



Universidade Federal de Sergipe  
CCET – DCOMP  
Disciplina: *Sistemas de Apoio à Decisão*  
Professor: *Gilton José Ferreira da Silva*  
Aluno: *Matheus Mendonça Menezes*

## Atividade 5

a) A partir do site [<http://dados.gov.br/>], utilize as informações advindas dos sistemas do Governo para servir de carga

**dados.gov.br**  
PORTAL BRASILEIRO DE DADOS ABERTOS

**Pesquisa ...** PESQUISAR  
em 1124 conjuntos de dados com 8313 recursos (o que é isto?)

**Dados em destaque**

- Reclamações do Consumidor.gov.br**  
O Consumidor.gov.br é um novo serviço público para solução alternativa de conflitos de consumo por meio da internet, que permite a interlocução ...
- Lista de Eleitores Filiados aos Partidos Políticos**  
Conforme [provisamento nº 04/2012](<http://www.justicaeleitoral.jus.br/arquivos/tse-provimento-n-4-2012-cge>) (formato PDF), a publicação, na Internet, ...
- Imóveis Dominiais da União**  
Esta seção reúne informações relativas aos imóveis da União que não são ocupados pela Administração Pública, mas sim cedidos a terceiros por meio dos ...

**Publicações mais recentes**

Conjunto de dados	Autor	Quando
Banco de Variáveis de Ciência, ...	Coordenação -Geral de ...	29 Jul
Indicadores Nacionais de Ciência, ...	Coordenação -Geral de ...	28 Jul
Cidades Digitais	Secretaria de Inclusão ...	28 Jul
REGIÃO DE INFORMAÇÃO DE VOO - FIR	Divisão de Operações	26 Jul
AEROVIAS	Divisão de Operações	26 Jul

**Dados por etiqueta**

INDE (923)	Geoespacial (789)	Geoserviço (789)
Cartografia (789)	PGI (182)	Indicador (182)
Previdência Social (101)	AEPS (100)	Benefícios (44)
Benefícios Previdenciários (43)	Contribuintes (21)	UF (19)
GFIP (19)	Equipamentos públicos (17)	INEP (15)
federal (15)	GPS (14)	Receitas (13)
IBGE (13)	SP (13)	

**Notícias**

- Ministério do Turismo publicará dados em formato aberto**  
Informações sobre visitas de turistas estrangeiros estarão disponíveis 12/08/2016  
O Ministério do Turismo acaba de publicar o seu Plano de Dados Abertos. ...
- Ministério da Transparência divulga Plano de Dados Abertos**  
Bases serão abertas pelo órgão até o fim de 2017 O Ministério da Transparência, Fiscalização e Controle (MTFC) publicou hoje (3) o seu Plano de ... 04/08/2016

Figura 1: Página inicial de <http://dados.gov.br>

Após acessar o site especificado na Figura 1, basta clicar em “Dados” e em seguida escolher dentre uma das bases de dados disponíveis para download. Conforme mostrado na Figura 2, os

dados são agrupados de acordo com as organizações que os disponibilizam, ou ainda por categorias como formatos, etiquetas, grupos e tipo de licença dos arquivos.

The screenshot displays the 'dados.gov.br' portal, the Brazilian Open Data Portal. The header includes navigation links for 'Participe', 'Acesso à informação', 'Legislação', and 'Canais'. The main navigation bar features a search bar and social media icons. The 'Conjuntos de dados' (Data Sets) section is active, showing a list of organizations on the left and search results on the right. The search results for 'Anuário Estatístico de Acidentes de Trabalho - AEAT' are displayed, including a description and download links for CSV and PDF formats.

**BRASIL** Serviços | Participe | Acesso à informação | Legislação | Canais

Ir para o conteúdo 1 Ir para o menu 2 Ir para a busca 3 Ir para o rodapé 4

**dados.gov.br**  
PORTAL BRASILEIRO DE DADOS ABERTOS

Pesquisar conjuntos de dados...

