NRC·CNRC
Institute for Information Technology

Connecting Legacy Code, Business Rules and Documentation

Erik Putrycz

Software Engineering Group, IIT, NRC

Outline

- Legacy Software & Modernization
- Extracting Business Rules
- Connecting documentation with business rules
- Conclusions

Outline

⇨ Legacy Software & Modernization

- Extracting Business Rules
- Connecting documentation with business rules
- Conclusions

Facts on Legacy Systems

- **Recent report "Legacy Applications Trend Report" released by Information and Communications Technology Council**
  - In Canada 60,000 employees are working on legacy systems = 10% of the 600,000 total ICT employment
- **In 2006, 70% of all transaction systems were written in COBOL**
- **490 companies of the Fortune 500 process more than 30 billion transactions or $1 trillion worth of business each and every day using legacy systems**

More facts on Legacy Systems

- **The average Fortune 100 Company maintains 35 million lines of legacy code, and adds about 10% each year for enhancements and maintenance**
- **In total, there are well over 200 billion lines of COBOL code in use today – the largest percentage of code in corporate business systems**
- **HR issue still not handled with retirements - "people in the C-Suite don't know they have a (HR) problem yet. Since they don't perceive the problem, there are few HR initiatives for it"**
  - CEO of MB Foster Associates Inc., a Chesterville, Ontario firm specializing in supporting HP legacy systems and data migration

Modernization

- **Any process for evolving a system**
  - Legacy system can be replaced by a new one, or
  - Interfaced with a new system
- **Motivations**
  - High cost to operate legacy system
  - Impossible to keep the legacy system up-to-date
  - Lack of qualified staff
- **New system or integrate legacy with new system**
  - Need for requirements
- **Many requirements buried in the source code**
- **Recovering business rules major issue**
  - Recent survey from Software AG: 51% of companies who have difficulties modernizing said that a major issue are "hard-coded and closed business rules"

Stakeholders

- Two main classes of the stakeholders are involved with business rules:
  - *Legacy system maintainers*
    - Fix bugs and implement new rules.
    - Need to understand the business rules they are affecting and the execution paths to a specific business rule
  - *Business analysts*
    - Involved in modernization of the legacy system
    - Business rules in legacy system used for validating new requirements or finding requirements
    - Often no background in technologies used in legacy system

Outline

- Legacy Software & Modernization

⇨ Extracting Business Rules

- Connecting documentation with business rules
- Conclusions

Context of this work

- **Large modernization project**
  - Old legacy system being replaced by new COTS-based system
- **Old COBOL system will still be used for several years**
  - Lot of maintenance required
- **New COTS based system**
  - Need for requirements
  - No precise documentation on business rules used in legacy system
- **Small part of the system current studied**
  - ~1 million lines of the COBOL source code and 4000 documents

Objectives

- **Extract business rules:**
  - If <conditions> then <consequence>
  - <conditions> and <consequence> as easy to understand as possible
- **Use “business terms” instead of programming language constructs**
- **Focus on calculations, branching and exceptions**
- **Implementation for COBOL legacy software but process is generic and applicable to other languages**

From Source code to Business Rules

End user data

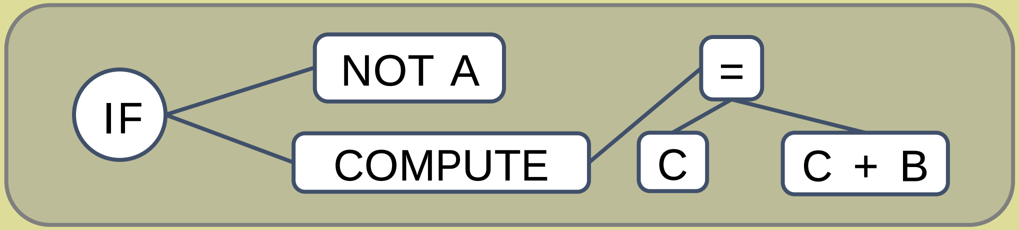
*When **Employee On Leave** is not **true**, **Total Salary** = **Total Salary** + **Union Fees***

Business Rule

NOT A

Calculate C = C + B

Abstract Syntax Tree



Source Code

```
011115 IF NOT A THEN
011116    COMPUTE C = C + B
011117 ENDIF
```

Outline

- Legacy Software & Modernization
- Extracting Business Rules

⇨ Connecting documentation with business rules

- Conclusions

Connecting documentation with business rules

- **Objective: Make the business rules understandable to business analysts**
- **How: translate identifiers used in business rules to non technical terms**
- **Assumption: Existing documentation on data fields**
  - Data = very valuable resource vs. code
- **Other works in the literature: focus on connecting code to technical documentation**
  - Legacy systems rarely have a technical documentation but the data is documented

Example of Data Document

Axx Indicator

Technical Name:AXX YYY IND

Definition: AXX YYZZs Indicator within YYZZs Codes Control File

Model Status:System Information/Skip: Indicates whether a particular YYZZs can appear in an AXX transaction

System(s): System1 System2 System3

Element Type: Business

Data Type: Base

Data Structure:1 character, alphanumeric

System1

Notes: Synonym is: OL-YYZZ-AXX V1-YYZZ-AXX V2-YYZZ-AXX

Valid Values:Y, N

Input Forms:N/A

Element Name:YYZZ-AXX

- subordinate to:GO-YYZZ GO**-**YYZZSES GO**-**DATA GOSS

Picture:PIC X(01)

Subordinate

Elements:N/A

File ID/Records

Description

MM200-XXXX-YYYY-LR logical record used by input/output module

MM401-SB-XXXX-YYYY-MMMMMM logical record used to build online screen

Example of Data Document (2)

