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

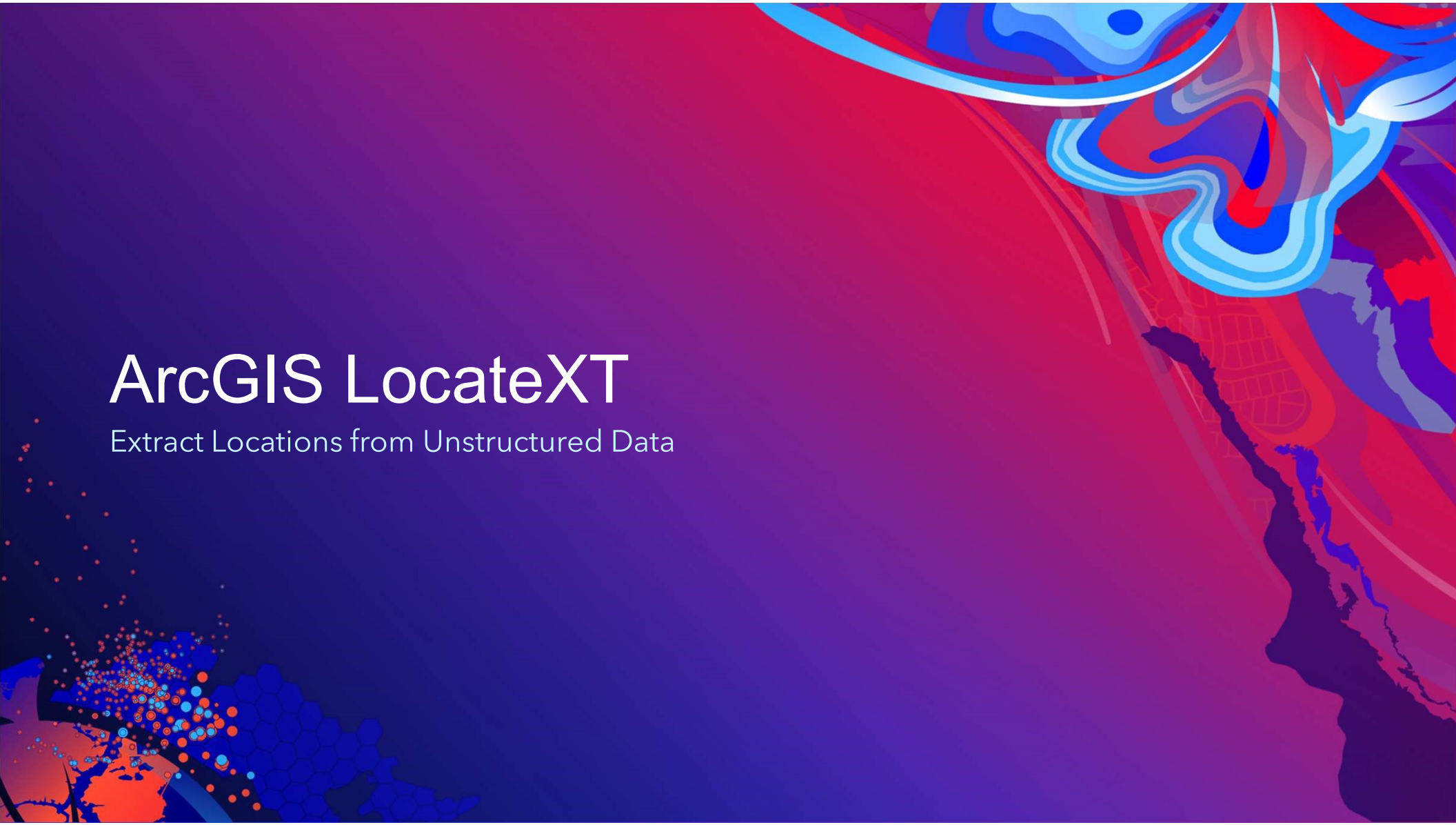
ArcGIS Pro for Intelligence

ArcGIS Enterprise w/ LocateXT

What is the best tool?

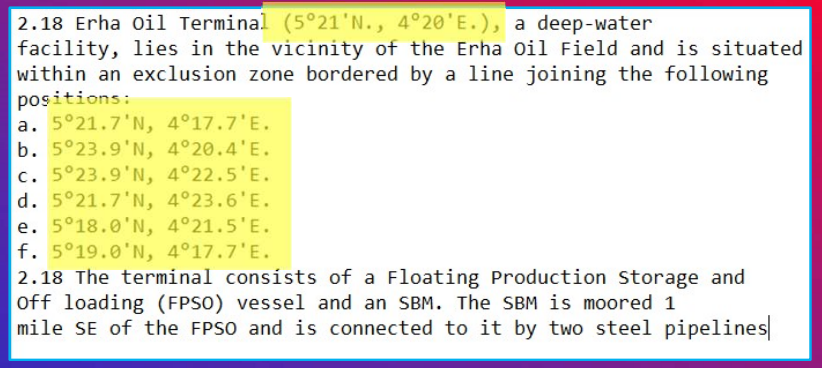
- Data is at least somewhat understood
- Data benefits from identifiable and repeating patterns
- Little to no programming experience available/needed

