Automate your Search
The Potential of STN Scripts Demonstrated by some Selected Examples

Ernst Aichinger

STNext®

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CAS®
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

Agenda

- STN scripts explained
- Use case I: Monitoring of a substance in DCR and retrieval of DWPI records
- Extensions to use case I
  - Specify transcript name and save transcript in folder
  - Increase limits of ECHO
  - Highly automated version of script
  - Interactive version of script
- Use case II: Monitoring of a large number of substances with a masterscript
- Use case III: ANALYZE indexed DCR substances of an answer set

Powerful scripting feature unique on STN

- **What is a STN script?**

  STN Scripting = simple programming language
  Scripts include the full Messenger (STN retrieval language) functionality
  Scripts includes script features: variables, operators, statements, and conditions

- **How to work with a STN script?**

  1) Prepare script offline or in STNext
  2) Start script manually
  3) The script can run highly automated or interactive

- **What are three key benefits?**

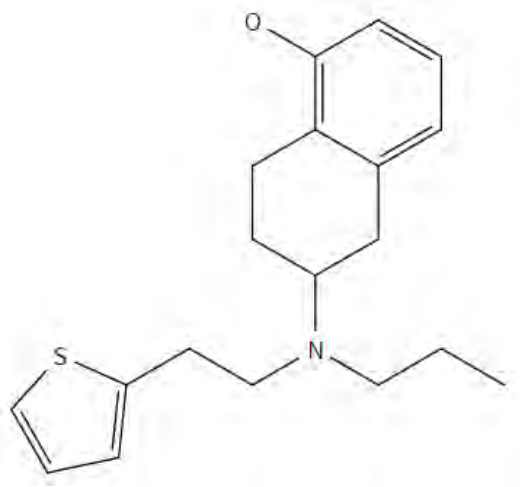
  – Set up sophisticated monitoring solutions.
  – Save time by encoding recurring search concepts.
  – Modular design: Scripts can be built of re-usable elements and shared among colleagues.

Use case I: Monitoring of a substance in DCR and retrieval of corresponding DWPI records

Rotigotine – used for treating Parkinsons disease – should be monitored once a month.

A structure search in DCR combined with a keyword search in DWPI should be performed and corresponding patent documents should be displayed manually depending on number of retrieved records.

Monitoring can be performed using update codes (e.g. UP) or with stored answer sets. The example here uses saved answer sets.



Comments start with ***** and are not executed. Include version number or date!

```
\* Monitoring rotigotine
\* version 1.0 20201014
```

Comments start with ***** and are not executed. Include version number or date!

Lines with STN commands have to be prefixed with a prompt **=>**

```
1  \* Monitoring rotigotine
2  \* version 1.0 20201014
3
4  => FIL WPINDEX
```

Comments start with ***** and are not executed. Include version number or date!

Lines with STN commands have to be prefixed with a prompt **=>**

Answer sets are assigned to variables by **\>** and variable names have to start with **_**

```
\* Monitoring rotigotine
\* version 1.0 20201014

=> FIL WPINDEX
=> S ROTIGOTIN? OR NEUPRO OR N0923 OR N(W)0923 OR SPM962 OR SPM(W)962 \>_patentsDWPI
```

Comments start with ***** and are not executed. Include version number or date!

Lines with STN commands have to be prefixed with a prompt **=>**

Answer sets are assigned to variables by **\>** and variable names have to start with **_**

Upload structure from MyFiles or a subfolder and assign structure to variable. **UPLOAD LNUM _variable <folder/name>**

Conditional statement with **IF**: IF there are new additional records retrieved, then add and save them. IF/ELSE statement are marked with **BEGIN** and **END**.

STN system question prompts require a colon **:**

ECHO displays information on screen and transcript.

Track the progress of the script with output messages. Option: LOG H

```
\* Monitoring rotigotine
\* version 1.0 20201014

=> FIL WPINDEX
=> S ROTIGOTIN? OR NEUPRO OR N0923 OR N(W)0923 OR SPM962 OR SPM(W)962
\>_patentsDWPI
UPLOAD LNUM _uploadedstr <monitoring/rotigotine_script>
=> S _uploadedstr FAM FULL \>_structuresDCR
=> S _structuresDCR/DCR \>_patentsDCR
=> S _patentsDWPI OR _patentsDCR \>_allpatents

=> ACT rotigscript/A \>_activated

=> S _allpatents NOT _activated \>_newpatents

IF #_newpatents > 0 BEGIN
    => S _activated OR _newpatents \>_mustbesaved
    => SAV _mustbesaved rotigscript/A
    :y
    ECHO "New patents retrieved."
    ECHO "The new records are in _newpatents and part of saved answer set
rotigotine/A."
END
ELSE BEGIN
    ECHO "No new records retrieved."
END

ECHO "Script finished"
```