Dados | Aplicativos | Perguntas frequentes | Contato | Sobre o portal

**/ Conjuntos de dados**

**Organizações**

- Ministério da Saúde... (65)
- Ministério do Traba... (42)
- Ministério da Justi... (30)
- Instituto Brasileir... (18)
- Ministério do Plane... (14)
- Agência Nacional do... (9)
- Fundação Nacional d... (5)
- Ministério do Desen... (4)
- Ministério da Ciênc... (4)
- Tribunal de Contas ... (3)

**Mostrar mais Organizações**

**Grupos**

- Governo e Política (32)
- Saúde (27)
- Educação (20)
- Pessoa, Família e S... (19)
- Equipamentos Públicos (16)

Pesquisar conjuntos de dados...

**1.122 conjuntos de dados encontrados**

Ordenar por: Relevância

**Anuário Estatístico de Acidentes de Trabalho - AEAT**

Anuário Estatístico de Acidentes de Trabalho - AEAT, documento público, com edição e publicação anuais em meio papel e versão digital acessível por meio do sítio do Ministério...

**Domínios Gov.br**

Informações sobre os domínios Gov.br registrados no Registro.br e autorizados pelo Ministério do Planejamento. Contém todos os domínios autorizados e seus respectivos...

**CSV PDF**

**Indicadores sobre UPA - Unidade de Pronto Atendimento**

Série histórica de indicadores dos resultados governamentais relativos à UPA - Unidade de Pronto Atendimento. PGI é um repositório de informações estruturadas (indicadores) à...

**XML JSON**

**Indicadores sobre Unidade de Atendimento do INSS**

Série histórica de indicadores dos resultados governamentais relativos à Unidade de Atendimento do INSS. PGI é um repositório de informações estruturadas (indicadores) à feição...

**XML JSON**

Figura 2: Parte dos dados disponíveis

Para este trabalho foi efetuado o download do arquivo CSV “Cidades Digitais” disponível a partir da etiqueta “Internet”, conforme Figura 3.



Figura 3: Arquivo selecionado para este trabalho

**b) Inicie o PDI Pentaho e crie um pequeno ETL:** ENTRADA: XLS ou CSV  
SAÍDAS: JSON, XLS, TXT e XML

O PDI Pentaho é executado no Mac OSX a partir do terminal, conforme Figura 4.

The image shows a terminal window titled 'data-integration — java — 80x24'. The terminal output shows the following commands and responses:  
Last login: Wed Aug 17 21:21:41 on console  
macmini:~ matheus\$ cd Downloads/Mac\_apps/  
macmini:Mac\_apps matheus\$ cd data-integration/  
macmini:data-integration matheus\$ ./spoon.sh  
Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=256m; sup  
port was removed in 8.0  
21:37:06,626 INFO [KarafInstance]  
\*\*\*\*\*  
\*\*\* Karaf Instance Number: 1 at /Users/matheus/Downloads/Mac\_apps/data-inte \*\*\*  
\*\*\* gration/./system/karaf/caches/spoon/data-1 \*\*\*  
\*\*\* Karaf Port:8802 \*\*\*  
\*\*\* OSGI Service Port:9051 \*\*\*  
\*\*\*\*\*  
21:37:06,626 INFO [KarafBoot] Checking to see if org.pentaho.clean.karaf.cache  
is enabled  
ago 17, 2016 9:37:24 PM org.apache.karaf.main.Main\$KarafLockCallback lockAcquired  
INFORMAÇÕES: Lock acquired. Setting startlevel to 100  
2016/08/17 21:37:30 - cfgbuilder - Warning: The configuration parameter [org] is  
not supported by the default configuration builder for scheme: sftp  
2016-08-17 21:37:49.493:INFO:oejs.Server:jetty-8.1.15.v20140411  
2016-08-17 21:37:50.663:INFO:oejs.AbstractConnector:Started NIOSocketConnectorWr  
apper@0.0.0.0:9051  
ago 17, 2016 9:37:54 PM org.apache.cxf.bus.blueprint.NamespaceHandlerRegisterer  
register

Figura 4: Execução do PDI Pentaho via terminal

Após efetuar o download e executar o PDI Pentaho, deve selecionar o método de input para correta utilização da ferramenta, como o arquivo baixado foi no formato CSV, foi escolhido input para o mesmo tipo de arquivo, conforme pode ser visto na Figura 5.

