RAIL

MOVING AMERICA FORWARD

Improvements in Data Collection and Linkage for Risk Modeling in Railroad Safety Inspection Programs

Andrew LaBounty, Federal Railroad Administration, USDOT
Patrick Johnson, Federal Railroad Administration, USDOT
Ruby Li, Federal Railroad Administration, USDOT
Tony Ye, Federal Railroad Administration, USDOT

U.S. Department of Transportation
Federal Railroad Administration

October 15, 2024

Disclaimer

The opinions, findings, and conclusions expressed in this presentation are those of the authors and not necessarily those of the United States Department of Transportation, the Federal Railroad Administration, or the Bureau of Transportation Statistics. The United States Government assumes no liability for its contents or use thereof.

Introduction

Challenges

1. Accurate risk model outputs (i.e., risk scores) depend on many inputs from a variety of sources, such as recent incidents and inspections.

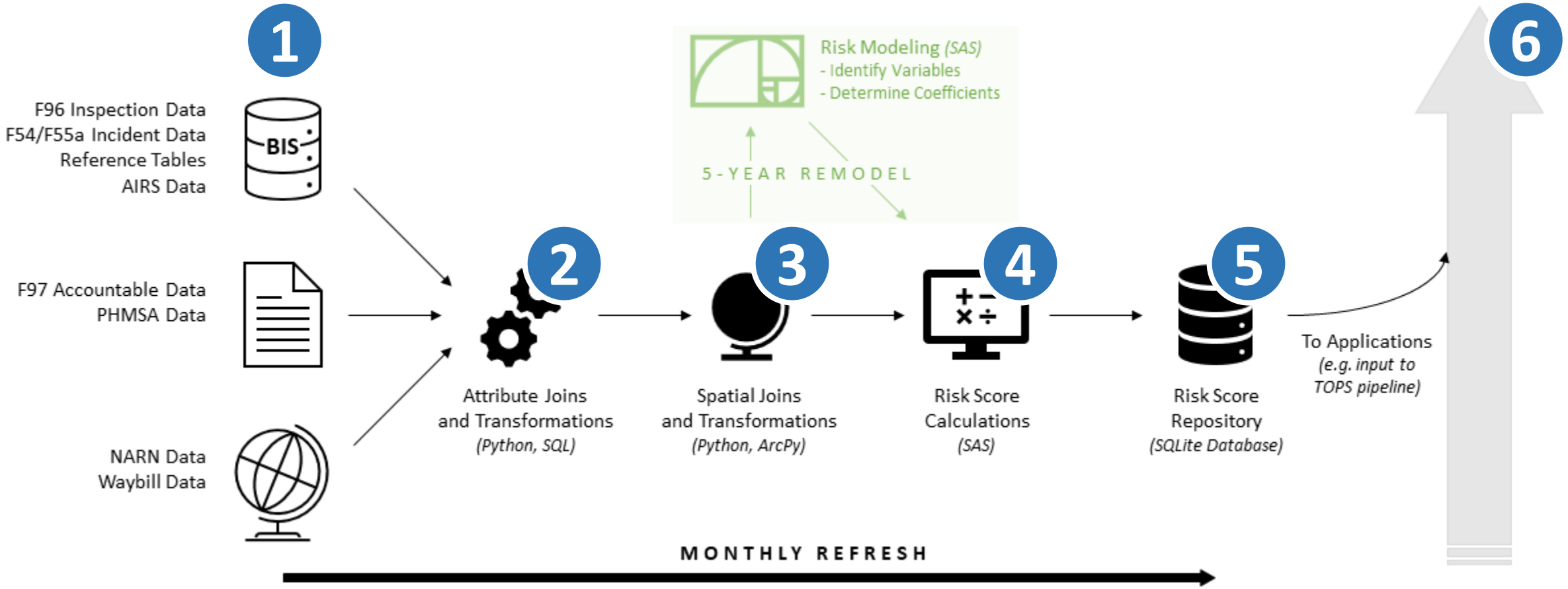
2. Identifying and recording trends in risk requires continuous, regular input of new data.

