

Talend Open Studio for Data Integration

Créer une métadonnée de connexion à un fichier délimité

Paramétrer des schémas de fichier délimité dans le Repository.

Ce tutoriel montre comment utiliser l'assistant "Delimited File" lorsque vous devez traiter des formats de fichiers complexes. Vous pouvez créer des schémas spécifiques correspondant à chacun de vos besoins.

Par exemple, vous pouvez imaginer un schéma contenant une "adresse personnelle" et un autre contenant une "adresse de livraison" mais correspondant tous les deux au même fichier.

Pré-requis :

Pour suivre ce tutoriel, vous avez besoin d'extraire et d'installer les fichiers customer.csv et state.txt



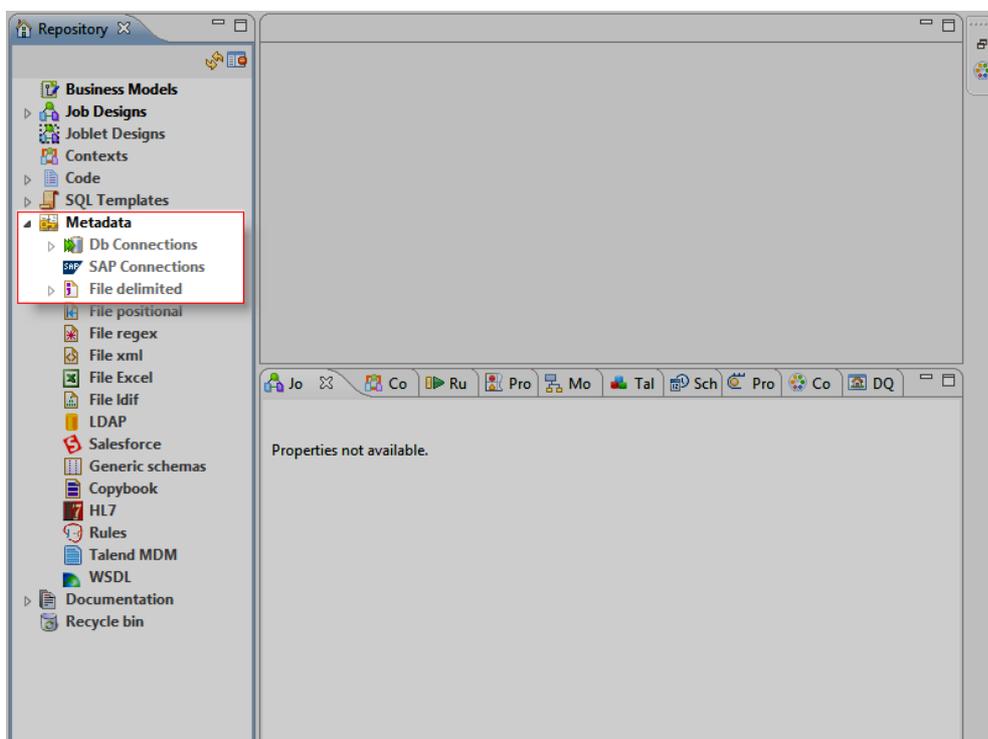
Créer la métadonnée *customers*

Dans le Repository sur la gauche de Talend Open Studio :

Développez le nœud Metadata.

Cliquez-droit sur File delimited.