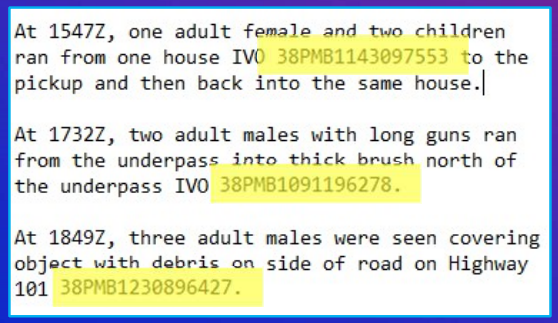
How is it best used?

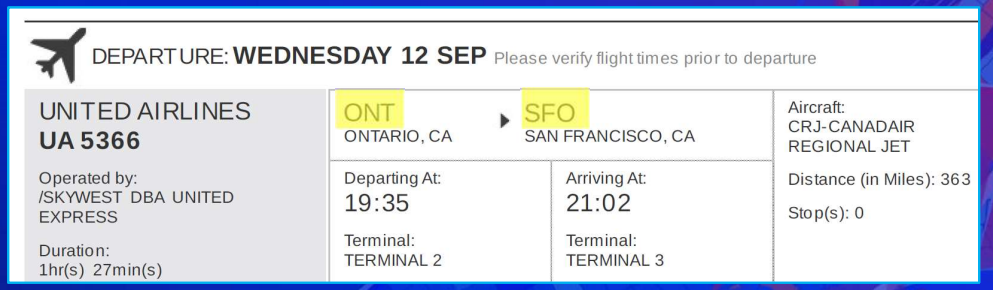

ArcGIS LocateXT
Extract Locations from Unstructured Data

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)

