

**DATA+AI**  
SUMMIT 2022

# Using Computer Vision to Mitigate Wildfire Risks for Electric Utilities



**Dexter Lewis**

Principal Technical Leader, EPRI



**Nick Lee**

Partner Solutions Tech Lead, Labelbox

ORGANIZED BY  databricks

**EPRI**

ELECTRIC POWER  
RESEARCH INSTITUTE

Talk sponsor



**Labelbox**

# Discover & Organize Unstructured Data in your Lakehouse

The screenshot displays the Labelbox web interface. At the top, there are navigation tabs for 'Projects', 'Models', 'Catalog', and 'Schema'. The 'Labelbox' logo is prominently displayed in the center. On the right side, there is a user profile for 'Nick Lee' with a 'Public demo' label. Below the navigation, a search bar is present with the text 'Search dataset name'. A sidebar on the left lists various datasets, with 'distribution-circuit1-16' selected, showing it has 29,704 data rows. The main area shows a grid of 16 images of utility poles with green bounding boxes and labels for 'Insulator'. Above the grid, there are filter controls: 'Add attribute filters...', '1 applied', 'Clear all', and a filter for 'Insulator'. On the right side of the grid, there are buttons for 'Display', 'Select all', 'Sample', and 'Created At'.

Projects Models Catalog Schema

Labelbox

Nick Lee  
Public demo

Search dataset name

All datasets

| Dataset Name                 | Count     |
|------------------------------|-----------|
| distribution-circuit1-16     | 29,704    |
| Open images by Go...         | 8,237,684 |
| Environmental sound cl...    | 2,000     |
| Video actions (youtube ...   | 3,149     |
| WikiNEuRal dataset           | 1,225,050 |
| Pomological Watercolors      | 1,523     |
| Movie Plots                  | 1,409     |
| People Clothing Segme...     | 1,000     |
| xVIEW overhead imager...     | 1,127     |
| Mapillary traffic signs 1... | 1,587     |
| SAR dataset (chipped)        | 9,000     |
| Retail grocery (synthetic)   | 999       |
| COCO Data                    | 20,000    |
| Geospatial Vessel Detection  | 798       |
| Mapillary Vistas             | 24,982    |
| customer-success-tick...     | 48,549    |

distribution-circuit1-16 (17,791 data rows)

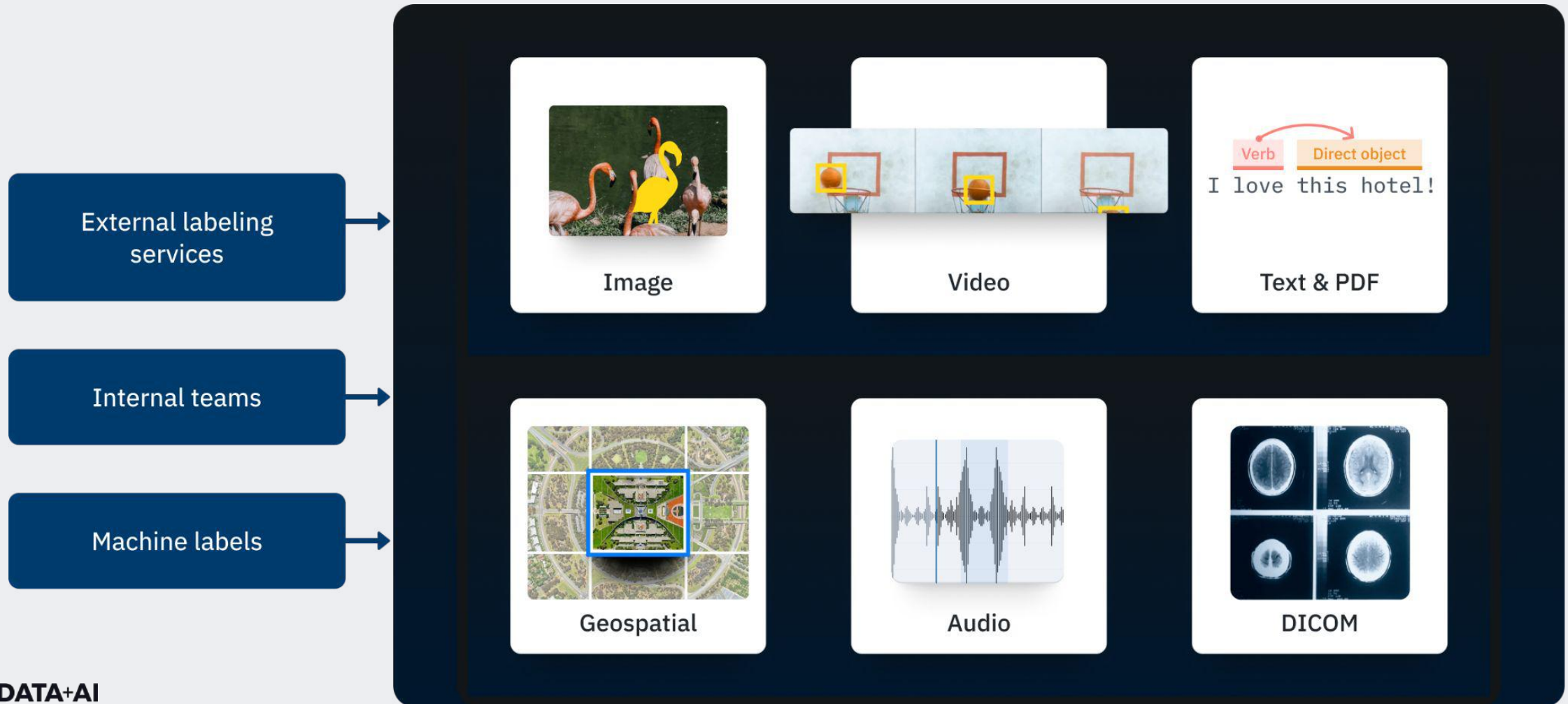
Add attribute filters...

1 applied Clear all

Annotation is Insulator or

Display Select all Sample Created At

# Easily Prepare Unstructured Data for AI Training



# Using Computer Vision to Mitigate Wildfire Risks for Electric Utilities

The logo for the Electric Power Research Institute (EPRRI) features the acronym 'EPRRI' in a bold, blue, sans-serif font. The letters are stylized with rounded edges and a consistent thickness.

ELECTRIC POWER  
RESEARCH INSTITUTE

# Electric Power Research Institute

EPRI is a non-profit, international organization that performs research to advance safe, reliable, affordable, and clean energy for the public benefit.

