

Chipping Away at the Paywall:

Leveraging APIs to Promote
Green Open Access

Dolsy Smith
Leah Richardson
GWU

25% of published scholarly content is available through OA

75% is still behind the paywall

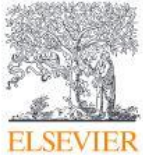
700+ institutionally mandated OA policies

\$3 billion spent annually by libraries for access to top 5 publishers' content

In order to develop a culture of OA awareness across the university

- ❖ Get a bird's-eye-view of publishing @ GW
- ❖ Highlight publisher's green OA policies
- ❖ Outreach to faculty centered on their needs re impact & tenure
- ❖ Provide liaison librarians with OA data by department
- ❖ Democratize tech tools through peer learning

Project Plan



GW-affiliated publication data



Publisher copyright and self-archiving policies

What is an application-programming interface (API)?

Another way to interact and retrieve information from a database or website

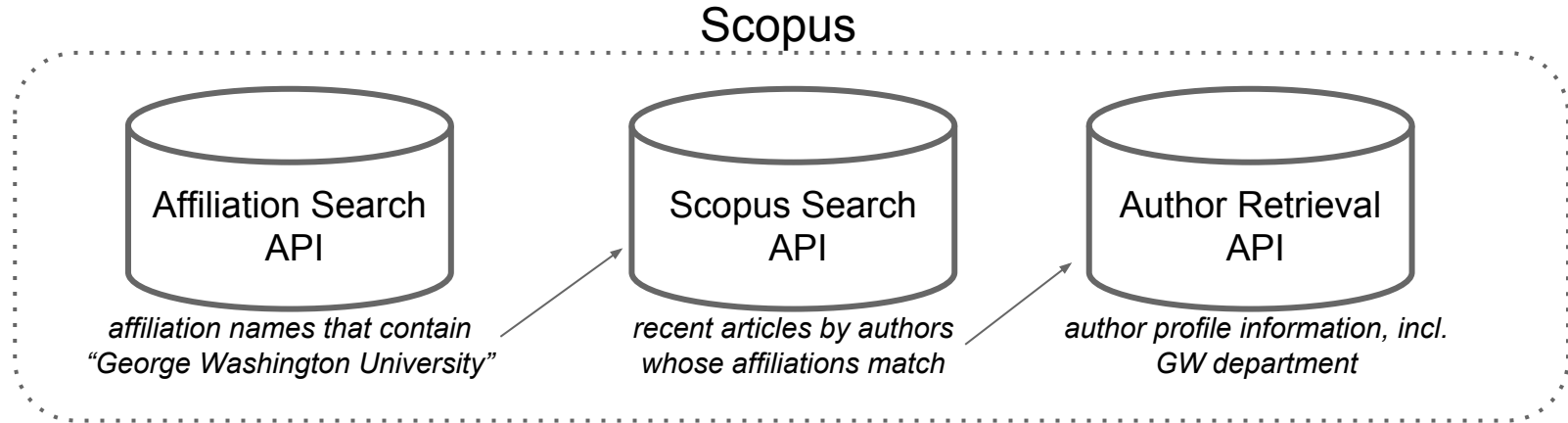
Good for bulk data retrieval

Traditionally the domain of administrators

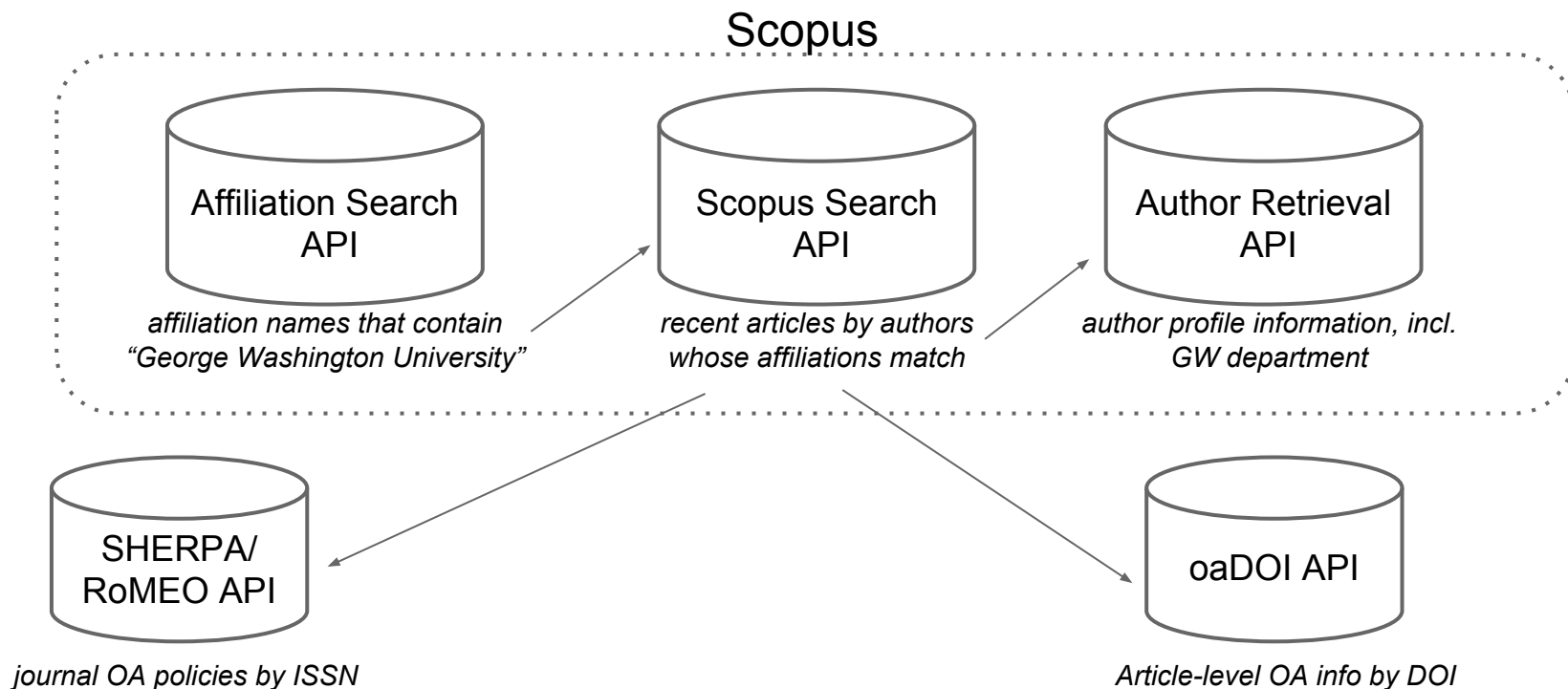
Anyone with a little programming know-how can access large data sets