Axx Indicator

Technical Name:AXX YYY IND

Definition: AXX YYZZs Indicator within YYZZs Codes Control File

Model Status:System Information/Skip: Indicates whether a particular YYZZs can appear in an AXX transaction

System(s): System1 System2 System3

Element Type: Business

Data Type: Base

Data Structure:1 character, alphanumeric

System1

Notes: Synonym is: OL-YYZZ-AXX V1-YYZZ-AXX V2-YYZZ-AXX

Valid Values:Y, N

Input Forms:N/A

Element Name:***YYZZ-AXX***

- subordinate to:GO-YYZZ GO**-**YYZZSES GO**-**DATA GOSS

Picture:PIC X(01)

Subordinate

Elements:N/A

File ID/Records

Description

MM200-XXXX-YYYY-LR logical record used by input/output module

MM401-SB-XXXX-YYYY-MMMMMM logical record used to build online screen

NRC·CNRC
Institute for Information Technology

Connecting identifiers and data documents

- **Achieved by locating identifiers in data documentation**
- **Translation accuracy:**
  - Some identifiers might appear in many documents
  - All documents have a similar structure that we can use
  - Transitive connections are possibly less accurate
- **Accuracy measurement:**
  - Number of documents where the identifier is found
  - Location of the identifier in the document
    - Documents = Sections + Fields
    - Section name and field title are important
  - Transitive connections

Identifying temporary identifiers and state analysis

- In COBOL, certain identifiers are directly connected to data elements
- Developers often use temporary identifiers in operations with the following pattern:

```
Load value into identifier A from database
Temporary identifier Ta = A
. . .
Calculate Ta = . . .
. . .
Set A = Ta
Save A in database
```

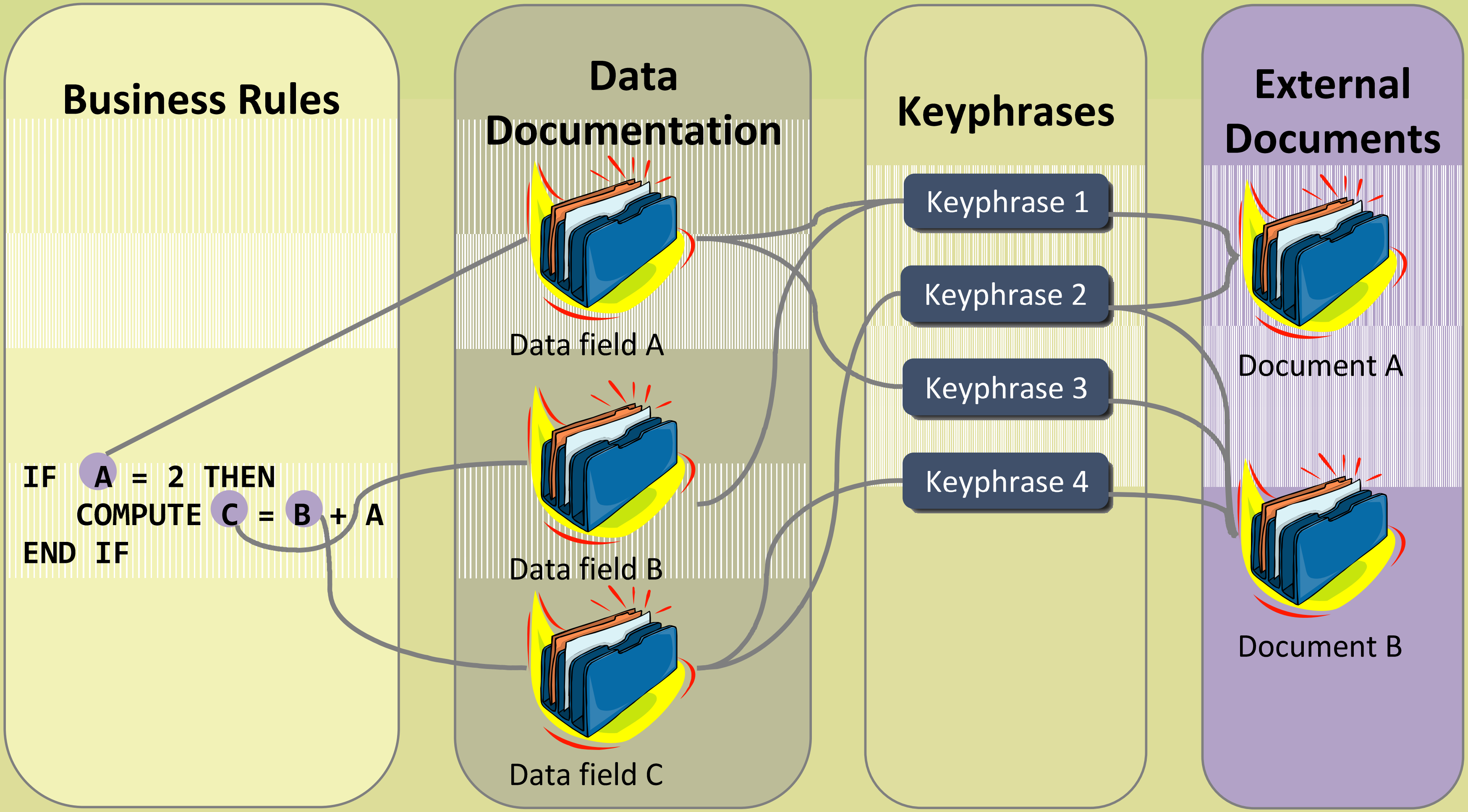
Connecting external documents to business rules

- **Business Analysts: “I’d like to know all the business rules related to this document - data fields are too low level”**
- **Direct path from document to code impossible**
- **Solution: reverse path possible**
  - Code -> Business Rules -> Data field documentation -> External documents
- **Connecting data field documents to external documents**
  - Keyphrase extraction: extract keyphrases from data documentation and use the keyphrases to connect external documents

Keyphrase extraction

- **Keyphrase: list is a short list of phrases (typically 5 to 15 noun phrases) that capture the main topics discussed in a given document**
- **Often referred as keywords**
- **Based on several feature values**
  - **TFxIDF:** measure describing the specificity of a term for a document under consideration, compared to all other documents in the corpus. Candidate phrases that have high TFxIDF value are more likely to be keyphrases.
  - **First occurrence:** computed as the percentage of the document preceding the first occurrence of the term in the document. Terms that tend to appear at the start or at the end of a document are more likely to be keyphrases.
  - **Length of a phrase:** the number of its component words. Two-word phrases are usually preferred by human indexers.

