CASE STUDY

Newsroom:

El Vocero de Puerto Rico

About the newsroom

El Vocero de Puerto Rico is a newspaper serving all of Puerto Rico. It is owned by Publi-Inversiones de Puerto Rico. The paper has a robust digital presence. During hurricane season, the staff keep a watchful eye on the National Hurricane Center (NHC) for developing storms in the Atlantic Ocean. When a strong tropical system is headed toward Puerto Rico, the team delivers updates on their digital platforms as soon as they're published by the NHC.

Project goals

The newspaper wants to publish hurricane updates as quickly as possible on its digital platforms. They proposed a project to automatically deliver weather alerts issued by the NHC or NWS into their BLOX CMS.

Why is this project important to the newsroom?

The faster an update can be published, the higher the chances potentially lifesaving information can be acted on by Puerto Rico residents.

The project was pitched to replace a manual process which involves a journalist sitting on the NHC webpage and constantly refreshing for updates every three hours. The NHC forecast public advisories are published first in English. A team from the San Juan office of the National Weather Service (NWS) translates those updates into Spanish when they are able to. To reduce the delay, the El Vocero team translates the content themselves and rewrites it into a brief story and alert.

Engineering process

The project was split into three segments: Parsing the NHC and NWS data feeds; crafting a brief story in Spanish; and integrating multiple components.

In the first segment, different parsers were developed to handle the NHC and NWS data feeds as they publish their content differently.

For the second segment, El Vocero's team provided story templates to be filled in for various weather situations, such as the issuance of a Hurricane Warning or Tropical Storm Warning by the NHC. A third-party AI translation service, DeepL, converts narratives provided by the meteorologists into Spanish from English. Key data elements from the bulletin are extracted and then input into the templates created.

In the third segment, the developed system integrated the incoming weather feeds, with story generation and translation, along with delivery into the BLOX CMS.

Were the goals met?

The goal of automatically writing publishable weather alerts and delivering them into the CMS was met. As designed, the stories were meant to be reviewed by a journalist before publication. However, El Vocero is considering bypassing the human review for the sake of getting critical information to audiences even faster.

Major challenges

The data feeds from NHC come in a substantially unstructured format, which led to a lot of development time to be spent on the parser. Additionally, it was difficult getting test data from the NHC and NWS to use for the project.

Future work

There is no user interface. Creating a UI to maintain settings such as the Spanish language glossary, and other operational configurations, would potentially make the tool easier to adopt.

Link to repository

github.com/associatedpress/local-ai-el-vocero

STAKEHOLDER REACTION

"We are very pleased with the results of the Weather Bot created for El Vocero that will be soon available as an open source to every newsroom that needs it. Its capability of producing a complete news article in just seconds will definitely complement the work carried out by our digital newsroom. This will allow us to be more strategic and efficient when allocating resources and assigning reporters to work on specific weather news updates after the AI generated content is published."

RAFELLI GONZALEZ COTTO
DIRECTOR DE MULTIPLATAFORMA DIGITAL

Development team

This project was led by the Northwestern University Medill Knight Lab under the guidance of Professor Jeremy Gilbert.

For El Vocero: Juan Muniz, Carlos Otero, Rafelli Gonzalez, Hector Vazquez, Maria Arce

For Knight Lab: Jeremy Gilbert, Joe Germuska, Scott Bradley, Mame Coumba Ka, Maria Aragon

For The Associated Press: Aimee Rinehart, Ernest Kung

Core components of the system

Input: National Weather Service, National Hurricane Center watches and warnings
Code: Python, HTML
Integrations: NWS via API, NHC via XML, DeepL via API, BLOX via API
Output: BLOX CMS, Email
Hosting: Amazon Web Services

