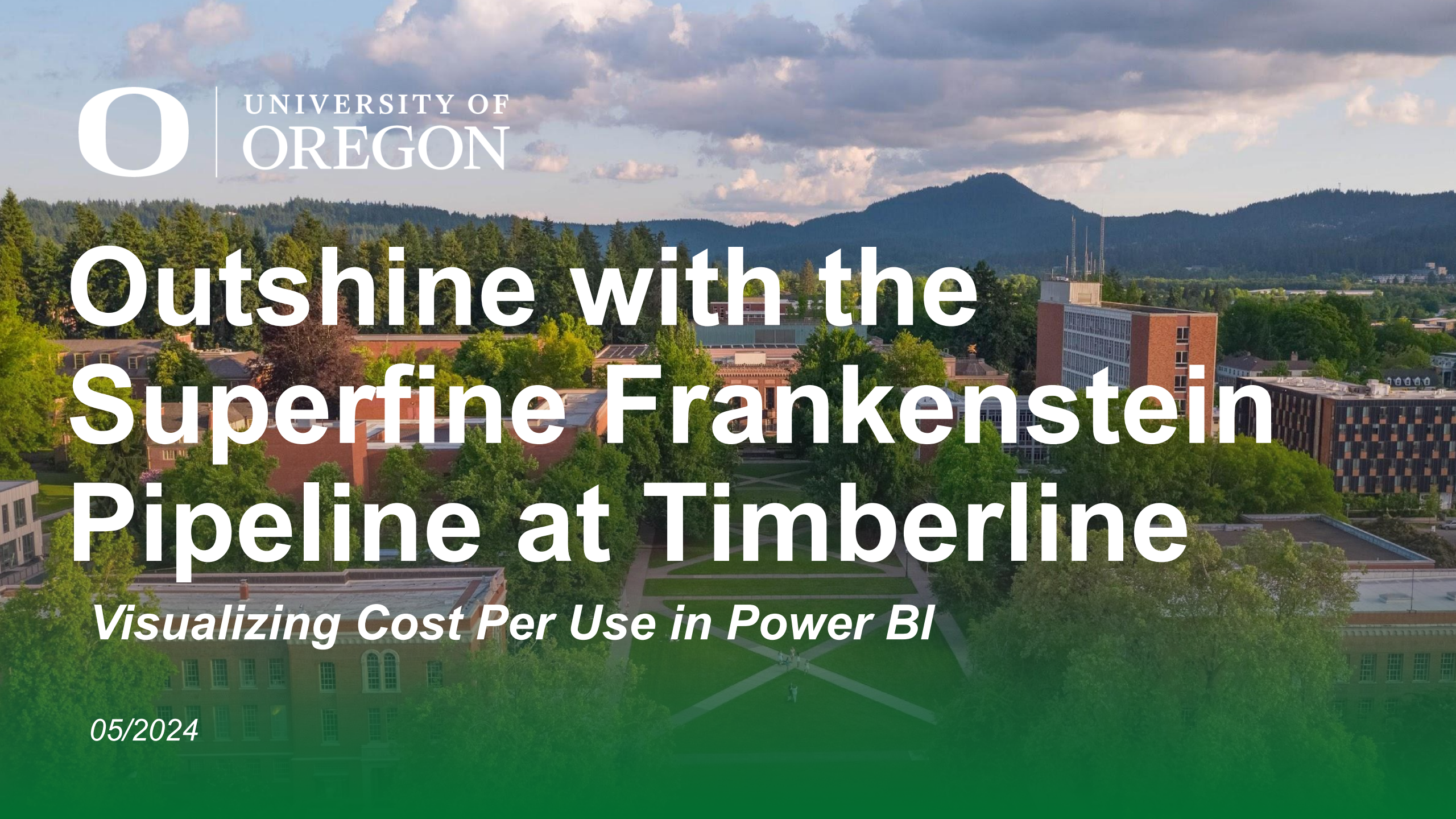


Outshine with the Superfine Frankenstein Pipeline at Timberline

Visualizing Cost Per Use in Power BI

05/2024

Introductions

- **Lydia Harlan**

Budget Analyst for Collections, Discovery, and Digital Strategy

- **Kristin Buxton**

Head of Science Liaisons

- **Gabriele Hayden**

Head of Data Services

UNIVERSITY OF OREGON

Roadmap

- Why we built a cost per use model in Power BI
- Our iterative process
- Views from our dashboard
- Ongoing work
- Discussion and Questions

UNIVERSITY OF OREGON

Discussion

- How do you gather usage data?
- Do you do it in house or through a vendor?
- How do you share it?
- Are any of you using Power BI or another data visualization tool?
- If so, what information are you gathering and presenting?

What Did We Do and Why Did We Do It?

Why: No comprehensive cost per use data, manual usage stats collection, and little analysis

What: Could we use Power BI to create a cost per use model?

Who: “Change will not come if we wait for some other person or some other time. We are the ones we've been waiting for. We are the change that we seek.” (Thanks Obama.)

When: May 2022 – May 2023 but also some pre-work starting November 2021.

UNIVERSITY OF OREGON

A Technical Adventure

We began this project not knowing what technology we needed. We each brought existing skills in Excel, plus

- Lydia: SQL queries in our ILS reporting module (Alma Analytics), access to previous Cost Per Use spreadsheets
- Kristin: former software engineer (mostly C++), a "customer" for the finished project
- Gabriele: basic Power BI and Python

We each contributed based on our skill set.