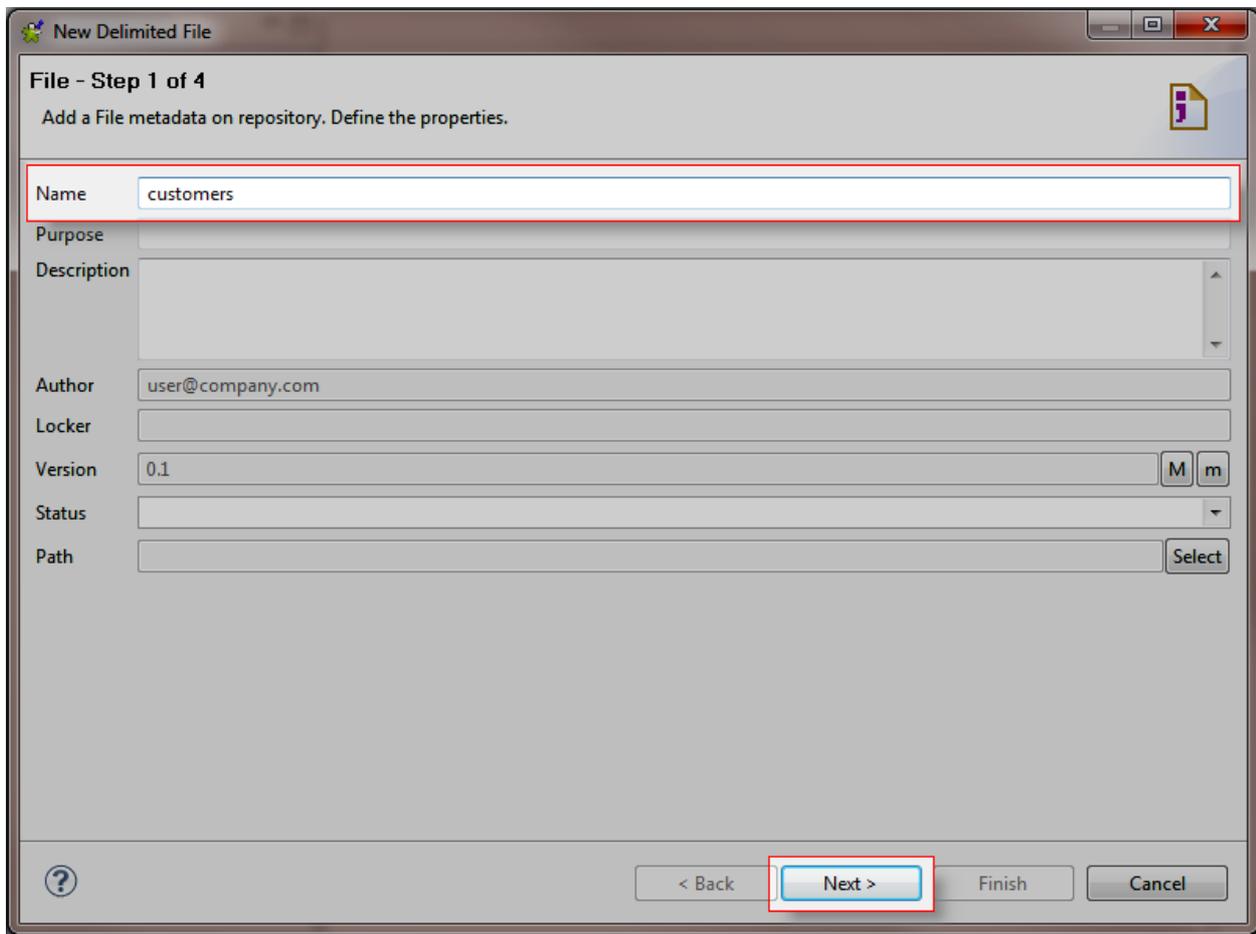
Dans le menu contextuel, cliquez sur Create file delimited. L'assistant New Delimited File s'ouvre.



Dans l'assistant New Delimited File :

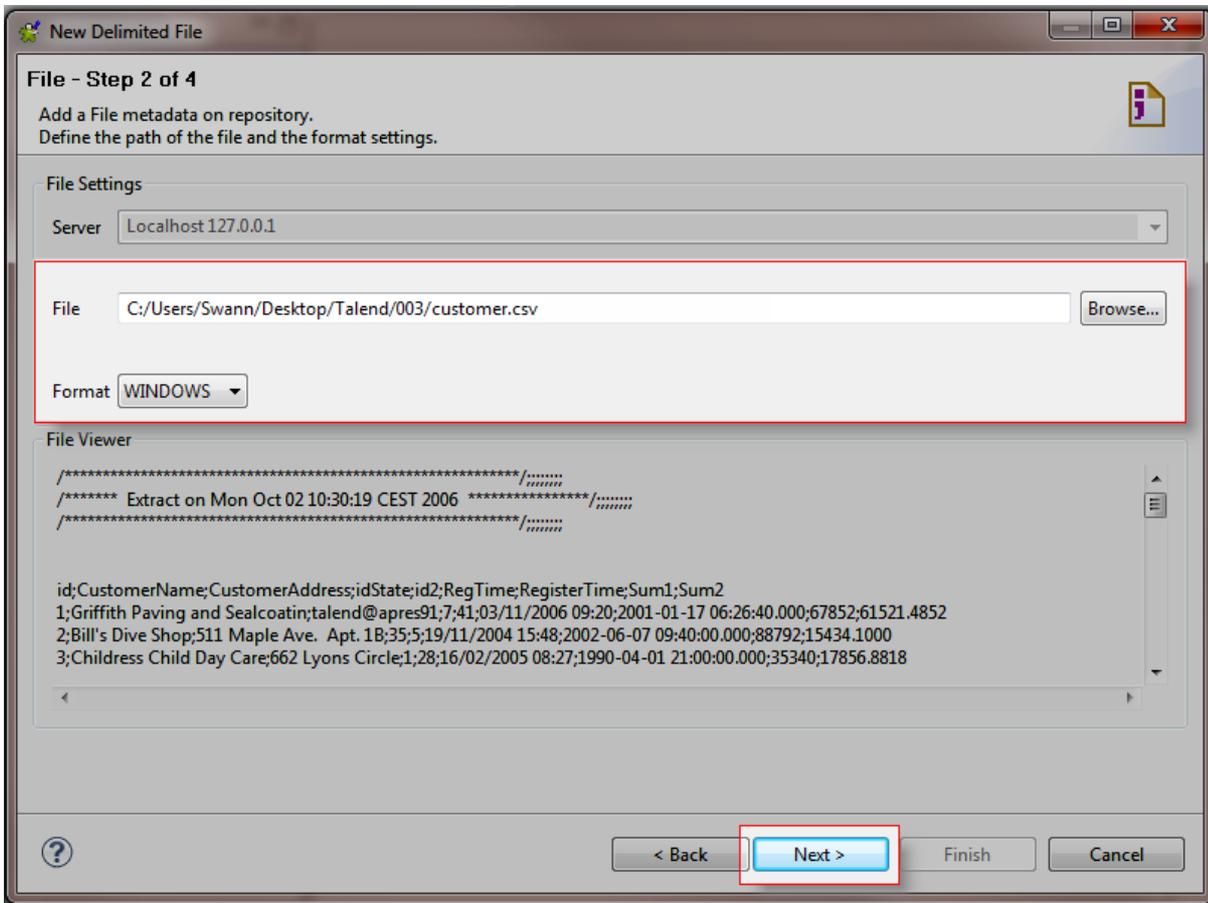
Dans le champ Name, saisissez le nom de la métadonnée : customers.