## Advance technologies

Develop, evaluate, and accelerate solutions that provide value



**Provide value to utilities and society**



## Work objectively

Follow scientifically based research and reporting

## Collaborate with industry

Work with utilities, academia, private industry, and the public



# Electric infrastructure failures can be catastrophic



**It is challenging to inspect vast infrastructure.**



**Distribution**

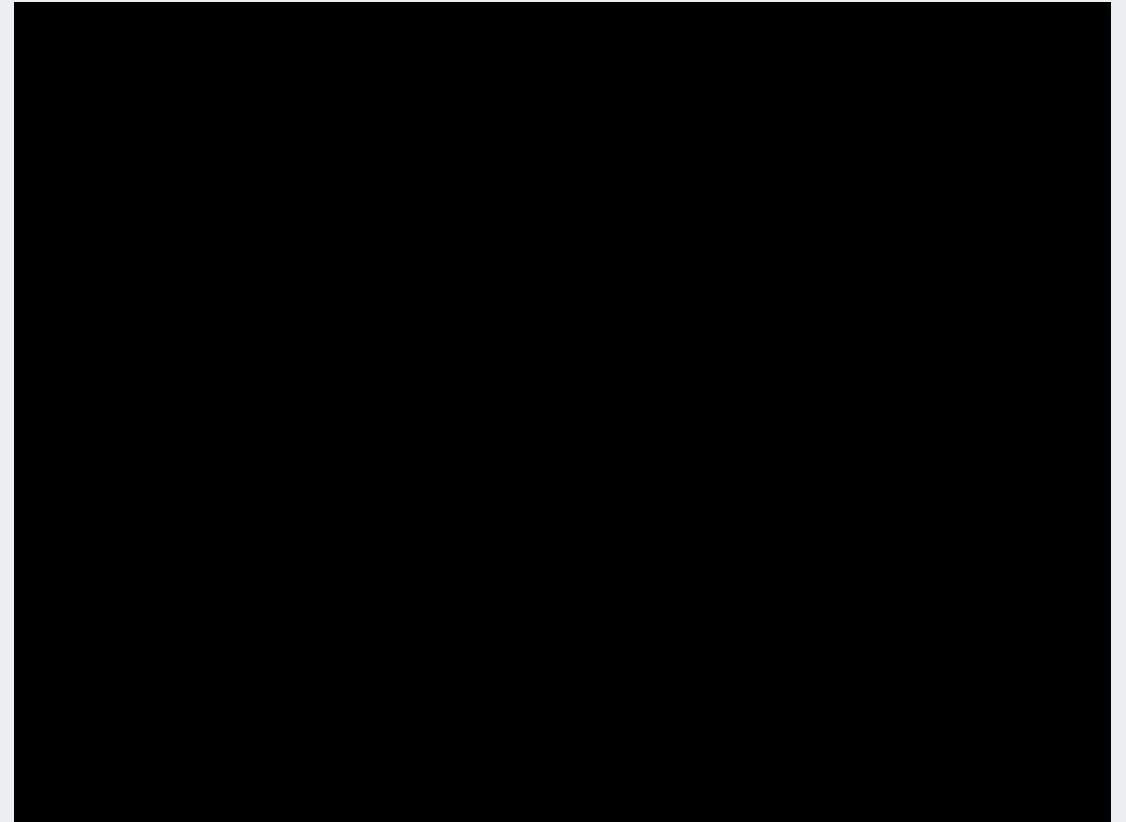


**Transmission**

# The goal is to optimize the inspection workflow.

*Autonomous Data Predictions*

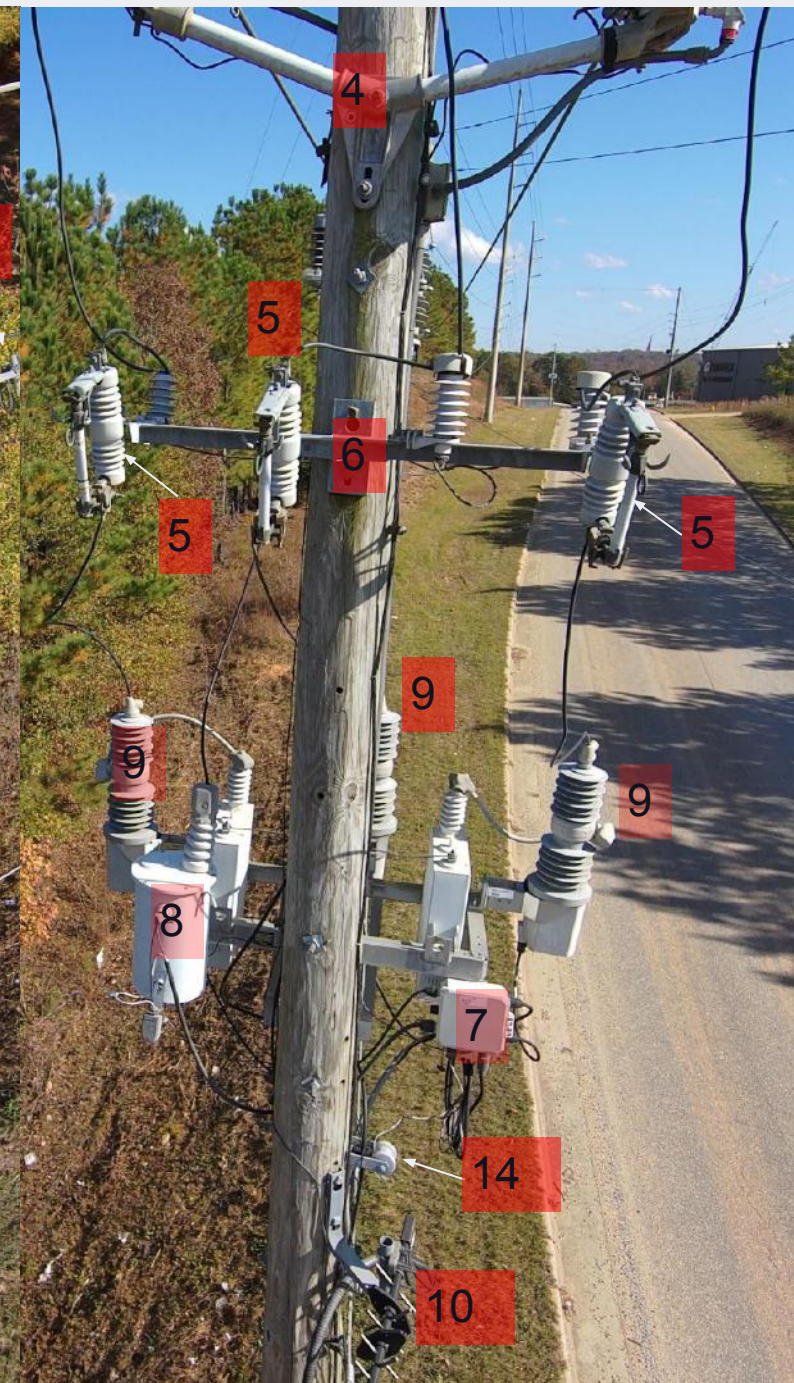
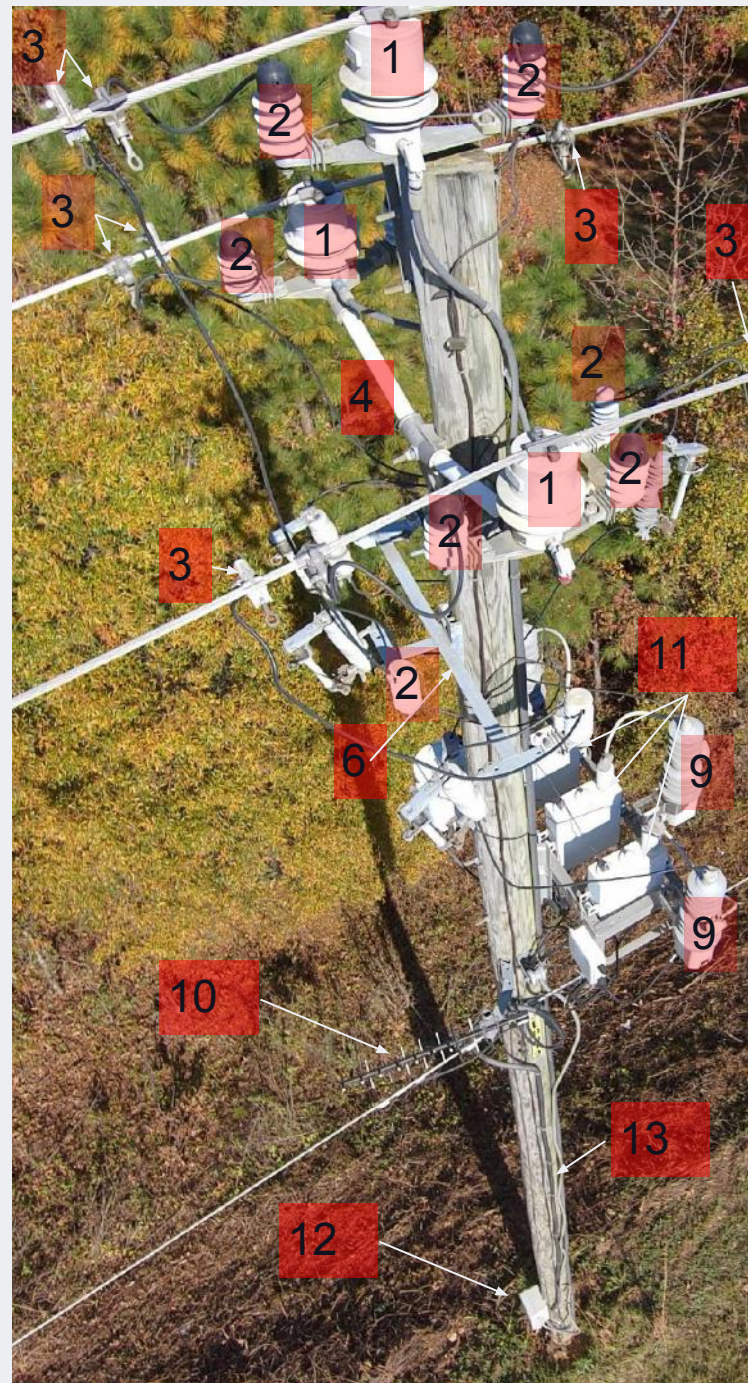
*Autonomous Data Capture*