UNIVERSITY OF OREGON

Along the way we learned

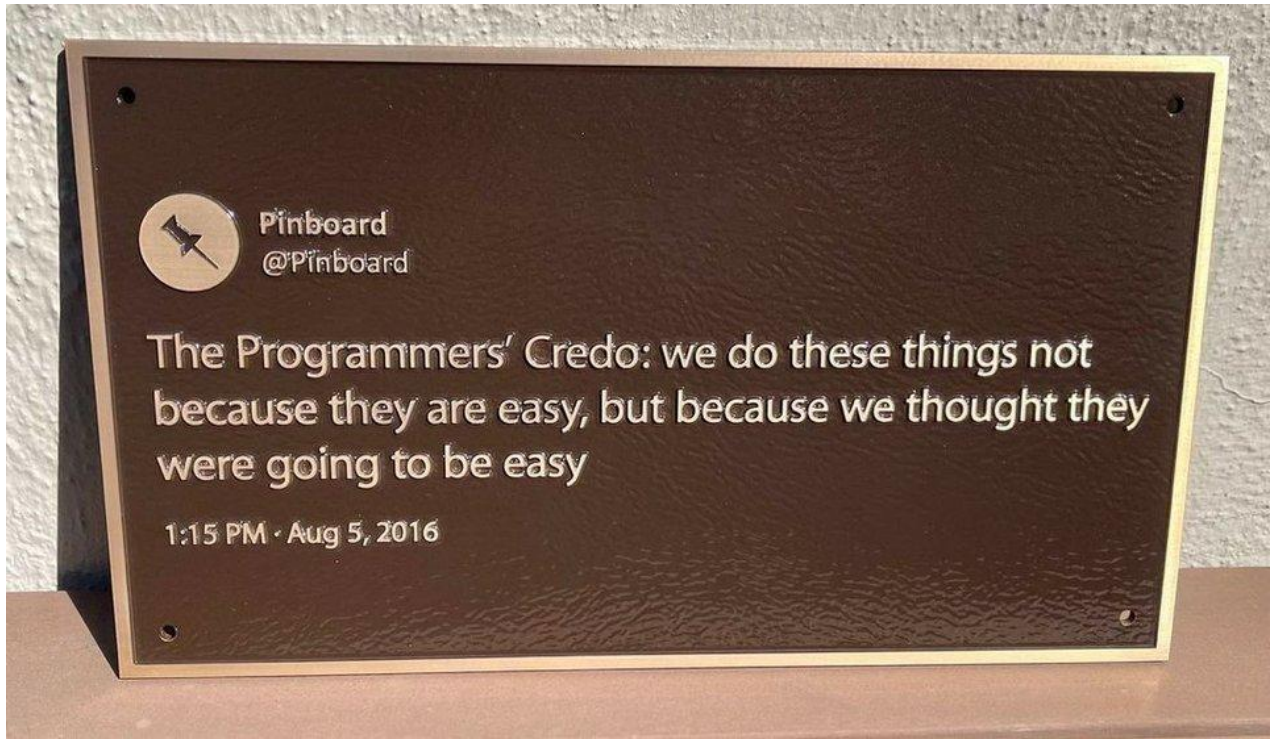
- How APIs work, the SUSHI API protocol
  - reading articles and documentation

- Power BI languages Power Query (M) and DAX
  - 16 hours of formal training, reading documentation, troubleshooting

Plus individual skill development in Python (Gabriele) and Alma Analytics (Lydia)

How

- Get cost
- Get use
- Join, transform data into Cost Per Use
- Turn static data loads into API queries
- Make it look pretty
- Share it
- Use it



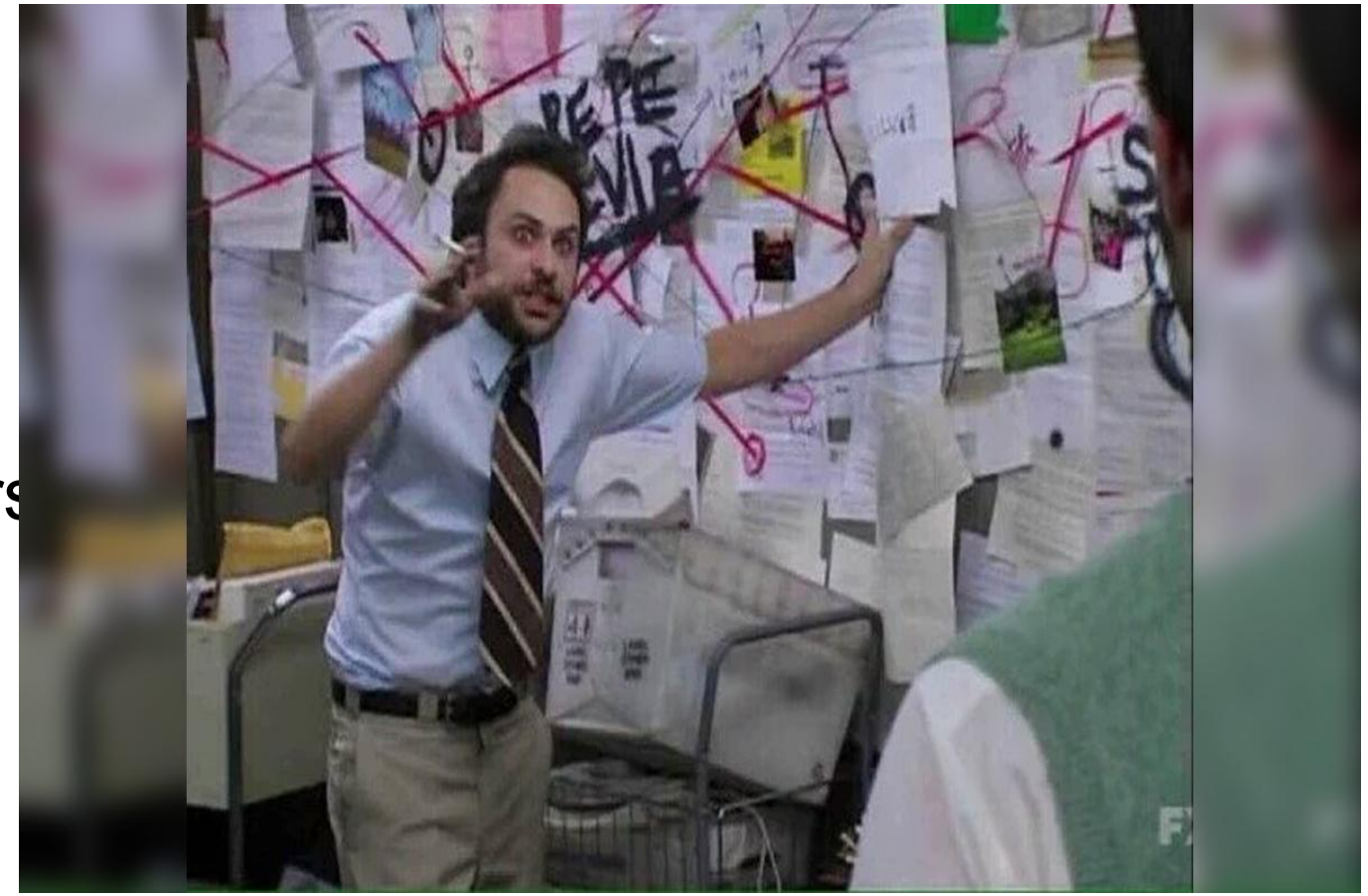
Wait, how?