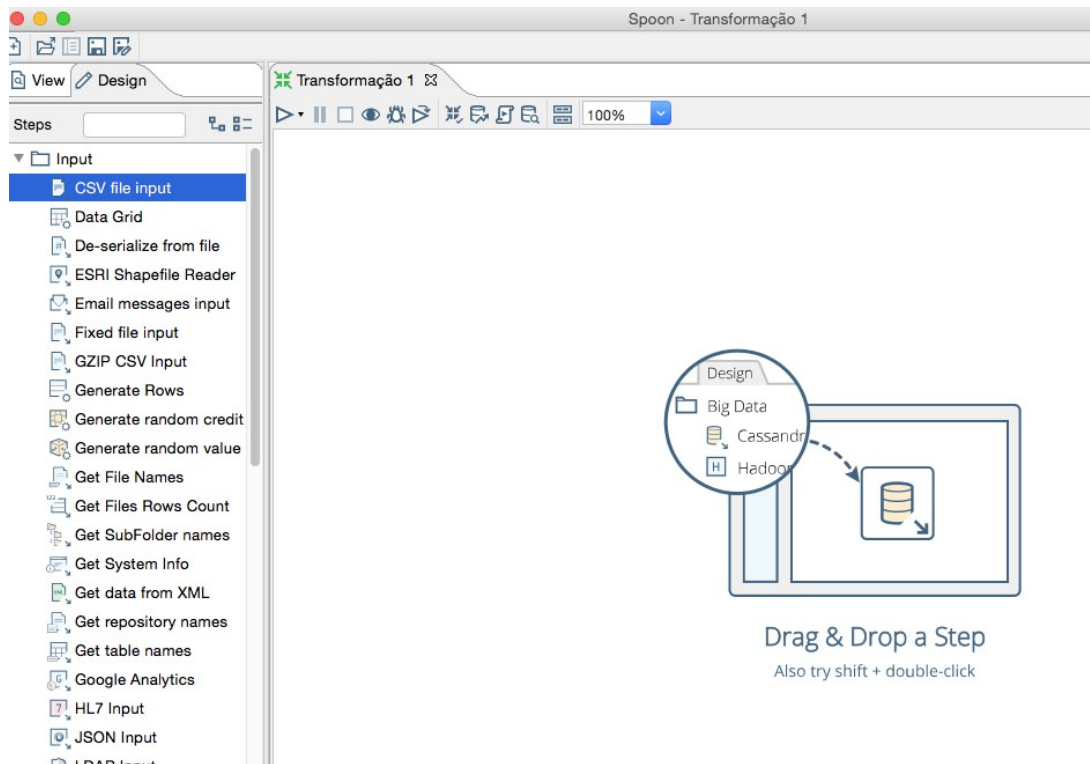


Figura 5: Escolha de input no PDI Pentaho

Depois basta clicar duas vezes no tipo de arquivo, escolher o arquivo baixado e clicar em “obtem campos” para que o arquivo possa ser processado pelo PDI, conforme ilustrado na Figura 6.