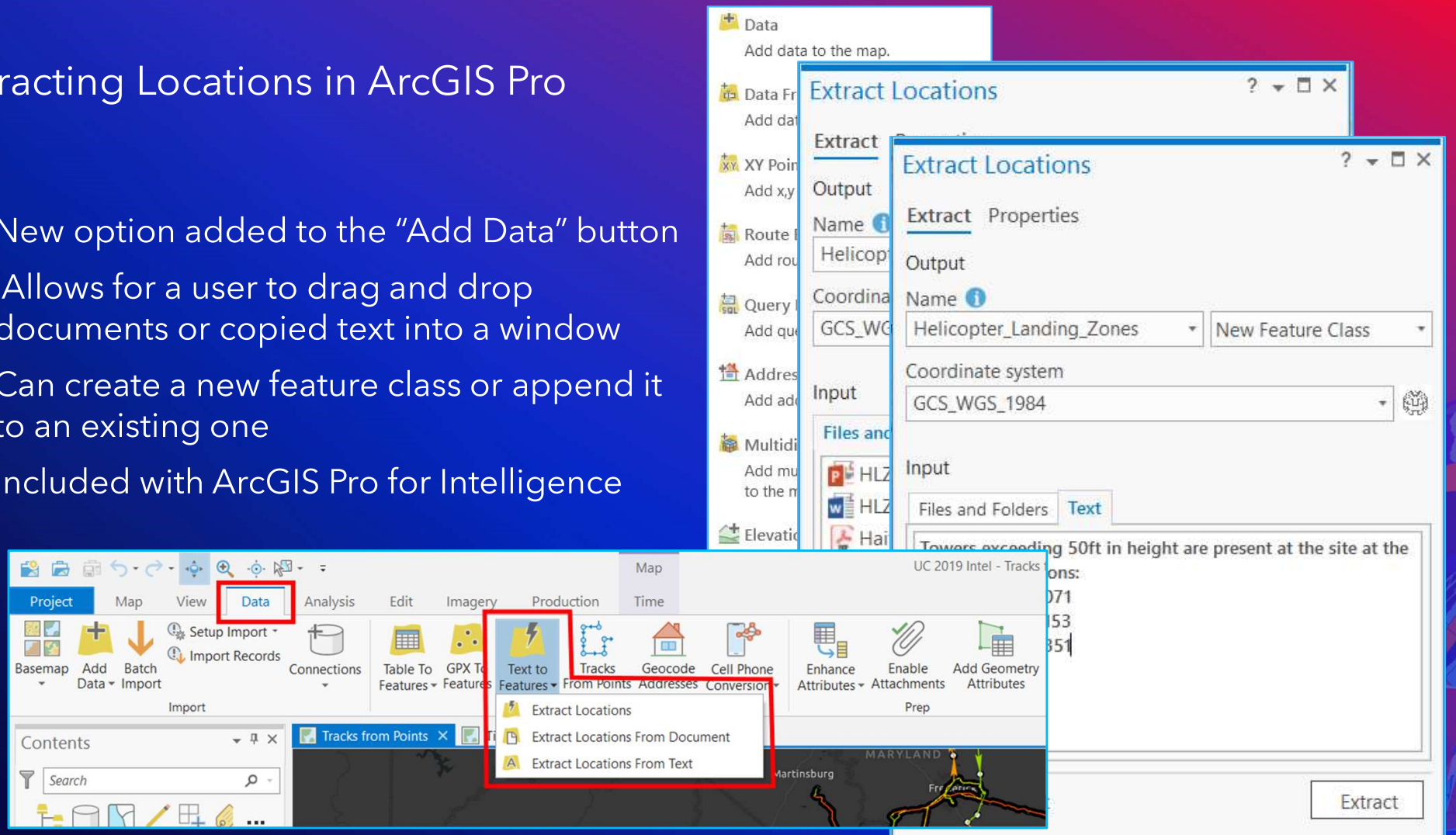






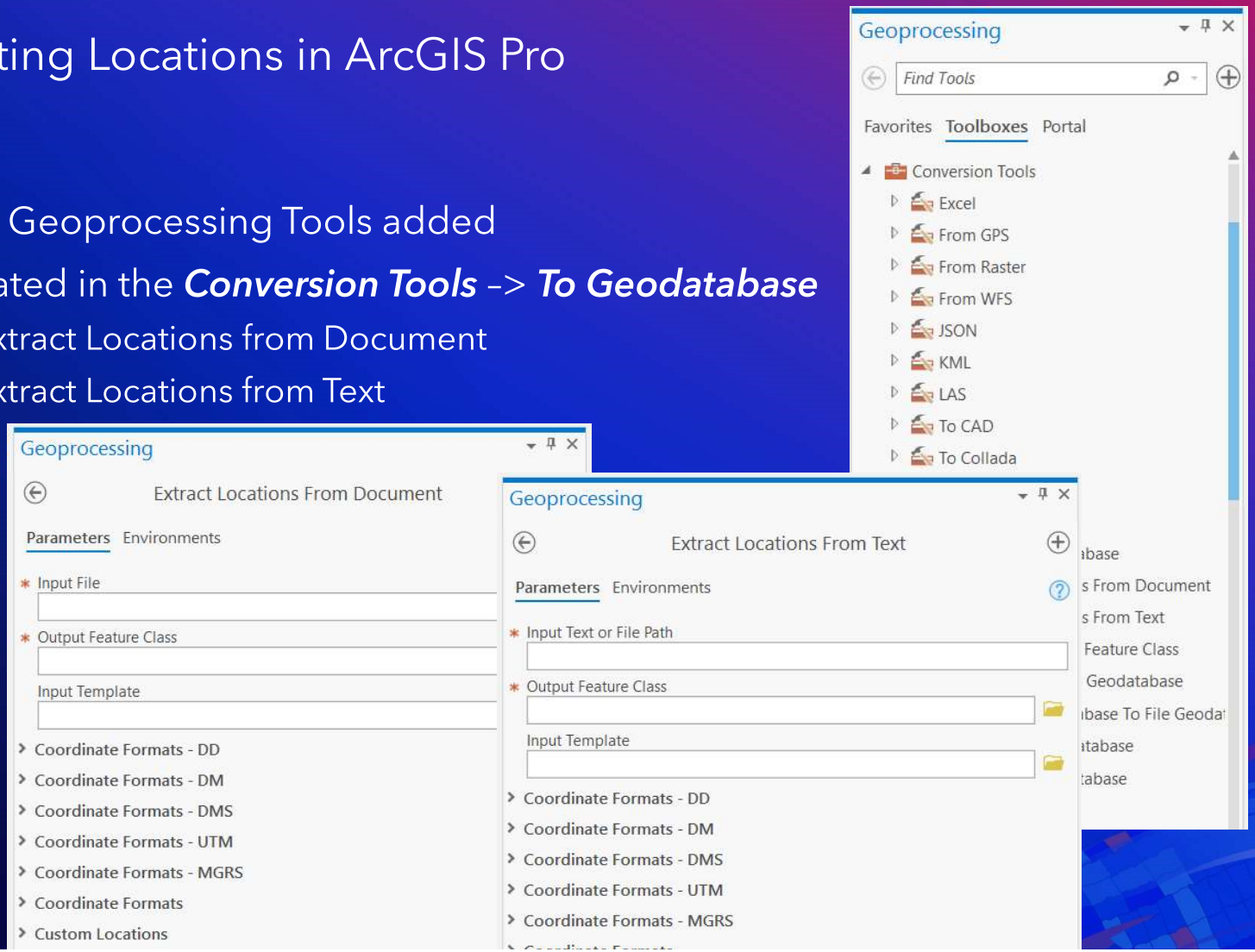
Extracting Locations in ArcGIS Pro

- 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



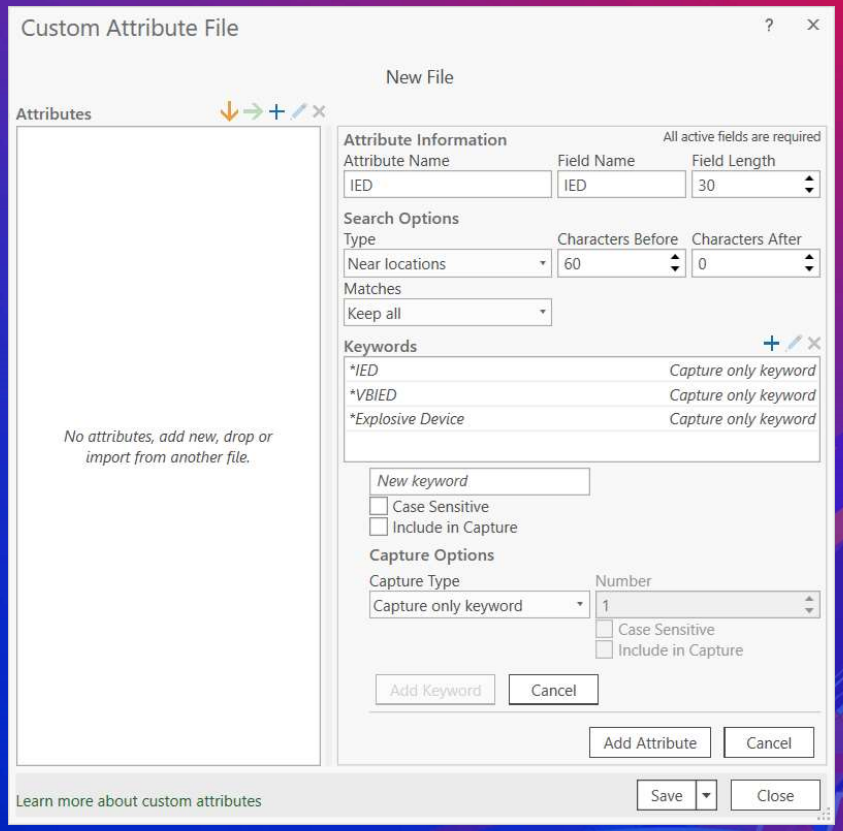
Extracting Locations in ArcGIS Pro