Cliquez sur Next pour continuer.



The screenshot shows a window titled "New Delimited File" with a subtitle "File - Step 1 of 4". Below the subtitle is the instruction "Add a File metadata on repository. Define the properties." The form contains several fields: "Name" (containing "customers"), "Purpose", "Description", "Author" (containing "user@company.com"), "Locker", "Version" (containing "0.1" with "M" and "m" buttons), "Status", and "Path" (with a "Select" button). At the bottom, there are four buttons: "< Back", "Next >", "Finish", and "Cancel". The "Name" field and the "Next >" button are highlighted with red boxes.

Cliquez sur le bouton Browse... et sélectionnez le fichier customer.csv dans l'assistant qui s'ouvre.
Dans la liste Format, sélectionnez le système d'exploitation de votre ordinateur.
Cliquez sur Next.



Dans la zone Rows To Skip, cochez la case Header et saisissez le nombre de lignes d'en-tête à ignorer : 5.

Dans la zone Preview en bas de l'assistant, cochez la case Set heading row as column names pour récupérer les noms des colonnes du fichier.

Cliquez sur le bouton Refresh Preview pour rafraîchir l'aperçu de la structure et des données du fichier.

Cliquez sur Next.



Lorsque vous cochez la case **Set heading row as column names**, le nombre que vous avez saisi dans le champ **Header** est incrémenté.

New Delimited File

File - Step 3 of 4
Add a File metadata on repository. Define the setting of the parsed job.

File Settings

Encoding: US-ASCII

Field Separator: Semicolon Corresponding Character: ";"

Row Separator: Standard EC Corresponding Character: "\n"

Escape Char Settings

CSV Delimited

Escape Char: Empty

Rows To Skip

If any rows must be ignored, specify the following parameters

Header 6

Footer

Limit Of Rows

If the number of lines must be limited, specify this number

Limit

Preview Output

Set heading row as column names Refresh Preview

id	CustomerName	CustomerAddress	idState	id2	RegTime	RegisterTime	Sum:
1	Griffith Paving and Sealcoat	talend@apres91	7	41	03/11/2006 09:20	2001-01-17 06:26:40.000	6785.
2	Bill's Dive Shop	511 Maple Ave. Apt. 1B	35	5	19/11/2004 15:48	2002-06-07 09:40:00.000	8879.