Start the script under MyFiles/Scripts

Workflow and some extensions

1 - Keyword search in WPINDEX

2 - Uploading structure and structure search in DCR

3 - Combining answer sets (L3)

4 - Activating saved answer set from previous monitoring run (L4)

5 - Subtracting: L3 NOT L4

6 - If there are new answers, save them

Workflow and some extensions (1/4)

0 - Rename transcript file and save in folder

1 - Keyword search in WPINDEX

2 - Uploading structure and structure search in DCR

3 - Combining answer sets (L3)

4 - Activating saved answer set from previous monitoring run (L4)

5 - Subtracting: L3 NOT L4

6 - If there are new answers, save them

Transcript names can be specified, also by variables. Transcripts can be saved in new folders or existing folders.

Screenshots from script run

```
4 GET _update LABEL = "Date and remarks for transcript file name"
5 CAPTURE ON <rotigotinemonitoring/_update rotigotine run>
```

Workflow and some extensions (2/4)

0 - Rename transcript file and save in folder

1 - Keyword search in WPINDEX

2 - Uploading structure and structure search in DCR

3 - Combining answer sets (L3)

4 - Activating saved answer set from previous monitoring run (L4)

5 - Subtracting: L3 NOT L4

6 - If there are new answers, save them

7 - Summarize information in the transcript file

ECHO limits the output to 140 characters.

Summarize information in the transcript file

To display more than 140 characters, assign text to variables and output them with ECHO. Variables can be combined, e.g. _alltext = _text1 + _text2.

```
22    _text1 = "The name of this transcript is _update rotigotine run and it is
stored under MyFiles/Transcripts/rotigotinemonitoring."
23    _text2 = "New patents retrieved. The new records are in _newpatents and part of
saved answer set rotigotine/A."
24    ECHO "_text1 _text2"
```

Screenshot transcript

```
The name of this transcript is 20201014 rotigotine run and it is stored under
MyFiles/Transcripts/rotigotinemonitoring. New patents retrieved. The new records are in L72 and part of saved
answer set rotigotine/A.
```

Workflow and some extensions (3/4)

…

6 - If there are new answers, save them

7 - Automate the display of results

Automatic decisions specified in the script:

```
18 IF #_newpatents > 0 BEGIN
19     => S _activated OR _newpatents \>_mustbesaved
20     => SAV _mustbesaved rotigscript/A
21     :y
22     IF #_newpatents < 100 BEGIN
23         => D _newpatents 1-TOT IFULL
24     END
25     ELSE BEGIN
26         => D _newpatents 1-TOT TI PA PI
27     END
```

If there are less than 100 new records, records are displayed with the IFULL format. Otherwise, records are displayed in short with TI, PA and PI.

This example is intended to show that automatic decisions can be integrated, to enable a workflow tailored to your needs.

Workflow and some extensions (4/4)

0 - Rename transcript file and save in folder

1 - Keyword search in WPINDEX

2 - Uploading structure and structure search in DCR

3 - Combining answer sets (L3)

4 - Activating saved answer set from previous monitoring run (L4)

5 - Subtracting: L3 NOT L4

6 - If there are new answers, save them

7 - Automate the display of results or **include interactive steps to request input from the user**

8 - Summarize information in the transcript file

Interactive script design

OPTION 1: The GET command asks for user input. The records to be displayed and the display format can be specified:

```
22    GET _display LABEL = "#_newpatents new records retrieved. How to display?"
23    => D _display
```

Interactive script design

OPTION 1: The GET command asks for user input. The records to be displayed and the display format can be specified:

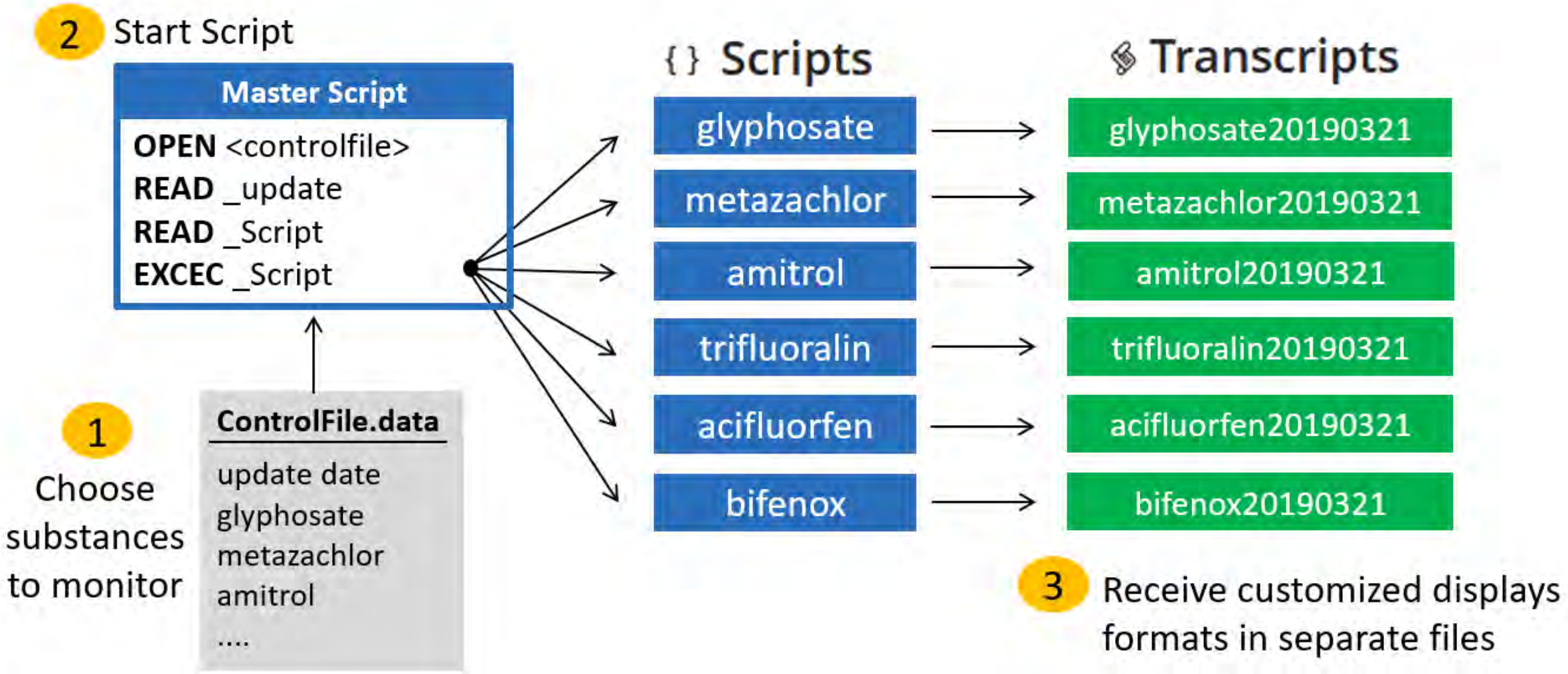
```
22    GET _display LABEL = "#_newpatents new records retrieved. How to display?"
23    => D _display
```

OPTION 2: EDIT ON pauses the script and allows to modify the command specified in the script:

```
22    EDIT ON
23    => D _newpatents 1-TOT IFULL
24    EDIT OFF
```

Screenshot from script run

Use case II: Monitoring a large number of substances



Case III: ANALYZE indexed DCR substances of an answer set

What are the 20 most common substances indexed in an answer set from DWPI?

The script analyzes the DCR-number of the field DCR.WRS (Chemical Resource Number, without Role, with DCR-prefix).

To ensure easy reuse of the workflow and to share it conveniently with colleagues, the procedure was wrapped into a script.

```
=> FIL WPINDEX

GET _lnumber LABEL = "Type in WPINDEX answer set to be analyzed"
EDIT ON
=> ANA _lnumber 1- DCR.WRS \>_analyzedDCR
EDIT OFF

=> EDI _analyzedDCR /DCR.WRS /AN.S
=> SET AUDIT OFF
=> TRA _analyzedDCR 1-20 \>_DCRrecords
=> DEL SEL Y
=> SEL _DCRrecords 1-20 CN.P
ECHO "Below is a list of the TOP 20 subtances indexed (alphabetical order):"
=> D SEL E1-20

ECHO "Analyzed DCR-numbers are stored in _analyzedDCR, the TOP 20 DCR records are
in _DCRrecords"
```

Results of analyzed substances indexed

Screenshots from final transcript

Below is a list of the TOP 20 subtances indexed (alphabetical order):

E#	FILE	FREQUENCY	TERM
E1	WPINDEX	1	ALECTINIB/CN.P
E2	WPINDEX	1	CABOZANTINIB/CN.P
E3	WPINDEX	1	CYCLOPHOSPHAMIDE/CN.P
E4	WPINDEX	1	DEXAMETHASONE/CN.P
E5	WPINDEX	1	DOVITINIB/CN.P
E6	WPINDEX	1	DOXORUBICIN/CN.P
E7	WPINDEX	1	EVEROLIMUS/CN.P
E8	WPINDEX	1	FLUOROURACIL/CN.P
E20	WPINDEX	1	VINBLASTINE/CN.P

=>

Analyzed DCR-numbers are stored in L10, the TOP 20 DCR records are in L12

Contact Us

CAS **help@cas.org**
www.cas.org

FIZ Karlsruhe **helpdesk@fiz-karlsruhe.de**
www.stn-international.de

Questions to these scripts: **ernst.aichinger@fiz-karlsruhe.de**