



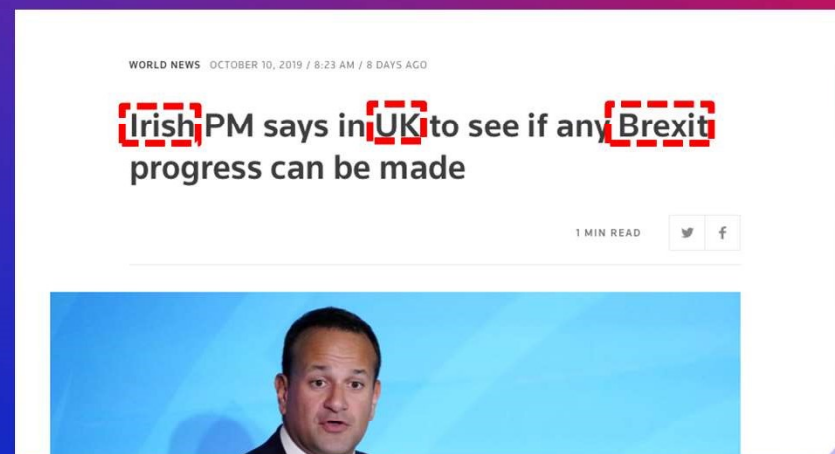
Unstructured Data Analysis in ArcGIS

Colby Neeb, Megan Banaski, Aryle Butler

2021 ESRI FEDERAL GIS CONFERENCE

What is Unstructured Data?

- Does not have a recognizable structure or is loosely structured
- Can be in a variety of formats and storage mechanisms
 - Word Documents
 - Email
 - Social Media Posts
 - PowerPoint
 - PDF
 - Share drive



**“Every two days we create as much
information as we did up to 2003”**

Eric Schmidt, 2010



What does that look like?

Every minute...

Twitter sees new 350,000 tweets



144 million e-mails are sent



15.2 million Text Messages are sent



Facebook has 510,000 comments posted,
293,000 statuses updated



954,000 new Microsoft Office documents
are created





Coordinates
Custom Locations
User defined keywords

What are you looking for?



ArcGIS Pro w/
LocateXT



ArcGIS Pro for
Intelligence



ArcGIS Enterprise w/
LocateXT

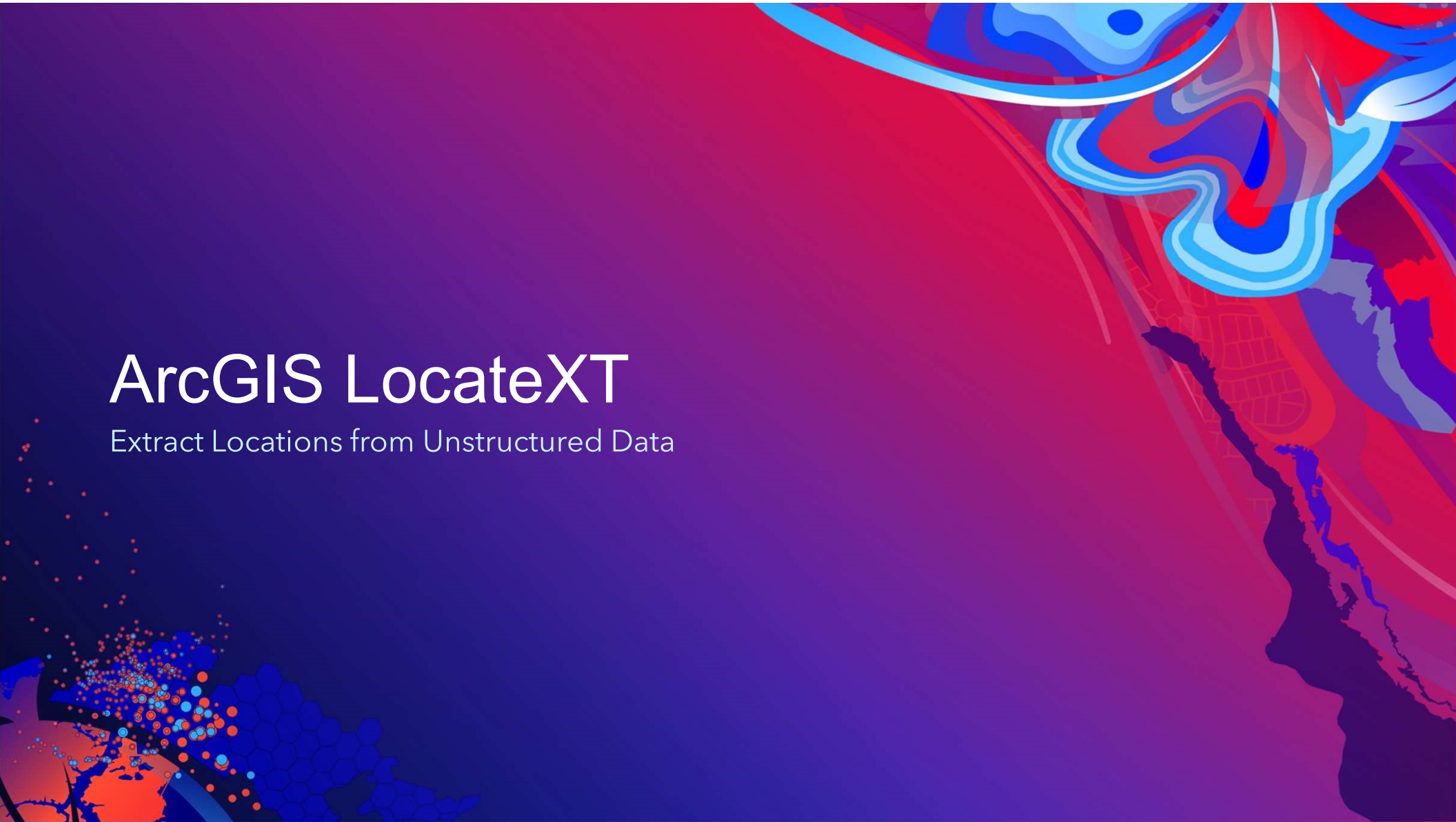
What is the best tool?