Solutions

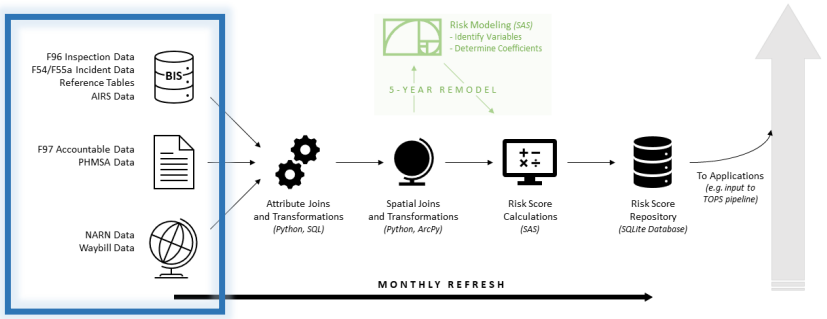
- FRA's data pipeline incorporates geospatial processes to quickly associate all available data per location.

- FRA's data pipeline repeatedly and regularly brings current safety data into the risk models.

Refined Process and Agenda



1 Centralized Data Sources

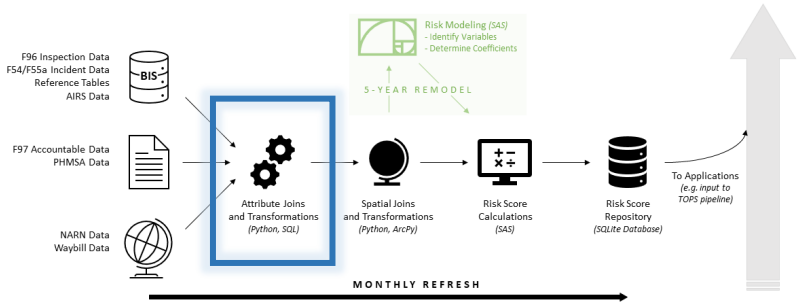


- FRA's current data lake ingests nightly backups from each system of record.
- Risk models benefit from these feeds to regularly update the resulting scores (monthly).

Inputs

- Filed **inspection** reports
- Reported **incident** data (equipment and illness/injury)
- Asset Inventory of Railroad Shippers (**AIRS**) – inspection location base data
- **Accountable** incidents (do not meet regulatory reporting threshold)
- **Hazardous material** release incidents (PHMSA NAR)
- North American Rail Network (**NARN**) base data
- **Waybill** sample, flowed on the NARN

2 Query, Import, and Transform (Build) Current Data



- Python-scripted queries and transformations extract the data relevant to **each discipline**.
- **Close engagement** with specialists and inspectors in each discipline eliminates irrelevant or erroneous input data.

Example: Operating Practices (OP)

- Equipment Incidents (Form 54)
  - Focused on reducing **human factor-caused** reportable equipment incidents.
  - Risk considers the rate of **cause-code 'H'** incidents.

Cause-Code H

- Exceptions (if any)

- Accountables (Form 97)
  - Incidents of a certain cause code *may* be later determined to be human-factor and are **added** from F97.

+ Accountables (subset)

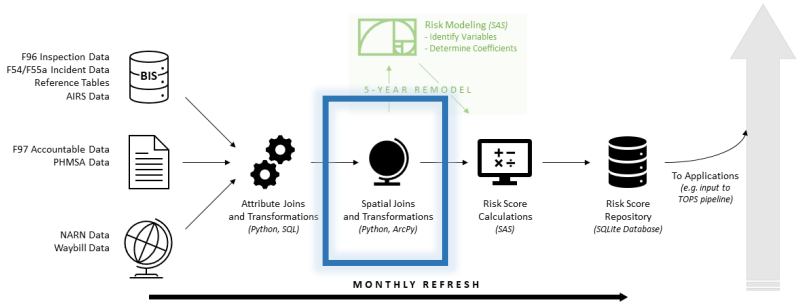
- Inspections (Form 96)
  - Operating Practices inspectors are **type 'O'.**
  - But **activities** of certain types are *not* relevant to risk (excluded).
  - A subset of **critical defects** (inspection results) are elevated in weight.