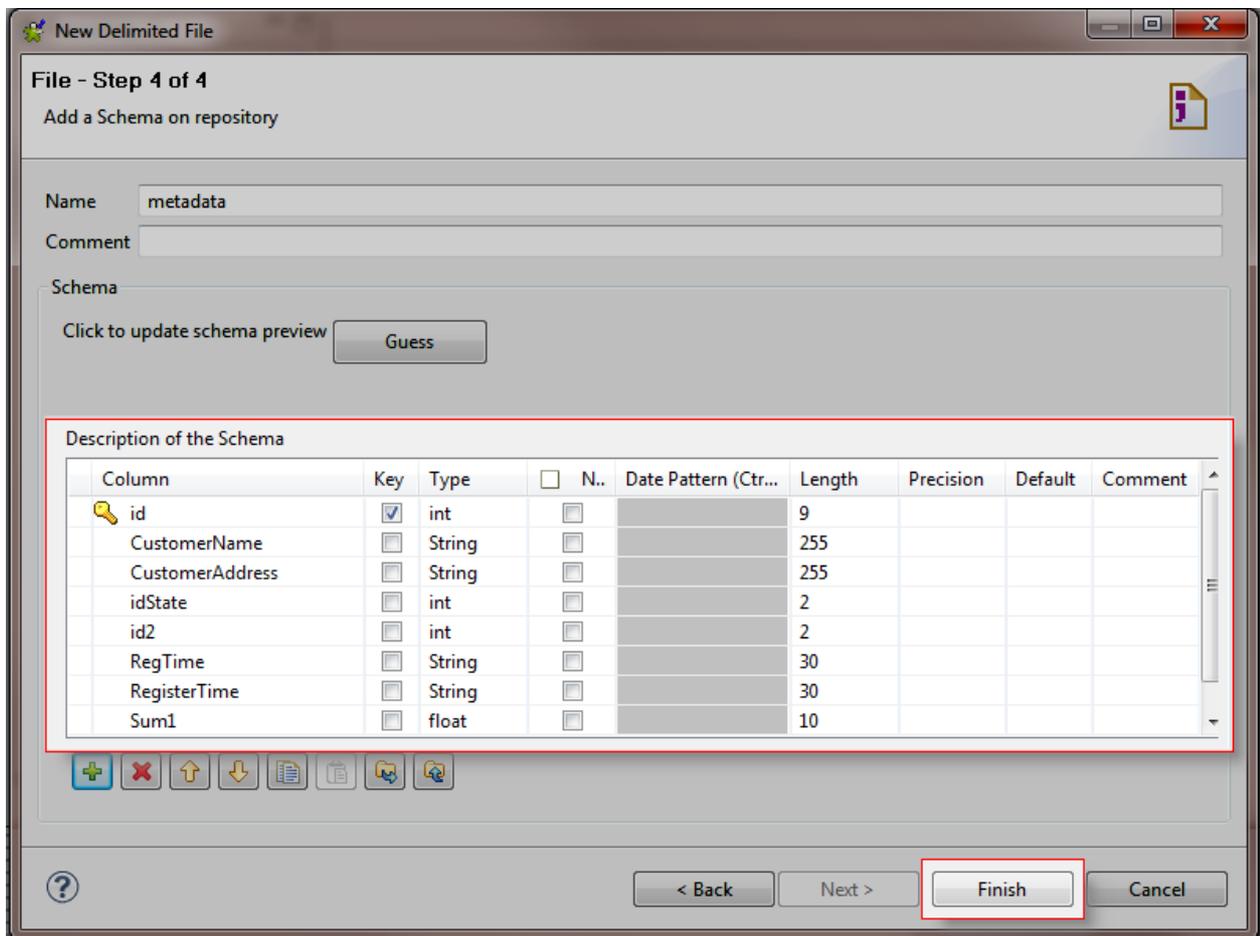
Export as Context Revert from Context

< Back **Next >** Finish Cancel

Dans le tableau Description of the schema, paramétrez la clé, le type et la longueur des colonnes comme indiqué sur la capture d'écran.

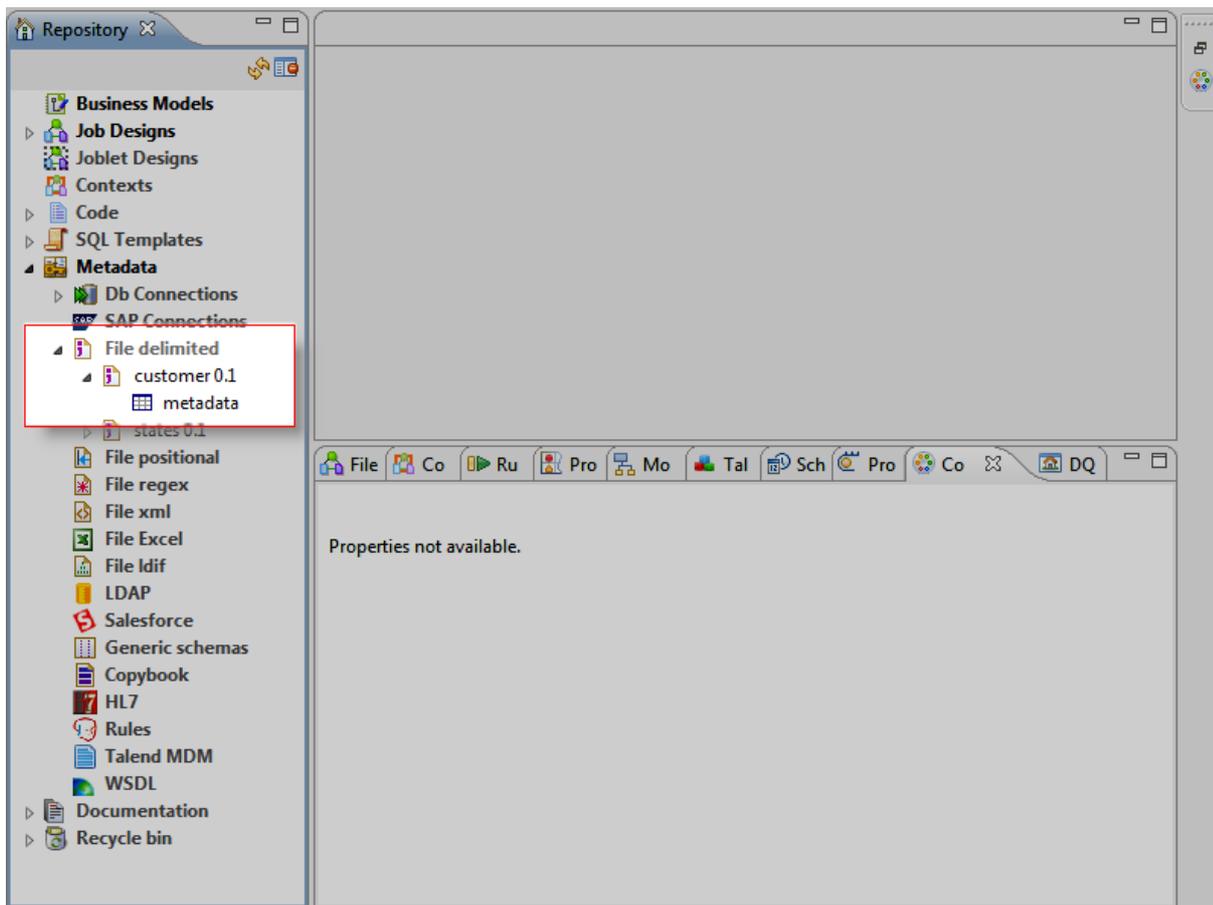
Les colonnes Sum1 et Sum2 sont paramétrées de la même manière.