It’s all connected…somehow.

Disparate documentation.

Documentation is for developers, not us.

Multiple dependencies.



Approaches to the Pipeline

UNIVERSITY OF OREGON

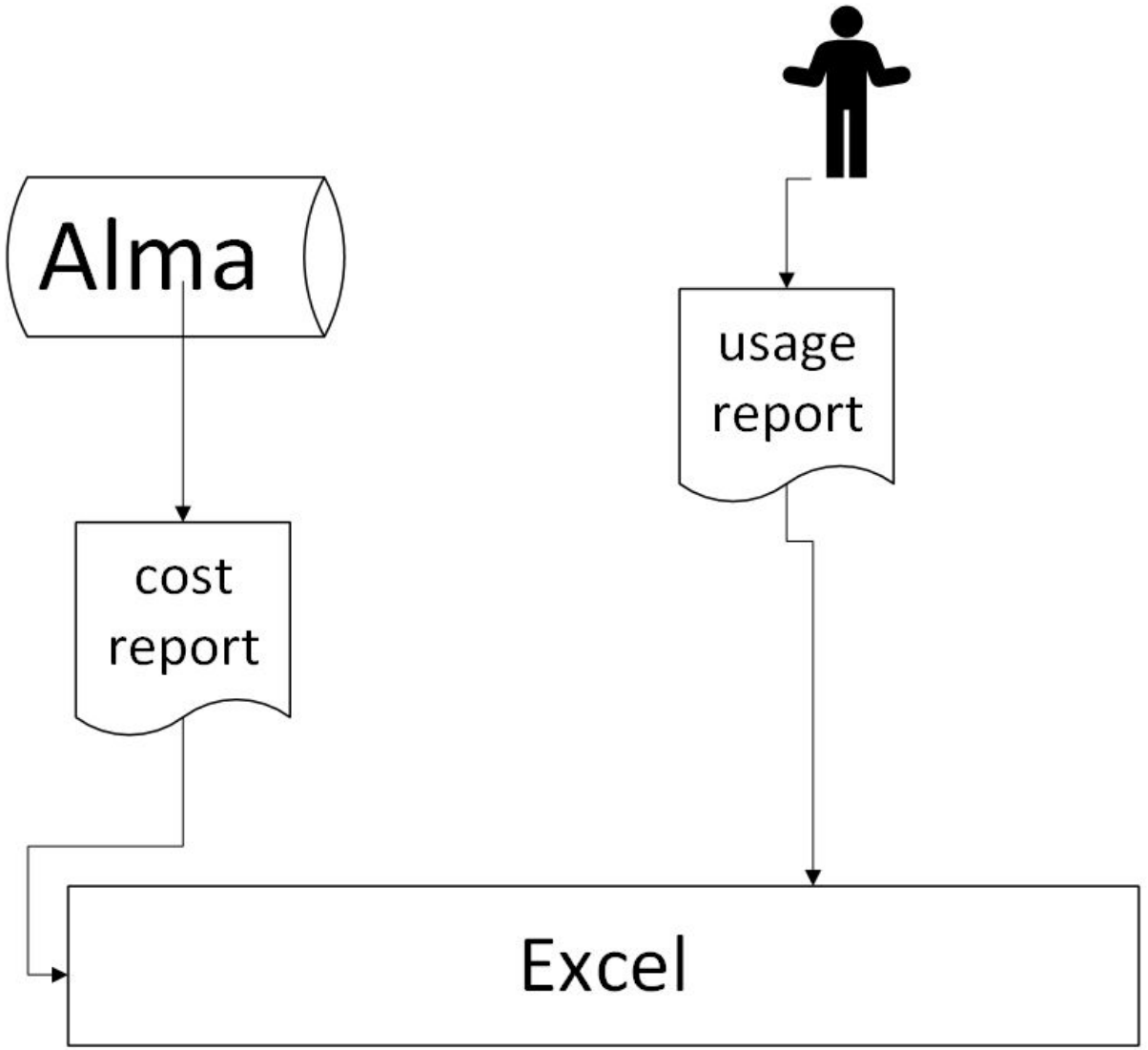
Technical Definitions

- **COUNTER 5**
- **SUSHI**
- **API**

Technical Definitions

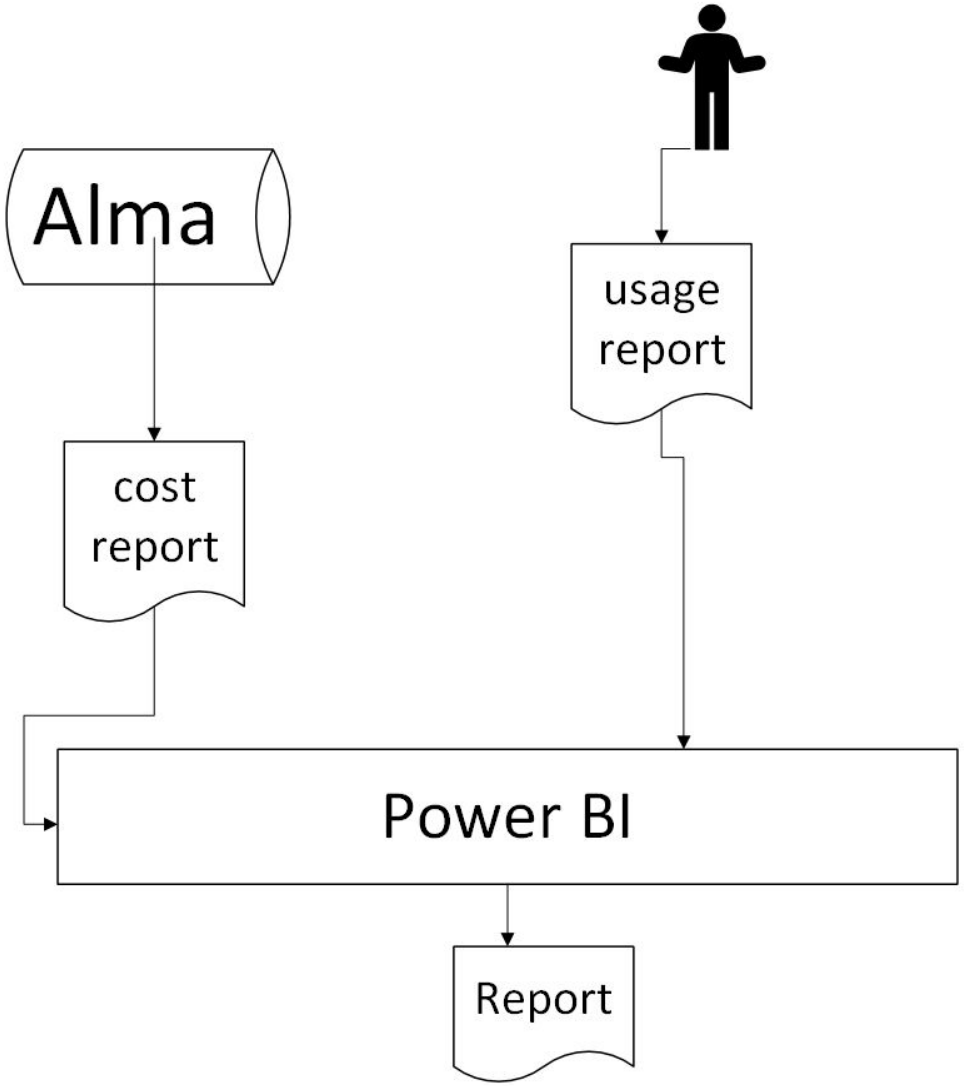
- **COUNTER 5** – Standard for reporting usage statistics for electronic resources
- **SUSHI** – An API for retrieving COUNTER 5 usage reports
- **API** – Application programming interface. A way to exchange data between computers/servers programmatically