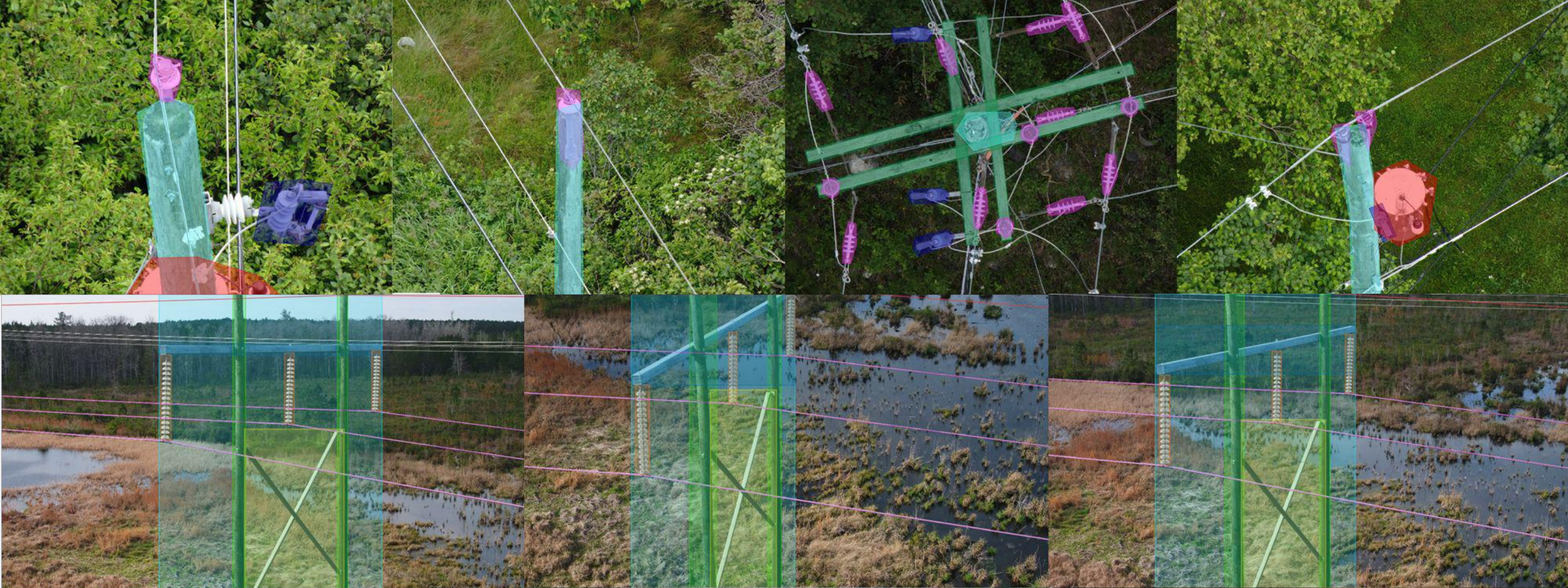




# The taxonomy is huge, and the datasets don't exist.

1. Piedmont voltage/current sensor
    - Healthy, Burnt
  2. Sensor arresters with wildlife cap
    - Healthy, Burnt, Wildlife Cap (on and off), Blown Isolator
  3. Hotline clamp(s)
    - Healthy, Loose
  4. Fiberglass arm
    - Healthy, Broken, Loose
  5. Capacitor overcurrent fuses and cutouts
    - Open, Closed, Broken, Loose, Contaminated, Cracked
  6. Tri-Mount bracket
    - Healthy, Corroded, Loose
  7. Control junction box
    - Healthy, Defective
  8. Service transformer with secondary arrester
    - Healthy, Leaking, Corroded, Loose, Wildlife Cap (on and off)
  9. Solid dielectric vacuum switch
    - Healthy, Burnt, Damaged
  10. SCADA antennas
    - Healthy, Defective
  11. Capacitor can(s) and rack
    - Healthy, Swollen, Broken Bushing, Corroded, Loose
  12. Control cabinet with SCADA radio
    - Healthy, Defective
  13. Control cable and conduit
    - Healthy, Defective
  14. Neutral spool insulator
    - Healthy, Defective
- Not shown (15. grounds, 16. leads/jumpers, 17. conductors)
- Healthy, Connected, Loose, Broken, Arcing Damage

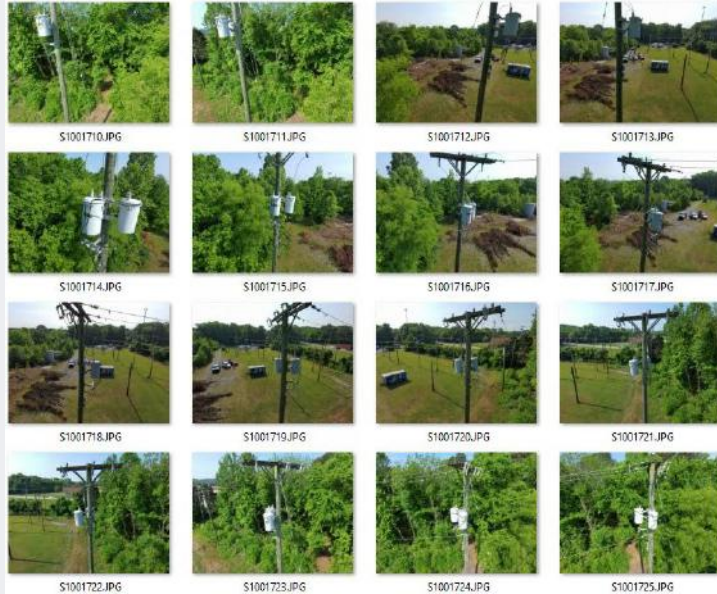