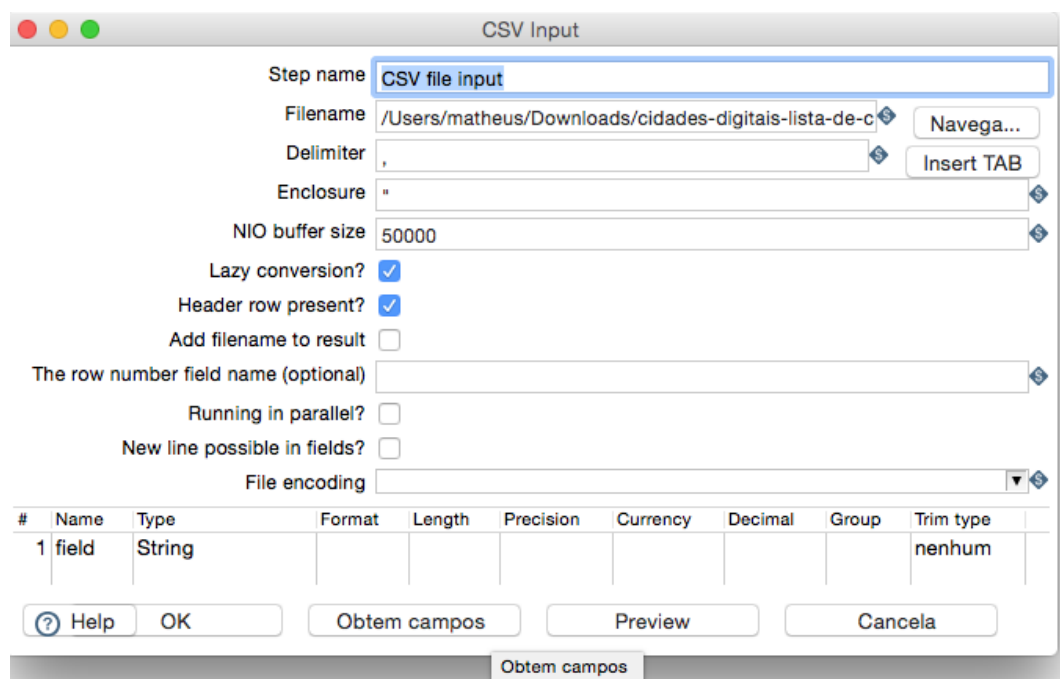


Figura 6: Adicionando o arquivo de input



Depois de selecionar o arquivo de entrada deve-se escolher os tipos de saída desejados também na barra lateral e em seguida interligá-los, como mostrado na Figura 7. Neste trabalho foram selecionados os seguintes outputs: XML; TXT; XLS e JSON.

