581001

Sandia National Laboratories

Operated for the United States Department of Energy
by National Technology and Engineering Solutions
of Sandia, LLC.

Albuquerque, New Mexico 87185-0101
Livermore, California 94551-0969

date: January 16, 2024

to: SNL WIPP[1] Geochemistry, SNL WIPP Records Center
Sandia National Laboratories
4100 National Parks Highway, Building A, Carlsbad, New Mexico 88220, U.S.A.

from: Martin B. Nemer *Signature on File*
Sandia National Laboratories
1515 Eubank Blvd SE, Albuquerque, New Mexico 87111, U.S.A.

subject: Steps involved in the initial stages of the literature screening

1 Overview

This memo outlines the semi-automated approach that was utilized in 2023 for an initial screening of publicly available scientific literature, for determining the actinide-solubility uncertainty distributions, as currently used in Performance Assessment (PA). For an explanation of the downstream application of this literature search, we refer the reader to Domski [1].

Several web-based literature-search tools now allow for a degree of automation. This includes both Web of Science (WOS) [2] and Office of Scientific and Technical Information (OSTI) [3] web-based search engines. Google Scholar currently does not allow automation via an application programming interface (API).

In this literature-screening exercise, we attempted to automate the initial steps of the process for the purpose of increasing traceability/verifiability. Regardless of the method, a consistent-repeatable protocol for automated literature searches remains a work in progress. No method of literature search can guarantee that we obtain all relevant literature, however automated/logged approaches are at least traceable and repeatable.

The Web of Science (WOS) API is described in [2]. The OSTI API is described in [3]. In both APIs, references are requested via an HTML string, with the response being either an XML or JSON

[1] SNL WIPP: Sandia National Laboratories at the Waste Isolation Pilot Plant (WIPP). WIPP is located ~26 miles southeast of Carlsbad, New Mexico, U.S.A., built 2,150 feet underground in the Salado Formation, a Permian evaporite deposit in the northern Delaware Basin.

WIPP:4.2.1:PA:QA-L:574938

Information Only

human-readable database. In both APIs, there is a limit on the number of references that can be obtained in a single call. For WOS that number is 100. For OSTI that number is at least 100 (exact number couldn't be found in the documentation). Both APIs appear to be based on each other, perhaps utilizing the same basic underlying engine. As an aside, it was found that JSON worked better than XML as the receiving format, because both OSTI and WOS produced records with direct (un-escaped) Unicode symbols in their fields, which are not allowed in the XML standard [4].

The initial literature search and screening was a team effort involving most of the Carlsbad Program Group (CPG) Sandia Geochemistry team. Some researchers had more-or-less experience with programming and APIs. As such, in the end we utilized both manual and automated searches for WOS and OSTI.

All of the retained files, scripts, etc. are located in cvs in the directory */data/cvs/CVSLIB/WIPP_EXTERNAL/litscreen2023*. Paths are given in the remainder of this document relative to this path.

2 Methodology

The specific details of the methodology used to download all citations before an initial screening is given in Appendix A. In summary, researchers downloaded all the citations they could find using generic search terms for the element (e.g. Pu, U, Nd,...) that was assigned to them. Those citations were then imported into an EndNote library. That EndNote library was then exported as an EndNote XML file. The EndNote XML file was then parsed with a Python script using regular expressions [5] in three stages (herein called passes); an example of the Python script is given in the file Nd_litsearch.py within *./Actinide Uncertainty Search 2023/Python*. In the first pass (pass #1), references were selected if they contained one or more search terms within the string variable `patt1`. The search term(s) found for each reference that made it past this step were recorded within the EndNote *Notes* field for that reference as a JavaScript Object Notation (JSON) [6] structured-text object. In the second step (pass #2), references that made it past pass #1 were screened for un-favorable search terms contained in the string variable `patt2`. These references were then tagged as not having passed pass #2, again with the term found recorded within the EndNote *Notes* field for that reference. In the third step (pass #3), references that made it past pass #2 were screened against an additional set of favorable search terms located within the string variable `patt3`. Again, the favorable search term from pass #3 was stored within the EndNote *Notes* field.

The rational of separating the steps was to provide a set of human-scrutiny insertion points for checking the results of pass #1 through pass #3. References that didn't make it past the first pass were generally not examined further. References that made it past pass #1 but not pass #2 were for the most part quickly checked to make sure the search terms in pass #2 weren't overly restrictive. References that made it through to pass #3 needed a more thorough human

Information Only

examination to determine whether they contained relevant actinide test data, and whether pass #2 required additional negation terms.