Type 'O' Inspections

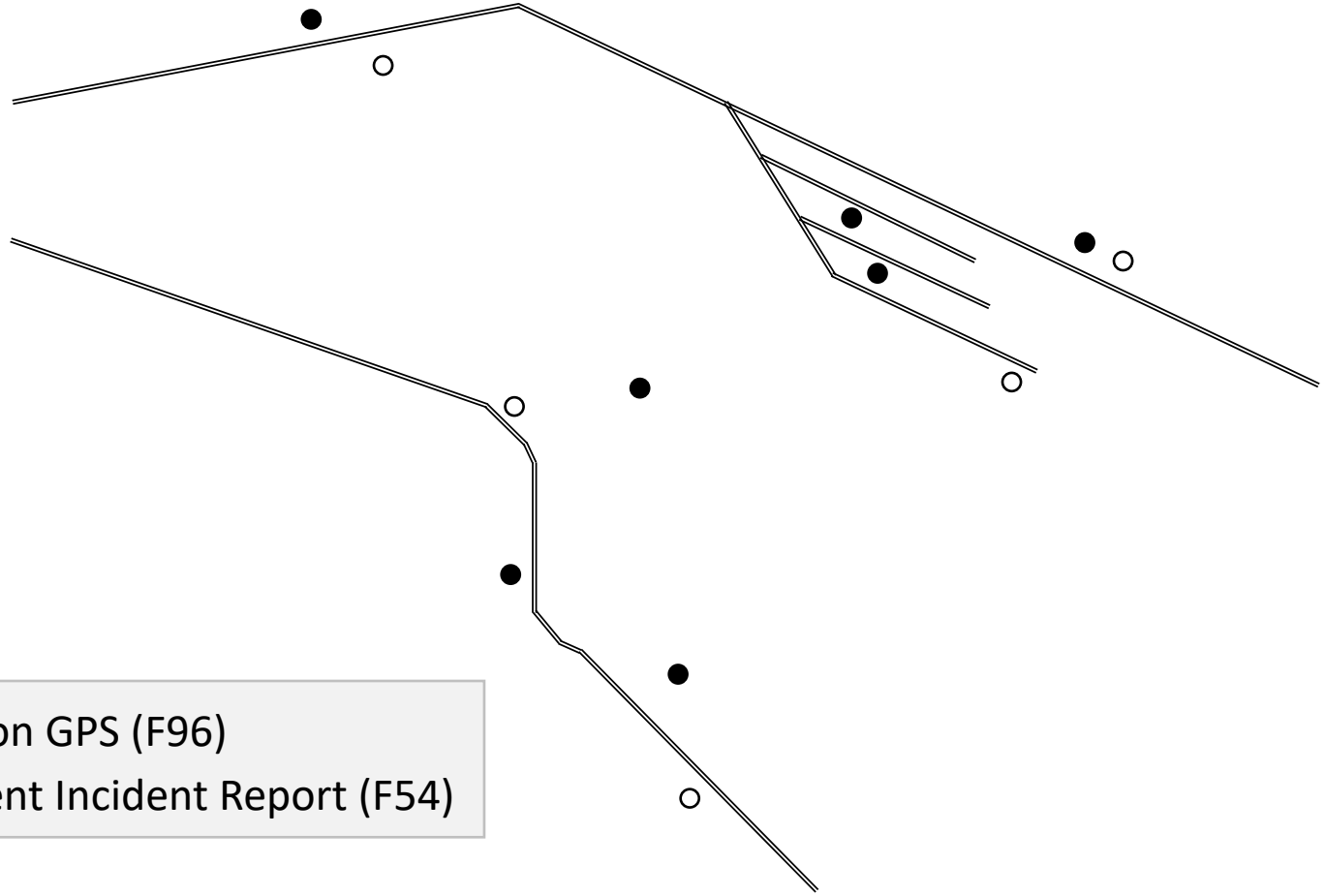
- Excl. Activities

Flag Critical Defects

3 Per-Discipline Spatial Analyses and Summaries

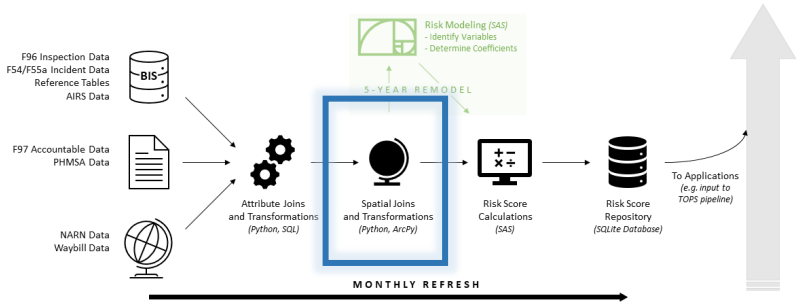


- Geospatial processes summarize **disassociated data**, maximizing the breadth of inputs.
- GIS **quickly and accurately** links multiple events (which do not reference any inventory) to a single geospatial asset.