EPRI labeled over 29,000 distribution images and 12,000 transmission images to support public, private, and academic innovation.

# The workflow begins with data and labeling preparation.

Step 1: Collect the data.



Step 2: Obscure sensitive information.



Step 3: Edit or remove metadata.



Every **H-Structure** in this dataset is similar.

- All structures will contain:
  - 2x Poles
  - 1x Crossarm
  - 3x Insulators
  - 3x Conductors
  - 2x Shield Wires
- Some may contain:
  - 1x X-Brace

**Polylines and Bounding Boxes**

- Polyline Classes
  - Conductor
  - Shield Wire
- Bounding Box Classes
  - H-Structure
  - Pole
  - Crossarm
  - Insulator
  - X-Brace

**Bounding Box Guidance**

The bounding box should include the entire object.

A series of images showing bounding boxes around utility poles and structures, with labels 'Correct' and 'Wrong' indicating proper and improper box placement.

Step 4: Create the labeling taxonomy and guidance.

# Identify a labeling software and workforce solution.



Images files remain in Azure Blob, but the image is rendered in Labelbox's web interface.



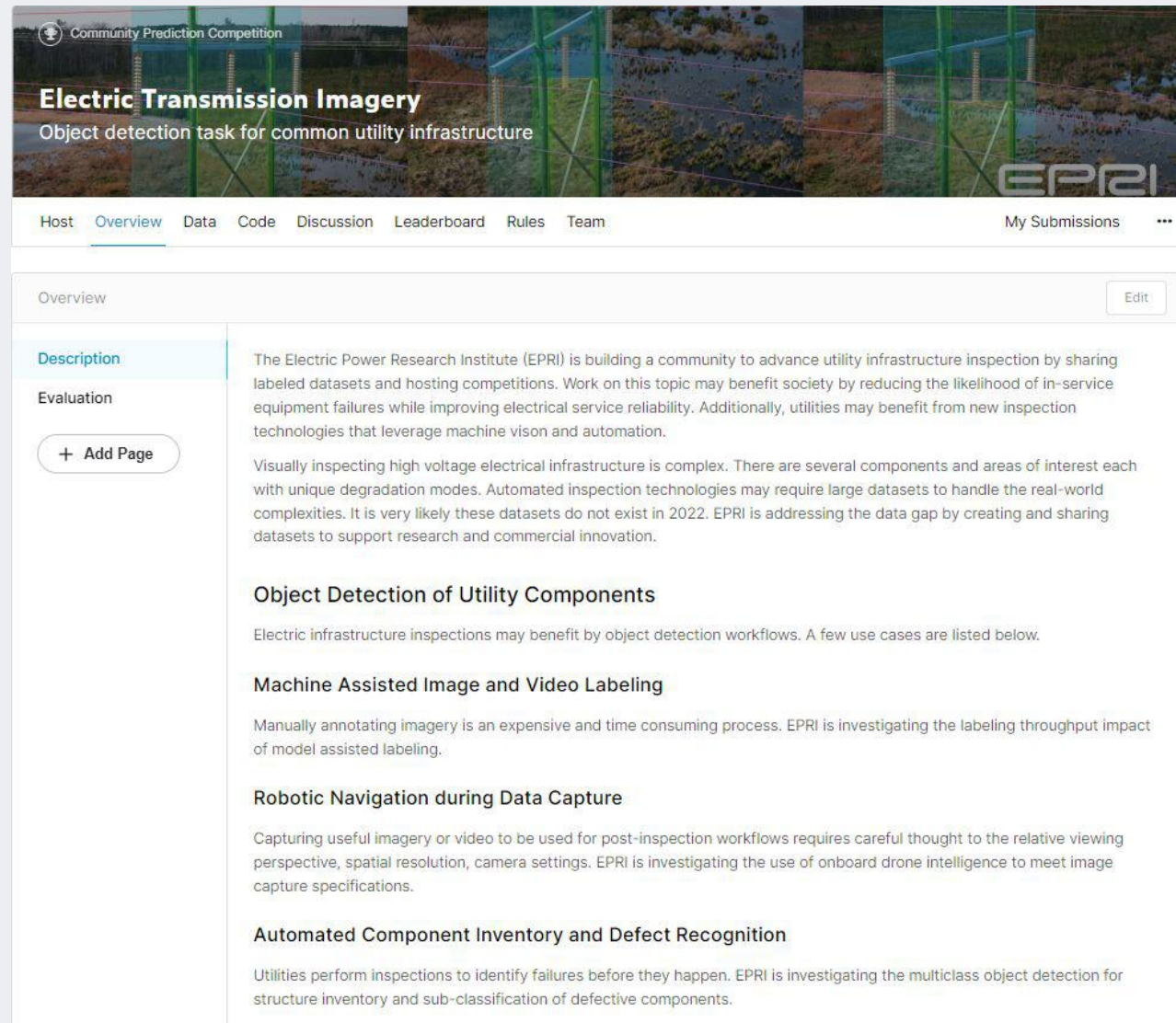
EPRI hosted the 'cleaned' imagery in Microsoft Azure.