Proof of concept in Excel



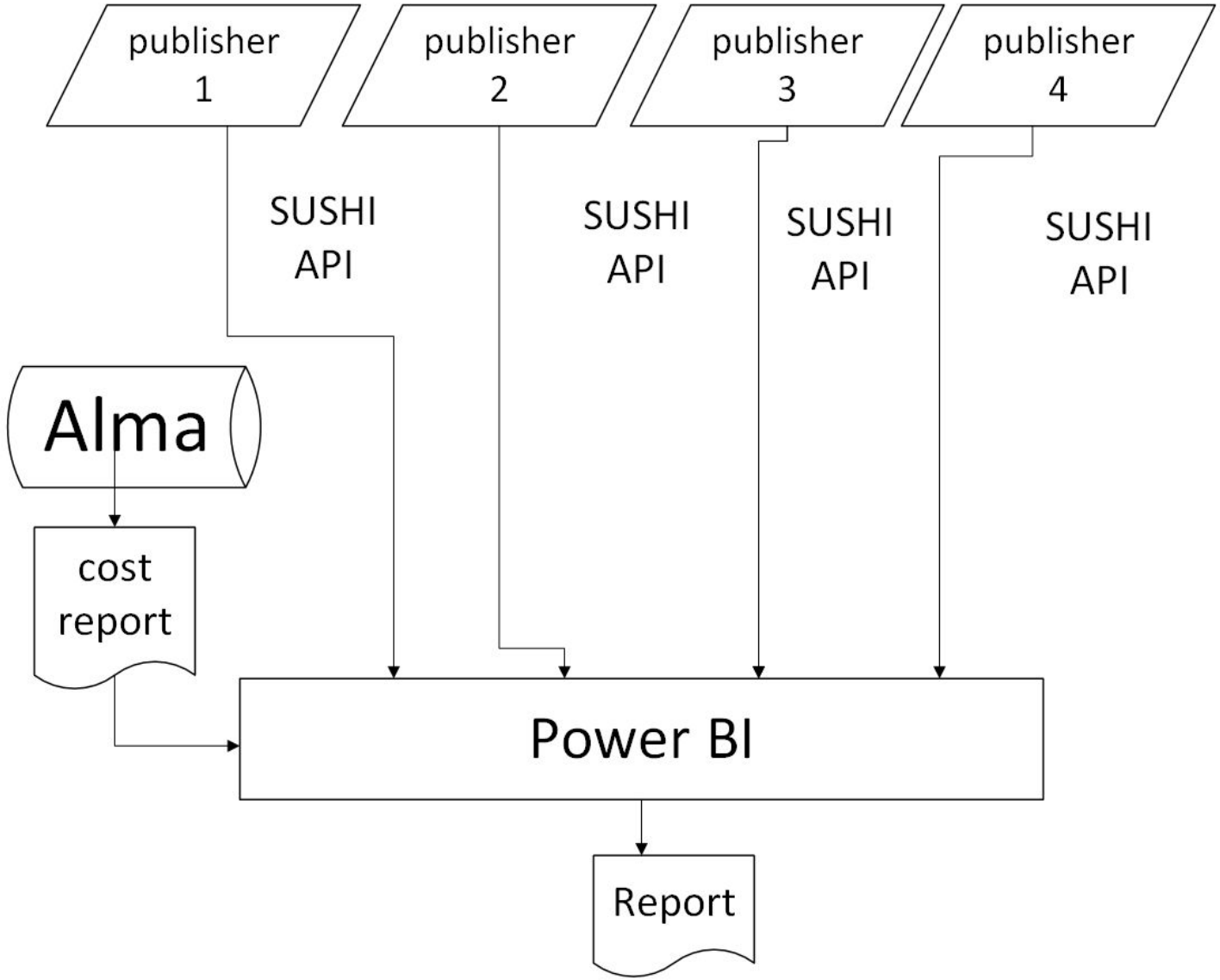
Proof of concept in Power BI