Appendix

Figure 1. Process map

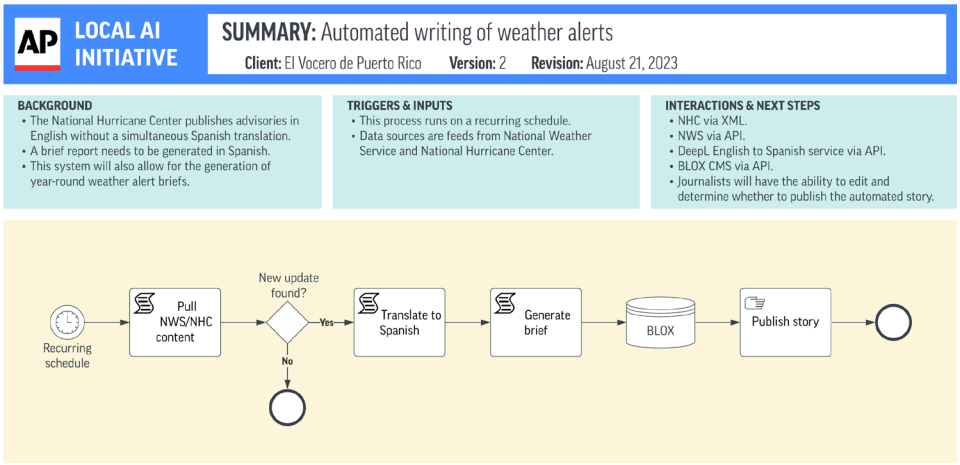


Figure 2. Story draft in CMS

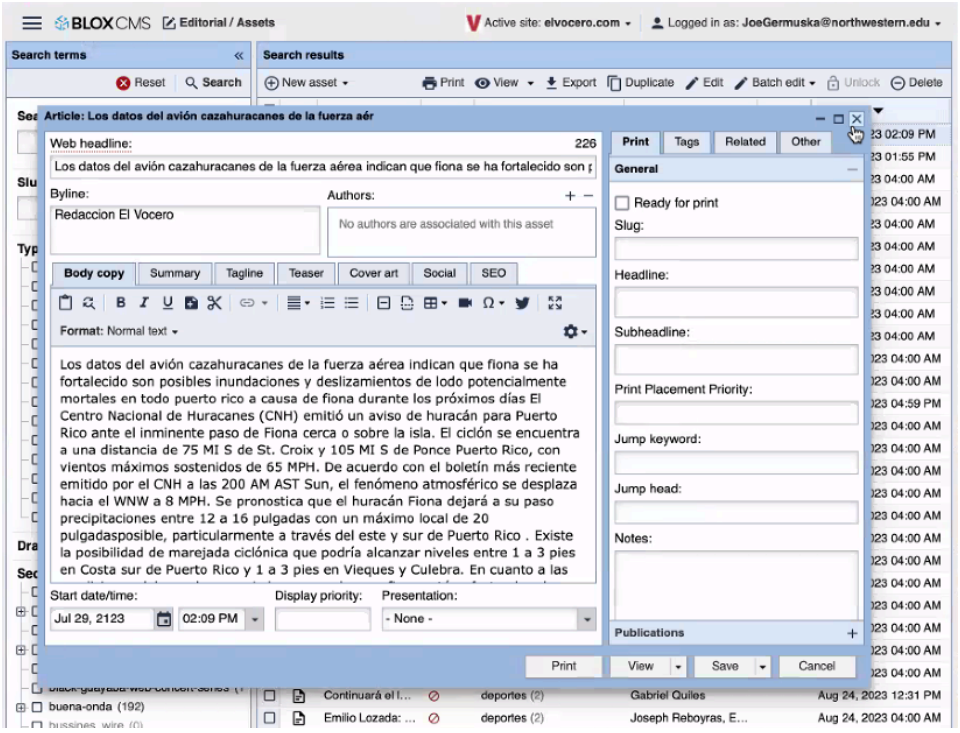
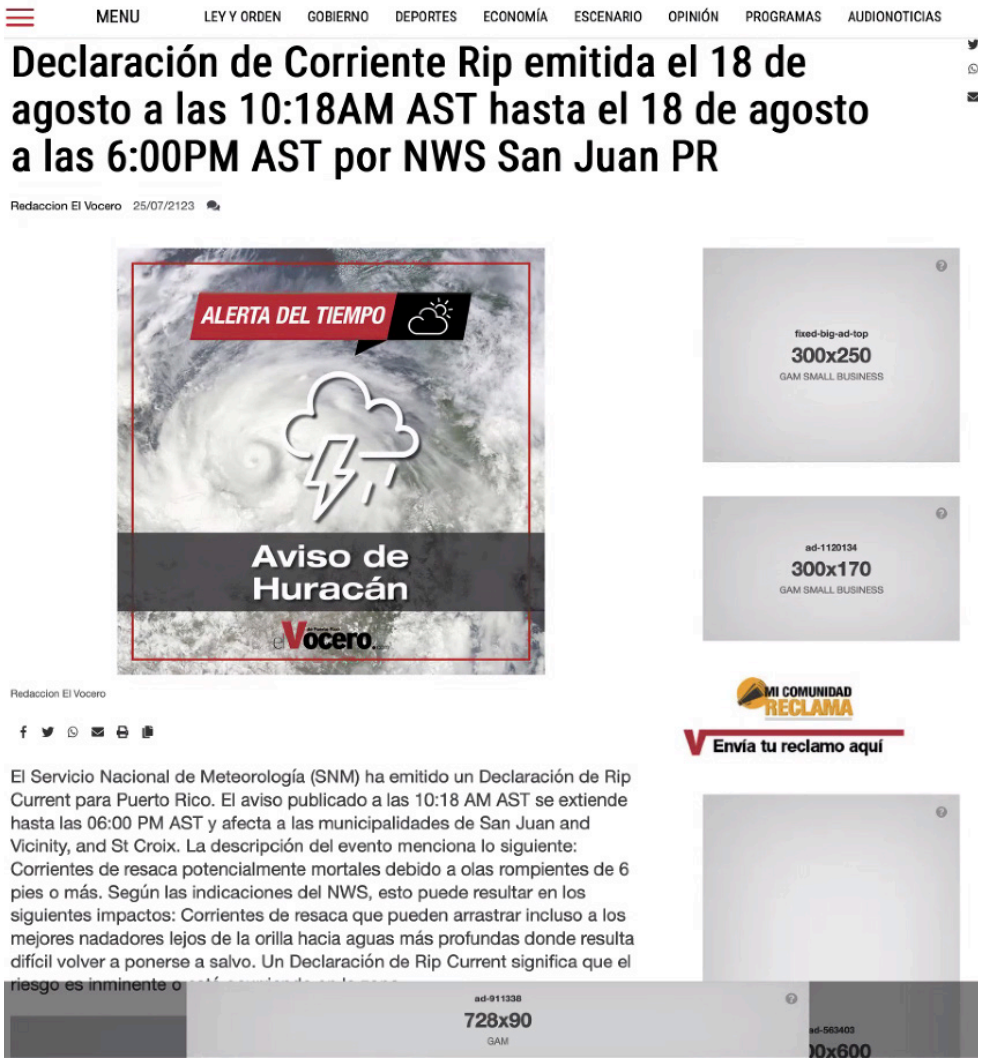


Figure 3. CMS preview with test image



THE ASSOCIATED PRESS

For more information, go to ap.org/ai