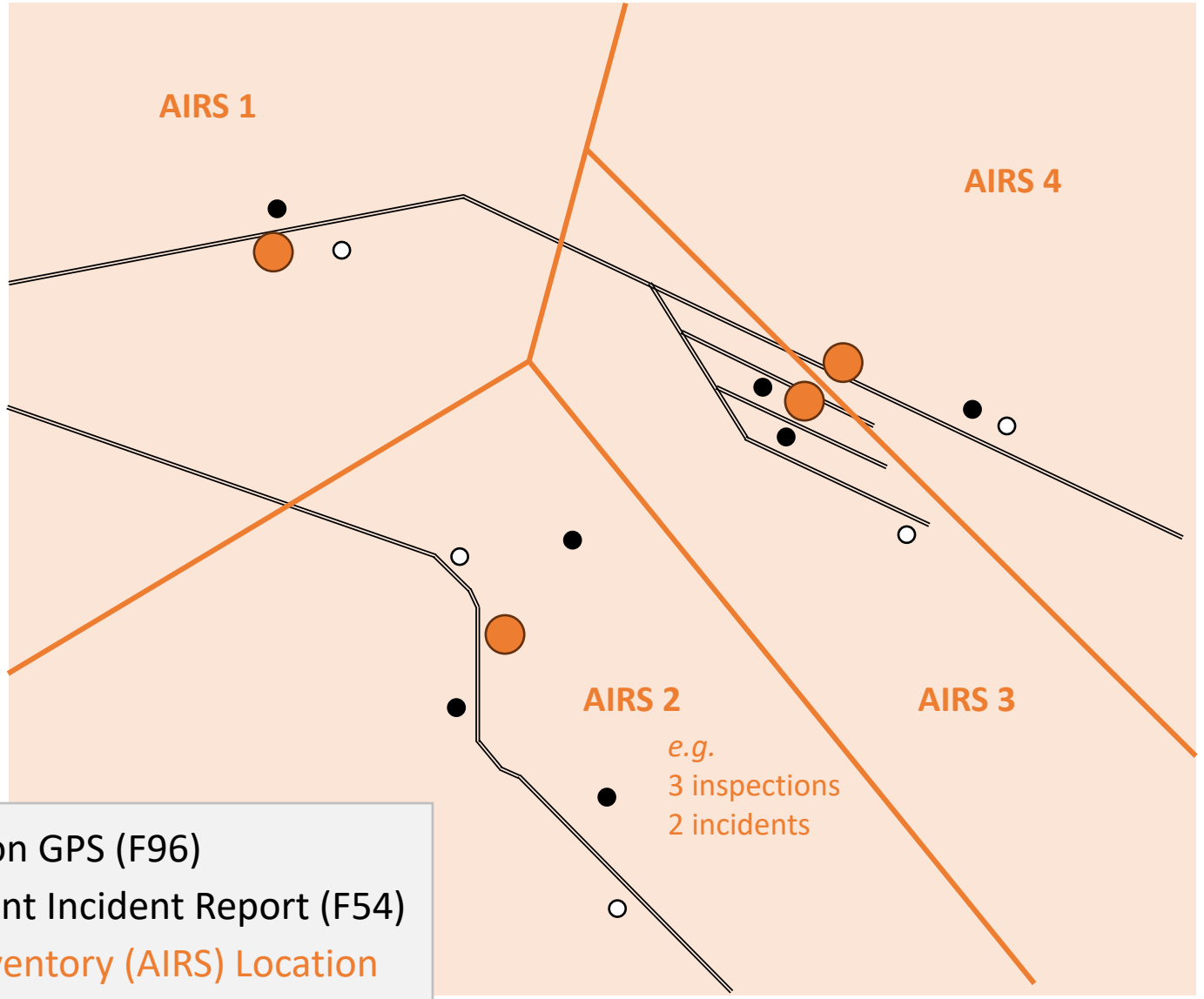


- ● Inspection GPS (F96)
- ○ Equipment Incident Report (F54)

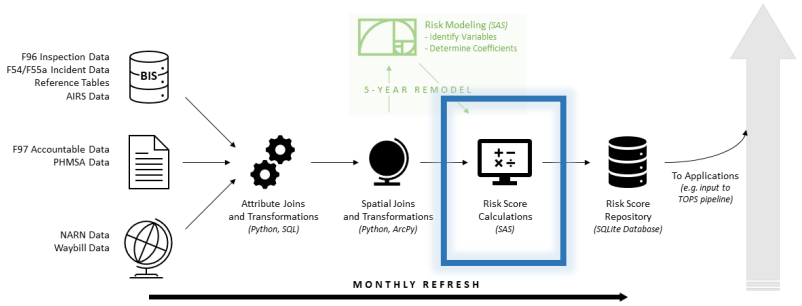
3 Per-Discipline Spatial Analyses and Summaries