Once the references were filtered according to the above schema, the EndNote XML file was then re-imported into EndNote to produce a final "screened" database. Ideally, all of the collected references would be re-imported into EndNote, for human cross-checking using EndNote tools. However, in some cases, the size of the database was too large to allow for this.

As will be discussed in the next sub sections, the description here is general. There was variability between researchers as to how the literature search was performed. As such, the details of what was done for each actinide element or analog is described in the next sub sections.

2.1 Americium

Americium data in the +III state was searched for using WOS and OSTI. The search terms used were: *Americium III*, *Americium 3+*, *Am III*, *Am 3+*, and *Americium*. The citations from each of these terms were placed into separate EndNote library files. The Python script was then used for the initial screening as described in Section 2. The search terms for the three passes were:

```
patt1 = r'solub|dissol|formation constant|dissociation constant'

patt2 =
r'molten|arctic|modular|reactor|serum|mantle|crust|cave|marine|w
eather|monaz|glacial|separation|recovery|seawater'

patt3 = r'Am\(III\)|americium'
```

Note that the strings listed above follow the general rules of regular expressions [5]. The lack of a "\s" for spaces between words in multi-word search terms above appears to work in Python using the `re` package for regular expressions. It should be noted the re.IGNORECASE option was used for each pass, such that case (lower, upper) was ignored for each search pass.

Python codes used for the initial screening are located in *./Actinide Uncertainty Search 2023/Am /Codes and Instructions*. The original and parsed XML files are located in *./Actinide Uncertainty Search 2023/Am/Web of Science* and *./Actinide Uncertainty Search 2023/Am/Osti* for WOS and OSTI respectively. The final EndNote libraries produced for Americium contain only those references that were screened in by pass #1. These files are located in *./ Actinide Uncertainty Search 2023/Am/Endnote libraries/*.

After pass #3, the EndNote libraries were compiled into two files: OSTI combined, WOS combined. Additional human screening was then performed within these EndNote libraries, boldfacing text from the abstract as the reasoning. That reasoning formed the basis for two EndNote Groups [7]: screened *in_initial* and screened *out_initial*. Finally for those papers that needed further investigation, the full paper was read and the final screening results were placed in the EndNote groups: *in_final* and *out _final*.

Information Only

2.2 Neodymium

Neodymium (III) was searched in both WOS and OSTI. The initial downloaded references are located at *./Actinide Uncertainty Search 2023/Neodymium*. The assembled EndNote libraries and Python Parsed libraries are located in *./Actinide Uncertainty Search 2023/Neodymium/OSTI*, *../WOS Nd*, and *../WOS Nd(III)*. The search terms used for the Python screening were

```
patt1 = r'solub|dissol|formation constant|dissociation constant'

patt2 = r'
leach|extract|molten|arctic|modular|reactor|alloy|polymer|magnet
|dope|hydrothermal|magma|marine|tissue|seawater|serum|mantle|cru
st|cave|weather|monaz|glacial|separation|recovery|glass'

patt3 = r' Nd\(III\)|neodymium'
```

A final compilation of the parsed WOS and OSTI libraries is given in *./Actinide Uncertainty Search 2023/Neodymium/final library*. In addition to the three groups in EndNote that the researcher created for the three Python-script passes (pass #1, pass #2, pass #3), the final EndNote library contains additional groups generated by the researcher with screened in and screened out references. Table 1 includes explanations of the additional researcher-generated groups.

Table 1. Additional Groups used in EndNote for Neodymium screening.

Group	Notes
"0 references to screen"	This was a placeholder group. The references remaining in this group were all screened out.
"manually screened in"	These references were screened in after reading the abstract and/or the paper.
"manually screened out"	These references were screened out after reading the abstract.
"no access"	References that Sandia did not have access to. These references were all screened out based on the title, abstract, and year of publication.
"screened out and need a write up"	These references came close to being screened in but were screened out after reading the paper.

2.3 Plutonium

Plutonium references were downloaded from both WOS and OSTI. For WOS, the original .cis files are located in *./Actinide Uncertainty Search 2023/Pu/Reference Downloads/Web of Science*. For OSTI, the references were downloaded as *.json* and *.ris* files and placed into *../OSTI/*. For both WOS and OSTI, the search terms used were: *Plutonium*, Plutonium *(*)*, *Plutonium (III)*, *Plutonium (IV)*, *Pu*, *Pu (*)*, *Pu(III)* and *Pu(IV)*. These references were loaded into EndNote in the directory *./Actinide Uncertainty Search 2023/Pu/Endnote Libraries*. The *.ris* files from OSTI were loaded into EndNote and then exported together to create one single *.xml* file. The initially Python-

Information Only

parsed EndNote libraries in these folders are *WoS References.enl* and *OSTI references.enl* for WOS and OSTI, respectively. Note that both files are close to 1Gb, which is close to the limit for EndNote. The OSTI file contains over 560K references, which is an example of where use of scripting was useful. Unfortunately, there are also a significant number of duplicate references in these files.

