Unleashing the Potential: How Automated Metadata Creation Can Improve Your Data Quality and Accessibility

Elena Robu | Astun Technology

What is Metadata anyway? 🤔

What is Metadata anyway? 🤔

What is Metadata anyway? 🤔

In simple terms, metadata is data about data.

What is Q-FAIR all about? 🎡

What is Q-FAIR all about? 🎡

Quality

What is Q-FAIR all about? 🎡

Quality

- is just a hyphen

What is Q-FAIR all about? 🎡

Quality

Findable

- is just a hyphen

3.3

What is Q-FAIR all about? 🎡

Quality

Findable

- is just a hyphen

Accessible

What is Q-FAIR all about? 🎡

Quality

Findable

Interoperable

- is just a hyphen

Accessible

What is Q-FAIR all about? 🎡

Quality

Findable

Interoperable

- is just a hyphen

Accessible

Reusable

Background

Background

Why are we doing this? 🤷

Background

How do we do this now? 🕰️

Background

How do we do this now? 🕰️

Metadata Crawler open source plugin for Talend Open Studio (aka Crawler)

What are these fields? 📜

What are these fields? 📜

Abstract

Alternative Title

Bounding box

Constraints (usage and licensing)

Contact Details

Date

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

What are these fields? 📜

- Abstract
- Alternative Title
- Bounding box
- Constraints (usage and licensing)
- Contact Details
- Date
- Geographic Extent (keyword)
- INSPIRE Keywords (controlled)

- Keywords (free text or controlled)
- Lineage
- Spatial Reference System
- Resolution
- Temporal Extent
- Title
- Topic Category (controlled)
- Update Frequency

What are these fields? 📜

- Abstract
- Alternative Title
- Bounding box
- Constraints (usage and licensing)
- Contact Details
- Date
- Geographic Extent (keyword)
- INSPIRE Keywords (controlled)

- Keywords (free text or controlled)
- Lineage
- Spatial Reference System
- Resolution
- Temporal Extent
- Title
- Topic Category (controlled)
- Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Constraints (usage and licensing)

Contact Details

Date

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Lineage

Spatial Reference System

Resolution

Temporal Extent

Title

Topic Category (controlled)

Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Constraints (usage and licensing)

Contact Details

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Lineage

Spatial Reference System

Resolution

Temporal Extent

Title

Topic Category (controlled)

Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Constraints (usage and licensing)

Contact Details

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Lineage

Resolution

Temporal Extent

Title

Topic Category (controlled)

Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Constraints (usage and licensing)

Contact Details

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Lineage

Resolution

Temporal Extent

Topic Category (controlled)

Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Constraints (usage and licensing)

Contact Details

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Lineage

Resolution

Temporal Extent

Topic Category (controlled)

Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Contact Details

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Lineage

Resolution

Temporal Extent

Topic Category (controlled)

Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Lineage

Resolution

Temporal Extent

Topic Category (controlled)

Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Resolution

Temporal Extent

Topic Category (controlled)

Update Frequency

8.3

What are these fields? 📜

Abstract

Alternative Title

Keywords (free text or controlled)

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Temporal Extent

Topic Category (controlled)

Update Frequency

8.4

What are these fields? 📜

Abstract

Alternative Title

Keywords (free text or controlled)

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Topic Category (controlled)

Update Frequency

What are these fields? 📜

Abstract

Alternative Title

Keywords (free text or controlled)

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Topic Category (controlled)

What are these fields? 📜

Abstract

Alternative Title

Geographic Extent (keyword)

INSPIRE Keywords (controlled)

Keywords (free text or controlled)

Topic Category (controlled)

How do we automate these? 🧑‍💻

How do we automate these? 🧑‍💻

spaCy

How do we automate these? 🧑‍💻

spaCy

"industrial strength open source NLP in Python"

Geographic Keywords 🌍

Geographic Keywords 🌍

NER model trained on OS Open Names

Geographic Keywords 🌍

NER model trained on OS Open Names

Pattern recognition model (postcodes)