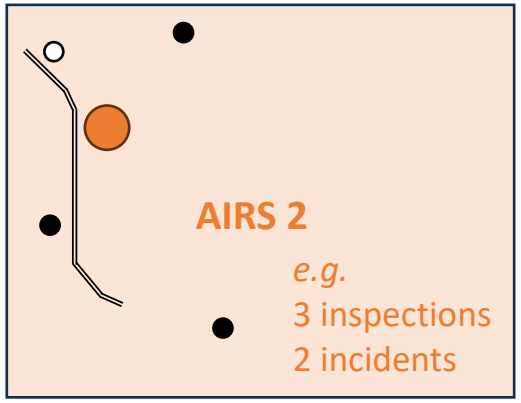
- Geospatial processes summarize **disassociated data**, maximizing the breadth of inputs.
- GIS **quickly and accurately** links multiple events (which do not reference any inventory) to a single geospatial asset.



4 Apply Risk Model to Refreshed Summary Data



- **SAS-scripted calculations** take spatially summarized data per AIRS location as input…



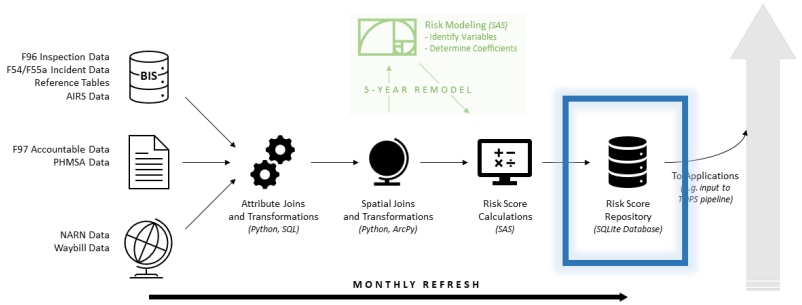
- …and output an **updated risk score** (0-100) reflecting probability and severity of an incident at each AIRS location.

AIRS 2 Risk Score: **78**

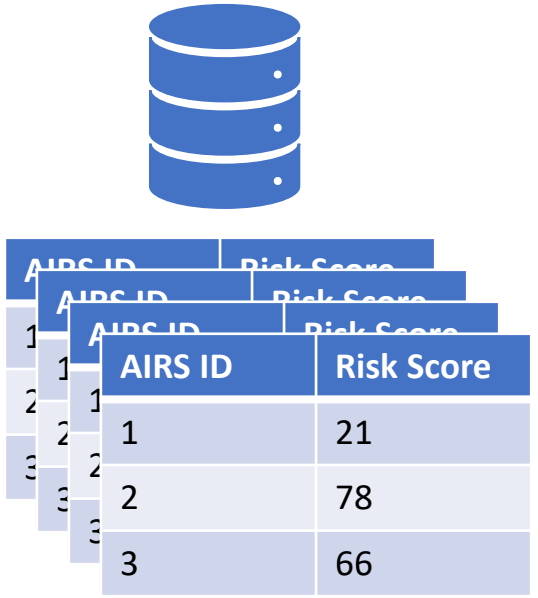
- **Driving inspections**, which ideally result in improvements that reduce incidents and therefore risk.

AIRS 2
5 inspections
1 incident

5 Write Refreshed Risk Scores to Repository



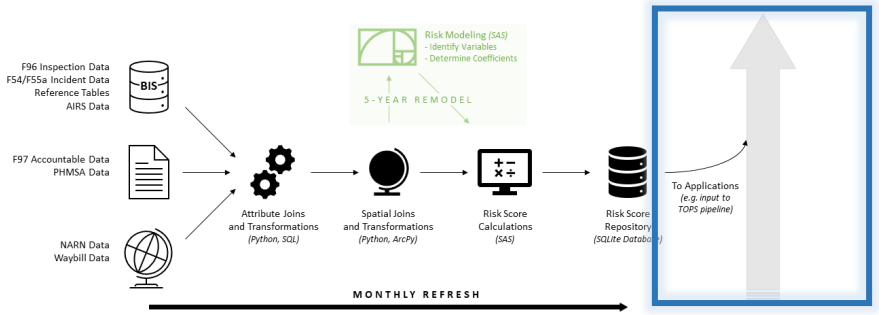
- After calculating refreshed results, risk scores are saved to a **risk repository database**.
- Keeps a **history** of risk scores per AIRS point.
- Makes current scores **available** to other applications.