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



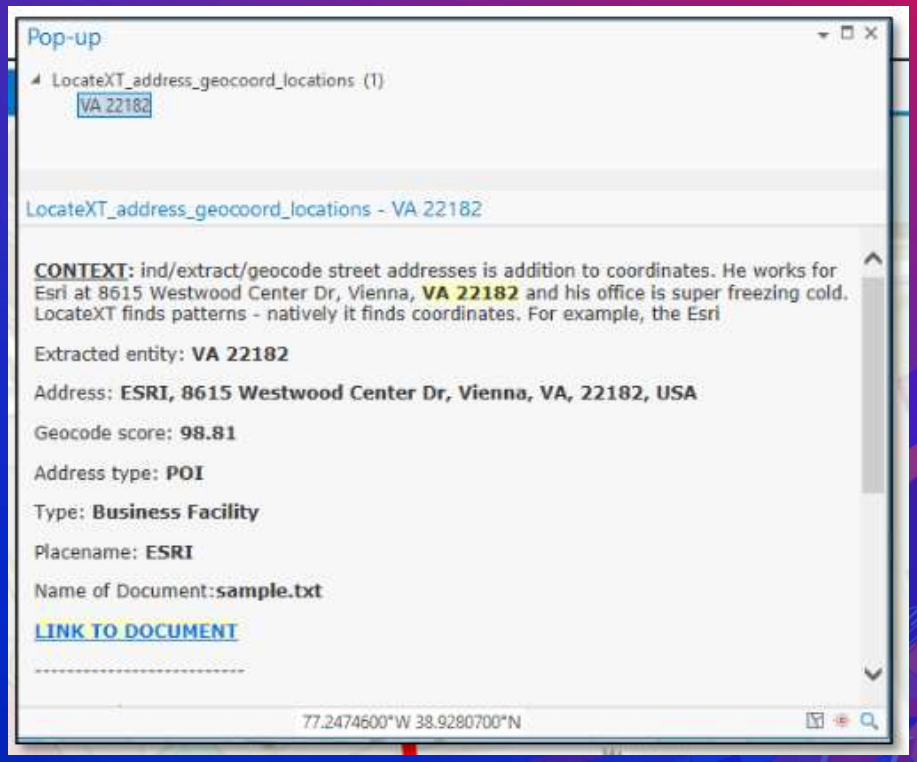
Extracting Custom Attributes

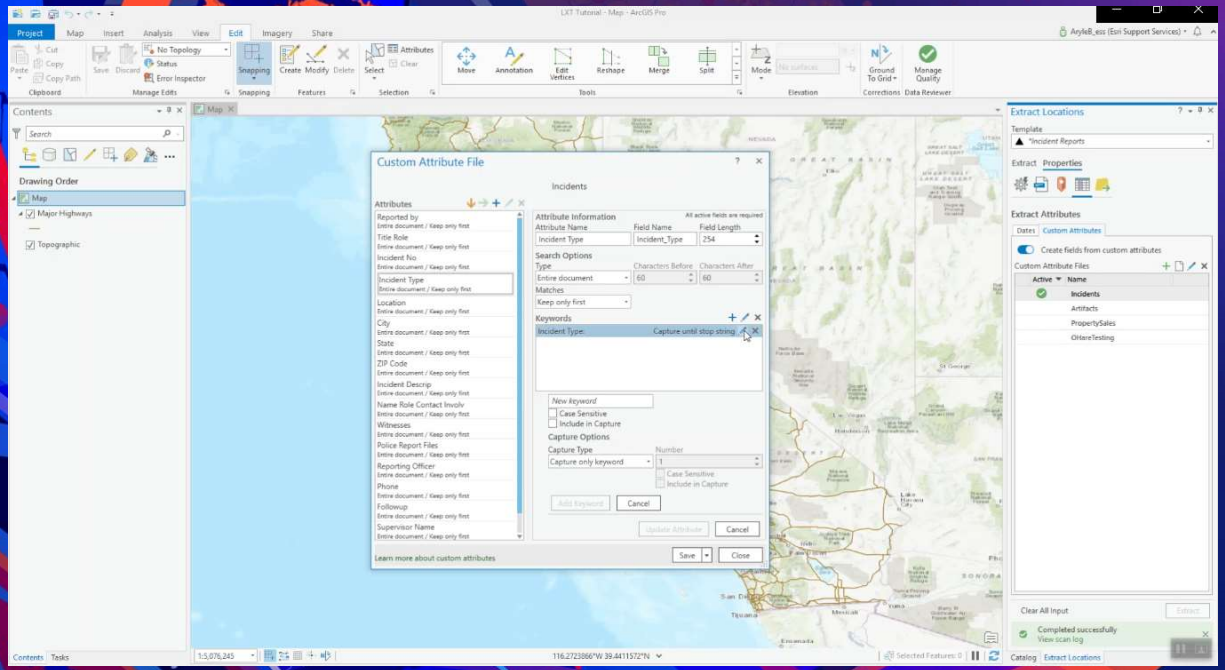
- Ability to create custom attributes based on content within document or near a location
  - Triggered by location extraction
- Based on keywords
  - Tag locations based on keywords
  - Scrape/harvest portions of document based on keywords
- Ability to extract based off of:
  - Number of characters/words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