The regular-expression search terms for pass #1 (`patt1`), pass #2 (`patt2`), and pass #3 (`patt3`) are given below:

```
patt1 = r'solub|dissol|formation constant|dissociation constant'

patt2 =
r'molten|arctic|modular|reactor|serum|mantle|crust|cave|marine|w
eather|monaz|glacial|separation|recovery|seawater|health|biology
|food|corrosion|leach|sorption|dosimetry|transport'

patt3 = r'Pu\(IV\)|Pu\(III\)|plutonium'
```

The search terms *biology* and *food* were required as part of the pass #2 search terms (`patt2` above) as many papers focused on the biological effects on the food chain and on the workers themselves when working with plutonium. The search term *dosimetry* was also needed in pass #2 because the obtained papers that discussed dosimetry focused on the biological impacts rather than its solubility in connection to the WIPP-relevant conditions.

EndNote Groups were not used within the WOS EndNote library. Groups were used in the OSTI EndNote library. EndNote groups within the OSTI file were used to reduce the original list of references down to those that passed pass# 1 – pass #3 from the Python script. Each initial plutonium search term (*Plutonium*, Plutonium *(*), Plutonium (III), Plutonium (IV), Pu, Pu (*), Pu(III)* and *Pu(IV))* has a group with all of the original citations. These groups are given by the name of the initial search term followed by *Parsed*. E.g., for *Plutonium* the original references are stored in the group *Plutonium Parsed.* References that passed through pass# 1 – pass #3 of the Python script were placed into a secondary group given the name of the initial search term with the addition of *Parsed Search*, *Plutonium Parsed Search.*

The only other group in the OSTI EndNote library was *All parsed Search references* which is where all of the final parsed citations were placed from all of the search terms, allowing for easy reference to the reduced set of citations for further human scrutiny.

2.4 Thorium

Both WOS and OSTI were searched for thorium. The thorium (IV) search was performed by manually downloading the citations from OSTI and WOS. The search terms used to pull the initial citations from WOS and OSTI were: *Thorium IV*, *Thorium 4*, *Th IV*, *Th 4*, *Th(IV)*. The original downloaded citations from OSTI and WOS are located in *./Actinide Uncertainty Search 2023/Th/original files*. The compiled EndNote library prior to Python parsing is located in *./Actinide Uncertainty Search 2023/Th/Endnote* file before parsing. All citations that were

Information Only

downloaded and Python parsed are located in the EndNote file *My EndNote Library-cut.enl* located in *./Actinide Uncertainty Search 2023/Th*. The Python search terms used in each pass were:

```
patt1 = r'solub|dissol|formation constant|dissociation constant'

patt2 =
r'molten|arctic|modular|reactor|serum|mantle|crust|cave|marine|w
eather|monaz|glacial|separation|recovery|seawater'

patt3 = r'Th\(IV\)|thorium'
```

The EndNote library contains a number of EndNote groups. The first EndNote group called *New Group* contains all references that passed pass #1 from the Python script. Likewise, the EndNote group *search 2 passed* contains all references that passed pass #2. Finally, the EndNote group *inc. Thorium* contains all references that passed pass #3; this group contains about 160 results.

The references from the EndNote group *inc. Thorium* were manually examined and filtered further into a new *group Final* by deleting citations that contain in the title or abstract: *liquid-vapor*, *sulfosalicylate*, *phenyl*, *chelator*, *mining*, *hair*, *diet*, *leaching*, *molten metal*, *molten salt*, *sorption*, *corrosion* and many more. These citations then were exported into an Excel spreadsheet with a detailed screening as described in Appendix A. The Excel document *exp_final.xlsx* is located in ./Actinide Uncertainty Search 2023/Th .

2.5 Uranium

For Uranium both WOS and OSTI were searched. However, because of the large number of references on uranium at OSTI, the OSTI references had to be downloaded via the API. An example of the Python script used to download the references is given in the file OSTI_search_example.py within the directory *./Actinide Uncertainty Search 2023/Python*.