Cliquez sur Finish pour fermer l'assistant.



Dans le Repository situé à gauche :

La métadonnée customers est affichée sous le nœud Metadata > File delimited.



La métadonnée customers est créée.

Maintenant, créez la métadonnée states.

2

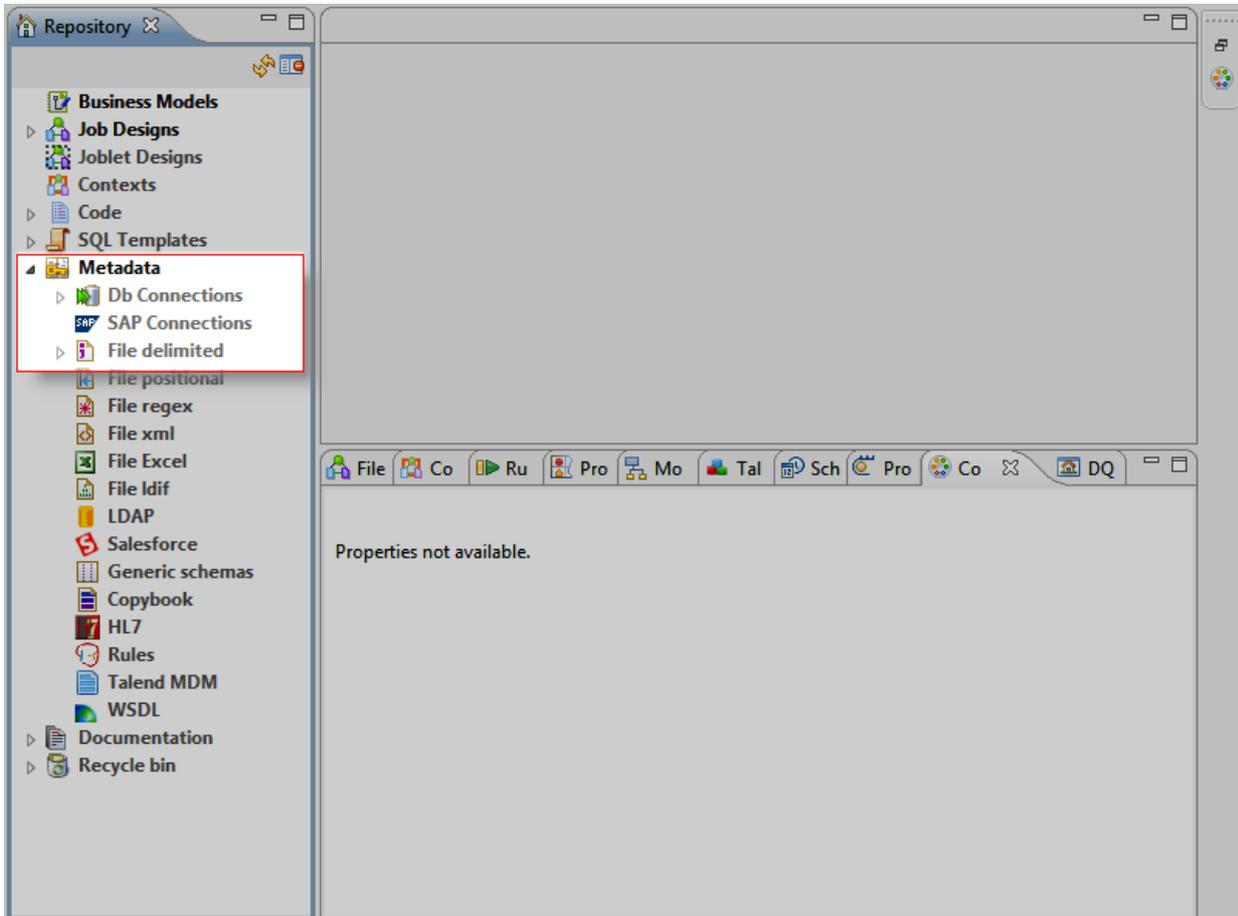
Créer la métadonnée *states*

Dans le Repository sur la gauche de Talend Open Studio :

Développez le nœud Metadata.

Cliquez-droit sur File delimited.

Dans le menu contextuel, cliquez sur **Create file delimited**. L'assistant **New Delimited File** s'ouvre.



Dans l'assistant New Delimited File :