Extracting Addresses

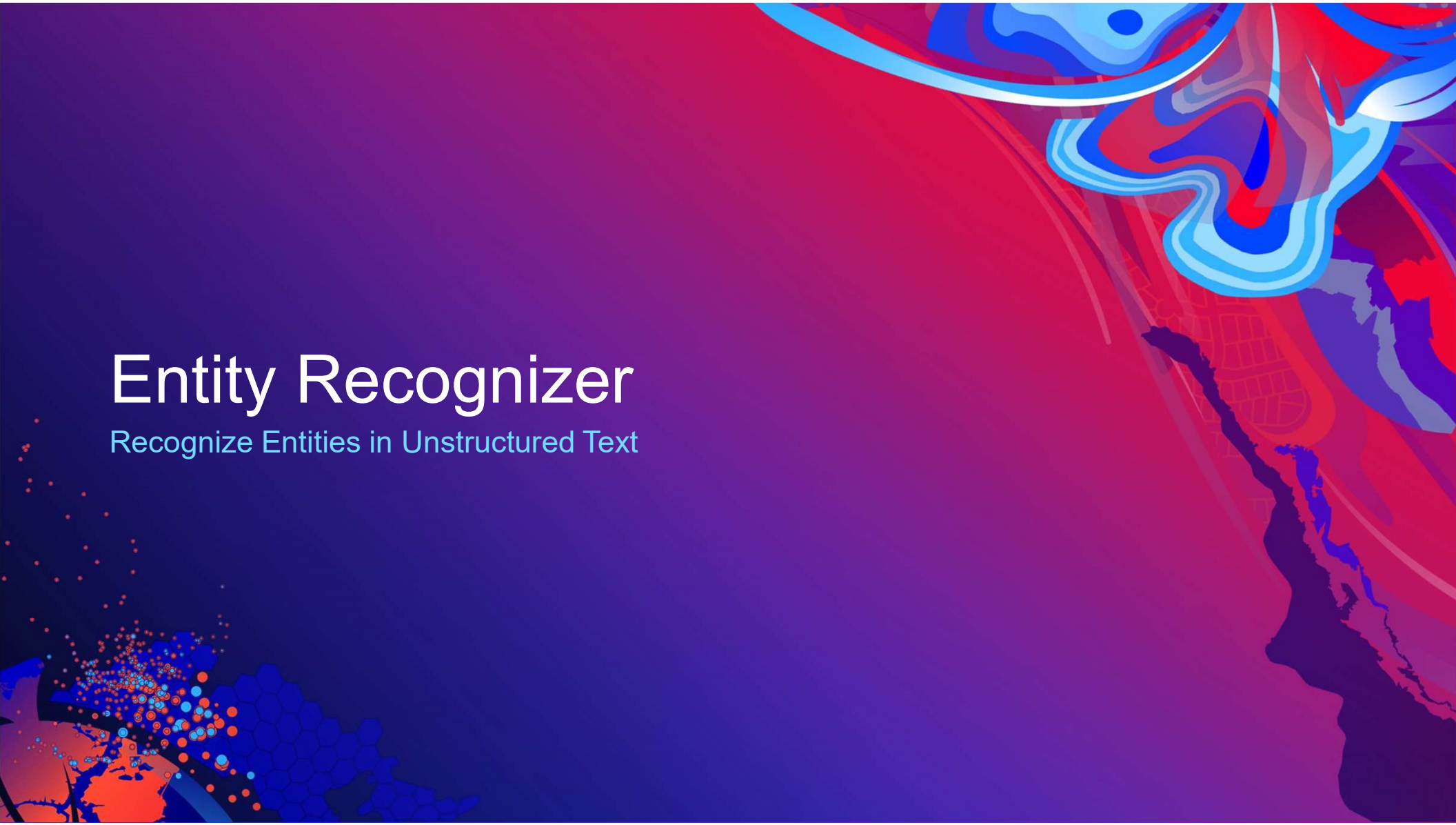
- Ability to extract addresses from documents based on combination of:
  - State
  - Zipcode
    - Ex. **VA 22182**
- The combination of **extracted text** and **pre-text** is geocoded





Police Reports in California

LocateXT


Entity Recognizer
Recognize Entities in Unstructured Text

What are you looking for?

Coordinates
Custom Locations
User defined keywords

Locations
People/Organizations
Events
Dates
Relationships

What is the best tool?

ArcGIS Pro w/ LocateXT

ArcGIS Pro for Intelligence

ArcGIS Enterprise w/ LocateXT

ArcGIS API for Python
Entity Recognizer

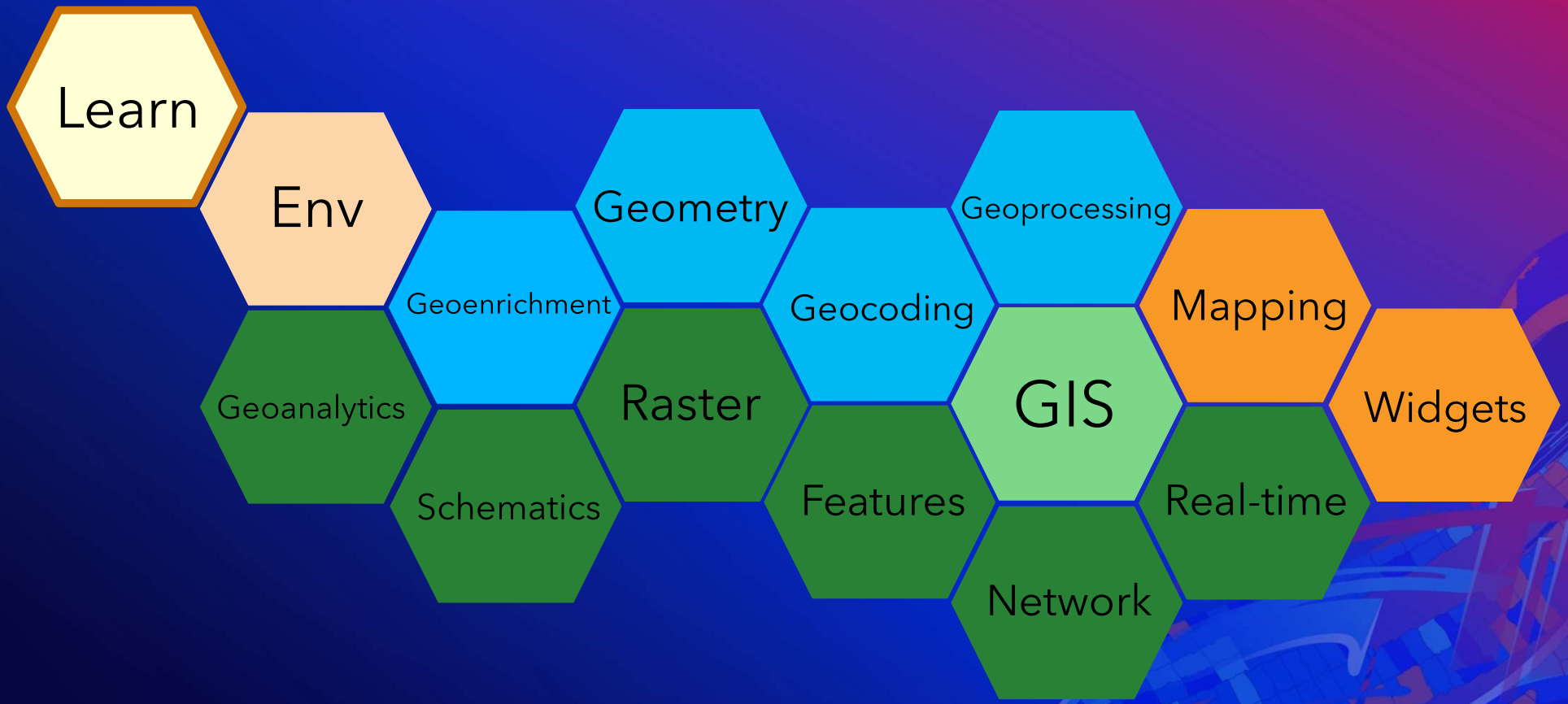
How is it best used?

