



Service Public
Fédéral
FINANCES

OUR EXPERIENCE WITH KNIME AS E-AUDITORS

FEDERAL PUBLIC SERVICE FINANCE BELGIUM

SPECIAL TAX INSPECTORATE, 5TH DIRECTORATE

Pieter Vercruysse

Pascal Beaujeant



INTRODUCTION (1)

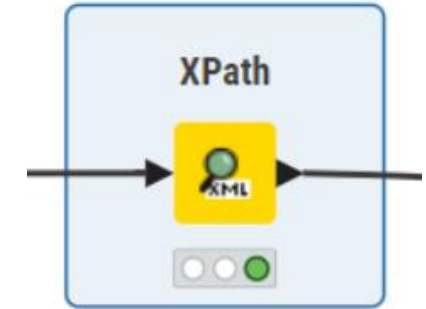
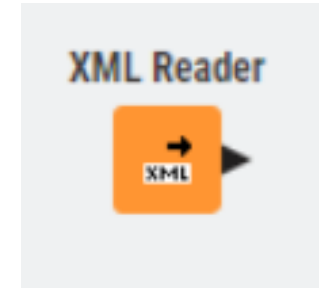
- E-audit team in Belgium
- Support for tax auditors
- Usage of the open source desktop version of KNIME: the KNIME Analytics Platform
- We use KNIME mainly for analyzing the following input files:
 - Excel and CSV files
 - PDF files
 - XML files
 - Outlook Mailboxes (PST files)



KNIME NODES WE USE REGULARLY (1)

1) XML Reader node + XPath node

- Our goal: Processing of financial documents



Use namespace:
2020-02-01

COMPTES ANNUELS

BILAN APRÈS RÉPARTITION

	Ann.	Codes	Exercice	Exercice précédent
ACTIF				
FRAIS D'ÉTABLISSEMENT		20		
ACTIFS IMMOBILISÉS		21/28	763.216	783.245
Immobilisations incorporelles	6.1.1	21		
Immobilisations corporelles	6.1.2	22/27	752.319	773.026
Terrains et constructions		22	752.319	692.056
Installations, machines et outillage		23		
Mobilier et matériel roulant		24	0	80.970
Location-financement et droits similaires		25		
Autres immobilisations corporelles		26	0	0
Immobilisations en cours et acomptes versés		27		
Immobilisations financières	6.1.3	28	10.896	10.219
ACTIFS CIRCULANTS		29/58	450.390	427.133
Créances à plus d'un an		29		



KNIME NODES WE USE REGULARLY (2)

Dialog - 3:134 - XPath (Use namespace:)

File

Settings Namespace Flow Variables Job Manager Selection Memory Policy

XML column: ☒ XML

☒ Remove source column.

- XPath summary

Column name	XPath query
EntityCurrentLegalName	/xbrli:xbrl/pfs-gcd:EntityInformation/pfs-gcd:EntityName/pfs-gcd:EntityCurrentLegalName
IdentifierValue	/xbrli:xbrl/pfs-gcd:EntityInformation/pfs-gcd:EntityIdentifier/pfs-gcd:IdentifierValue
PeriodStartDate	/xbrli:xbrl/pfs:PeriodsCovered/pfs:PeriodStartDate[@contextRef='CurrentInstant' or @contextRef='c12' or @contextRef='c2']
PeriodEndDate	/xbrli:xbrl/pfs:PeriodsCovered/pfs:PeriodEndDate[@contextRef='CurrentInstant' or @contextRef='c12' or @contextRef='c2']
FormationExpenses	/xbrli:xbrl/pfs:FormationExpenses[@contextRef='CurrentInstant' or @contextRef='c12' or @contextRef='c2']
FixedAssetsFormationExpensesExcluded	/xbrli:xbrl/pfs:FixedAssetsFormationExpensesExcluded[@contextRef='CurrentInstant' or @contextRef='c12' or @contextRef='c2']
IntangibleFixedAssets	/xbrli:xbrl/pfs:IntangibleFixedAssets[@contextRef='CurrentInstant' or @contextRef='c12' or @contextRef='c2']
TangibleFixedAssets	/xbrli:xbrl/pfs:TangibleFixedAssets[@contextRef='CurrentInstant' or @contextRef='c12' or @contextRef='c2']
LeadBuildings	/xbrli:xbrl/pfs:LeadBuildings[@contextRef='CurrentInstant' or @contextRef='c12' or @contextRef='c2']

Selected XPath:

Selected XML element is not a tag nor an attribute.