Connecting external documents



```
IF  A = 2 THEN
   COMPUTE C = B + A
END IF
```

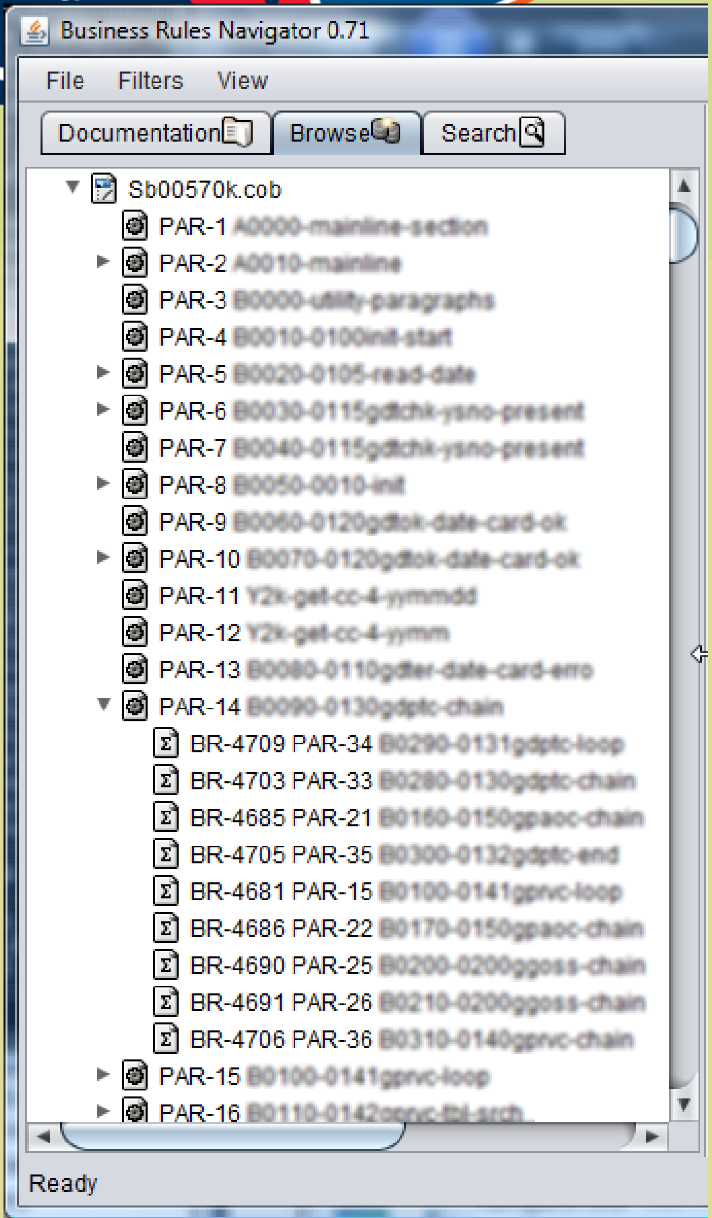
Connecting external documents (2)

- **Currently:**
  - Keywords extracted from external documents
  - Matching of similar keywords
  - Results generated with KEA – Open Source Tool for Keyphrase extraction
- **Results:**
  - Set 1: 352 documents, 207 keyphrases
  - Set 2 (data documentation): 3603 documents, 1427 keyphrases
  - 4106 keyphrases (73%) have only one document matched in each set and thus are not useful for grouping documents
  - 329 documents in the set 1 (93%) are connected with 1941 in set 2 (53%) through 156 keyphrases
- **Better approach planned:**
  - Using search engine ranking method to connect keywords from data documentation with external documents
  - Use Extractor developed by Peter Turney