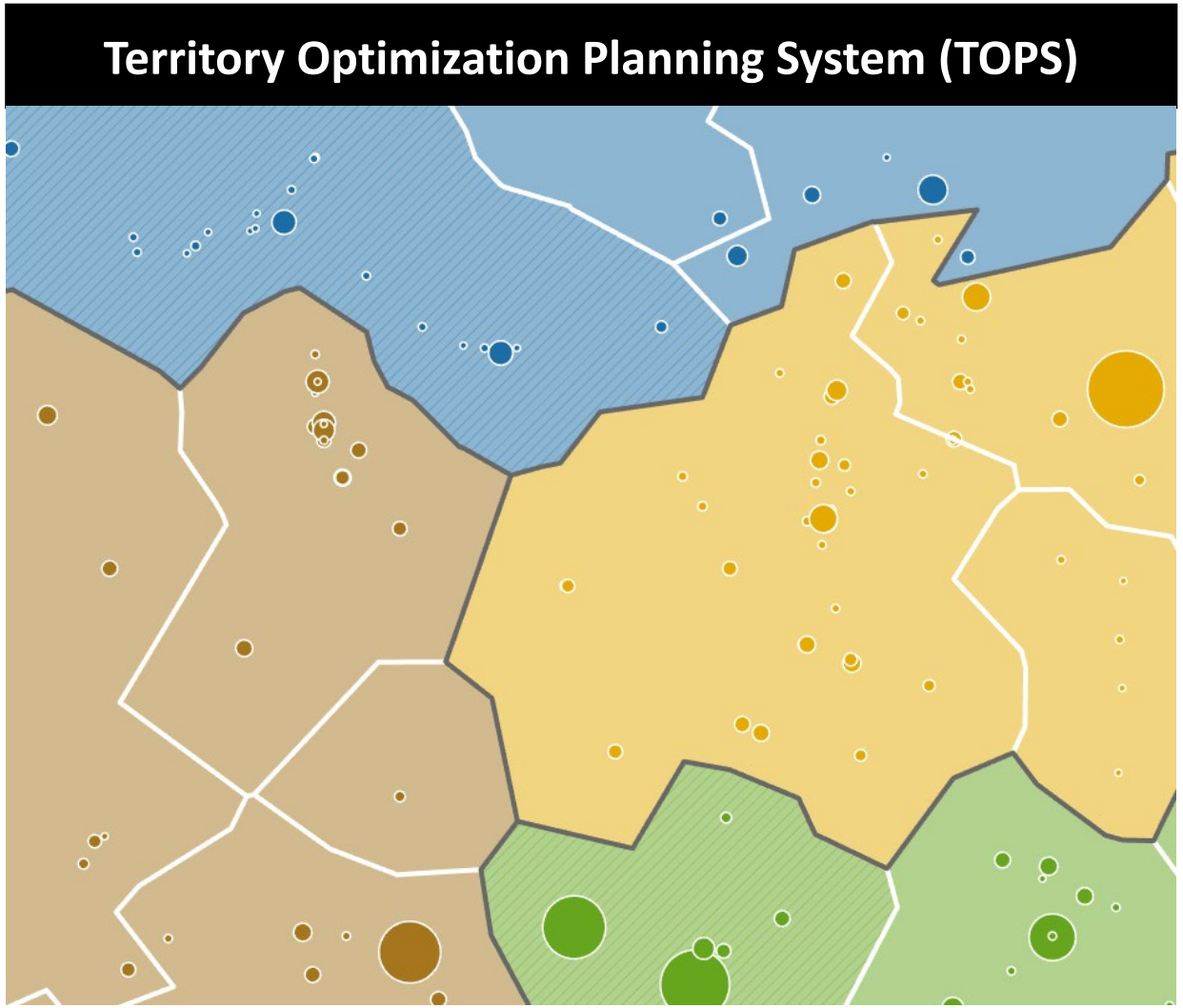
AIRS ID	Risk Score
1	21
2	78
3	66

- **Mapping**
- **Analysis**
- **Other Databases**
- **Reporting**

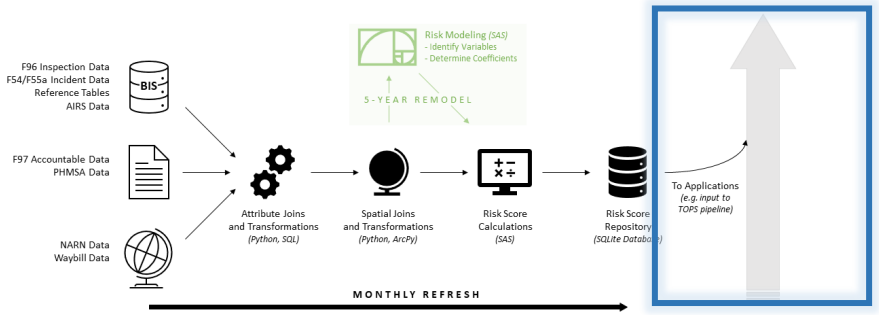
6 Presentation through Web Map Applications



- In TOPS, AIRS points are updated monthly, sized by relative risk, and provided to the inspectors.
- Larger points have a higher likelihood / greater severity of incidents.