Non-Geographic Keywords 🏷️

Non-Geographic Keywords 🏷️

Non-Geographic Keywords 🏷️

Non-Geographic Keywords 🏷️

INSPIRE Keywords and Topic Categories 🚩

INSPIRE Keywords and Topic Categories 🏷️
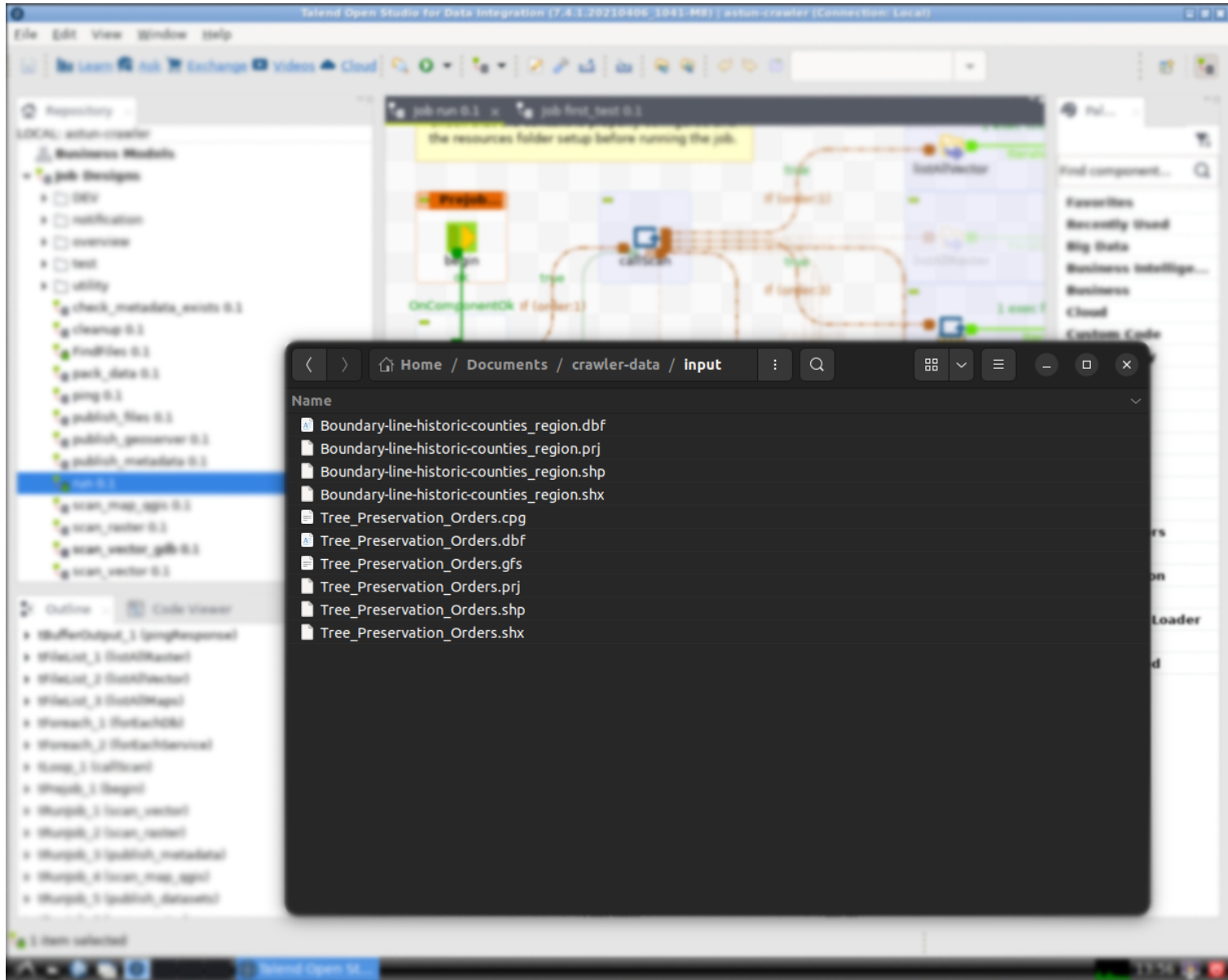
Text Classification

Everything Else

Everything Else

USE CHATGPT
FOR ALL THE THINGS!!

How will this work IRL? 🤔

Format   Data   Tools   Extensions   Help

change_time	organisation	inspire_keyword	topic_category	geographic_area	keywords	record_abstract	title
23-07-2013	Astun Technology	Transport networks	society	Aberdeen, Scotland	Aberdeen, Cults	On July 23, 2013, Astun Technology created a descriptive dataset on transport networks in Aberdeen, Scotland. The dataset focuses on the smallest geographic area that includes all locations in Aberdeen, including Cults and Tree.	Transport Networks in Aberdeen, Scotland: Cults Tree
23-07-2013	Astun Technology	Protected sites	society	United Kingdom	Historic County, East Lothian Historic County	This dataset provides information on protected sites in the United Kingdom, specifically focusing on East Lothian Historic County. Created on 23-07-2013, it includes data on the smallest geographic area that encompasses all the listed locations. The dataset also incorporates information on historic counties and boundary lines.	Protected Sites in the United Kingdom's East Lot Historic County

17

🔍 Back to search   ⬇ Download ▾   ◉ Display mode ▾

🗄 Transport Networks in Aberdeen, Scotland: Cults and Tree
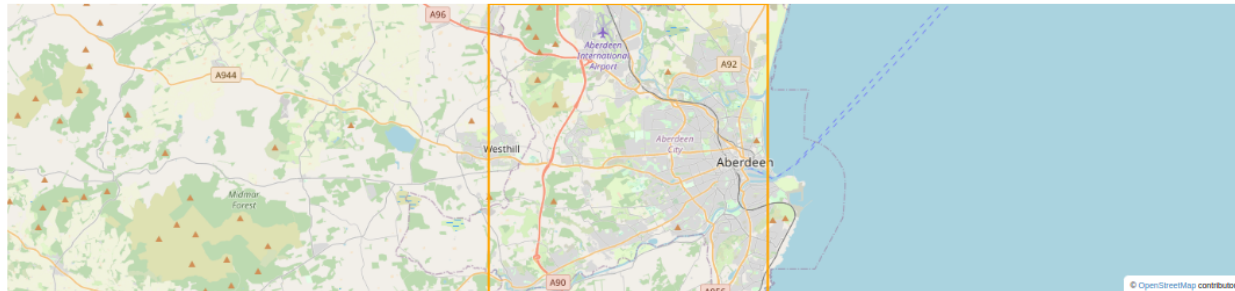
📅 **Temporal extent**
01-01-2000

On going

Also named: Tree_Preservation_Orders
On July 23, 2013, Astun Technology created a descriptive dataset on transport networks in Aberdeen, Scotland. The dataset focuses on the smallest geographic area that includes all locations in Aberdeen, including Cults and Tree.

Discover data

© OpenStreetMap contributors.

Download

🗄 **Description**

This dataset is published in the database **"/data/crawler/input/Tree_Preservation_Orders.shp"** in table **"Description"**.

🌐 **Format**

ESRI Shapefile

18

Questions? 🙋

🔍Photo by Hans-Jurgen Mager on Unsplash

Thank you! 😊

𝕏 **@AlexaSmile15** in **elenarobu**

✉ **elenarobu@astuntechnology.com**

astuntechnology.com/contact-us