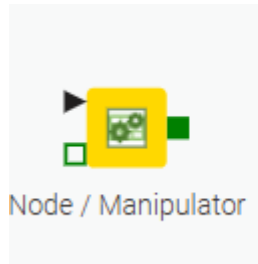
- XML-Cell Preview

```
93         <pfs-v1:XCode_LanguageCode_FR contextRef="D0">FR</pfs-v1:XCode_LanguageCode_FR>
94     </pfs-gcd:DocumentLanguage>
95 </pfs-gcd:DocumentInformation>
96 <pfs:PeriodsCovered>
97     <pfs:PeriodStartDate contextRef="I0">2019-07-01</pfs:PeriodStartDate>
98     <pfs:PeriodEndDate contextRef="I0">2020-06-30</pfs:PeriodEndDate>
99     <pfs:PrecedingPeriodData contextRef="D0">true</pfs:PrecedingPeriodData>
100    <pfs:AccountsApprovalDateGeneralAssembly contextRef="I0">2020-11-27</pfs:AccountsApprovalDateGeneralAssembly>
101    <pfs:Schema>
102        <pfs-v1:XCode_SchemaCode_40 contextRef="D0">40</pfs-v1:XCode_SchemaCode_40>
103    </pfs:Schema>
104 </pfs:PeriodsCovered>
```



KNIME NODES WE USE REGULARLY (3)

2) Continental Nodes for KNIME: to embellish Excel sheets



XLS Sheet Properties

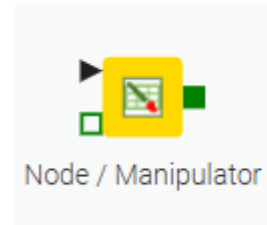
Community Nodes

Continental

XLS Formatter

+1

Node / Manipulator



XLS Background Colorizer

Community Nodes

Continental

XLS Formatter

+1

Node / Manipulator

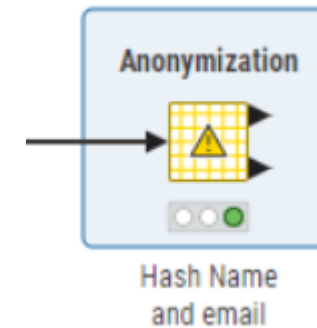
3) Redfield nodes: (3 nodes)

- PST Reader node: to read Outlook mailboxes



KNIME NODES WE USE REGULARLY (4)