Data collection



Data collection



Data collection

- **Python**

- Use *requests* library for API calls
- Need to batch calls to the Scopus API because of retrieval limits
- Handles JSON well (Scopus & oaDOI)
- Use `ElementTree` library for parsing XML (SHERPA/RoMEO)

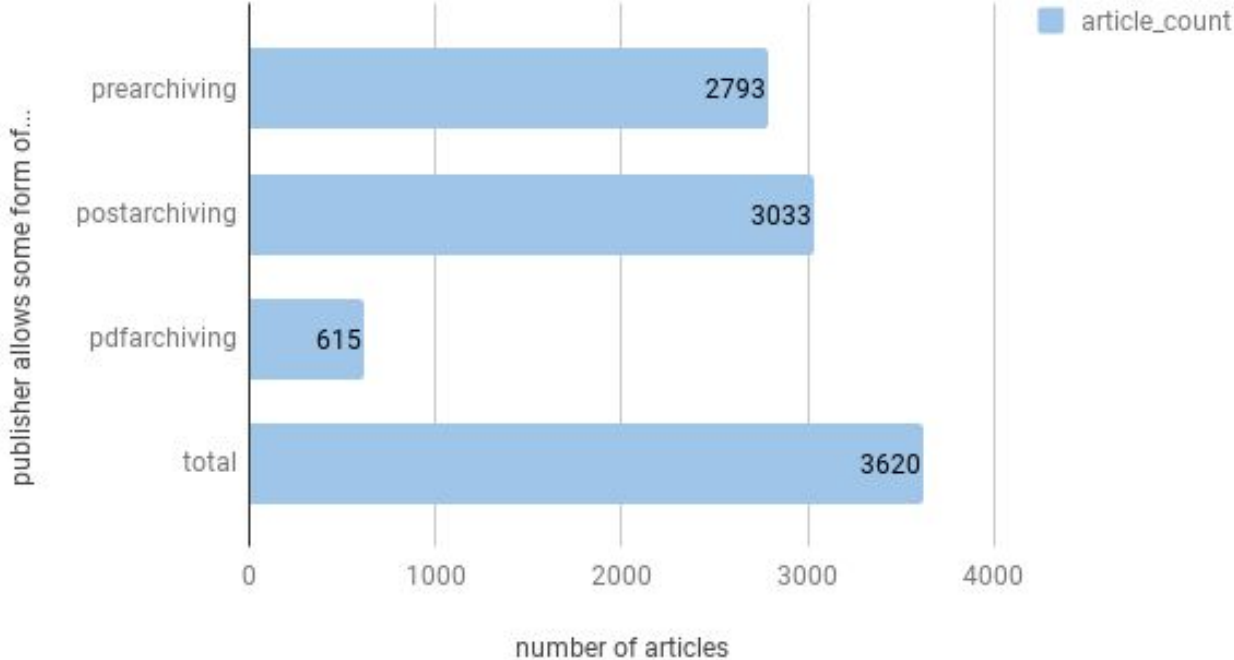
- **postgreSQL database**

- Store raw data in tables
- Create derivative tables for data analysis

Initial Results

77% Eligible for pre-archiving
84% Eligible for post-archiving
17% Eligible for PDF archiving

Articles by GW Authors, 2016-2017



Current status and next steps

- ❖ Collaboration with Division of IT/university data warehouse
 - Getting a canonical list of faculty members & their department affiliations
 - This dataset will allow us to clean up our Scopus data and make it usable for outreach.

- ❖ oaDOI API
 - Where are faculty archiving pre- and post-prints of their work?
 - Potential input for GW's IR, ScholarSpace