Outline

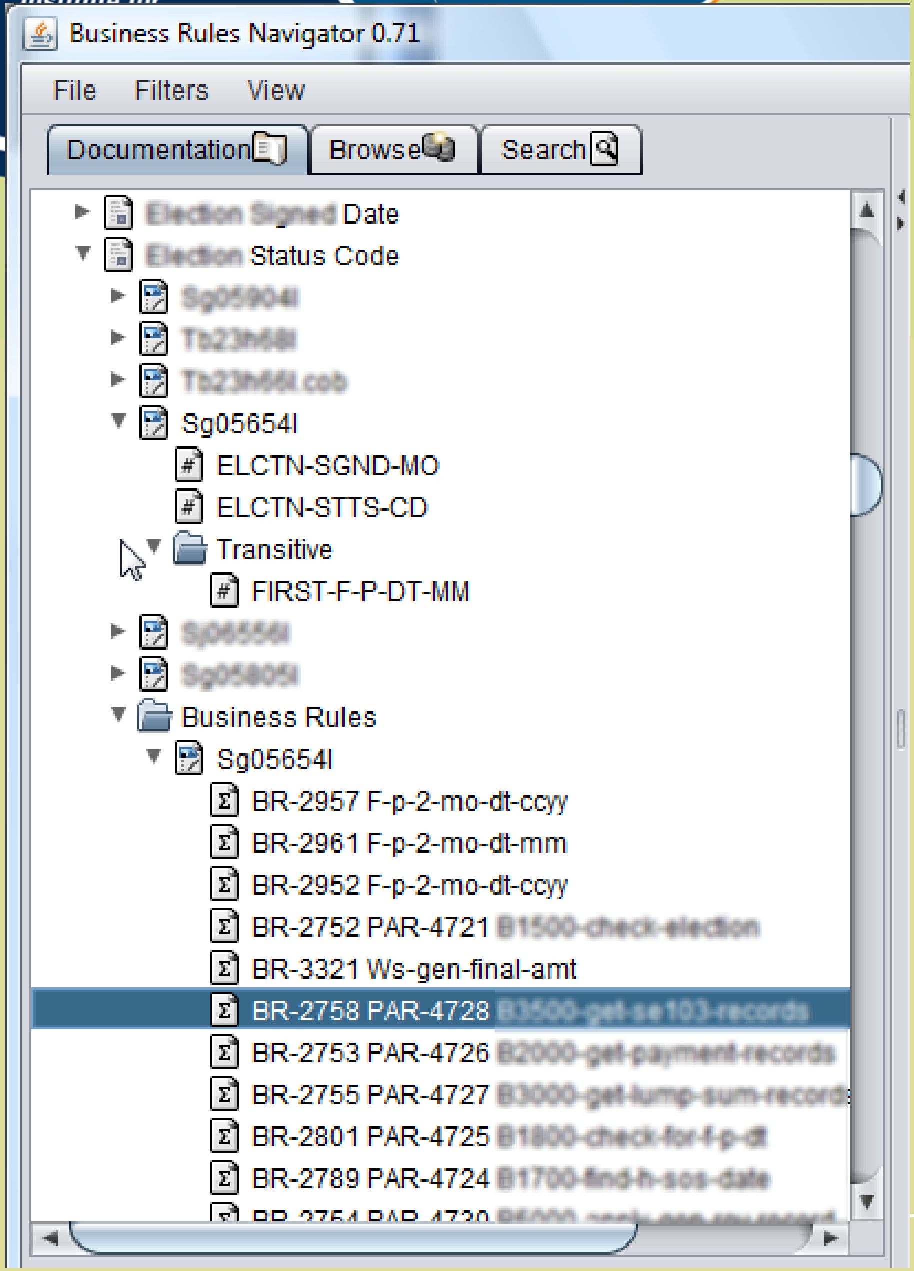
- Legacy Software & Modernization
- Extracting Business Rules
- Connecting documentation with business rules
- Conclusions

Linear navigation



Sb00570k.cob
PAR-1 [redacted]
PAR-2 [redacted]
PAR-3 [redacted]
PAR-4 [redacted]
PAR-5 [redacted]
PAR-6 [redacted]
PAR-7 [redacted]
PAR-8 [redacted]
PAR-9 [redacted]
PAR-10 [redacted]
PAR-11 [redacted]
PAR-12 [redacted]
PAR-13 [redacted]
PAR-14 [redacted]
BR-4709 PAR-34 [redacted]
BR-4703 PAR-33 [redacted]
BR-4685 PAR-21 [redacted]
BR-4705 PAR-35 [redacted]
BR-4681 PAR-15 [redacted]
BR-4686 PAR-22 [redacted]
BR-4690 PAR-25 [redacted]
BR-4691 PAR-26 [redacted]
BR-4706 PAR-36 [redacted]
PAR-15 [redacted]
PAR-16 [redacted]