- The Redfield Anonymization node

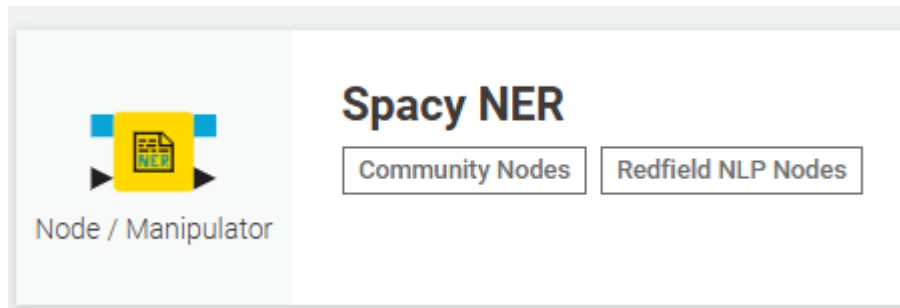


Name <i>String</i>	Name_anonymized <i>String</i>	email <i>String</i>	email_anonymized <i>String</i>
Pieter	ec0a7fb5e54734c9bccf9f827892f5e14fddd953	pieter.vercruysse@minfin.fed.be	4217fb44dc94fb0205ed2625ea739aad6cb9789e
Pascal	6777af32af33a53db4fe75fa96694a4d37be1cf2	pascal.beaujeant@minfin.fed.be	38e1f9ff494525a91ce3e3340ffdc198c9cc1742
Pieter	ec0a7fb5e54734c9bccf9f827892f5e14fddd953	pieter.vercruysse@minfin.fed.be	4217fb44dc94fb0205ed2625ea739aad6cb9789e
Pieter	ec0a7fb5e54734c9bccf9f827892f5e14fddd953	<input checked="" type="checkbox"/> pieter.vercruysse@minfin.fed.be	4217fb44dc94fb0205ed2625ea739aad6cb9789e



KNIME NODES WE USE REGULARLY (5)

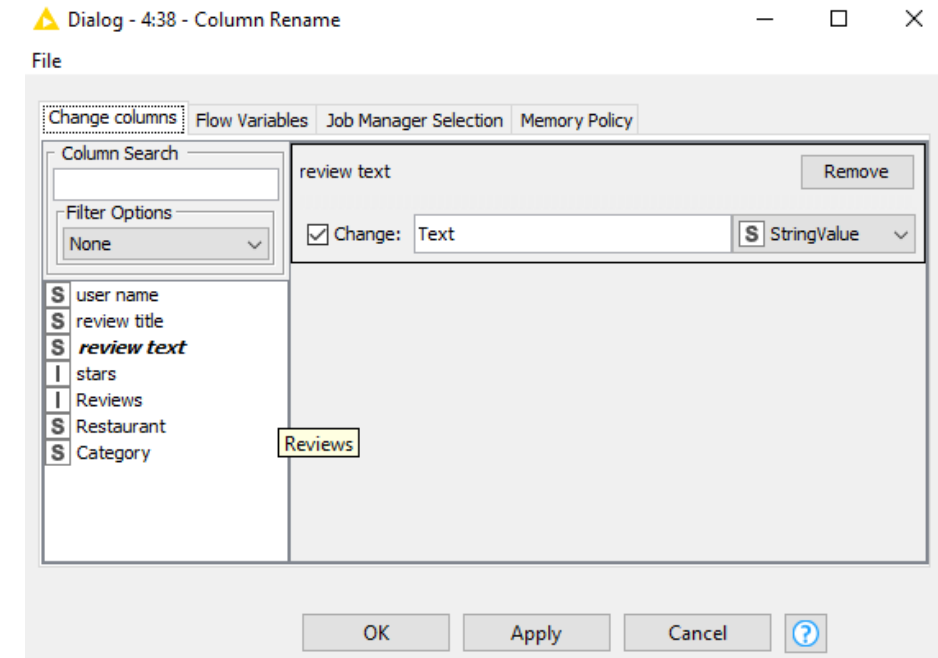
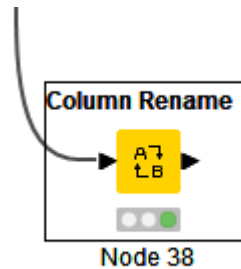
- The Spacy nodes for text processing
 - Based on a Python library
 - For instance NER node: Named Entity Recognition for finding persons, cities, ... in a text or body of an email, ...



