



Building Congressional Hearings Records with the govinfo API

(Library Services & Content Management) Library Technical Services Staff: built a computer script to connect to the govinfo API, obtain summary information on all newly ingested Congressional Hearings for the previous 24 hour period, request and retrieve the relevant MODS metadata for each hearing, convert the metadata into preliminary MARC records, which are used to produce Full level Congressional Catalog records in the Catalog of U.S. Government Publications and OCLC.

A New Process!

Old: Congressional Catalogers enter most data by hand

New: Script retrieves machine-built OCR data from the govinfo API and creates a provisional record **prepopulated** with Title, SuDoc, Physical Characteristics and Congressional Committee series information.

Future Plans: Currently, the script builds files and places them on the local machine, after which human intervention is required to place the catalog records into the cataloging tools in use at GPO (OCLC's Connexion Client).

Future versions of the script will skip Connexion Client entirely, and instead move machine-built files directly from govinfo to internal GPO fileshares or directly to OCLC.

Enhances

Python scripts run on a variety of systems. Script are also written in plain text which makes them easy to reconfigure and update as necessary.

Collections offered at <https://api.govinfo.gov>:

- ⇒ BILLS/BILLSTATUS: Congressional Bills and their status
- ⇒ CRPT: Congressional Reports
- ⇒ CPD: Compilation of Presidential Documents
- ⇒ HOB: History of Bills

Check GPO's GitHub repo:
<https://github.com/usgpo>

Expands

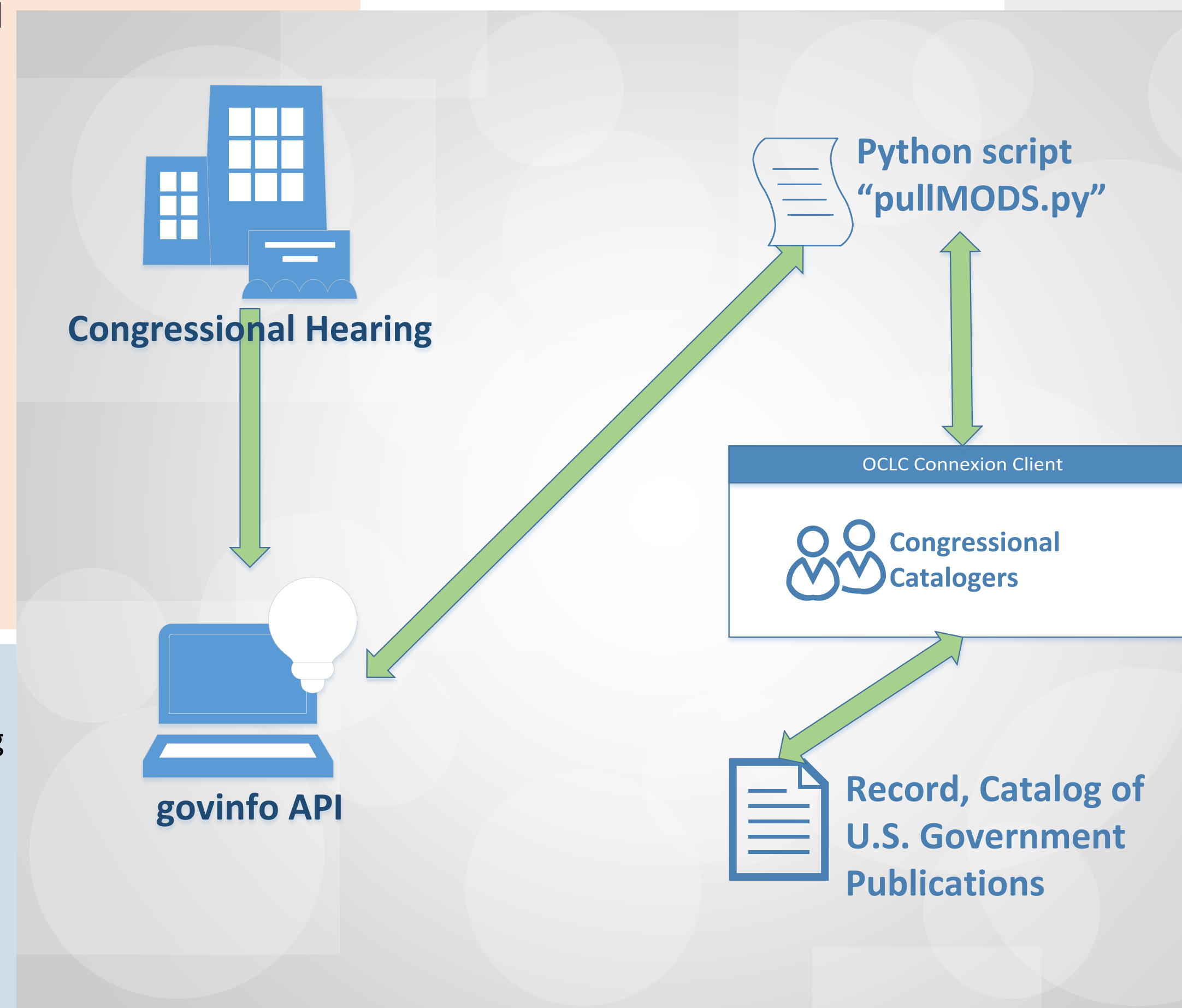
New fields added: 500 (govinfo accessID), 511 (Hearing witnesses), 518 (Date of hearing)
Script was first deployed in pilot project over the month of July 2019 with four Congressional Catalogers. From the pilot:

- 32 records created from 55 machine-built records (58% Completion rate)

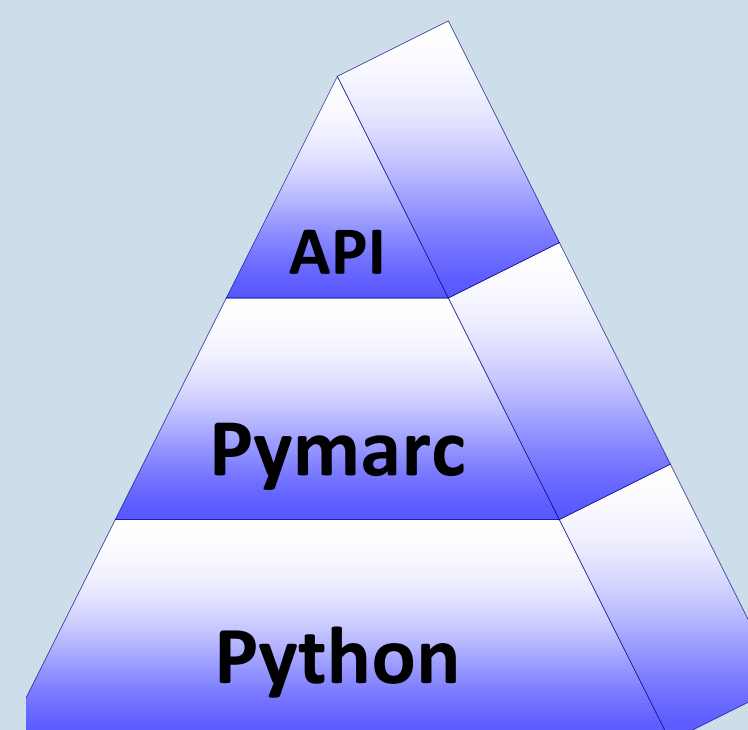
- An average daily production of 1.8 records.
- Metadata available, but not added to production records:
- "Public Law Citations"

What Metadata Would You Like To See In GPO Records?

Feedback, questions, concerns: <https://www.gpo.gov/askgpo>



Tech Pyramid



API: A programmatic process for exchanging data quickly and easily

Pymarc: Computer code developed by Ed Summers (@edsu) for handling MARC data with the Python programming language.

Python: "an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects."

Wikipedia contributors. (2019, October 3). Python (programming language). In Wikipedia, The Free Encyclopedia. Retrieved 18:56, October 3, 2019, from [https://en.wikipedia.org/wiki/index.php?title=Python_\(programming_language\)&oldid=919422996](https://en.wikipedia.org/wiki/index.php?title=Python_(programming_language)&oldid=919422996)

=500 \\\$aAccess ID (govinfo): CHR-116hrg35693.
=504 \\\$aIncludes bibliographical references.
=511 0\\\$aHearing witnesses: Garcia-Diaz, Daniel, Director, Financial Markets and Community Investment, U.S. Government Accountability Office (GAO).