Dans le champ Name, saisissez le nom de la métadonnée : states.

Cliquez sur **Next** pour continuer.

New Delimited File

File - Step 1 of 4

Add a File metadata on repository. Define the properties.

Name: states

Purpose:

Description:

Author: user@company.com

Locker:

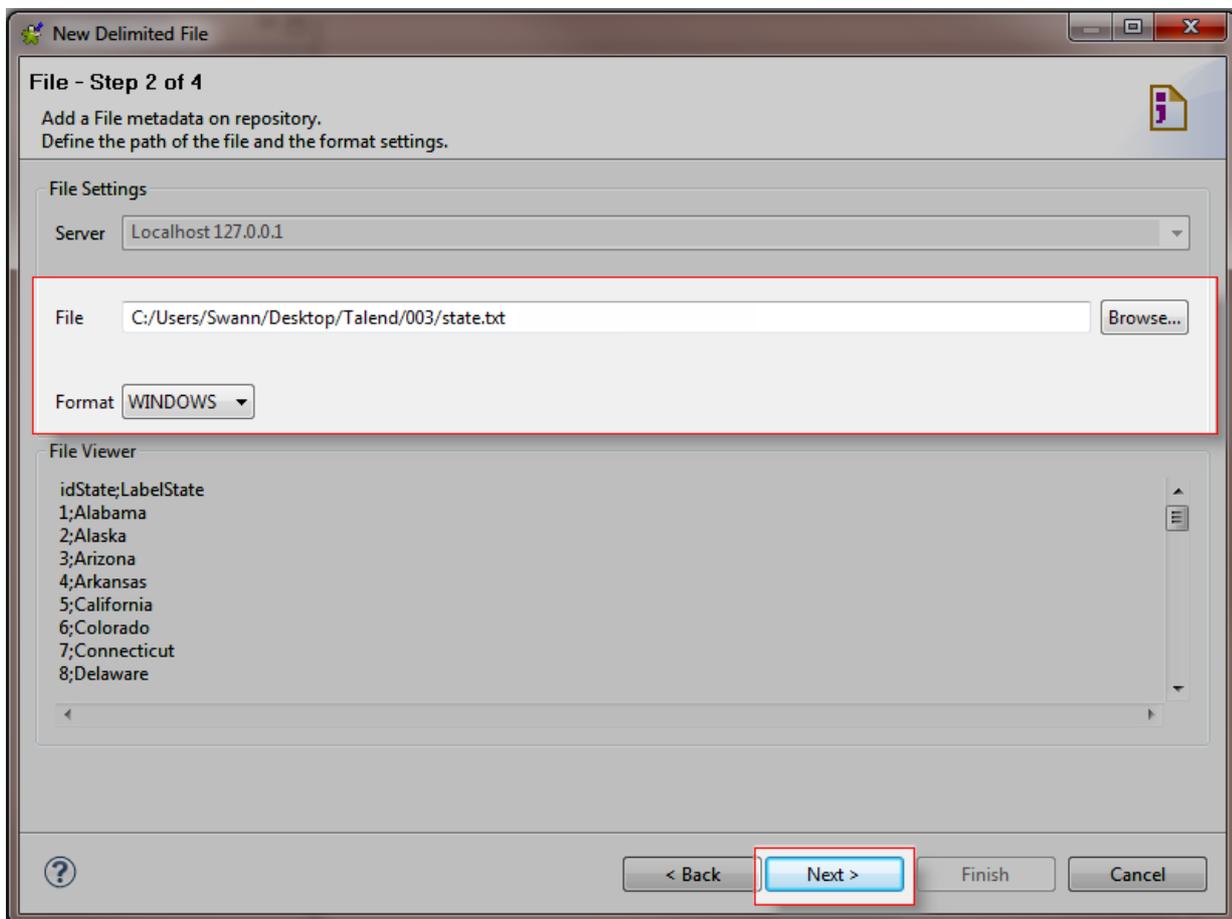
Version: 0.1 [M] [m]

Status:

Path: [Select]

< Back Next > Finish Cancel

Cliquez sur le bouton **Browse...** et sélectionnez le fichier `state.txt` dans l'assistant qui s'ouvre.
Dans la liste **Format**, sélectionnez le système d'exploitation de votre ordinateur.
Cliquez sur **Next**.