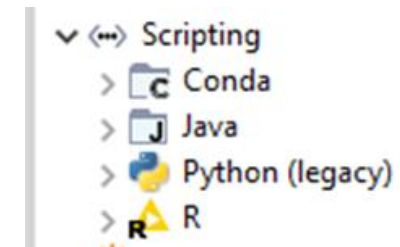


ADVANTAGES OF KNIME (1)

- Automation without using programming code



- If necessary, 4 programming languages can be used in KNIME:
even in the same workflow
 - Python and R, a statistical programming language
 - Java and JavaScript





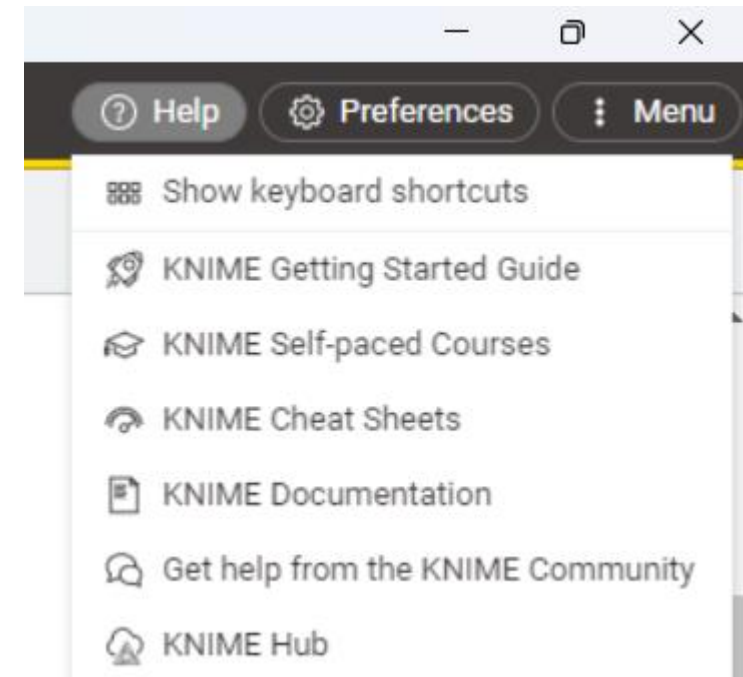
ADVANTAGES OF KNIME (2)

- Workflows are great for collaboration within an e-audit team
 - Everyone can use his preferred programming language
 - Everyone can work on his own component: a collection of nodes
 - Components can be assembled into one big workflow
- Workflows are great for “Total automation”
 - Possibly to verify all transactions
 - Possibly to work on hundreds of files: use of loops



ADVANTAGES OF KNIME (3)


- KNIME Documentation
 - See the Help function in KNIME
 - But also:
 - eBooks
 - KNIME TV on Youtube
 - KNIME Community Forum
 - For junior auditors: especially the self-paced courses are relevant





A REMARK ON THE R LANGUAGE

- R, the statistical programming language
 - Possible to use R in a virtual environment in combination with KNIME
 - Install R via a conda distribution
 - Installation via a yml file



```
environment.yml

Bestand  Bewerken

name: r-env06
channels:
  - conda-forge
dependencies:
  - r-base
  - r-broom
  - r-car
  - r-caret
```



THE LIFE OF AN E-AUDITOR

- Small tools repositories
- Reusable
- Adapted to Tax Officers
- Dedicated Sharepoint accessible to all