UNIVERSITY OF OREGON



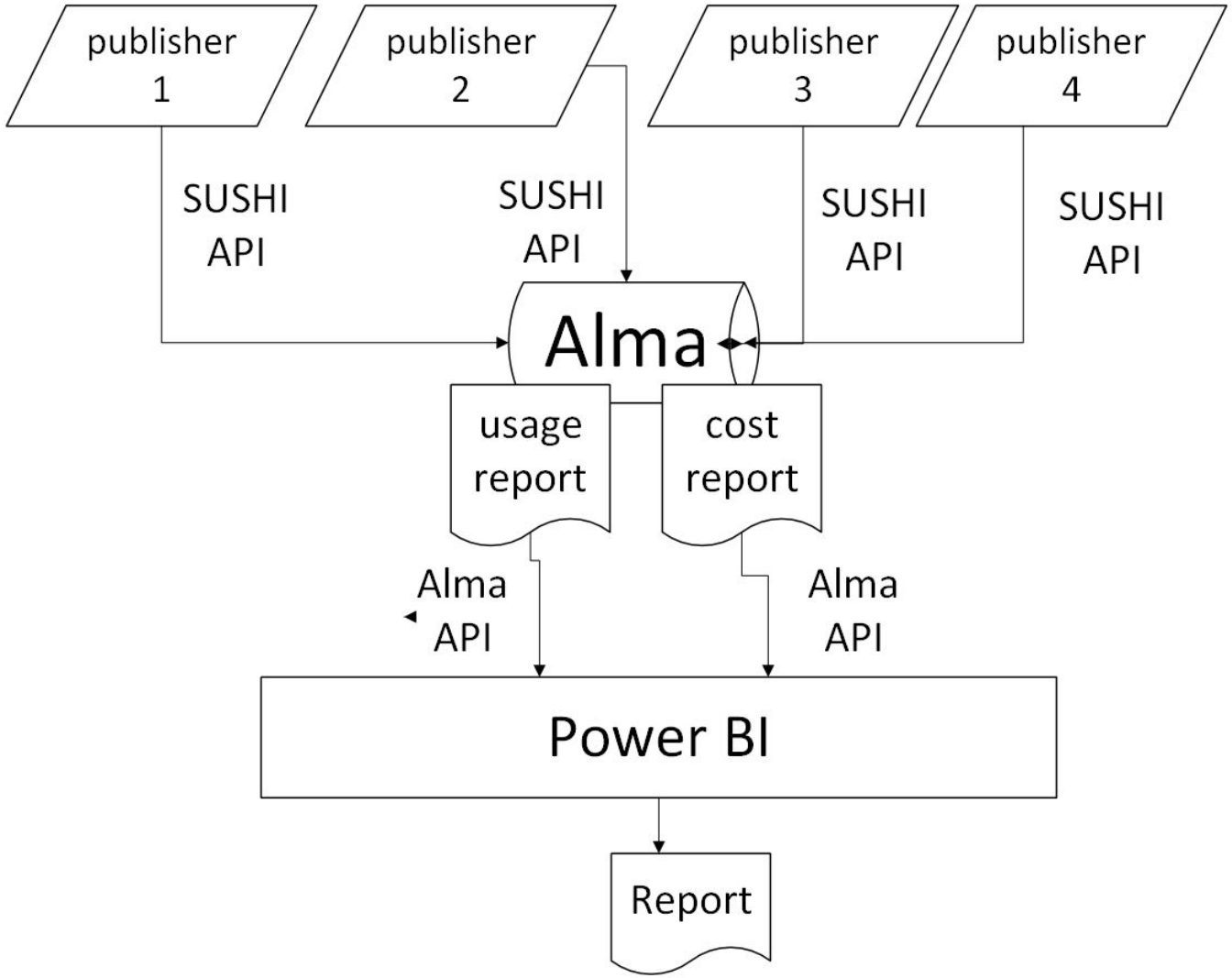
Query SUSHI from Power BI?