- Data is at least somewhat understood
- Data benefits from identifiable and repeating patterns
- Little to no programming experience available/needed

How is it best used?

ArcGIS LocateXT

Extract Locations from Unstructured Data



Extracting Locations with ArcGIS

- LocateXT Extension for ArcGIS Desktop and Enterprise
- Available for ArcMap 9.1 and later
- Available in ArcGIS Pro at 2.3
- 100% Feature function as ArcGIS Pro 2.4
- Uses pattern matching regular expressions (REGEX) to search for coordinates in a variety of formats
- Uses custom location list to match/extract other patterns (place names, codes, other terms)

2.18 Erha Oil Terminal (5°21'N., 4°20'E.), a deep-water facility, lies in the vicinity of the Erha Oil Field and is situated within an exclusion zone bordered by a line joining the following positions:


- 5°21.7'N, 4°17.7'E.
- 5°23.9'N, 4°20.4'E.
- 5°23.9'N, 4°22.5'E.
- 5°21.7'N, 4°23.6'E.
- 5°18.0'N, 4°21.5'E.
- 5°19.0'N, 4°17.7'E.

2.18 The terminal consists of a Floating Production Storage and Off loading (FPSO) vessel and an SBM. The SBM is moored 1 mile SE of the FPSO and is connected to it by two steel pipelines

At 1547Z, one adult female and two children ran from one house IVO 38PMB1143097553 to the pickup and then back into the same house.

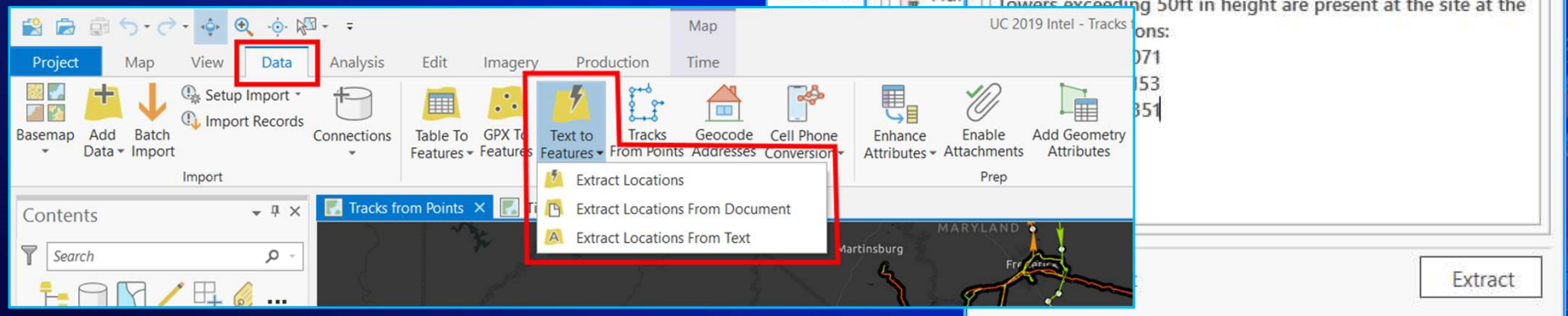
At 1732Z, two adult males with long guns ran from the underpass into thick brush north of the underpass IVO 38PMB1091196278.