**Labelbox**

EPRI subscribed to Labelbox and leveraged their Workforce Boost.

# EPRI is sharing these data in the public domain for free.



Community Prediction Competition

## Electric Transmission Imagery

Object detection task for common utility infrastructure

Host Overview Data Code Discussion Leaderboard Rules Team My Submissions

Overview Edit

**Description**

The Electric Power Research Institute (EPRI) is building a community to advance utility infrastructure inspection by sharing labeled datasets and hosting competitions. Work on this topic may benefit society by reducing the likelihood of in-service equipment failures while improving electrical service reliability. Additionally, utilities may benefit from new inspection technologies that leverage machine vision and automation.

Visually inspecting high voltage electrical infrastructure is complex. There are several components and areas of interest each with unique degradation modes. Automated inspection technologies may require large datasets to handle the real-world complexities. It is very likely these datasets do not exist in 2022. EPRI is addressing the data gap by creating and sharing datasets to support research and commercial innovation.

**Object Detection of Utility Components**

Electric infrastructure inspections may benefit by object detection workflows. A few use cases are listed below.

**Machine Assisted Image and Video Labeling**

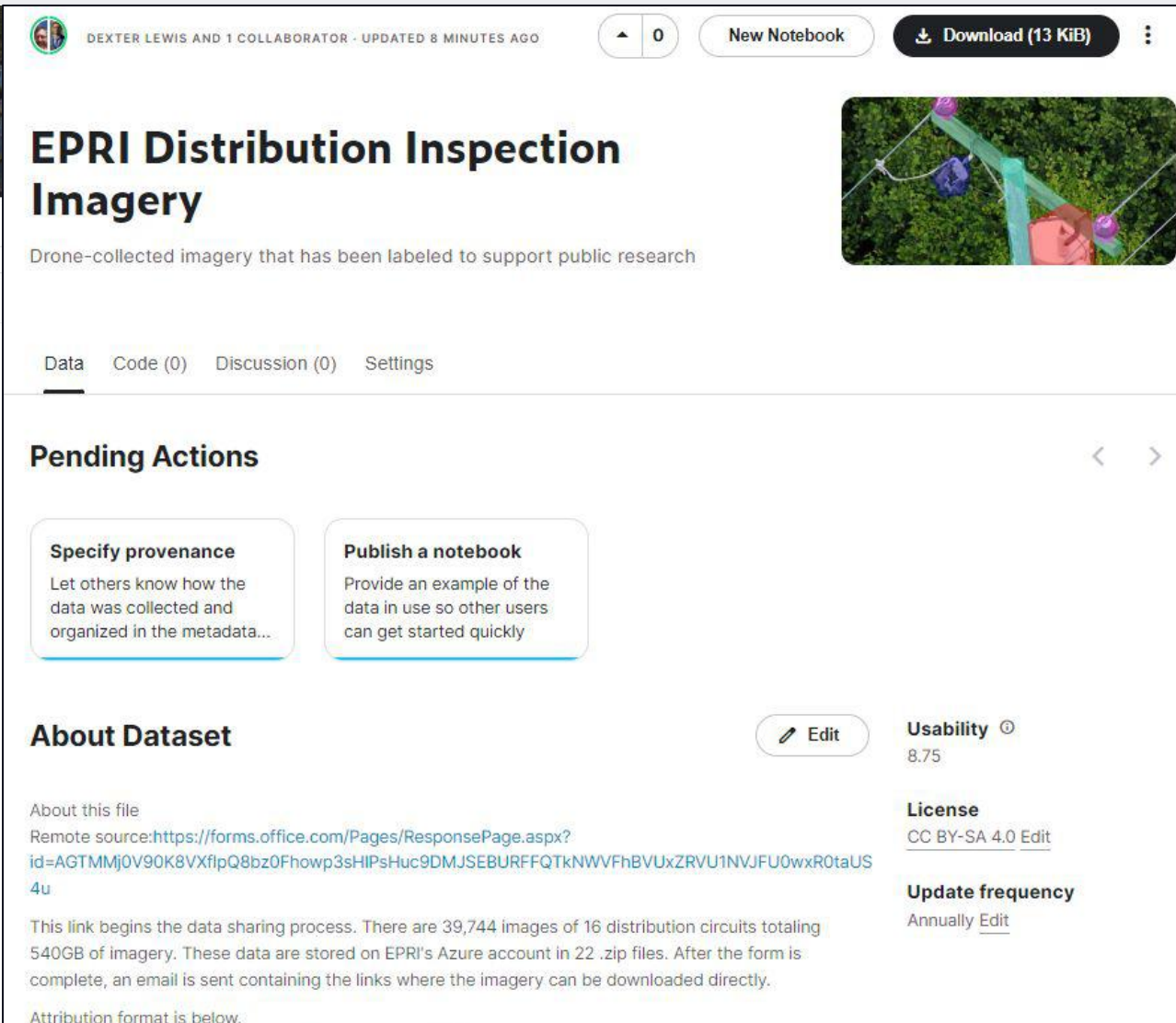
Manually annotating imagery is an expensive and time consuming process. EPRI is investigating the labeling throughput impact of model assisted labeling.