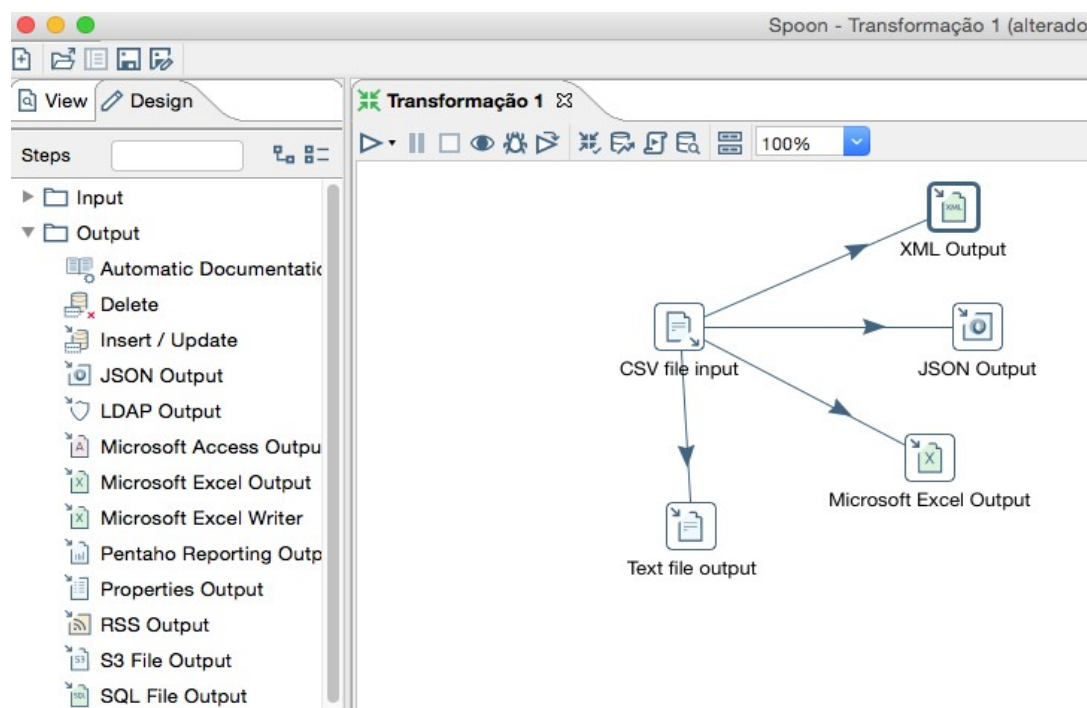


Figura 7: Seleção dos outputs

É importante também dar um duplo clique para cada output, de modo que seja possível escolher seus atributos como nome e local de salvamento, conforme Figura 8.