- **Based on program structure**
  - Program
  - Paragraphs
  - Business rules



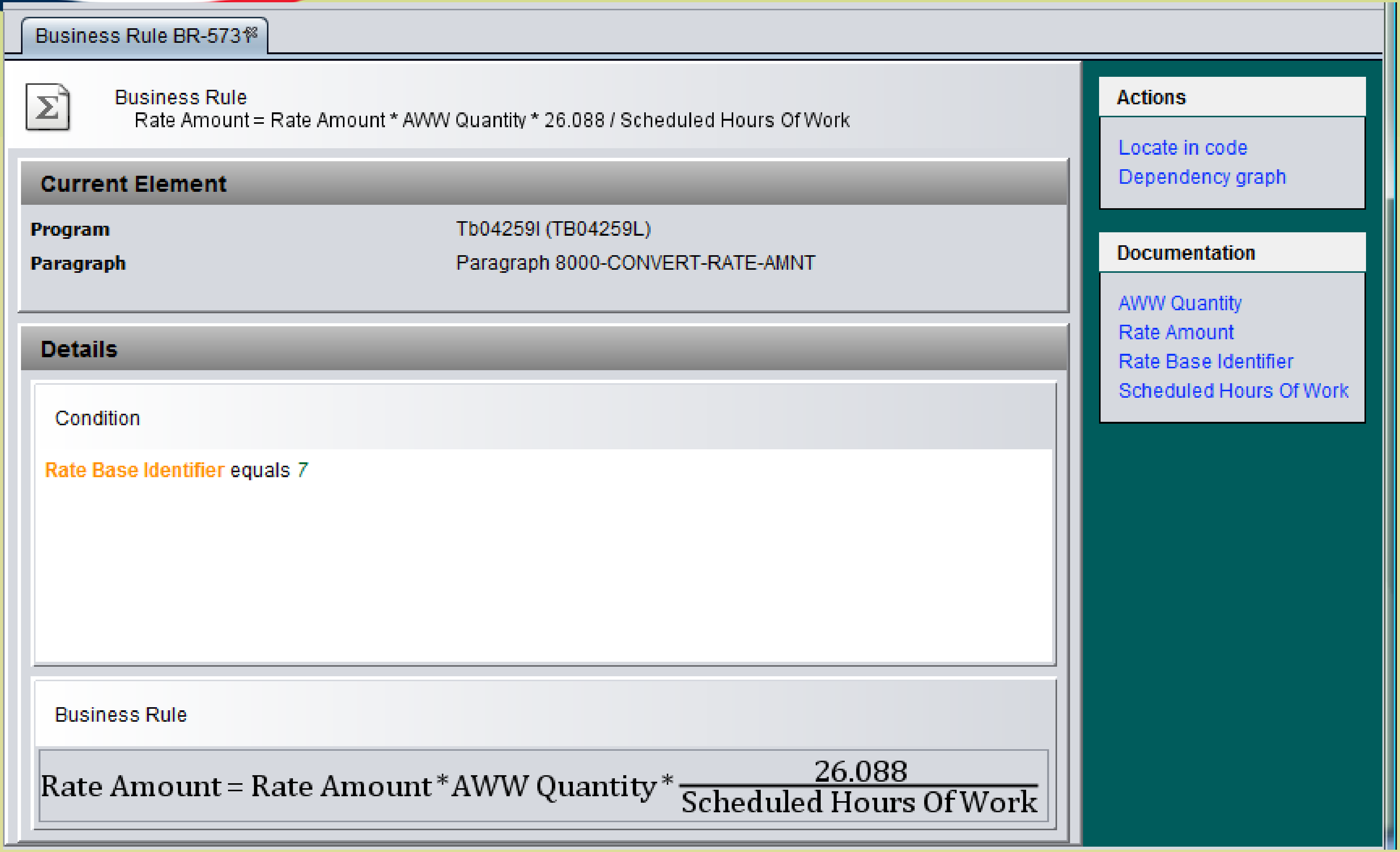
Business Rules Navigator 0.71

- [redacted] Date
- [redacted] Status Code
  - [redacted]
  - [redacted]
  - [redacted]
  - Sg05654I
    - ELCTN-SGND-MO
    - ELCTN-STTS-CD
    - Transitive
      - FIRST-F-P-DT-MM
  - [redacted]
  - [redacted]
  - Business Rules
    - Sg05654I
      - BR-2957 F-p-2-mo-dt-ccyy
      - BR-2961 F-p-2-mo-dt-mm
      - BR-2952 F-p-2-mo-dt-ccyy
      - BR-2752 PAR-4721 [redacted]
      - BR-3321 Ws-gen-final-amt
      - BR-2758 PAR-4728 [redacted]
      - BR-2753 PAR-4726 [redacted]
      - BR-2755 PAR-4727 [redacted]
      - BR-2801 PAR-4725 [redacted]
      - BR-2789 PAR-4724 [redacted]
      - BR-2764 PAR-4730 [redacted]

Data Document based navigation

- **Data Document**
  - Identifiers
  - Transitive connections
  - Business Rules

Business Rule Visualization



Business Rule
Rate Amount = Rate Amount * AWW Quantity * 26.088 / Scheduled Hours Of Work

Current Element

Program	Tb04259l (TB04259L)
Paragraph	Paragraph 8000-CONVERT-RATE-AMNT

Details

Condition

Rate Base Identifier equals 7

Business Rule

$$\text{Rate Amount} = \text{Rate Amount} * \text{AWW Quantity} * \frac{26.088}{\text{Scheduled Hours Of Work}}$$