**Robotic Navigation during Data Capture**

Capturing useful imagery or video to be used for post-inspection workflows requires careful thought to the relative viewing perspective, spatial resolution, camera settings. EPRI is investigating the use of onboard drone intelligence to meet image capture specifications.

**Automated Component Inventory and Defect Recognition**

Utilities perform inspections to identify failures before they happen. EPRI is investigating the multiclass object detection for structure inventory and sub-classification of defective components.



DEXTER LEWIS AND 1 COLLABORATOR · UPDATED 8 MINUTES AGO

0 New Notebook Download (13 KiB)

## EPRI Distribution Inspection Imagery

Drone-collected imagery that has been labeled to support public research

Data Code (0) Discussion (0) Settings

**Pending Actions**

- Specify provenance**  
Let others know how the data was collected and organized in the metadata...
- Publish a notebook**  
Provide an example of the data in use so other users can get started quickly

**About Dataset** Edit

Usability 8.75

License CC BY-SA 4.0 Edit

Update frequency Annually Edit

About this file  
Remote source: <https://forms.office.com/Pages/ResponsePage.aspx?id=AGTMMj0V90K8VXflpQ8bz0Fhowp3sHIPsHuc9DMJSEBURFFQtkNwVfHbVUxZRVU1NVJFU0wxR0taUS4u>

This link begins the data sharing process. There are 39,744 images of 16 distribution circuits totaling 540GB of imagery. These data are stored on EPRI's Azure account in 22 .zip files. After the form is complete, an email is sent containing the links where the imagery can be downloaded directly.

Attribution format is below.

# Data Collection and Labeling for AI



# Empowering AI Workflows with Unstructured Data

Key AI activities

Model deployment

Model creation and training

Data labeling and management



**databricks**



**Labelbox**



LabelSpark Demo - Data and AI Summit 2021 (Python)

Query Labelbox for Raw Annotations (Bronze Table)

```
1 client = Client(API_KEY) #refresh client
2 bronze_table = labelspark.get_annotations(client, "cokolzeshr7zsy0736w0usbxdj", spark, sc)
3 bronze_table.registerTempTable("street_photo_demo")
4 display(bronze_table.limit(2))
```

3 Spark Jobs

bronze\_table: pyspark.sql.dataframe.DataFrame = [Agreement: integer, Benchmark Agreement: integer ... 15 more fields]

| Label      | Labeled Data  |
|------------|---|
| 381g358fkl | https://storage.labelbox.com/ckmu2e5v/Street%20View%201.jpeg?Expires=162138&Signature=J5tzRocb5i4svO12IEwp1U... |

Showing all 2 rows.

TOOLS

Search or press /

- Pole 1
- Insulator 2
- Crossarm 3
- Conductor 4
- Other wire 5
- Cutouts 6
- Transformers 7
- Background Structure 8
- Image Quality 9

Select one

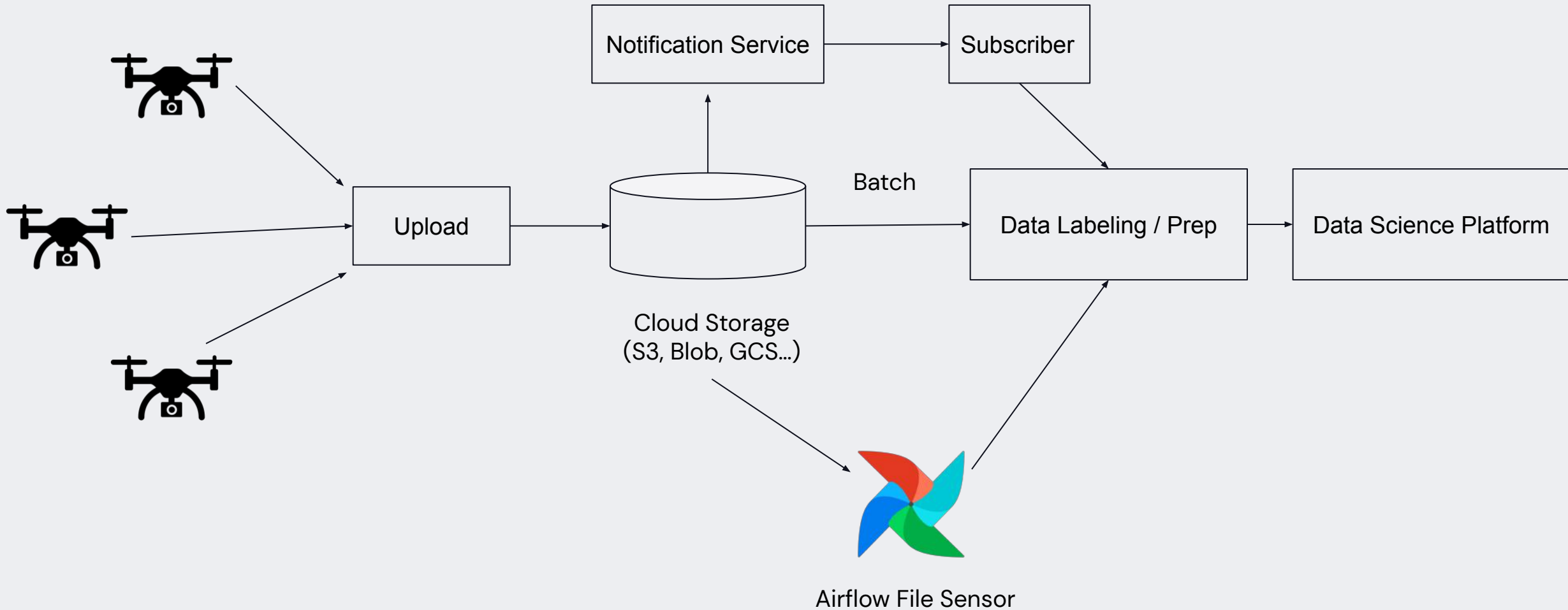
OBJECTS

- Pole 1
- Insulator (6)
- Crossarm (2)
- Conductor (3)
- Other wire 1
- Cutouts 1
- Transformers 1

