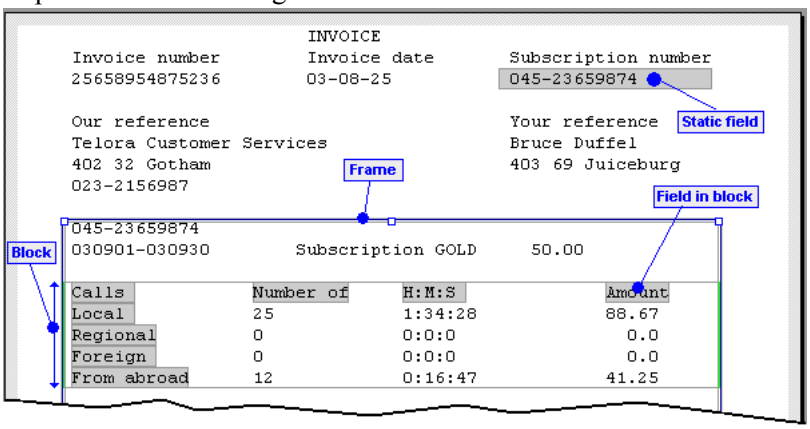
Used for: Navigating in, and configuring, the PageIN structure. Each block, field, and pattern in the Message view corresponds to a block, field, or pattern on the PageIN sheet.

The Message tree structure shown here will be displayed in the corresponding Process tool.

PageIN sheet

Used for: Creating and configuring the PageIN structure. You load the snap-shot of the actual input into this sheet. Then you draw the frames, blocks, fields, and patterns around the applicable text segments displayed on the sheet.

Each block, field, and pattern on the PageIN sheet corresponds to a block, field, or pattern in the Message view.



Properties view

Used for: Viewing and editing properties for patterns, blocks, and fields. Select the object (field etc.) in the Message view, or on the PageIN sheet, and edit the properties in this view.

Message properties

Used for: Editing Message properties.

Properties	
Label	Message name. Will be displayed in the Process tool.
Rule	Rule defining how to handle multiple patterns specified for the Message. See *Patterns* on page 11.
Use block sort priority	Select to enable sorting of data.

Pattern properties

Used for: Editing pattern properties. Select the pattern in the Message view, and edit the properties in this view.

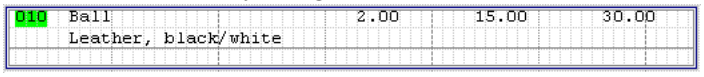
Properties	
Label	Pattern name.
Description	Textual description of the pattern.
Comment	Additional comment.
Left, Right, Top, and Bottom	Coordinates for the pattern match rectangle displayed on the PageIN sheet. The origin is the top-left corner. View only.
Match	Characters specifying the match.
Use wildcards	Select whether or not to allow wildcards in the definition of the Match property.
Ignore column	Select **Yes** to ignore the left and right borders of the pattern match rectangle. In this case, there can be a match even if the incoming data falls outside these borders. This option will affect performance.

Properties	
Ignore line	Select **Yes** to ignore the Top and Bottom of the pattern match rectangle. In this case, there can be a match even if the incoming data falls outside these borders. This option will affect performance.
Enabled	Select whether or not to use the pattern.

Wildcards	
`9`	Any digit (0-9). **Example** `A999` matches `A000` to `A999`
`$`	Any digit (0-9), comma (,), period (.), plus (+), and minus (-). **Example** `8$3` matches `803`, `8.3`, `8+3` etc.
`X`	Any alphabetical character (A-Z, a-z). **Example** `AX` matches `Aa`, `AA`, `Ab` etc.
`¤`	Any character, except space. **Example** `A¤` matches `A1`, `Aa`, `A2` etc.
`?`	Any character, including space. **Example** `?` matches anything.

Example 8 *Wildcards in block pattern*

Use `999` to match any 3-digit number.



Block properties

Used for: Editing block properties. Select the block in the Message view, and edit the properties in this view.

Properties	
Label	Block name. In the Process tool, you can select whether to display this label, or the `Description`.
Description	Textual description of the block. In the Process tool, you can select whether to display this description, or the block name.
Comment	Additional comment.
Left, Right, Top, and Bottom	Coordinates for the block rectangle displayed on the PageIN sheet. Origo is the top-left corner.
Lines	Number of lines in the block.
Rule	Rule defining how to handle multiple patterns specified for the block. See *Patterns* on page 11.
Block sort priority	Set the sort criterion for the block.
Use block sort priority	N/A

Field properties

Used for: Editing field properties. Select the field in the Message view – or PageIN sheet – and edit the properties in this view.

Properties	
Label	Field name. In the Process tool, you can select whether to display this label, or the `Description`.
Description	Textual description of the field. In the Process tool, you can select whether to display this description, or the field name.
Comment	Additional comment.
Left, Right, Top, and Bottom	Coordinates for the block rectangle displayed on the PageIN sheet. Origo is the top-left corner.
Sample data	An example of field content.

Properties	
Variable	Name of a variable that refers to the field. See *Field variables* on page 14.
Keep spaces	Select whether or not to keep leading spaces and trailing spaces defined in the field, when the field is used in the output data.
Class	Field class that can assist formatting in a PageOUT Process. For example, if you specify a font for a class in the Process, the font will be used for all fields belonging to this class. **Label** – For fields containing static data. **Dynamic** – For fields containing dynamic data. **Header** – For fields containing static header data.
Alignment	Specify alignment of data in the Process tool.
Input format	See *Numeric and date formats* on page 14.
Job ID	Select whether or not to assign an index to the content of the field to make it searchable in a Job ID repository.