UNIVERSITY OF OREGON



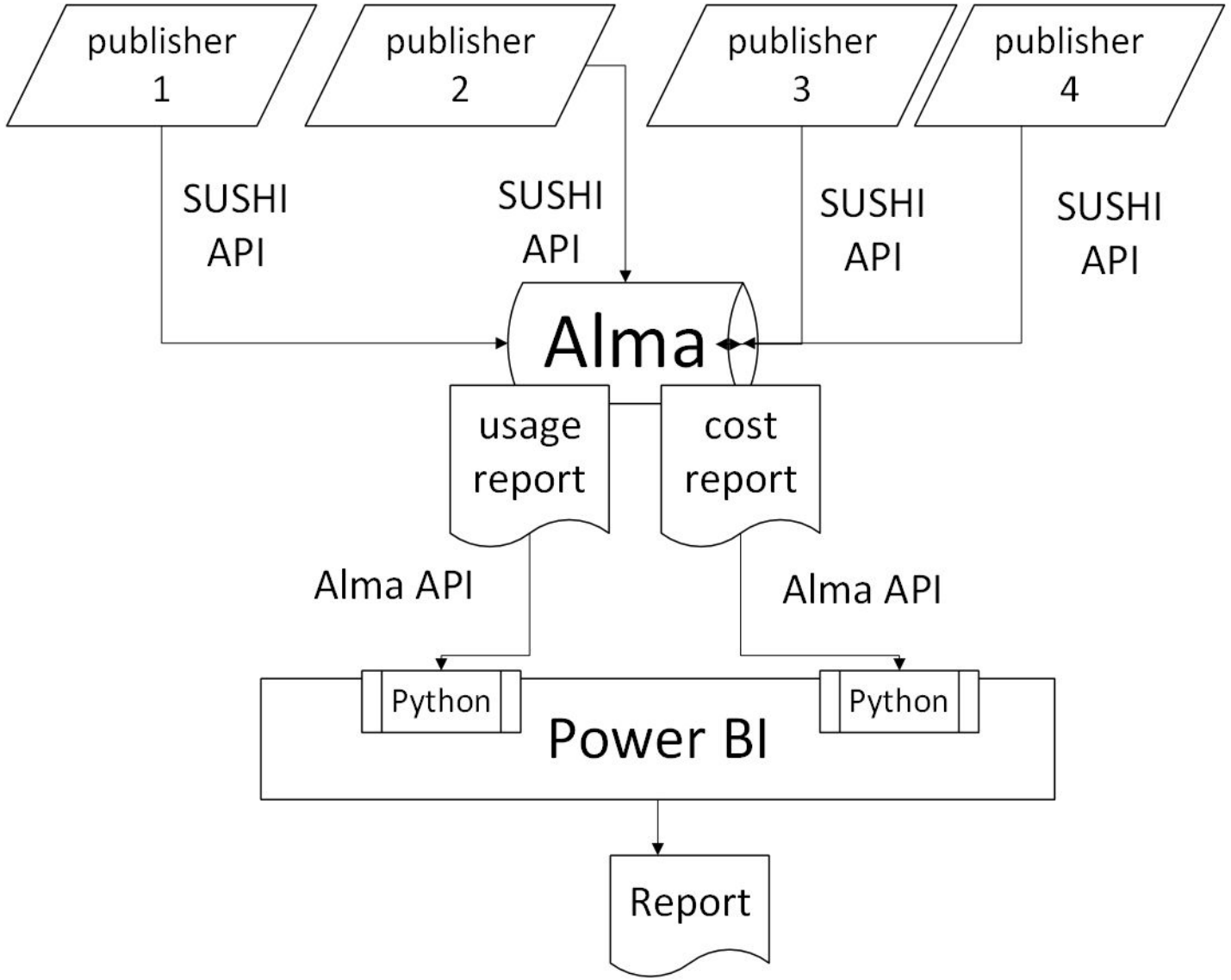
No! Query SUSHI from Alma

UNIVERSITY OF OREGON



Solution! Use Python for Power BI query

UNIVERSITY OF OREGON

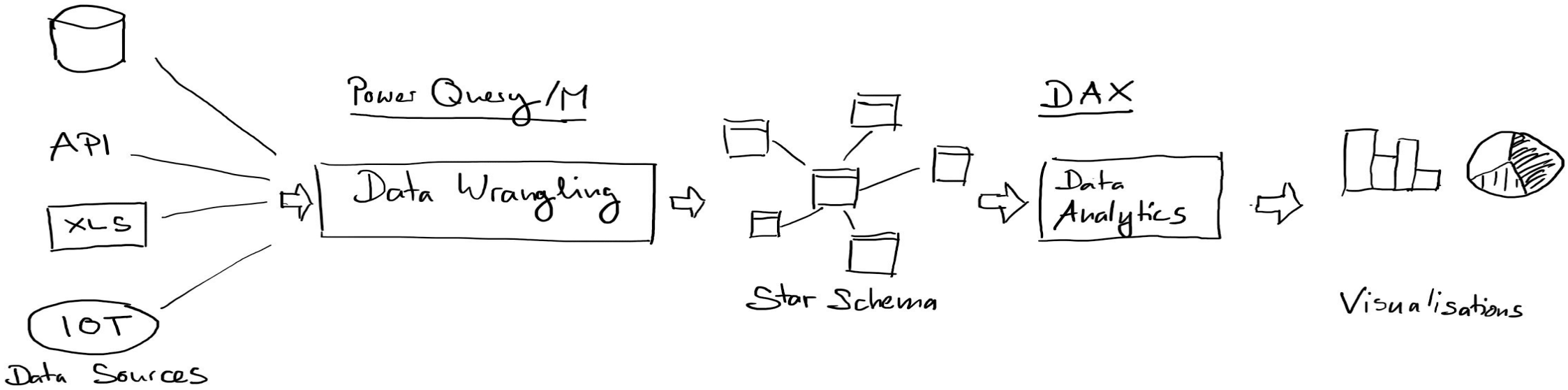


Our Dashboard

UNIVERSITY OF OREGON

The Guts of Power BI

UNIVERSITY OF OREGON



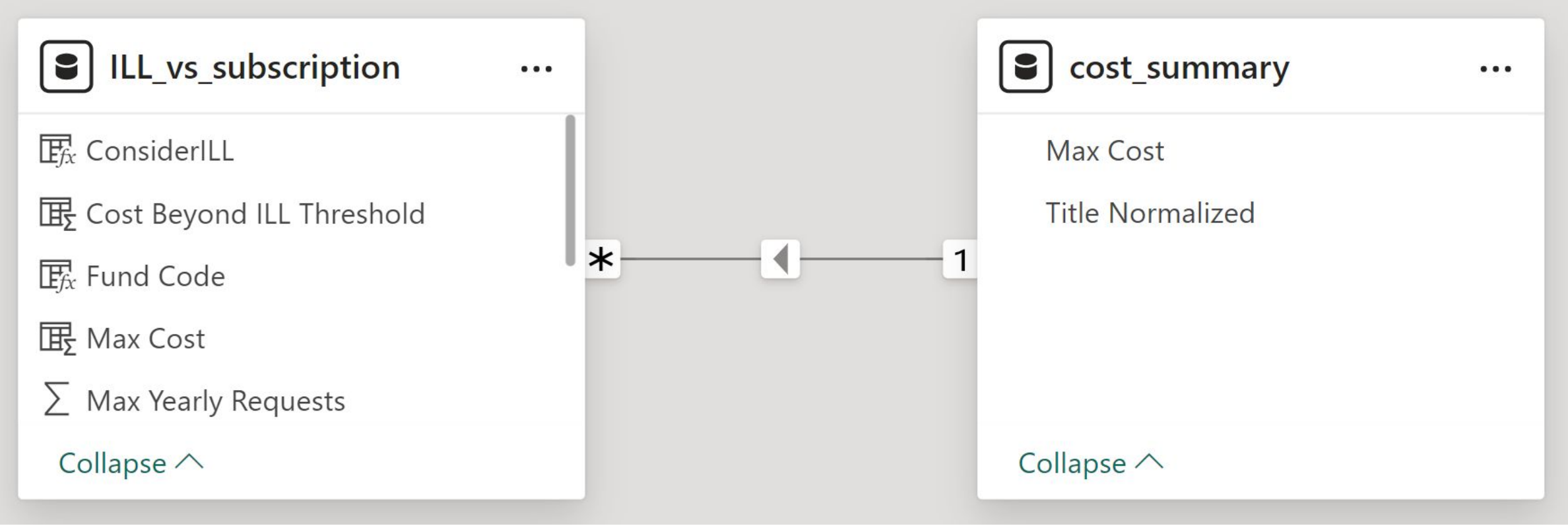
UNIVERSITY OF OREGON

Power Query Interface showing M Code

cost_data

```
let
    Source = Python.Execute("# packages#(lf)#(lf)import requests # for API calls#(lf)import xml.etree.ElementTree as ET # to par
    df1 = Source{[Name="df"]}[Value],
    #"Changed Type" = Table.TransformColumnTypes(df1,{{"0", Int64.Type}, {"ISSN", type text}, {"Title (Normalized)", type text},
    #"Removed Columns" = Table.RemoveColumns(#"Changed Type",{"0", " REPORT_SUM(Transaction Expenditure Amount BY Title (Normali
    #"Renamed Columns" = Table.RenameColumns(#"Removed Columns",{{"Title (Normalized)", "Title Normalized"}}),
    #"Replace Beginning the with nothing" = Table.ReplaceValue(#"Renamed Columns", each [Title Normalized], each if Text.StartsW
    #"Replace Beginning a with nothing" = Table.ReplaceValue(#"Replace Beginning the with nothing", each [Title Normalized], eac
    #"Replaced Value" = Table.ReplaceValue(#"Replace Beginning a with nothing","-","",Replacer.ReplaceText,{"ISSN"}),
    #"Split Column by Delimiter" = Table.SplitColumn(#"Replaced Value", "ISSN", Splitter.SplitTextByDelimiter("; ", QuoteStyle.C
    #"Changed Type1" = Table.TransformColumnTypes(#"Split Column by Delimiter",{{"ISSN.1", type text}, {"ISSN.2", type text}}),
    #"Renamed Columns1" = Table.RenameColumns(#"Changed Type1",{{"ISSN.1", "ISSN1"}, {"ISSN.2", "ISSN2"}}),
    #"Added Conditional Column" = Table.AddColumn(#"Renamed Columns1", "Fund Code Display", each if Text.Contains([Fund Code], "
    #"Filtered Rows" = Table.SelectRows(#"Added Conditional Column", each true)
in
    #"Filtered Rows"
```