At 1849Z, three adult males were seen covering object with debris on side of road on Highway 101 38PMB1230896427.

| | | | |
|--|-------------------------------|---------------------------------|--|
|  DEPARTURE: WEDNESDAY 12 SEP Please verify flight times prior to departure | | | |
| UNITED AIRLINES UA 5366 Operated by: /SKYWEST DBA UNITED EXPRESS Duration: 1hr(s) 27min(s) | ONT ONTARIO, CA | SFO SAN FRANCISCO, CA | Aircraft: CRJ-CANADAIR REGIONAL JET Distance (in Miles): 363 Stop(s): 0 |
| | Departing At: 19:35 | Arriving At: 21:02 | |
| | Terminal: TERMINAL 2 | Terminal: TERMINAL 3 | |

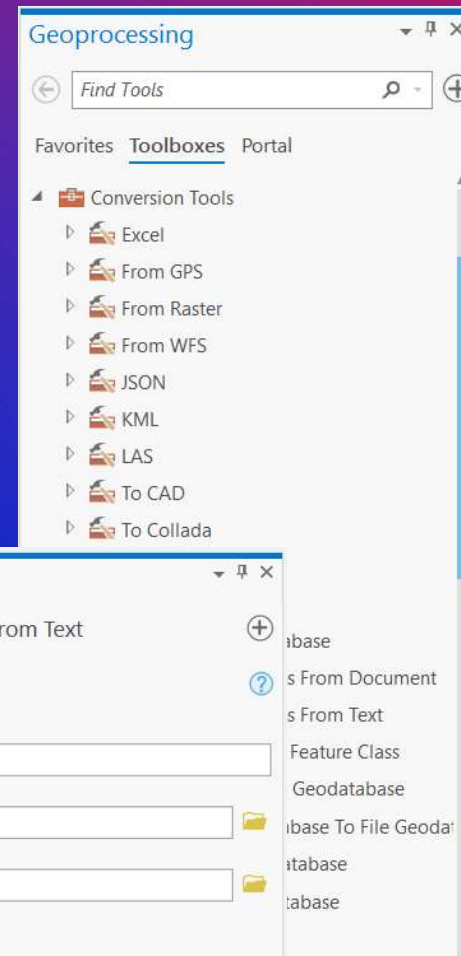
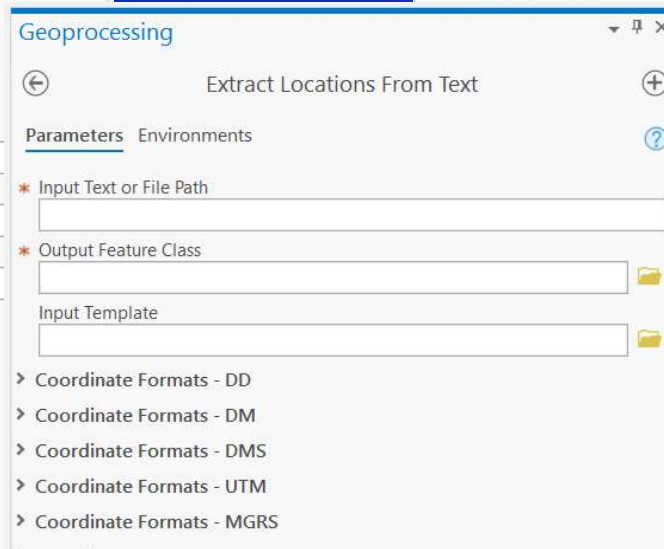
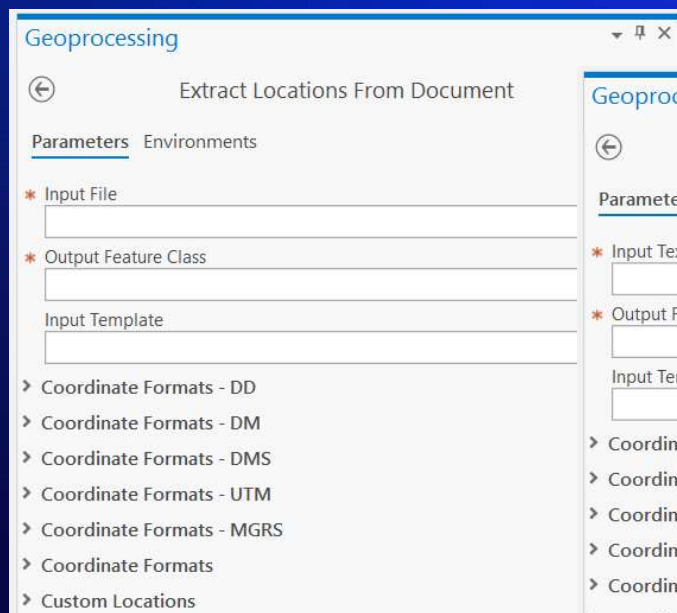
Extracting Locations in ArcGIS Pro

