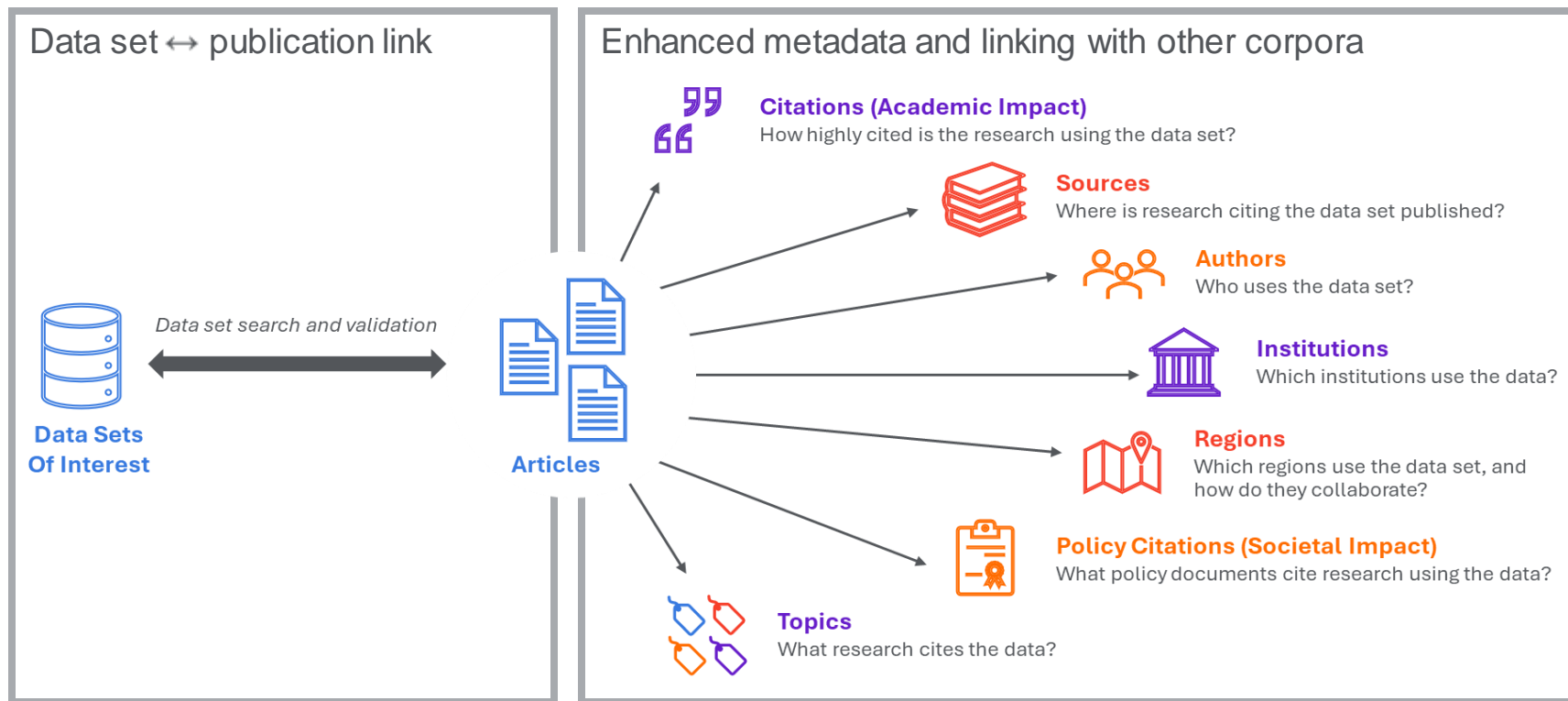


# Turning Visions Into Reality: Lessons Learned From Building a Search and Discovery Platform

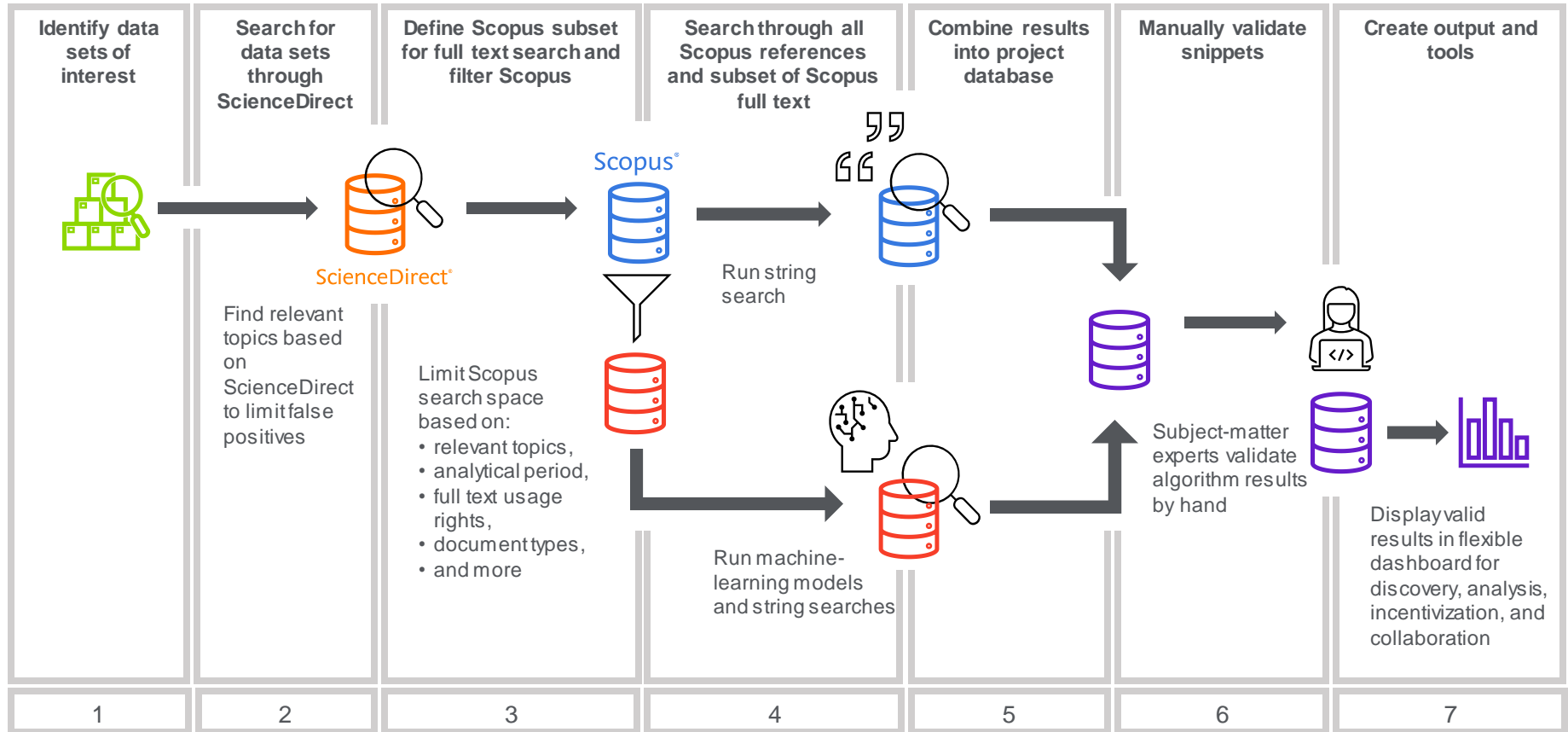


# Context: Search & Discovery Model



Articles: 96 million+ records in Scopus

# Process Description: Searching for Datasets



# Validation: Text Review by Subject Matter Experts

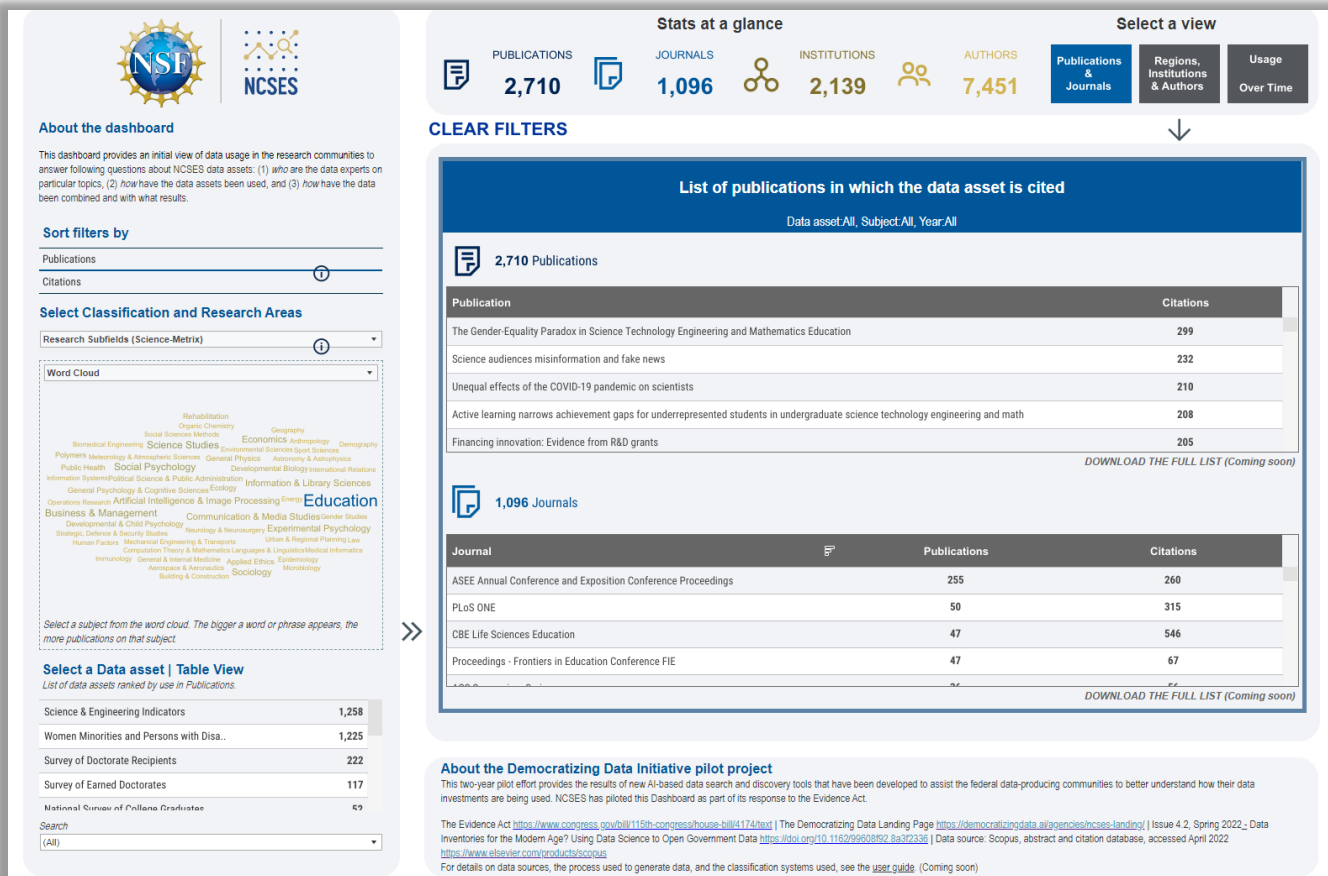
## Snippet Validation Example

<a href="#">Application Of An Innovative Alignment Optimisation Method To A Cross-Cultural Mean Comparison Of Teacher Self-Efficacy: A Cross-Country Study (2021)</a>		
Mention candidate: talis Alias: TALIS	Does the highlighted text in the snippet refer to a dataset?	Does the highlighted text in the snippet refer to the dataset below?
-efficacy: A cross-country study\nYusuf F. Zakariya*\nDepartment of Mathematical Sciences University of Agder Kristiansand Norway\nARTICLE INFO\nKeywords:\nAlignment optimisation method Latent mean comparison Scalar invariance <b>talis</b> 2018 Teacher self-efficacy\nABSTRACT\nTeacher self-efficacy is a crucial personal characteristic that is important not only for teachers' well-being but also for the overall teaching and learning. However the diffic	<input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unsure	Teaching and Learning International Survey <input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unsure