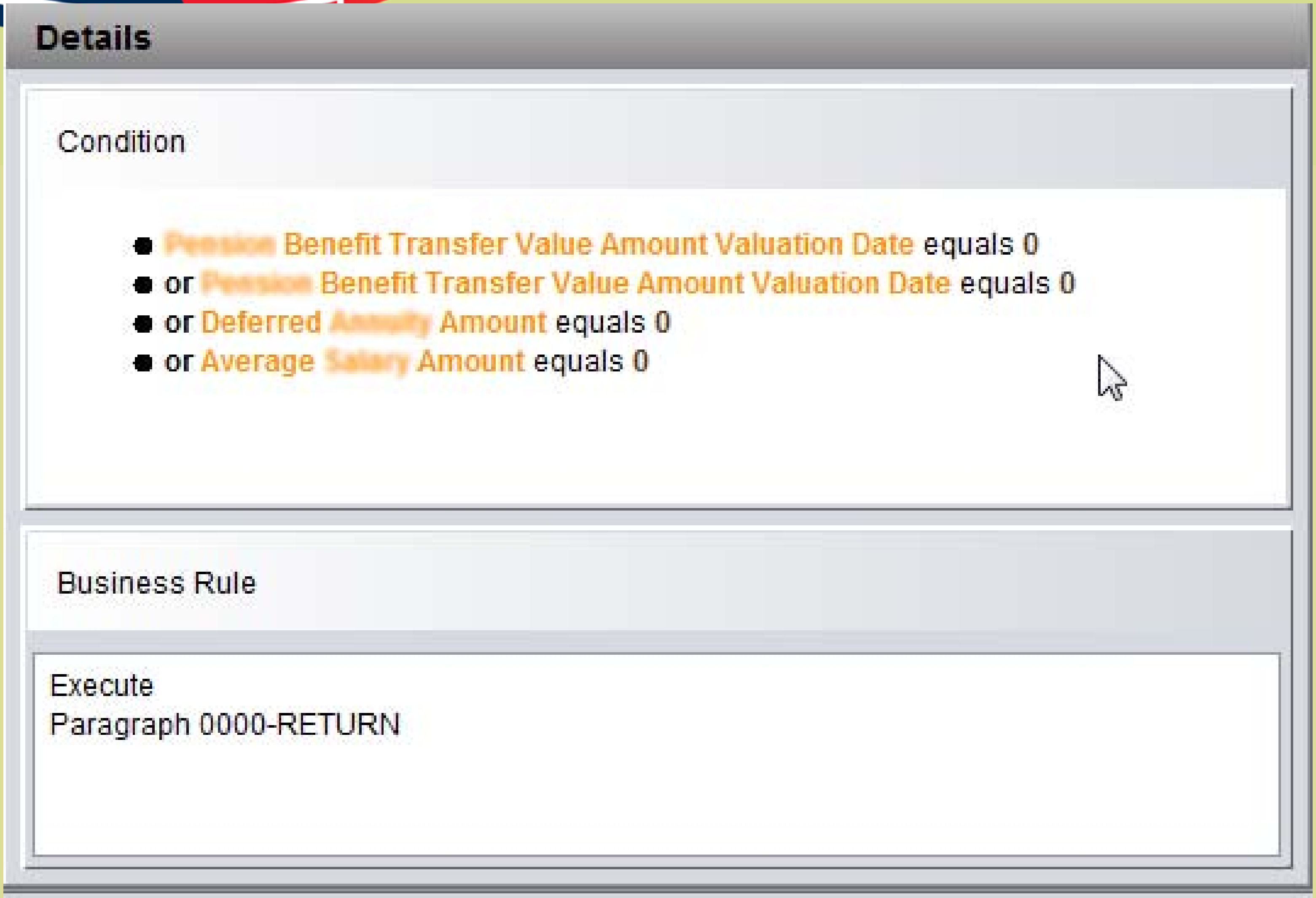
Actions

- Locate in code
- Dependency graph

Documentation

- AWW Quantity
- Rate Amount
- Rate Base Identifier
- Scheduled Hours Of Work

Conditions



Details

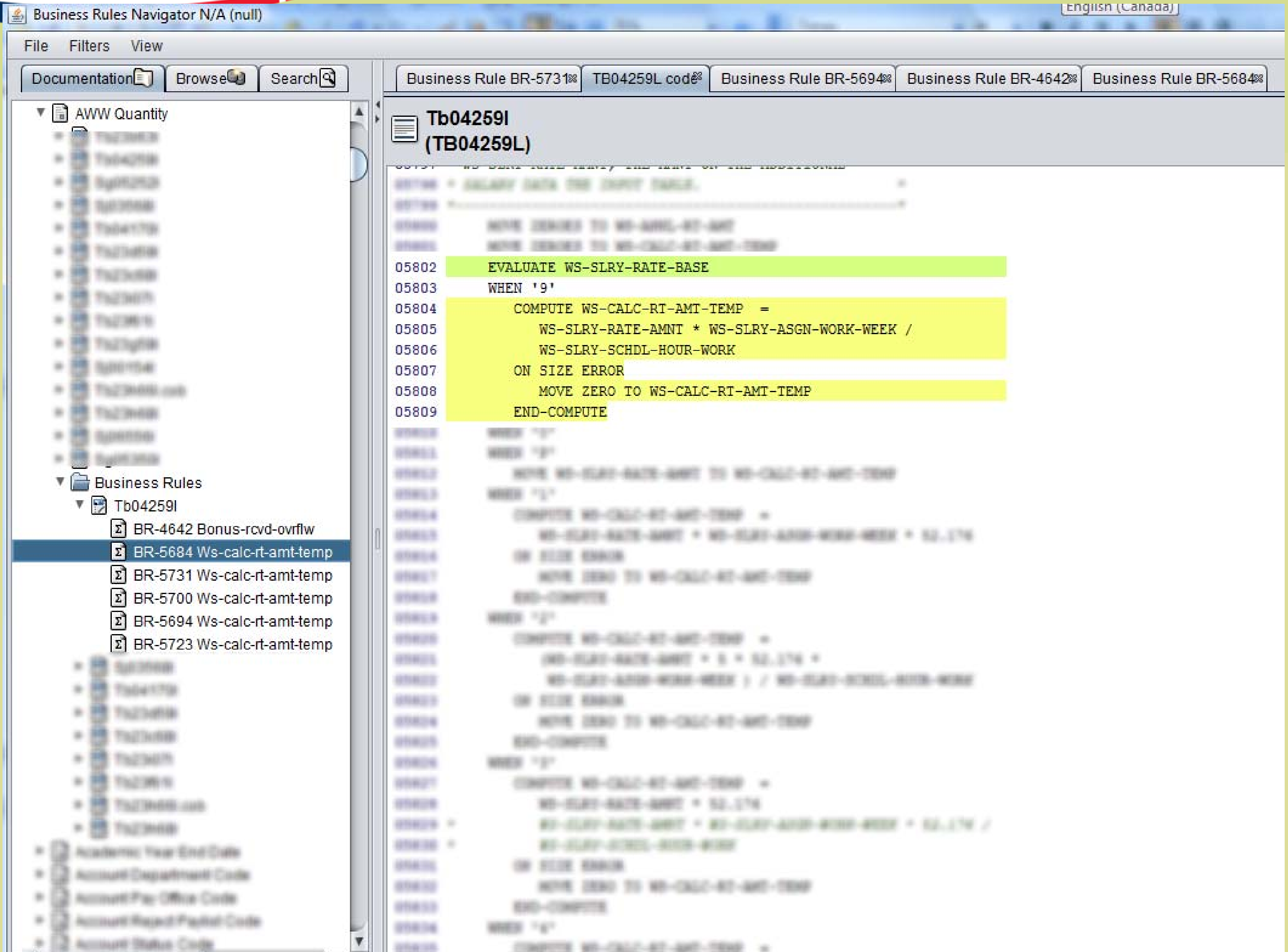
Condition

- [redacted] Benefit Transfer Value Amount Valuation Date equals 0
- or [redacted] Benefit Transfer Value Amount Valuation Date equals 0
- or Deferred [redacted] Amount equals 0
- or Average [redacted] Amount equals 0