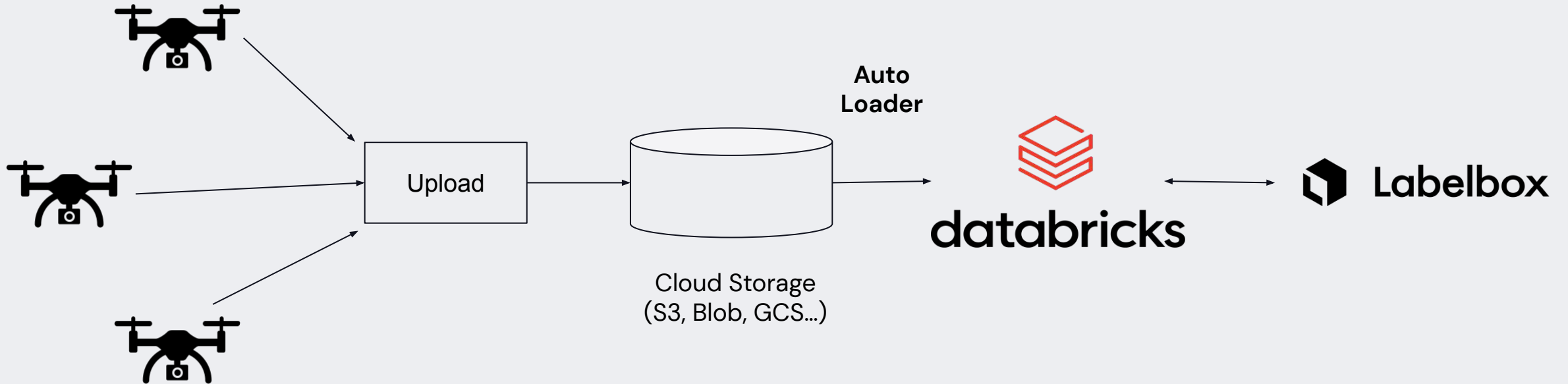




# Large Scale Field Data Collection for AI



# Large Scale Field Data Collection for AI

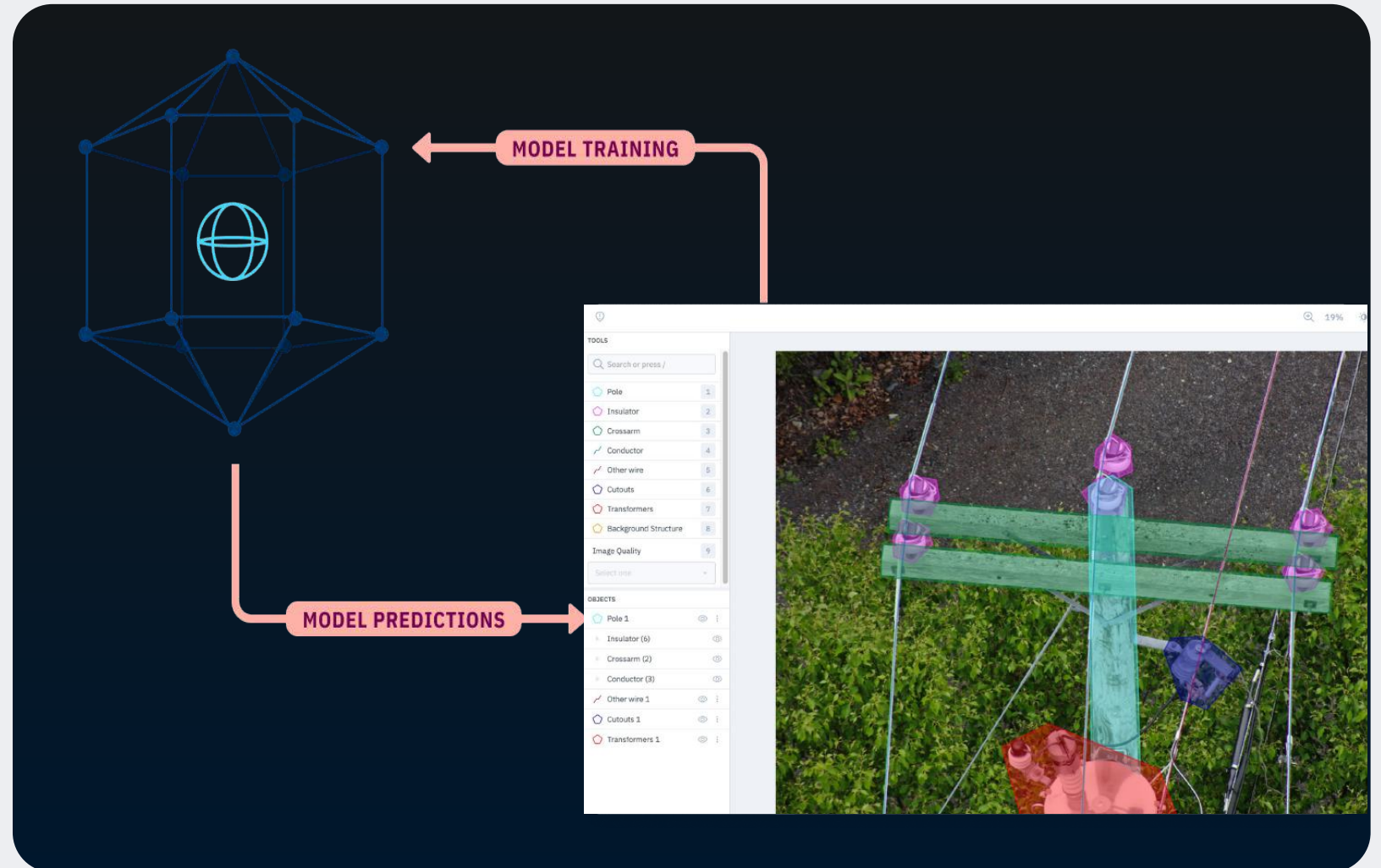


# Demo

# Save Time with Model Assisted Labeling

Use your own models to power automated annotation in Labelbox.

Make corrections and retrain with ease.



# Automate & Accelerate

Labelbox



Labelbox



Labelbox



Labelbox



# Root out all errors with Model Diagnostics

Visualize model errors and take action to improve performance faster. Immediately understand where model performance is weak so you can more effectively guide effort.