- New option added to the “Add Data” button
- Allows for a user to drag and drop documents or copied text into a window
- Can create a new feature class or append it to an existing one
- Included with ArcGIS Pro for Intelligence



Extracting Locations in ArcGIS Pro

- Two Geoprocessing Tools added
- Located in the **Conversion Tools** -> **To Geodatabase**
 - Extract Locations from Document
 - Extract Locations from Text



Extracting Custom Attributes

- Ability to create custom attributes based on content within document or near a location
 - Triggered by location extraction
- Based on keywords
 - Tag locations based on keywords
 - Scrape/harvest portions of document based on keywords
- Ability to extract based off of:
 - Number of characters/words
 - Number of lines/blank line
 - Stop string
- Previously built in separate LocateXT desktop application (until Pro 2.4)

Custom Attribute File

New File

Attributes

No attributes, add new, drop or import from another file.

Attribute Information

All active fields are required

| Attribute Name | Field Name | Field Length |
|----------------|------------|--------------|
| IED | IED | 30 |

Search Options

Type: Near locations

Characters Before: 60

Characters After: 0

Matches: Keep all

Keywords

- *IED Capture only keyword
- *VBIED Capture only keyword
- *Explosive Device Capture only keyword

New keyword

☐ Case Sensitive

☐ Include in Capture

Capture Options

Capture Type: Capture only keyword

Number: 1

☐ Case Sensitive

☐ Include in Capture

Add Keyword Cancel

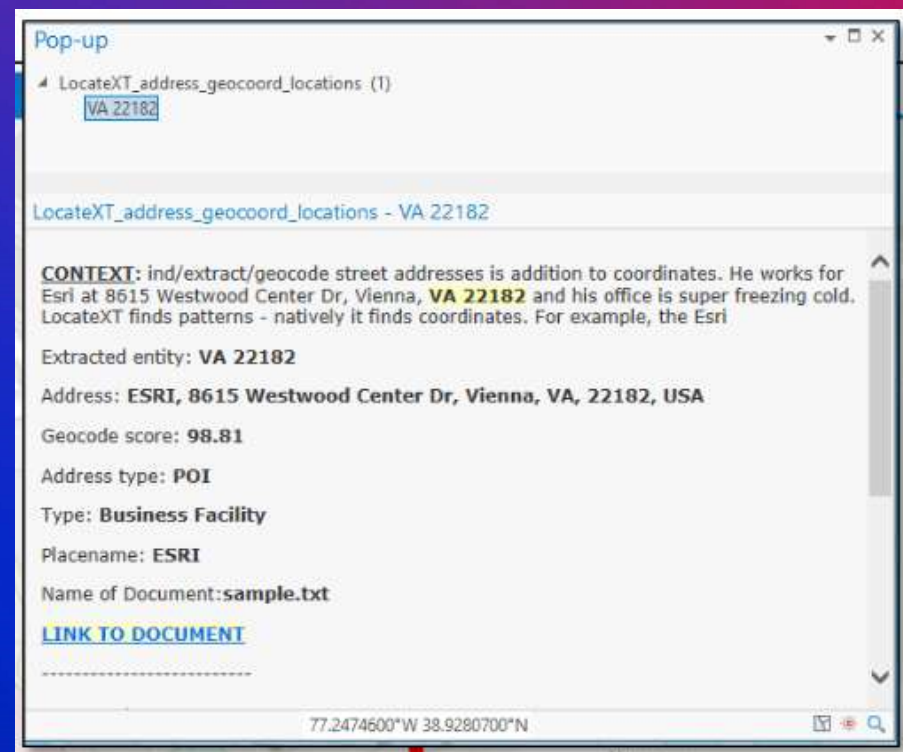
Add Attribute Cancel

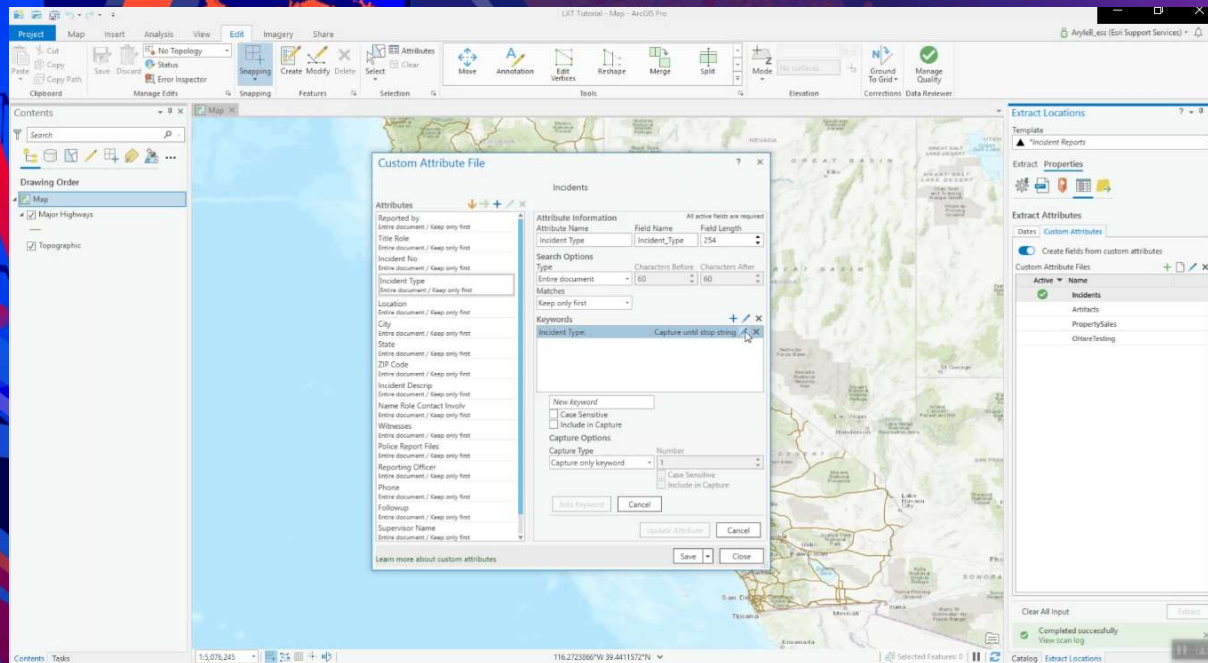
Learn more about custom attributes

Save Close

Extracting Addresses

- Ability to extract addresses from documents based on combination of:
 - State
 - Zipcode
 - Ex. **VA 22182**
- The combination of **extracted text** and **pre-text** is geocoded