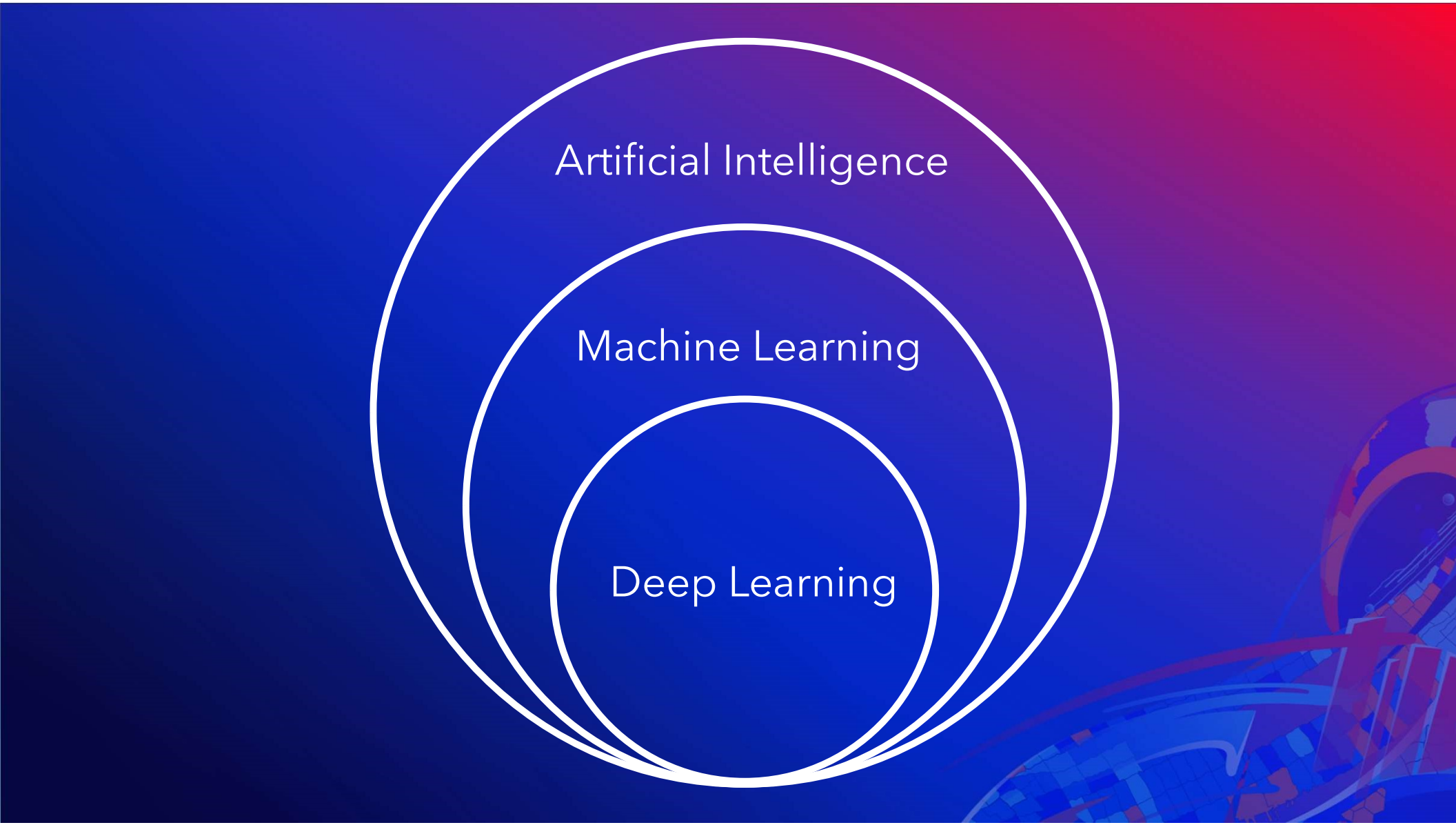
- Data is at least somewhat understood
- Data benefits from identifiable and repeating patterns
- Little to no programming experience available/needed