Works great with

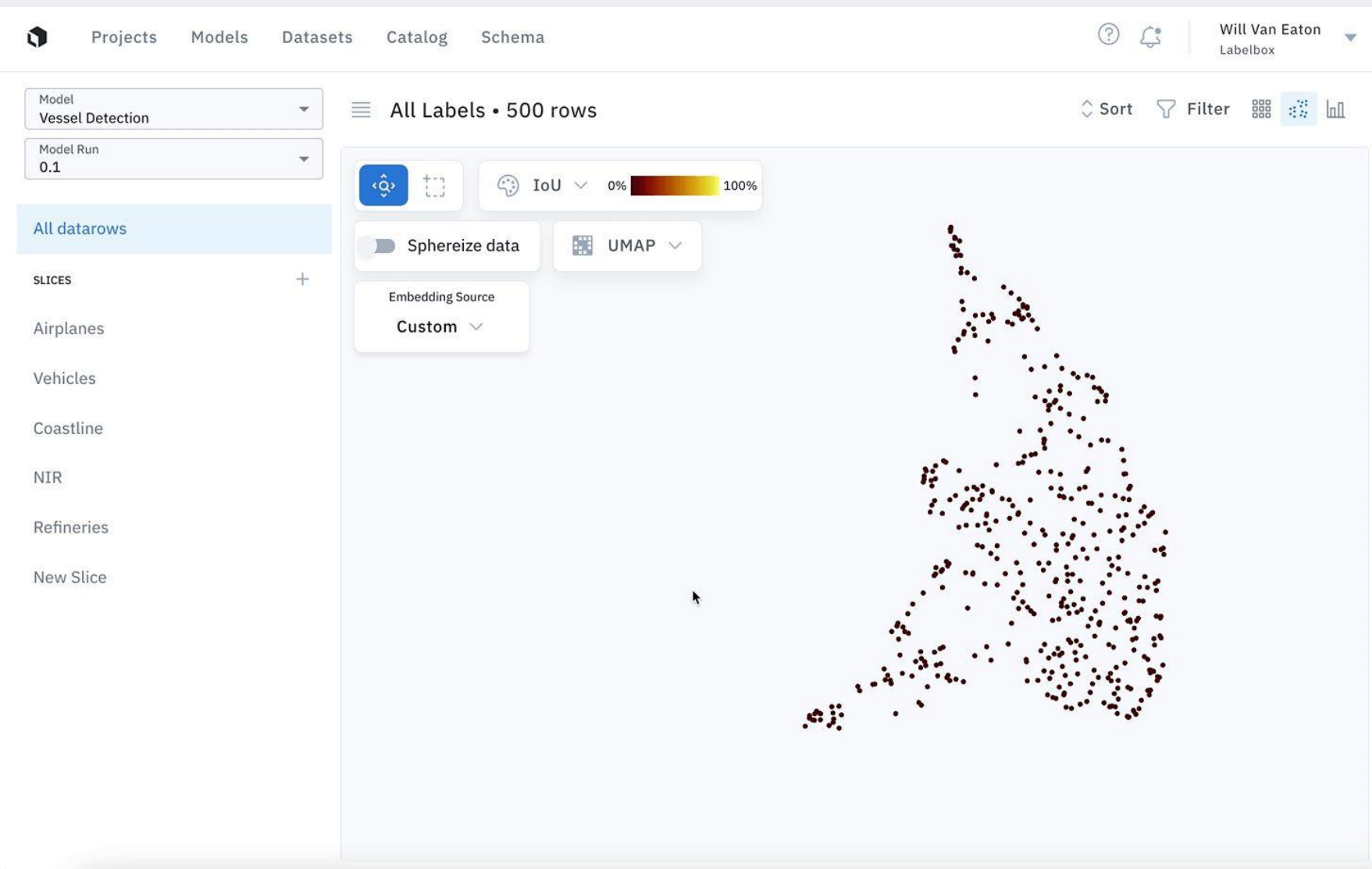


The screenshot displays the MLflow Model Diagnostics interface. At the top, there are navigation tabs for Projects, Models, Datasets, Catalog, and Schema. The user's name, Will Van Eaton, is visible in the top right corner. The main area shows a satellite image of a dark sea with several ships. Some ships are highlighted with red dashed bounding boxes, indicating areas of interest or error. Below the main image is a horizontal carousel of smaller images, each showing a different view or zoom of the satellite data with yellow bounding boxes.

On the right side, there is a sidebar with the following sections:

- DATA**:
  - Dataset: Geospatial vessel detection
  - Project: Vessel detection
  - Datarow ID: cktn817ni01ma0yu343ro6li2
  - External ID: positive\_image\_set/504.jpg
  - Label ID: cktn85cua51yj0yboc1fh273p
- LABEL**:
  - ship
- PREDICTION**:
  - ship
- METRICS**:

```
{  iou : {    value : 0.814  }  }
```



## Visualize trends in your data and identify outliers

Cluster similar data to simplify complex patterns, uncover trends in model performance, and find edge cases.

# Special Announcement

**EPRRI**  
ELECTRIC POWER  
RESEARCH INSTITUTE

 **Labelbox**



# Public Datasets



- Two Datasets
  - Distribution Lines
  - Transmission Structure
- 35,000+ utility images
- 300,000+ annotations
- Hosted for free on Labelbox
- Community Kaggle Competition
- Public-facing Kaggle Dataset

1 (1007).JPG

Display | Parsed | Raw

Data

|             |   |
|-------------|---|
| ID          | cd479a680e29079d80ccnw5   |
| External ID | 1 (1007).JPG  |
| Created     | about 6 hours ago   |
| Asset type  | Image   |
| Dataset     | distribution-circuit1-16  |
| Row data    | <a href="https://storage.googleapis.com/labelbox-datasets-16/1007.jpg">https://storage.googleapis.com/labelbox-datasets-16/1007.jpg</a> |

Ground truth

- EPRI Distribution Lab...
- Conductor - 1
- Conductor - 2
- Conductor - 3
- Crossarm (2)
  - Crossarm - 1
  - Crossarm - 2
- Cutouts

Date created | Similarity | Custom similarity

177,599 annotations

| Name         | Count | % Share |
|--------------|-------|---------|
| Insulator    | 53494 | 30.12   |
| Conductor    | 50966 | 28.70   |
| Pole         | 29924 | 16.85   |
| Crossarm     | 14283 | 8.04    |
| Other wire   | 10406 | 5.86    |
| Cutouts      | 9279  | 5.22    |
| Transformers | 8734  | 4.92    |

Kaggle Links:

<https://www.kaggle.com/competitions/electric-transmission-imagery>

<https://www.kaggle.com/datasets/dexterlewis/epri-distribution-inspection-imagery>

Labelbox Trial (with access to dataset): [www.labelbox.com](http://www.labelbox.com)

# DATA+AI SUMMIT 2022

EPRI Kaggle Competition:

<https://www.kaggle.com/competitions/electric-transmission-imagery>

EPRI Kaggle Dataset:

<https://www.kaggle.com/datasets/dexterlewis/epri-distribution-inspection-imagery>

Labelbox free trial: [www.labelbox.com](http://www.labelbox.com)

# Thank you



**Nick Lee**

rtner Solutions Tech Lead  
Labelbox



**Dexter Lewis**

Pr. Technical Leader