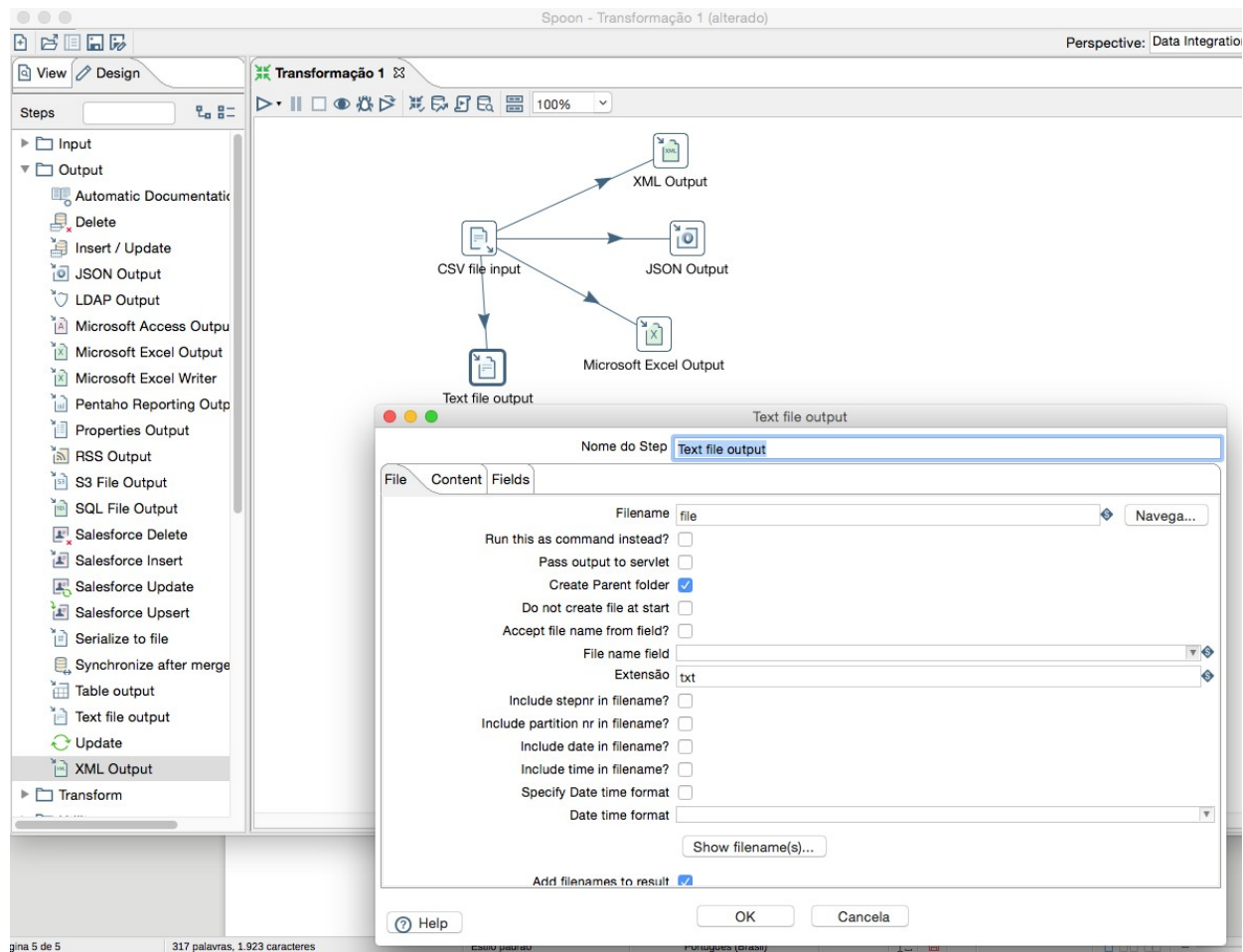


Figura 8: Detalhamento de output

Após clicar em RUN os arquivos são criados a partir do input, conforme figuras abaixo.