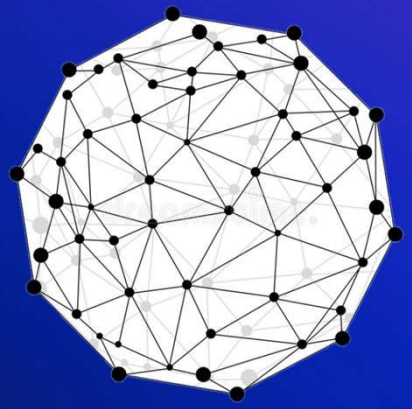
- Data is not well understood
- Data does not contain identifiable and/or repeating patterns
- Scripting needed for desired results

ArcGIS API for Python

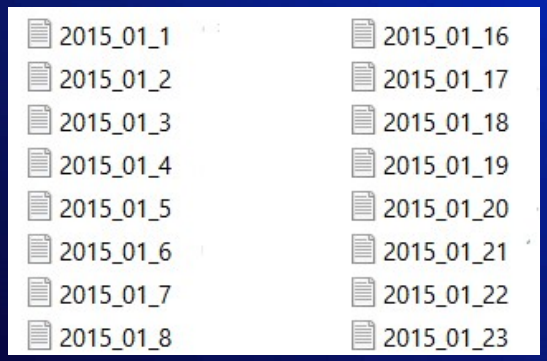
A powerful Python library for spatial analysis, mapping, and GIS




Artificial Intelligence
Machine Learning
Deep Learning

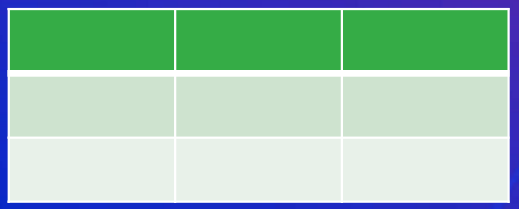


Trained EntityRecognizer Model

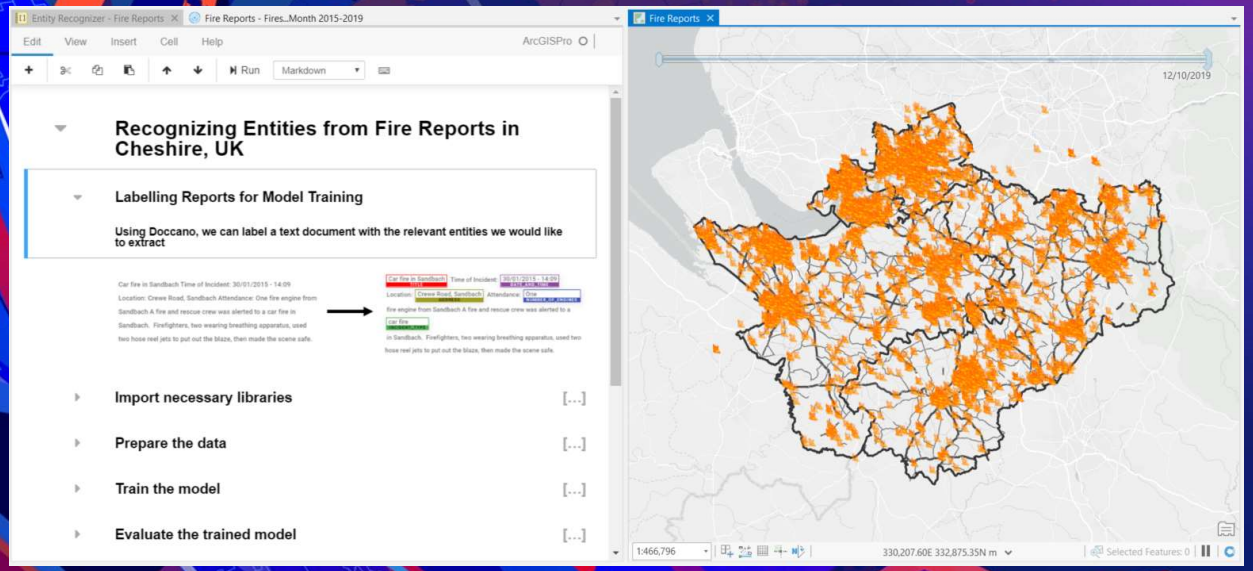


Unstructured Text Documents

ArcGIS API for Python



Structured Data



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

Copyright © 2021 Esri. All rights reserved.