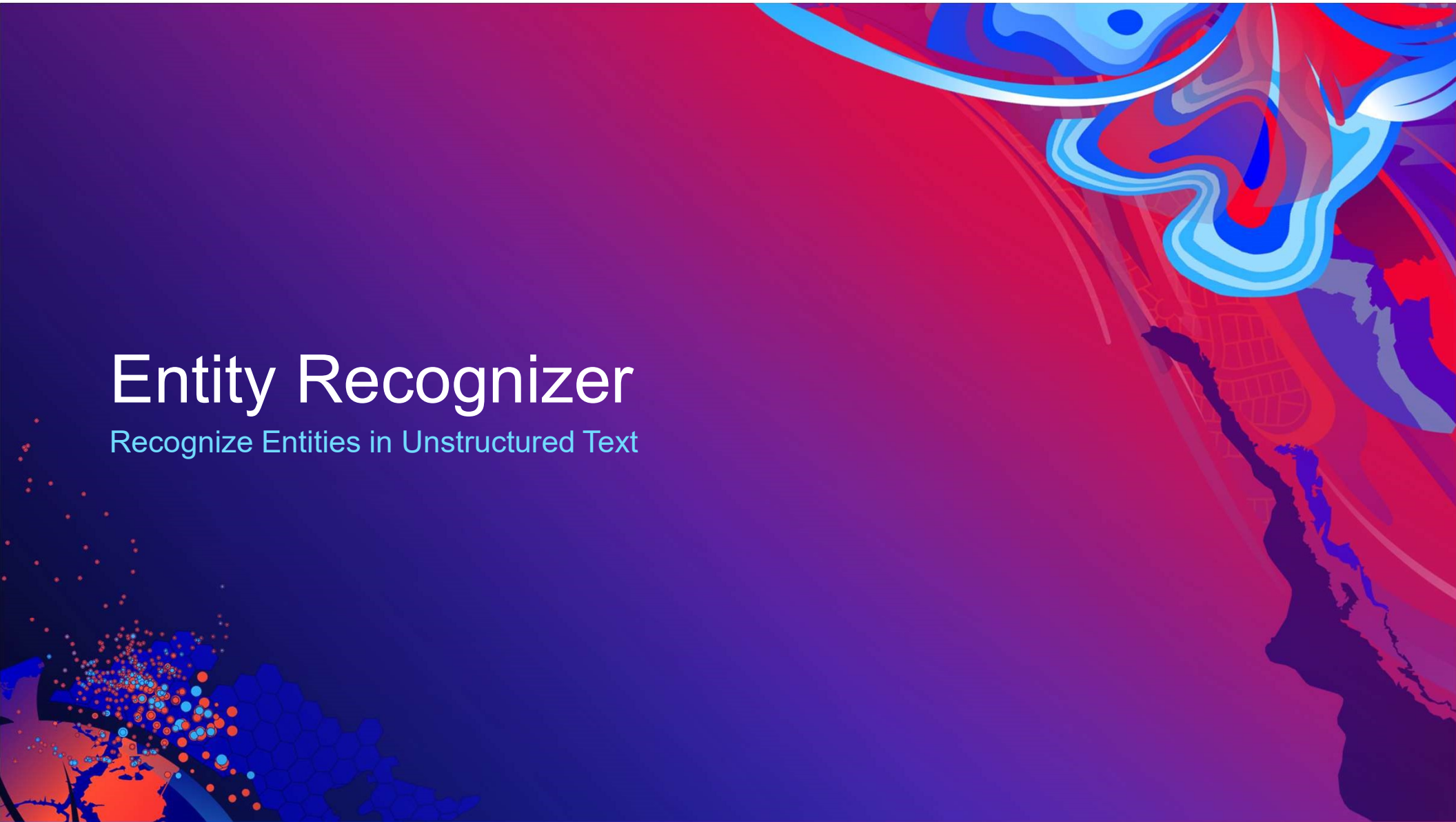


Police Reports in California

LocateXT

Entity Recognizer

Recognize Entities in Unstructured Text





Coordinates
Custom Locations
User defined keywords

What are you looking for?

Locations
People/Organizations
Events
Dates
Relationships



ArcGIS Pro w/
LocateXT



ArcGIS Pro for
Intelligence



ArcGIS Enterprise w/
LocateXT

What is the best tool?