## Validation Results Example

Data Asset / Alias name	Actively Validated	Positively Validated	Precision
Higher Education Research and Development Survey	42	42	100%
HERD	1308	8	<1%

# Access & Dissemination: User-Friendly Dashboards



# Access & Dissemination: Tools for Custom Analysis

- SciServer (cyberinfrastructure system with tools and big data)
- Application Programming Interface (API)

The screenshot shows a Jupyter Notebook with a single cell containing a SQL query. The query selects distinct external IDs and dataset aliases, joining various tables to filter for specific publication and dataset relationships. The output is a table with two columns: 'eid' and 'alias'.

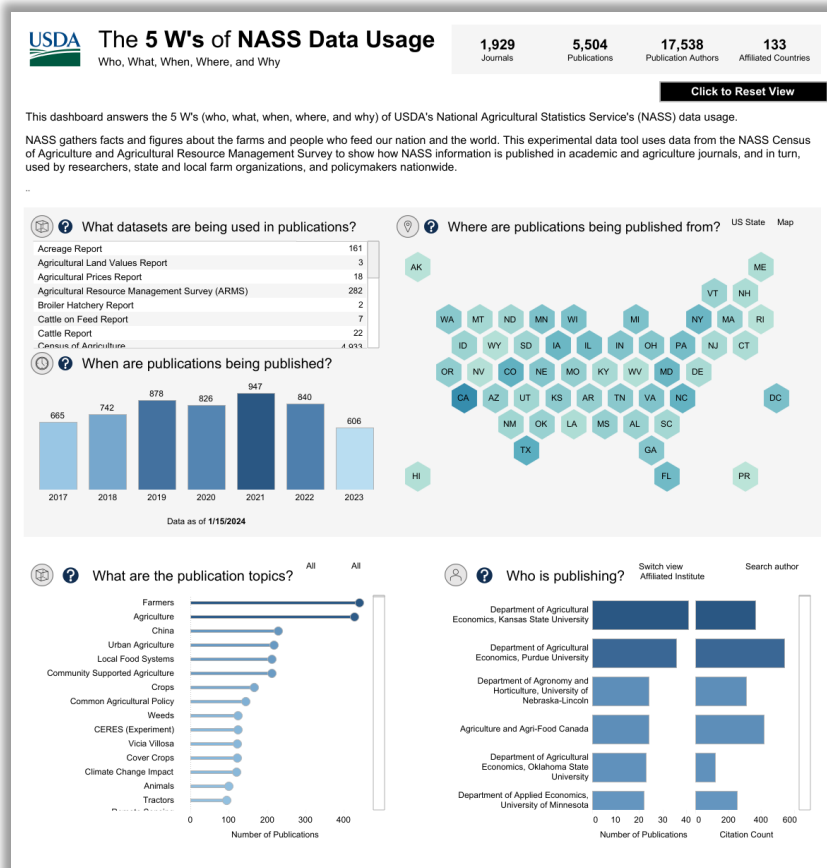
```

In [144]: sql="""
select distinct external_id as eid
, daa.alias
from publication p
left join dyad d on p.id=d.publication_id
left outer join dyad_model dm on dm.dyad_id=d.id and dm.model_id=5 -- hardcoded, id from refmatch in model table
left join dataset_alias da on d.dataset_alias_id=da.id
left join dataset_alias daa on da.parent_alias_id=daa.alias_id
left join dyadvalidated dv on p.id=dv.publication_id and d.dataset_alias_id=dv.dataset_id
left join snippet_validation sv on d.id=sv.dyad_id and sv.is_dataset_reference=1 and sv.agency_dataset_identified=1