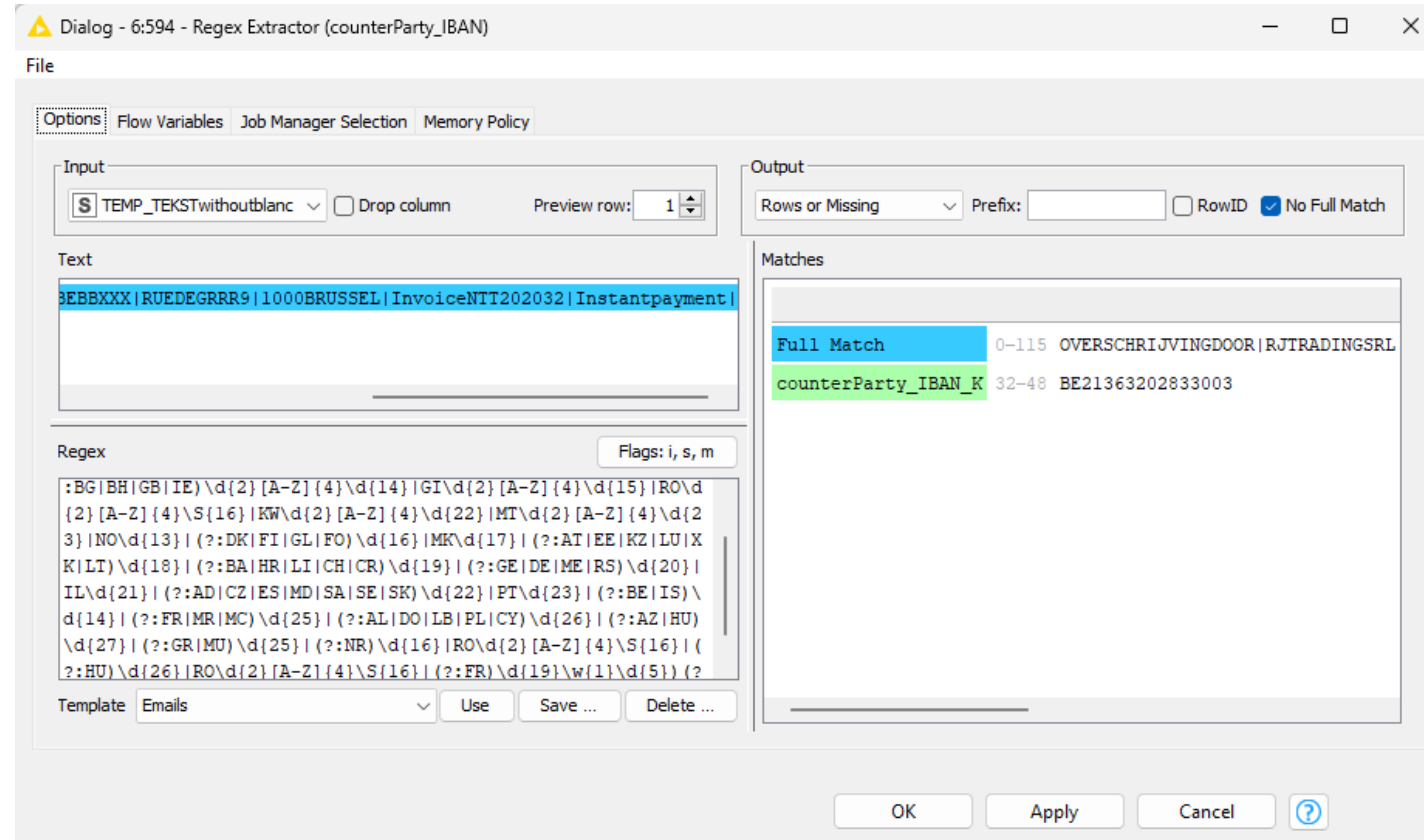
1: Extraction Result

#	RowID	TEKST	TEMP_TEKSTwit...	counterParty_IBAN
1	Row0	OVERSCHRIJVING DOC	OVERSCHRIJVINGDOO	BE21363202833003
2	Row1	Maandelijkse bijdrage	Maandelijksebijdragep:	?
3	Row2	Portkosten	Portkosten	?
4	Row3	OVERSCHRIJVING T.G.	OVERSCHRIJVINGT.G.\	NL15RABO014584025!
5	Row4	OVERSCHRIJVING DOC	OVERSCHRIJVINGDOO	BE06363201981322
6	Row5	OVERSCHRIJVING T.G.	OVERSCHRIJVINGT.G.\	NL70INGB0009181556



THE LIFE OF AN E-AUDITOR

- Easy to debug
- Easy to share with Tax Officers but also with other administrations and even countries
- Follow the Tax Officers needs
- Understandable by most of us (visual programming)