ArcGIS API for Python
Entity Recognizer

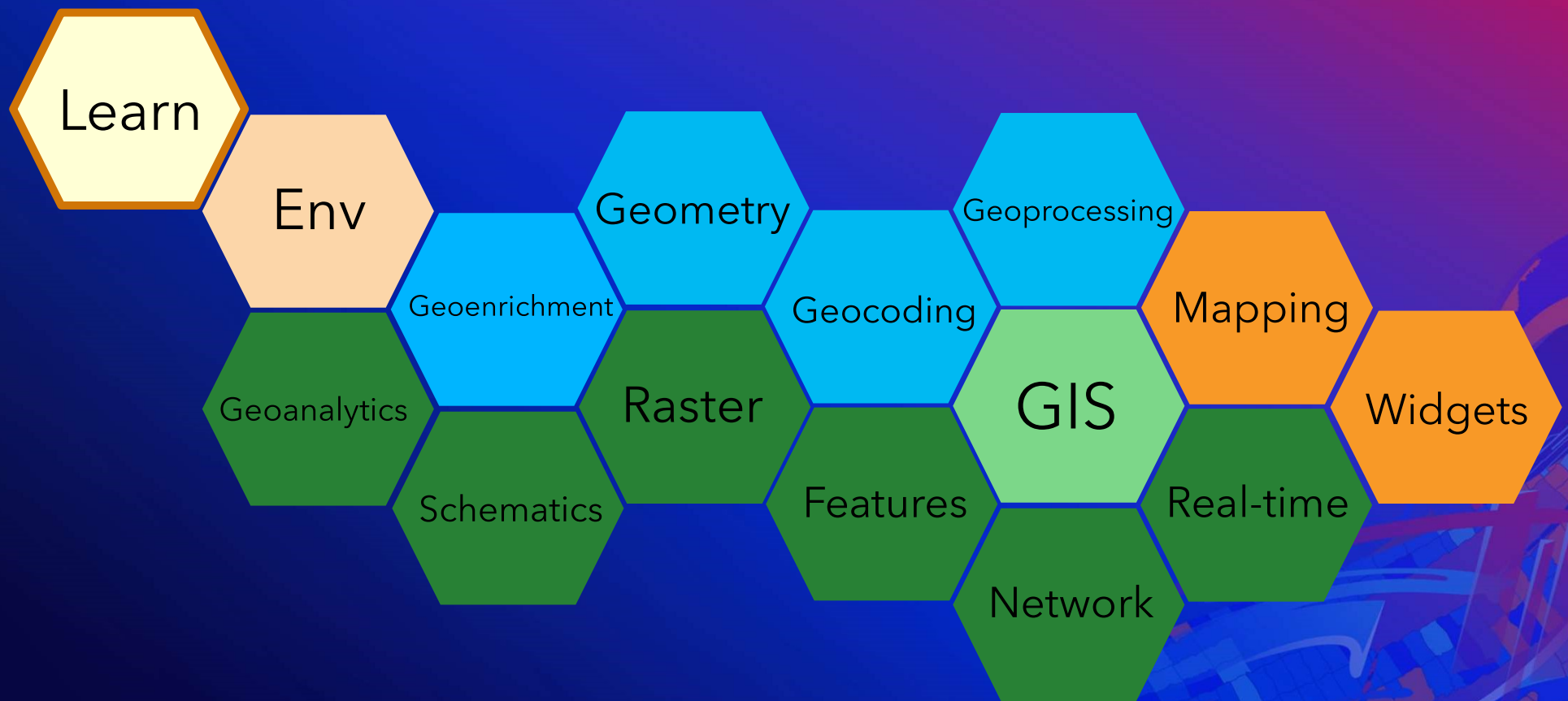
- Data is at least somewhat understood
- Data benefits from identifiable and repeating patterns
- Little to no programming experience available/needed

How is it best used?

- Data is not well understood
- Data does not contain identifiable and/or repeating patterns
- Scripting needed for desired results

ArcGIS API for Python

A powerful Python library for spatial analysis, mapping, and GIS



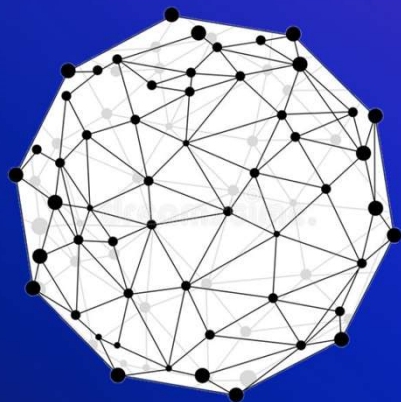


Artificial Intelligence

The diagram consists of three concentric circles. The outermost circle is labeled 'Artificial Intelligence'. Inside it is a circle labeled 'Machine Learning'. Inside the 'Machine Learning' circle is the innermost circle labeled 'Deep Learning'. The background is a gradient from dark blue on the left to red on the right, with a stylized graphic of a globe and a city skyline in the bottom right corner.