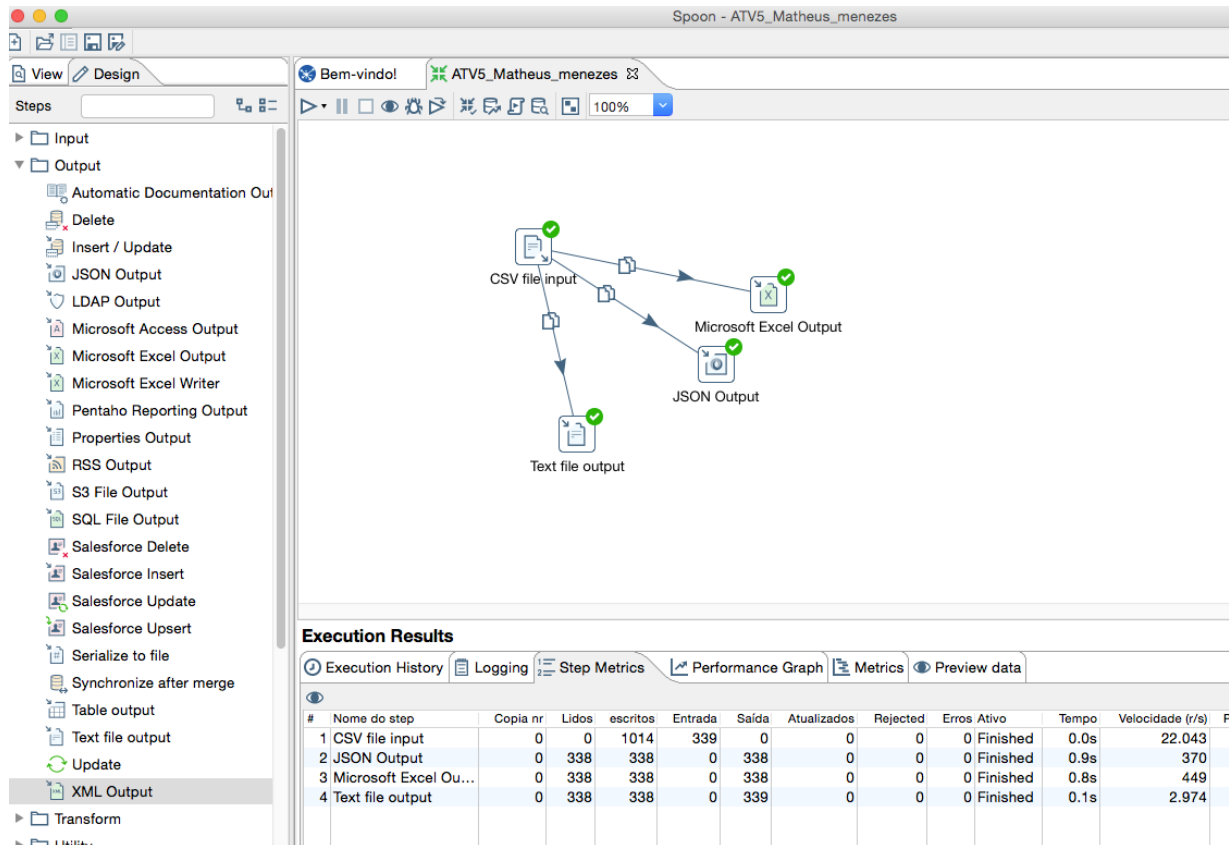


Figura 9: Execução da transformação

Houve um problema e o arquivo XML não foi criado, conforme erros ilustrados nas figuras abaixo.

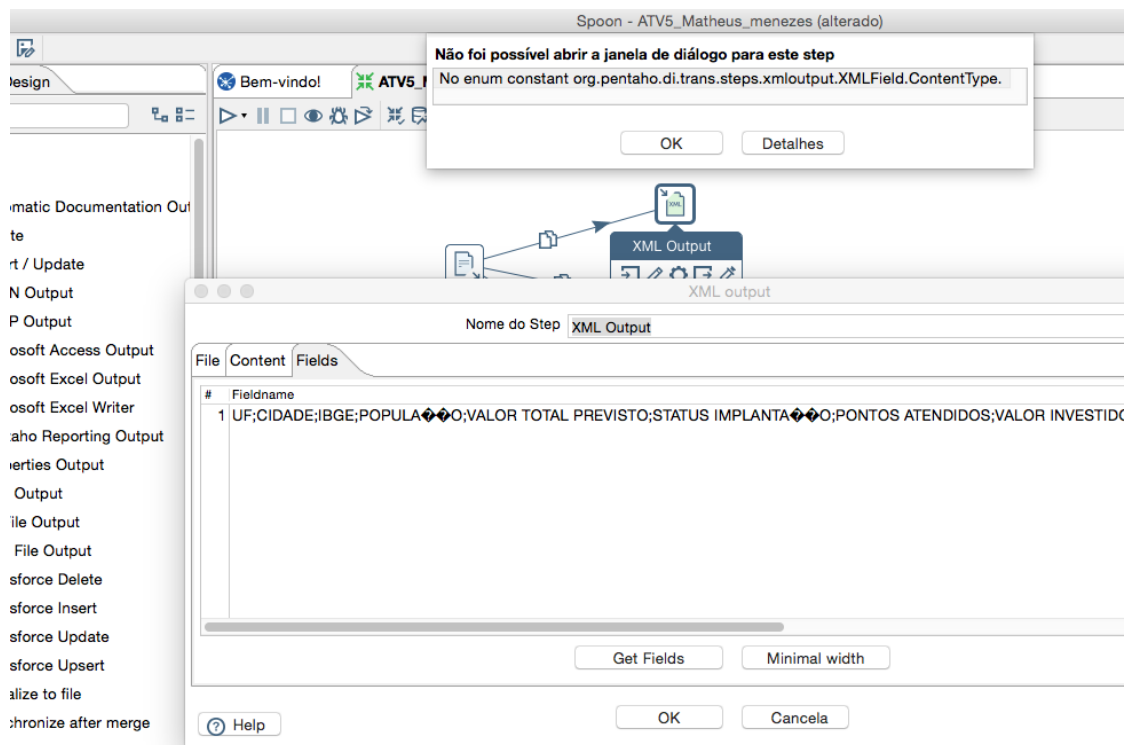


Figura 10: Erro no XML