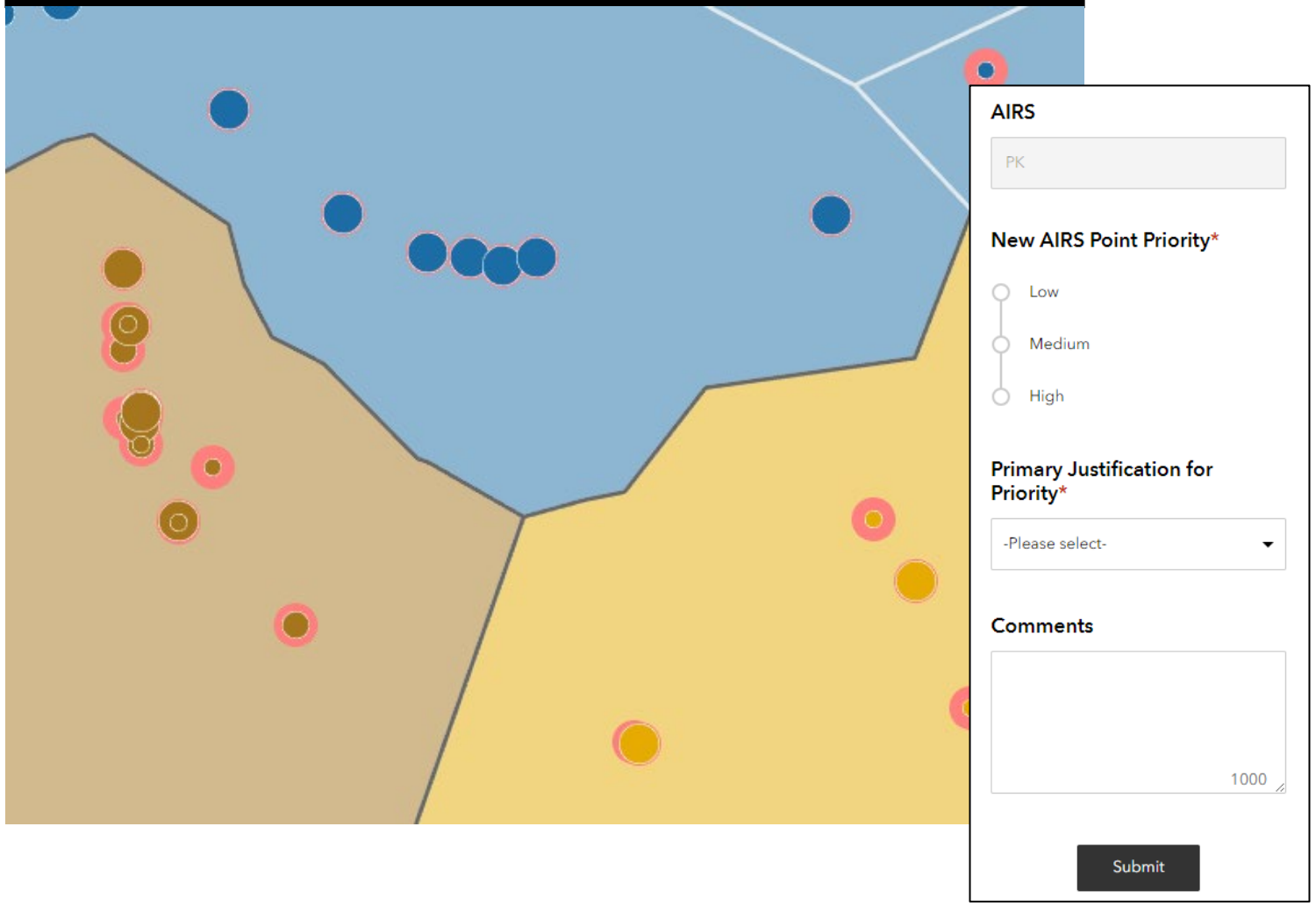


6 Presentation through Web Map Applications

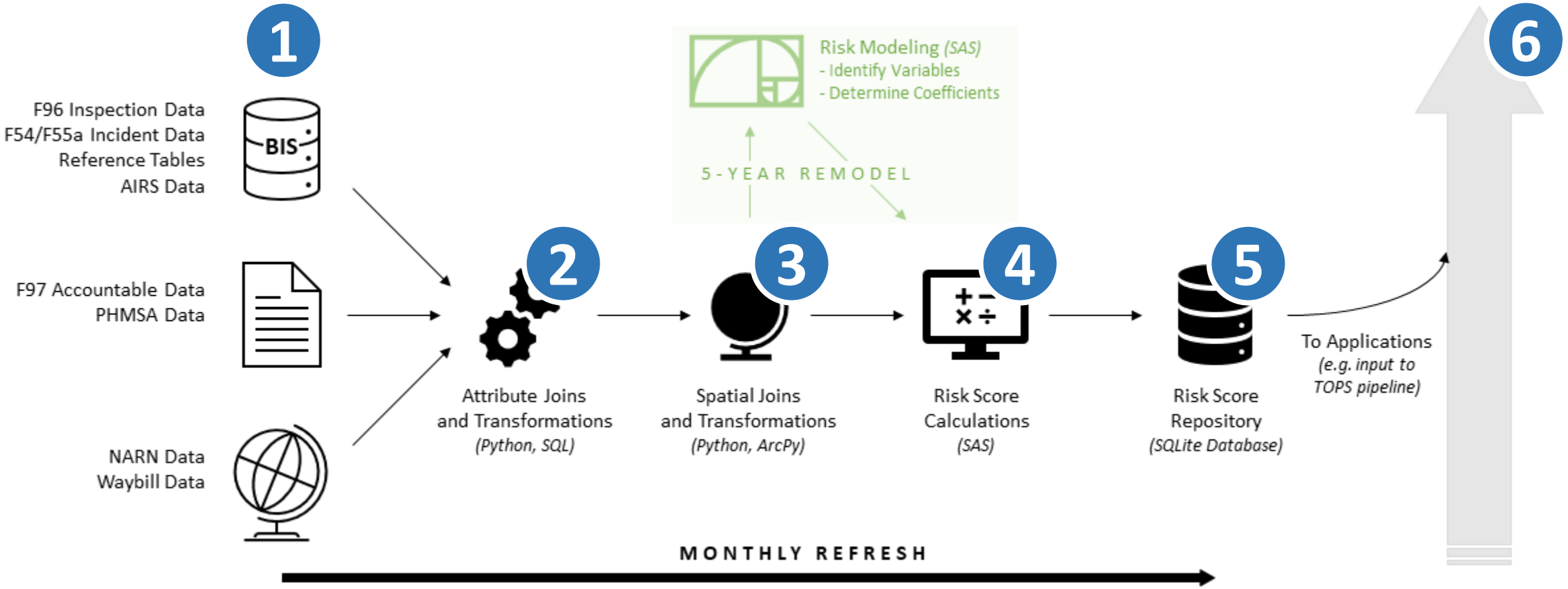


- Also through TOPS, inspectors can designate priorities, further modifying the risk score.
- AIRS points with red halos have a priority score: high, medium, or low – a data source for risk model validation and improvements.

Territory Optimization Planning System (TOPS)



Summary


1
F96 Inspection Data
F54/F55a Incident Data
Reference Tables
AIRS Data
BIS
F97 Accountable Data
PHMSA Data
NARN Data
Waybill Data
Risk Modeling (SAS)
- Identify Variables
- Determine Coefficients
5-YEAR REMODEL
2
Attribute Joins
and Transformations
(Python, SQL)
3
Spatial Joins
and Transformations
(Python, ArcPy)
4
Risk Score
Calculations
(SAS)
5
Risk Score
Repository
(SQLite Database)
To Applications
(e.g. input to
TOPS pipeline)
6
MONTHLY REFRESH

Contact Us

Federal Railroad Administration
1200 New Jersey Avenue, SE
Washington, DC 20590

Connect with us at USDOTFRA

U.S. Department of Transportation
Federal Railroad Administration

Andrew LaBounty
Phone: 331-758-4097
Email: andrew.labounty@dot.gov

Patrick Johnson
Phone: 771-233-1651
Email: patrick.johnson@dot.gov

Ruby Li
Phone: 202-493-8017
Email: ruby.li@dot.gov

Tony Ye
Phone: 202-493-0531
Email: jianqiang.ye@dot.gov