where p.run_id in (2,13,26,27,39,40) and (model_id=5 or agency_dataset_identified=1) and da.parent_alias_id in ('6000','6001')
"""
df=cj.executeQuery(sql,DATABASE)
df.convert_dtypes()

Out[144]:

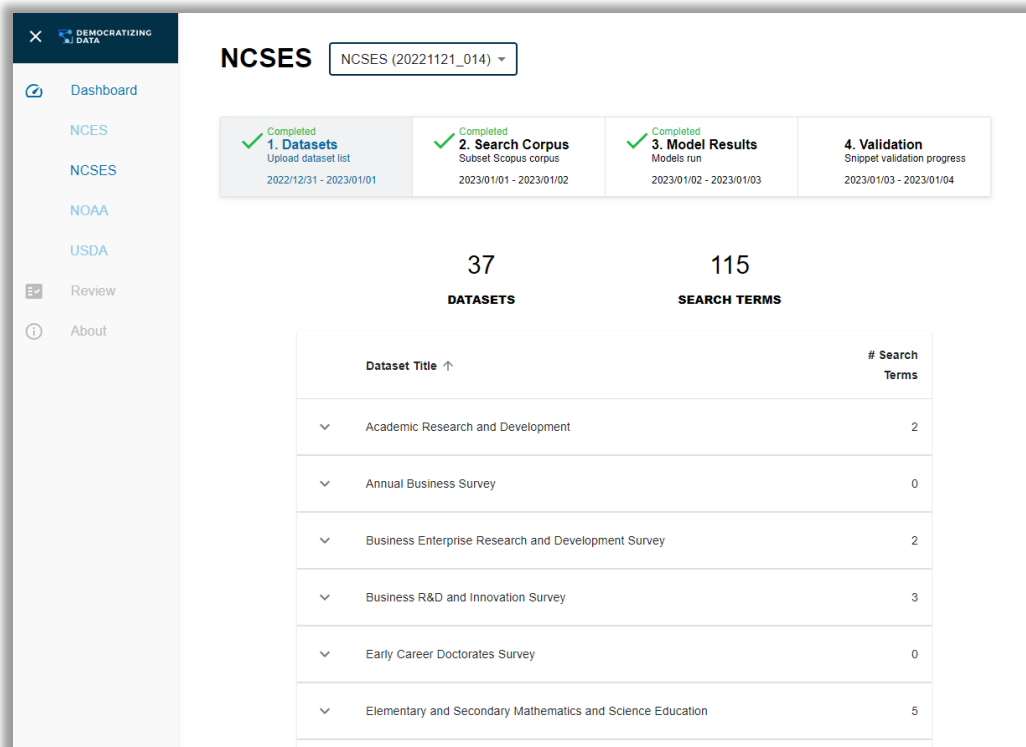
```

	eid	alias
0	2-42 0-5501563493	census of agriculture
1	2-42 0-5503031876	census of agriculture



# Future Directions

- Increase transparency and collaboration through administrative dashboard
- Expand to include other corpora (e.g., policy documents)
- Develop meaningful, tailored classifications through community engagement
- Improve data usage measures
- Refine user tools



**NCSES** NCSES (20221121\_014) ▾

Completed 1. Datasets Upload dataset list 2022/12/31 - 2023/01/01	Completed 2. Search Corpus Subset Scopus corpus 2023/01/01 - 2023/01/02	Completed 3. Model Results Models run 2023/01/02 - 2023/01/03	4. Validation Snippet validation progress 2023/01/03 - 2023/01/04
--	--	--	---

**37** **115**  
**DATASETS** **SEARCH TERMS**

Dataset Title ↑	# Search Terms
▼ Academic Research and Development	2
▼ Annual Business Survey	0
▼ Business Enterprise Research and Development Survey	2
▼ Business R&D and Innovation Survey	3
▼ Early Career Doctorates Survey	0
▼ Elementary and Secondary Mathematics and Science Education	5