Model view shows relationships between tables



UNIVERSITY OF OREGON

Python code written and tested outside of Power BI

```
87      # Extract the column names from the data
88      # Get the schema namespace (would it be better to pull this from the xml itself?)
89      ns = {'xsd': 'http://www.w3.org/2001/XMLSchema',
90            'saw-sql': 'urn:saw-sql'}
91
92      # Get the column headers (human-readable names) and
93      # names (the XML tag names) from the schema as a list
94      column_names = [elem.attrib['{urn:saw-sql}columnHeading']
95                      for elem in root.findall('.//xsd:element', ns)]
96      column_xml = [elem.attrib['name'] # elem.attrib doesn't have 'saw-sql' b/c not in xml
97                      for elem in root.findall('.//xsd:element', ns)]
```

Cost Per Use

UNIVERSITY OF OREGON

What's in here, how does it work

Power BI data pipeline matches cost and use data on ISSN and Title using M (Power Query) and DAX languages.

- Lots of details to get clean data.

Calculate cost per use in a fuzzy way.

- most expensive year and highest use year
- rounding
- holes in the data

Consider ILL

UNIVERSITY OF OREGON

Normalized Title	ConsiderILL	Max Yearly Requests	Max Cost	Cost Beyond ILL Threshold	Fund
[illegible]	True	11	$313.50	$0	Other
[illegible]	True	59	$2,420.79	$1,365	Social Sciences
[illegible]	True	7	$10,843.37	$0	Science
[illegible]	True	4	$8,505.30	$0	Science
[illegible]	True	4	$799.40	$0	Science
[illegible]	True	32	$861.05	$420	Social Sciences
[illegible]	True	2	$863.77	$0	Social Sciences
[illegible]	True	14	$452.10	$0	Design
[illegible]	True	10	$265.43	$0	Social Sciences
[illegible]	True	2	$1,277.89	$0	Social Sciences
[illegible]	True	15	$385.31	$0	Social Sciences
[illegible]	True	152	$10,572.72	$1,960	Science
[illegible]	True	7	$313.50	$0	Other
[illegible]	True	67	$1,645.03	$1,645	Social Sciences
[illegible]	True	11	$141.80	$0	Other
[illegible]	True	8	$689.70	$0	Science
[illegible]	True	4	$4,314.47	$0	Social Sciences
[illegible]	True	13	$359.33	$0	Other
[illegible]	True	15	$104.59	$0	Social Sciences
[illegible]	True	7	$668.90	$0	Other
[illegible]	True	5	$120.18	$0	Other
Total		**7763**	**$725,293.77**	**$141,365**	

Consider ILL Formula

This is our formula for Cost Beyond ILL Threshold, assuming documents through Reprints Desk cost $35 and we can request an item 20 times before having to pay for copyright:

Cost Beyond ILL Threshold = IF('cost_data'[Electronic Use MAX]-20<0,0, ('cost_data'[Electronic Use MAX]-20)*35)

The Consider ILL column is a simple True / False statement:

ConsiderILL = IF(AND('cost_data'[Cost Beyond ILL Threshold] < 'cost_data'[Transaction Expenditure Amount],cost_data[Electronic Use MAX] > 0), TRUE, FALSE)

Maintenance, Replication, and Future Work

UNIVERSITY OF OREGON

Ongoing maintenance

- Every month, after Alma performs the SUSHI harvest, load the Power BI model and refresh the data.
- Update vendor usernames and passwords, SUSHI credentials, API keys, and base URLs as needed.
- Respond to changes in platform functionality that can break report.
- Add new vendors as they enable COUNTER 5 reporting.

UNIVERSITY OF OREGON

Skills & tech you need to replicate this project

- Alma, Alma Analytics, COUNTER 5, APIs, SUSHI, Python, and Power BI.
- Power BI Desktop requires a Windows PC.
- Basic familiarity with programming with SQL in Alma Analytics, with M and DAX in Power BI, and with Python
- Infinite patience
- Friends

See our article coming soon!

Our code: github.com/uodataservices/cost_per_use

Future work

- Add more report types beyond tr_j1.
- Database evaluation is particularly tricky. Open to ideas.
- How to manage non-COUNTER compliant data?

UNIVERSITY OF OREGON

Discussion

What insights/data points/criteria do you use to make data driven decisions?

What are we missing in our current dashboard?

Where could you see a data dashboard helping conversations at your organization?

Questions?

(and cat tax)

Lydia Harlan, lharlan@uoregon.edu
Kristin Buxton, kbuxton@uoregon.edu
Gabriele Hayden, ghayden@uoregon.edu

UNIVERSITY OF OREGON

References

The New York Times. (2008, February 5). Remarks of Senator Barack Obama: 'A More Perfect Union'. *The New York Times*. https://www.nytimes.com/2008/02/05/us/politics/05text-obama.html

Programmer's credo [Digital image]. (2021). Retrieved from https://www.reddit.com/r/ProgrammerHumor/comments/nu3d5t/i_thought_same_somehow

Pepe Silvia [Digital image]. (2008). Retrieved from https://knowyourmeme.com/memes/pepe-silvia

Dax-vs-M[Digital image]. (2021). Retrieved from https://powerpartners.pro/en/power-query-vs-dax/

Thank you!

(and more cats)

UNIVERSITY OF OREGON