Machine Learning

Deep Learning



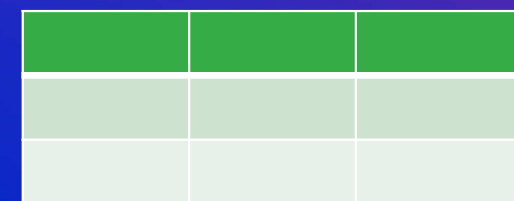
Trained EntityRecognizer Model

| | |
|-----------|------------|
| 2015_01_1 | 2015_01_16 |
| 2015_01_2 | 2015_01_17 |
| 2015_01_3 | 2015_01_18 |
| 2015_01_4 | 2015_01_19 |
| 2015_01_5 | 2015_01_20 |
| 2015_01_6 | 2015_01_21 |
| 2015_01_7 | 2015_01_22 |
| 2015_01_8 | 2015_01_23 |

Unstructured Text Documents



ArcGIS API for Python



Structured Data

Entity Recognizer - Fire Reports - Fires_Month 2015-2019

Edit View Insert Cell Help ArcGISPro

Recognizing Entities from Fire Reports in Cheshire, UK

Labelling Reports for Model Training

Using Doccano, we can label a text document with the relevant entities we would like to extract

Car fire in Sandbach Time of Incident: 30/01/2015 - 14:09
Location: Cresses Road, Sandbach Attendance: One fire engine from Sandbach A fire and rescue crew was alerted to a car fire in Sandbach. Firefighters, two wearing breathing apparatus, used two hose reel jets to put out the blaze, then made the scene safe.

Time of Incident: 30/01/2015 - 14:09
Location: Cresses Road, Sandbach Attendance: One fire engine from Sandbach A fire and rescue crew was alerted to a car fire in Sandbach. Firefighters, two wearing breathing apparatus, used two hose reel jets to put out the blaze, then made the scene safe.

Import necessary libraries [...]

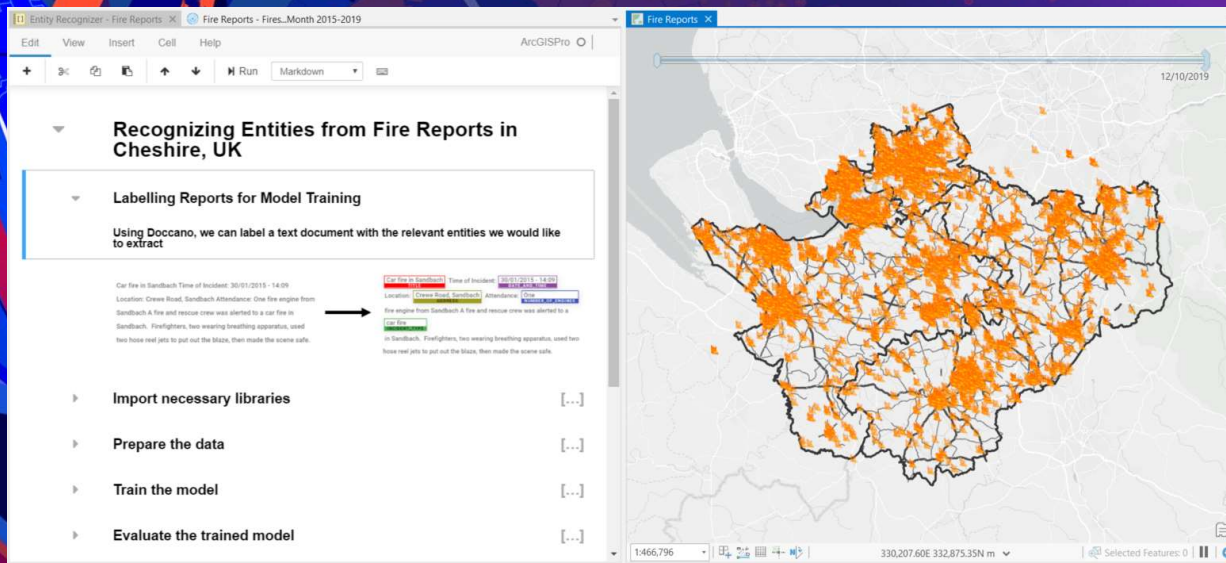
Prepare the data [...]

Train the model [...]

Evaluate the trained model [...]

1:466,796 330,207.60E 332,875.35N m Selected Features: 0

12/10/2019



Fire Reports in Cheshire, UK

Entity Recognizer



ArcGIS Pro w/ LocateXT



ArcGIS API for Python
Entity Recognizer

Questions?



Colby Neeb
cneeb@esri.com



Megan Banaski
mbanaski@esri.com



Aryle Butler
abutler@esri.com



esri®

THE
SCIENCE
OF
WHERE®

Copyright © 2021 Esri. All rights reserved.