1. **Problem/Question**
   1. Problem: It is difficult to track compliance with the public access directive
   2. Question: How compliant is the Agricultural Research Service with the USDA Public Access Implementation Plan?
2. **Process**

***Every script and dataset used is from the data in VIVO***

* 1. Original Spreadsheets from VIVO
     1. agid\_only.tsv
     2. doi\_agid.tsv
     3. doi\_only.tsv
     4. ARS-PROJECT-PUBLICATIONS.tsv
  2. Spreadsheets for Dataset 1
     1. all\_publications.xlsx
        1. A combination of the agid\_only.tsv, doi\_agid.tsv, and doi\_only.tsv
     2. all\_publications\_sample.xlsx
        1. A random sample of 1000 articles from the all\_publications spreadsheet
  3. Spreadsheets for Dataset 2
     1. pub\_doi\_agid.xlsx
        1. Sheet contains the Publication ID, DOI, and AGID
     2. ars\_projects\_pub.xlsx
        1. Sheet with the Project ID and Publication ID
     3. original\_pub\_proj.xlsx
        1. The combined spreadsheet of pub\_doi\_agid.xlsx and ars\_projects\_pub.xlsx
           1. pub\_proj.xlsx

Spreadsheet from original\_pub\_proj.xlsx that includes only Project IDs and Publication IDs that have an article title associated with them

* 1. General Scripts:
     1. Publication Info
        1. Uses Crossref to the publisher name and publication date for each DOI

1. Sheet Join
   1. Used for the Dataset 2
   2. Took two datasets and joined them on “PUBLICATION\_ID”
      1. Each article had a publication ID and project ID associated with it
      2. There are several publication IDs associated with one project ID

B. Order of Precedence when Tracking Compliance:

1. PubAg Script
   1. Searches through publications with AGIDs to see if there is full text in PubAg
   2. Challenges
      1. 1000 article limit
         1. Can request a limit increase

*If the article is not in PubAg, check PubMedCentral*

1. PubMedCentral Script
   1. Converts publication DOIs to PMCIDs (if applicable)
   2. Challenges
      1. 200 article limit
         1. Had to make several subsets
         2. Results change more frequently due to embargo

*If the article is not in PubMedCentral, check CHORUS*

1. CHORUS Script
   1. Checks the DOI against the Chorus API to determine is publications are available on CHORUS
   2. Challenges
      1. Results vary by day
      2. Had to incorporate the date and result limit for this script

*If the article is not in CHORUS, check Publisher site*

1. Public Access Journal Script
   1. Reads the DOIs to determine if they are associated with an open access journal or publisher
   2. Uses the Elsevier and SpringerNature APIs to check for accessible publications
   3. Challenges
      1. Some journals are hybrids (some articles are publicly accessible while others are not)
      2. Some Web pages could not be found

*If the article is not found, check manually*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PUBAG\_FULLTEXT** | **PUBMEDCENTRAL** | **CHORUS** | **PUBLISHER** | **FOUND BY** |
| NO/NO AGID | NO | NO | NO | **ALL SCRIPTS\*** |
| **YES** | N/A (already found in PubAg) | N/A (already found in PubAg) | N/A (already found in PubAg) | **PUBAG SCRIPT** |
| NO/NO AGID | **YES** | N/A (already found in PubMedCentral) | N/A (already found in PubMedCentral) | **PUBMEDCENTRAL SCRIPT** |
| NO/NO AGID | NO | **YES** | N/A (already found in CHORUS) | **CHORUS SCRIPT** |
| NO/NO AGID | NO | NO | **YES** | **PUBLIC ACCESS JOURNALS SCRIPT** |
| NO/NO AGID | NO | NO | (Blank) | **MANUALLY\*\*** |

\* ALL SCRIPTS: An article is considered not publicly accessible if each script reports NO

\*\*MANUALLY: An article has to be searched for manually if it reports NO for the first three scripts but cannot be found with the Public Access Journals Scripts

1. [**Dataset 1: Sample of 1000 ARS articles for 2008-2018 from VIVO**](https://docs.google.com/spreadsheets/d/1P14sFRoQByngWQ-U-2AH2JSQ66zIjIY5PwG7_kDLuRI/edit#gid=922169284)
   1. Created sample of 1000 articles
   2. Manually searched through each article to determine accessibility
   3. Used the completed scripts to search through the articles
      1. Compared script results to manual results
   4. Sheets within Dataset 1 (Manual)
      1. AGID Corrections
      2. Spreadsheets for Dataset 1
         1. Abstract from main SpreadSheet DATASET info (2008-2018)
            1. Tables from sample DATASET info (2008-2018)
   5. [Sheets within Dataset 1 (Automated)](https://docs.google.com/spreadsheets/d/14Y9nmCdoHqnhScV3Rz0OmOKeaY6ESppCY0jx-Jr1XHg/edit#gid=0)
      1. Tables
         1. Sheet that holds tables:
            1. For all articles that are/are not publicly accessible
            2. For the location of publicly accessible articles
   6. PMC Sheets
      1. Folder for the PubMedCentral spreadsheets
         1. Each spreadsheet had 200 or less articles in order to use the PubMed API
2. [**Dataset 2: Project Specific Article Availability for 2017 from VIVO**](https://docs.google.com/spreadsheets/d/1iMQ4198bXBmQLaLG-b3loS7pvgQjIIb2vOpyNXDAOhY/edit#gid=0)
   1. Created a sample of 1547 articles from 2017
      1. Calculated the number of projects based on the number of articles
   2. Coded scripts based on the information in this dataset
   3. Sheets within Dataset 2
      1. ALL
         1. Includes all the articles from 2017 that were found manually and by script
      2. Scripts Only
         1. Includes only articles found by script
      3. Tables:
         1. Sheet that holds tables:
            1. for articles that are/ are not publicly accessible for both manual and script work (ALL sheet)
            2. for articles that are/are not publicly accessible scripts only (Scripts Only sheet)
            3. for where articles are located (Scripts Only sheet)
      4. Tables for Projects
         1. Holds the tables for the projects that do/do not have publicly accessible articles
   4. PM Sheets
      1. Folder for the PubMedCentral spreadsheets
         1. Each spreadsheet had 200 or less articles in order to use the PubMed API

<https://github.com/bdalce/Accessible-Research-Compliance>