Business Rule

Execute
Paragraph 0000-RETURN

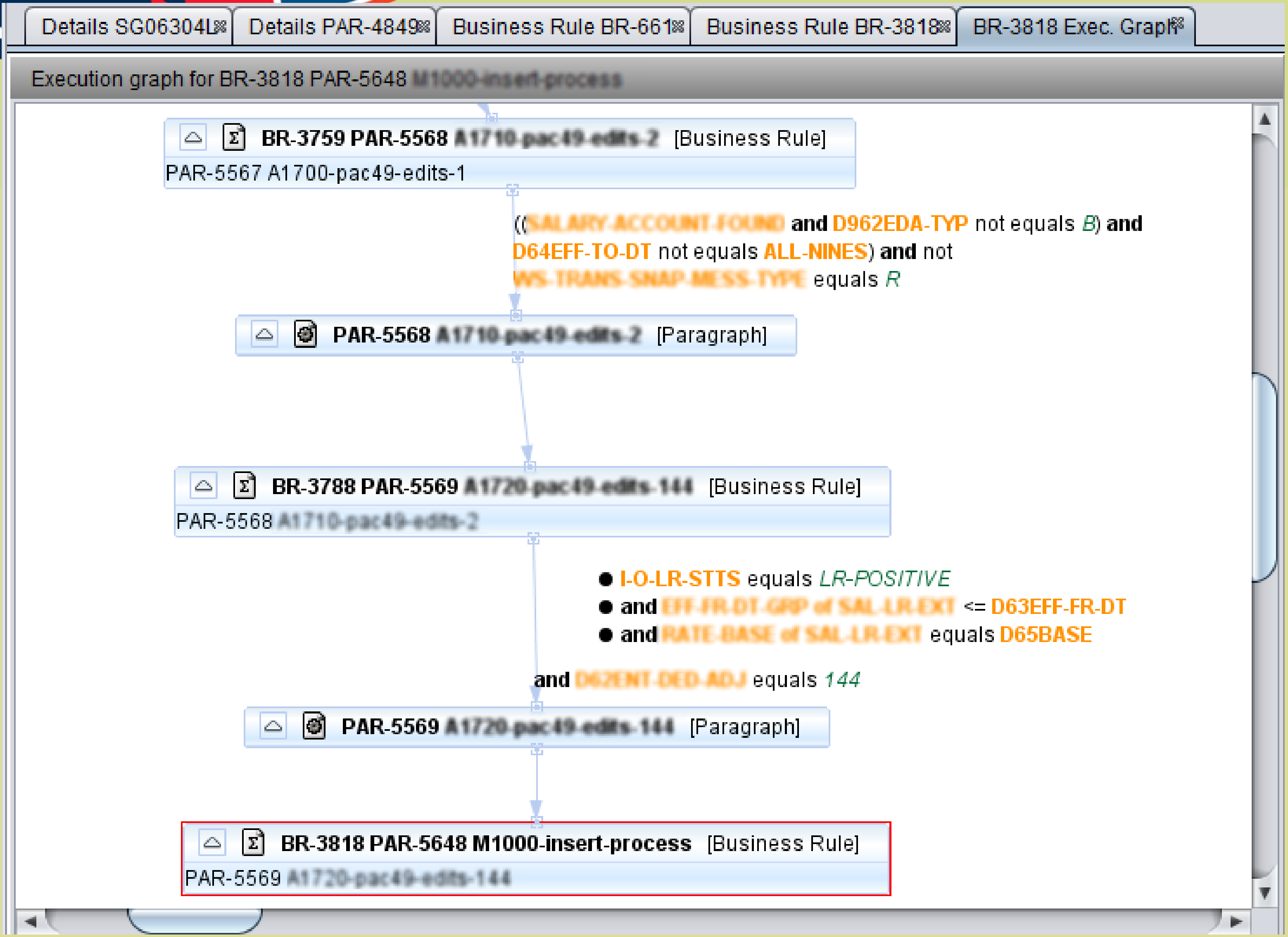
Linking artefacts to source code


Business Rules Navigator N/A (null)
English (Canada)
File
Filters
View
Documentation
Browse
Search
Business Rule BR-5731
TB04259L code
Business Rule BR-5694
Business Rule BR-4642
Business Rule BR-5684
AWW Quantity
Business Rules
Tb04259I
BR-4642 Bonus-rcvd-ovrflw
BR-5684 Ws-calc-rt-amt-temp
BR-5731 Ws-calc-rt-amt-temp
BR-5700 Ws-calc-rt-amt-temp
BR-5694 Ws-calc-rt-amt-temp
BR-5723 Ws-calc-rt-amt-temp

Tb04259I
(TB04259L)

```
       EVALUATE WS-SLRY-RATE-BASE
       WHEN '9'
          COMPUTE WS-CALC-RT-AMT-TEMP  =
             WS-SLRY-RATE-AMNT * WS-SLRY-ASGN-WORK-WEEK /
             WS-SLRY-SCHDL-HOUR-WORK
          ON SIZE ERROR
             MOVE ZERO TO WS-CALC-RT-AMT-TEMP
          END-COMPUTE
```

Global execution path



Execution graph for BR-3818 PAR-5648 [redacted]

BR-3759 PAR-5568 [redacted] [Business Rule]
PAR-5567 A1700-pac49-edits-1

(([redacted] **and** D962EDA-TYP not equals *B*) **and** D64EFF-TO-DT not equals ALL-NINES) **and** not [redacted] equals *R*

PAR-5568 [redacted] [Paragraph]