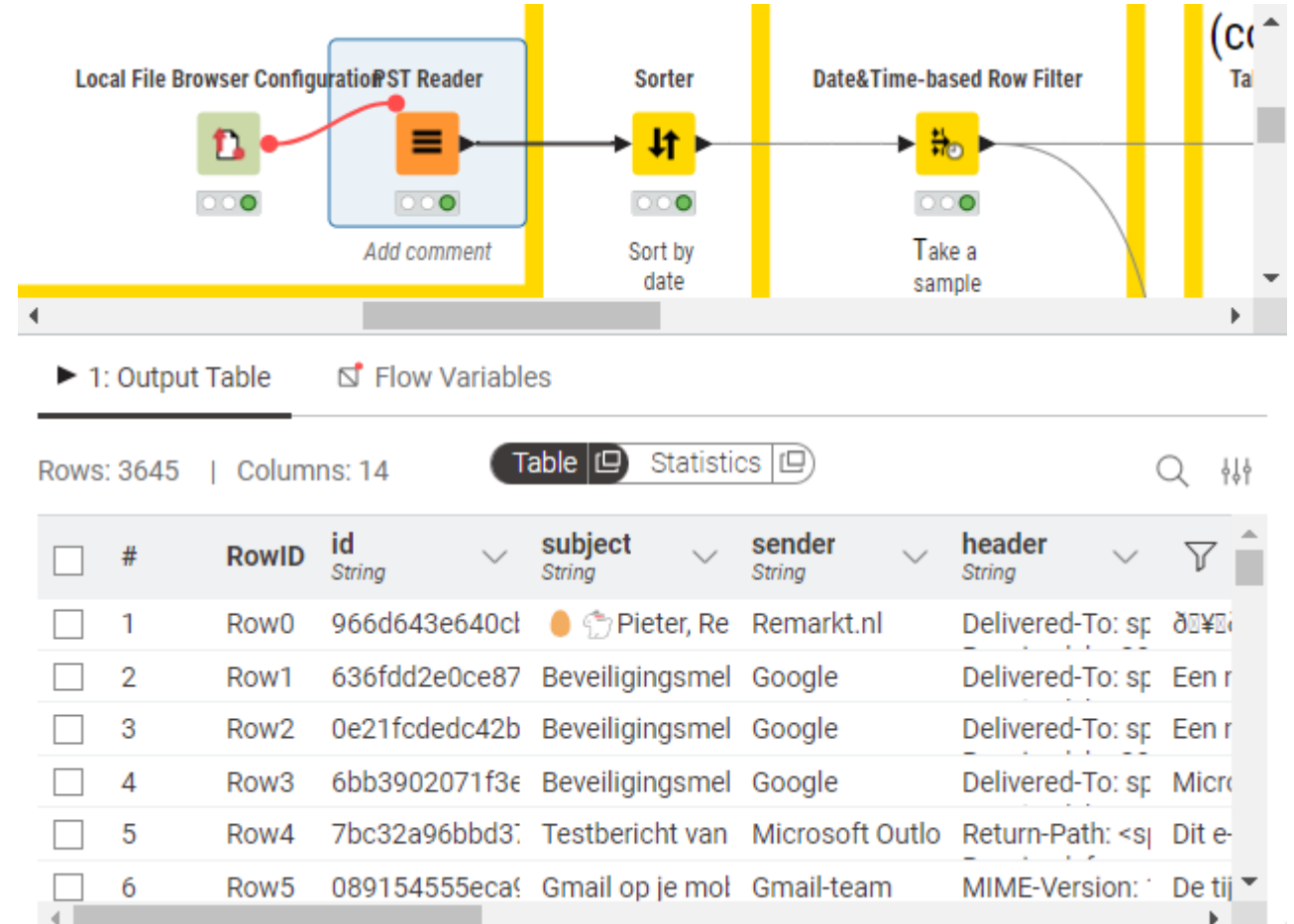




THE LIFE OF AN E-AUDITOR

- 20,000 members offering autonomous computing power
- Saving time through automation (week or month)

We learn much better and much faster, while having fun.





THE LIFE OF AN E-AUDITOR

- Empowering tax auditors
- Improves agent's knowledge in a fun way
- Capitalization of tools

Alone we may go
faster, together we go
further!