Dans la zone Preview en bas de l'assistant, cochez la case Set heading row as column names pour récupérer les noms des colonnes du fichier.

Cliquez sur le bouton Refresh Preview pour rafraîchir l'aperçu de la structure et des données du fichier.

Cliquez sur **Next**.



Lorsque vous cochez la case **Set heading row as column names**, le champ **Header** est incrémenté

File - Step 3 of 4
Add a File metadata on repository. Define the setting of the parsed job.

File Settings
Encoding: US-ASCII
Field Separator: Semicolon Corresponding Character: ";"
Row Separator: Standard EC Corresponding Character: "\n"

Escape Char Settings
 CSV Delimited

Rows To Skip
If any rows must be ignored, specify the following parameters:
Header: 1
Footer:

Limit Of Rows
If the number of lines must be limited, specify this number:
Limit:

Preview | Output

Set heading row as column names Refresh Preview

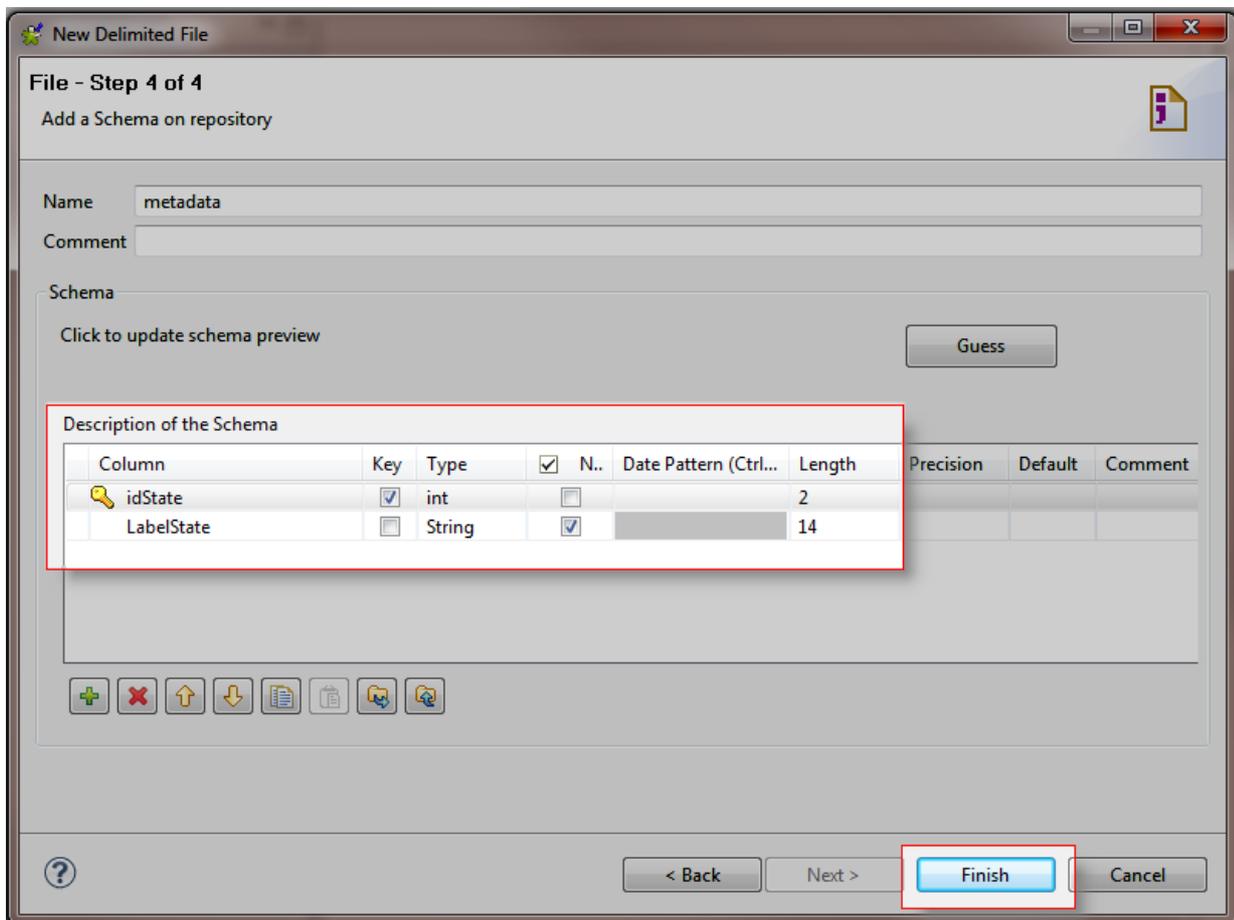
idState	LabelState
1	Alabama
2	Alaska

Export as Context Revert from Context

< Back **Next >** Finish Cancel

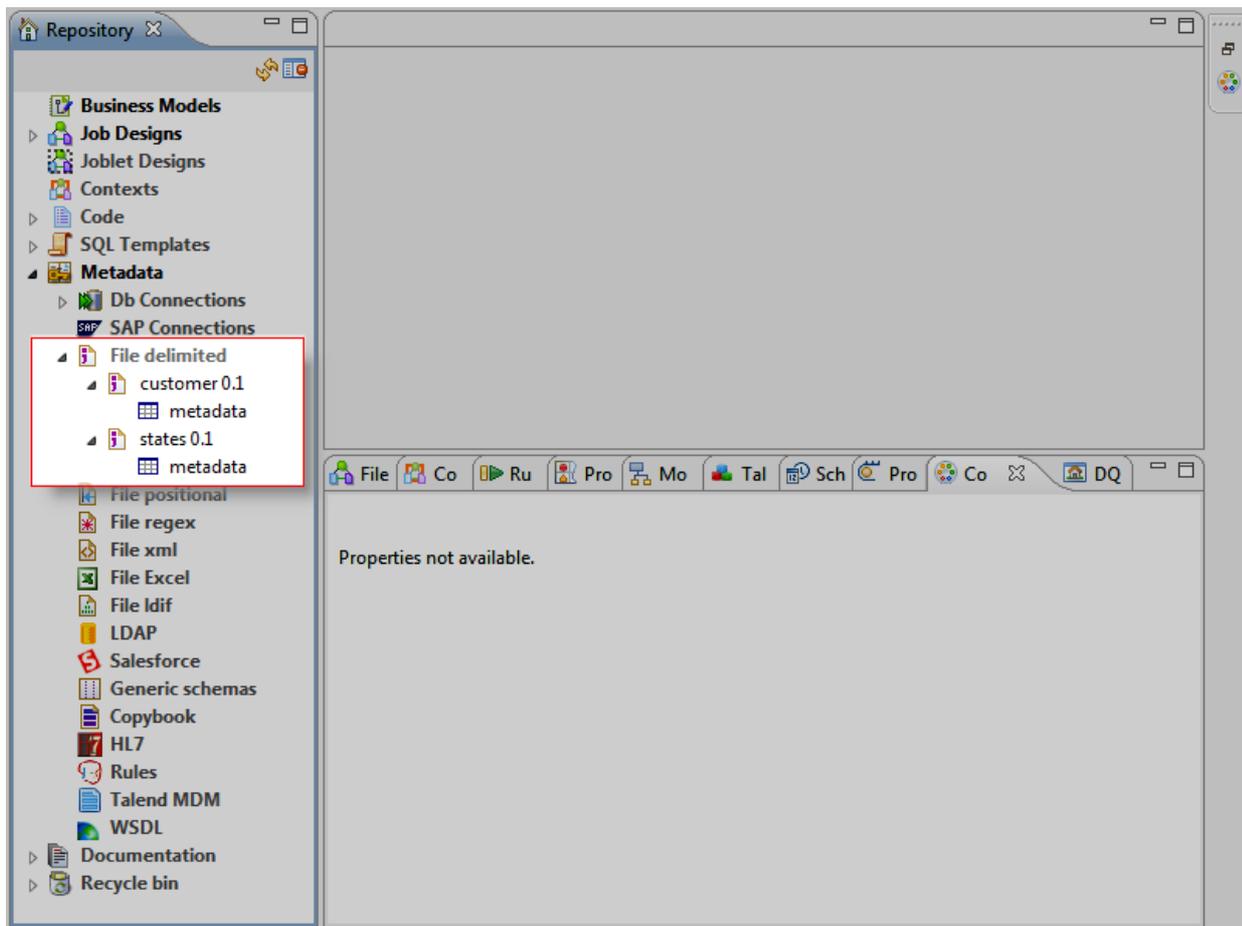
Dans le tableau Description of the schema, paramétrez la clé, le type et la longueur des colonnes comme indiqué sur la capture d'écran.

Cliquez sur **Finish** pour fermer l'assistant



Dans le Repository situé à gauche :

La métadonnée *states* s'affiche sous le nœud **Metadata**> **File delimited**.



Les deux métadonnées ont été créées. Vous pouvez maintenant les utiliser et réutiliser dans vos Jobs.