BR-3788 PAR-5569 [redacted] [Business Rule]
PAR-5568 [redacted]

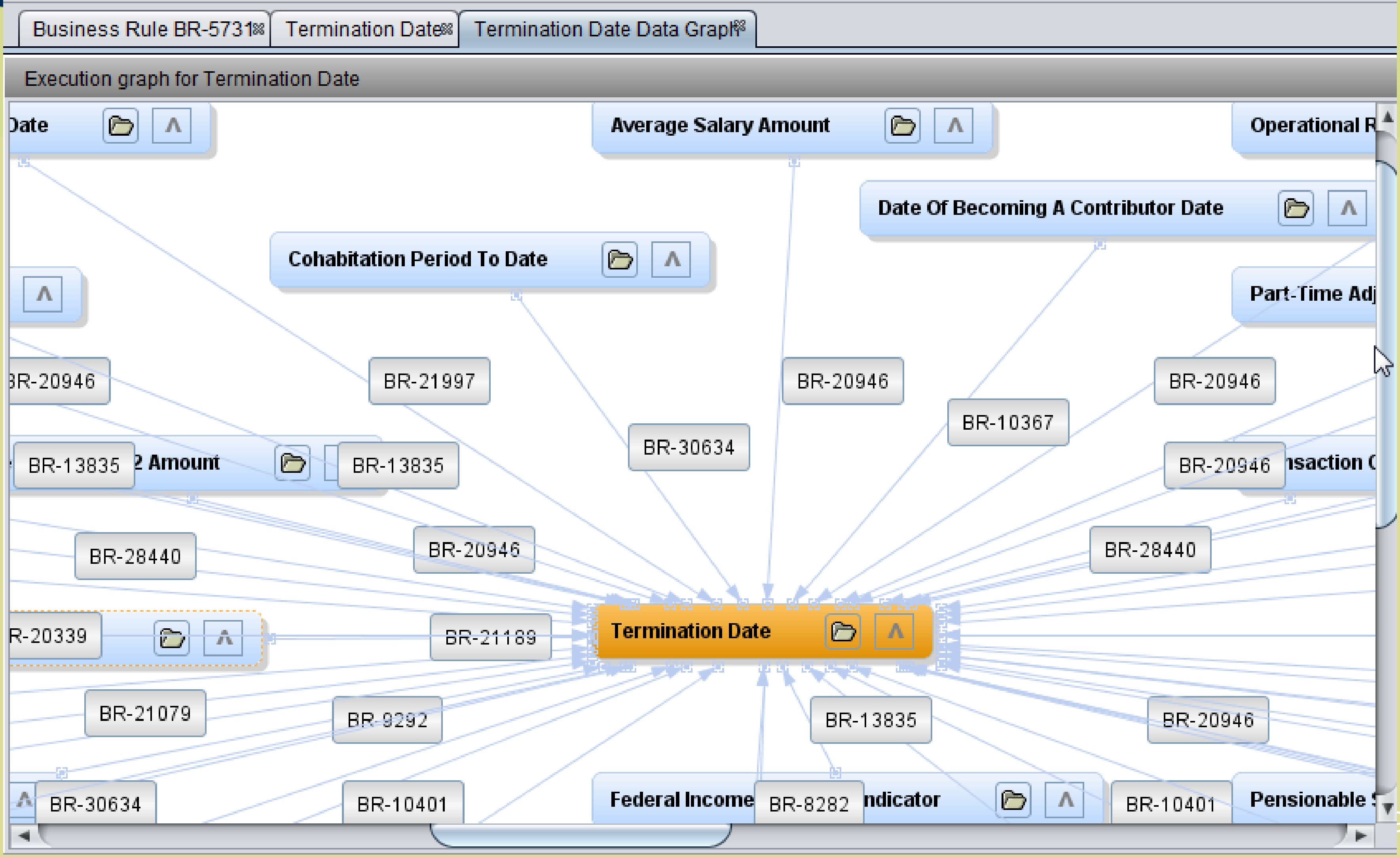
- I-O-LR-STTS equals *LR-POSITIVE*
- **and** [redacted] <= D63EFF-FR-DT
- **and** [redacted] equals D65BASE

and [redacted] equals *144*

PAR-5569 [redacted] [Paragraph]

BR-3818 PAR-5648 M1000-insert-process [Business Rule]
PAR-5569 [redacted]

Finding data dependencies



Outline

- Legacy Software & Modernization
- Extracting Business Rules
- Connecting documentation with business rules
- Conclusions

Conclusions

- **Business Rules: major element in legacy software modernization**
  - For system maintainers and business analysts
- **Possible to extract business rules from legacy source code**
- **Novelty**
  - output is targeted at business analysts
  - the business rules translated into non-technical terms
  - Business Rules are connected to existing documents using keyphrase extraction techniques
- **Use a formal model to represent the rules and enable complex transformations on extracted rules**

```
 Colonne . . . :    1  71            Editazione                  QGPL/QLBLSRC
 SEU==>                                                                VERIFY
 FMT CB ......-A+++B+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
        ***************** Inizio dati ***************************************
0001.00
0002.00
0003.00
0004.00
0005.00
0006.00
0007.00
0008.00                             QUESTIONS?
0009.00
0010.00
0011.00
0012.00
0013.00
0014.00
0015.00
0016.00

 F3=Fine   F4=Richiesta F5=Rivisual. F9=Duplicaz.  F10=Cursore  F11=Alternare
 F16=Ripetiz.ricerca   F17=Ripetiz.modifica        F24=Altri tasti
                                        (C) COPYRIGHT IBM CORP. 1981, 1996.
 1           »                                                          2/9
```

NRC·CNRC

Institute for
Information
Technology

Science
at work for
Canada

National Research
Council Canada

Conseil national
de recherches Canada

Canada