

Highlighter

Browser plugin that highlights entities on webpage

URL, language

Entities

News Entity Server

Flask API service that extracts entities from webpages

mediacloud-metadata

Date extractor

spacy

Age extractor

Add entities to
relevant stories

Email Alerts Server

Ruby-on-Rails app that manages projects, groups articles, and delivers results to users

Project configuration JSON

Relevant story metadata

Email

Daily project
reports

Web UI

Story
Grouper

Story Processor

Python code that ingests candidate stories, classifies them against models, and sends relevant stories to email alerts server.

Classifier
Worker

Story Queue
RabbitMQ

stories

Fetchers
run via cron

mediacloud-metadata

URL fetches

summaries

Email

Slack

story metadata

stories

Alerts DB
Postgres

Story DB
Postgres

Story data

Nightly backups

Binary models

Story Processor Dashboard

Streamlit app showing data visualizations to support system-level data debugging

Web UI

S3

Offsite backups and file exchange

DB backups

Trained Models

Manual Model Training

Jupyter notebook training processes

Wayback Machine

Media Cloud

newscatcher