The initial search terms used to download citations from WOS were: *Uranium*, *U(IV)*. The initial search term used for OSTI was: *Uranium*. The Python parsed EndNote libraries for each of these searches are located in the same directory as the un-parsed EndNote libraries. These are the directories *OSTI*, *WOS U(IV)*, and *WOS uranium*, located in *./Actinide Uncertainty Search 2023/Uranium*. The keyword *parsed* in the EndNote library file name within these directories indicates that the file was parsed by the Python script. The Python search terms used in each pass were:

```
patt1 = r'solub|dissol|formation constant|dissociation constant'

patt2 =
r'corrosion|sorption|dosimetry|modeling|transport|leach|molten|a
rctic|modular|reactor|serum|mantle|crust|cave|marine|weather|mon
az|glacial|separation|recovery|seawater|radon'

att3 = r'U\\(IV\\)|uranium'
```

Information Only

The OSTI parsed EndNote library contains a number of custom EndNote groups. These groups were further divided into EndNote Group Sets, which is an EndNote feature that allows the user to create a set of EndNote Groups. The EndNote Group Sets created in the OSTI file are *MY GROUPS* and *extra testing*. Table 2 - Table 3 include explanations of the groups. The EndNote group set *MY GROUPS* is self-explanatory from Table 2, however the group set *extra testing* requires additional explanation. The *extra testing* group set was included as an extra check of the references that hit the "modeling" or "seawater" search term during pass #2 (`patt2` above). For both group sets, references in the "manually unsure" or "manually screened in" groups were added to the final table for U(IV).

*Table 2. EndNote **Group Set: MY GROUPS** used for the uranium screening.*

Group	Notes
Search 1 Hit	References that passed the pass #1 screening.
Search 2 Hit: None	References that passed the pass #2 screening.
Search 3 Hit	References that passed the pass #3 screening.
abstract contains U(IV)	References where the abstract contained U(IV).
manually unsure	References that needed more screening/to read the paper.
manually screened out	References that were screened out after reading the abstract.
manually screened in	References that were screened in after reading the abstract.

*Table 3. EndNote **Group Set: extra testing** used for the uranium screening.*

Group	Notes
search 2 hit: seawater	References that contained "seawater."
search 2 hit: modeling	References that contained "modeling."
manually unsure	References that needed more screening/to read the paper.
manually screened out	References that were screened out after reading the abstract.
manually screened in	References that were screened in after reading the abstract.
abstract contains U(IV)	References where the abstract contained U(IV).

The parsed WOS U(IV) EndNote library contains the same group names and meanings as given in Table 2 - Table 3, except instead of *extra testing*, references that required additional screening were put into the EndNote group set *additional*. The *web of sci U(IV) parsed* EndNote database also contains a group set *sorption papers*. This group contains papers that didn't pass pass #2 because of the key word *sorption*, which the researcher felt required further human scrutiny.

The parsed WOS uranium EndNote library contained the same groups as the OSTI file with the addition of the EndNote group set *sorption* with a single group *search 2 hit*. This group is identical to the purpose of the *sorption papers* group set from the above paragraph for the WOS U(IV) library.

Information Only

Finally in the directory *./Actinide Uncertainty Search 2023/Uranium/Final libraries* is the EndNote library *all potential uranium Copy.enl* containing a final set of references that the researcher decided required additional human screening. The groups in this file are given in Table 4, all within the EndNote group set *MY GROUPS*.

*Table 4. EndNote **Group Set: MY GROUPS** used in the final uranium screening.*

Group	Notes
don't have access	References that we could not obtain
screened in	References that the team member felt needed further human investigation
screened out	References that the team member felt could be screened out without further human investigation
screened out and need a write up	References that the team member felt could be screened out but required a human justification

2.6 Curium and Neptunium

Curium and neptunium searches were combined into two EndNote libraries, one for WOS and one for OSTI. These searches did not utilize a Python script for initial screening and do not contain the JSON database generated by that initial screening Python script. All parsed references are located in the EndNote files *OSTI.enl* and *Web of Science.enl* located in *./Actinide Uncertainty Search 2023/Curium_Neptunium/Endnote Libraries*. Unfortunately, these files contain a large number of duplicate references.

2.6.1 Neptunium

The search terms used to download the initial citations from WOS and OSTI were: *Neptunium IV*, *Neptunium 4*, *Np IV*, *Np 4*, *Np (IV)*. Because the number of OSTI citations was too large to manually download reliably, a Python script was used to download the citations using the OSTI API [3]. The EndNote Groups in the *OSTI.enl* EndNote library have the following meaning: *NP Pass 1* are references containing the search terms *Neptunium 4*, *Neptunium IV*, *Np 4*, and *Np IV* and that were determined to potentially be solubility studies. The *Np Final* EndNote group contains references that manually appeared to pass criteria G2-G10 from [1]. No citations from OSTI were found to fit the criteria to be included within the *Np Final* EndNote group.

The EndNote groups in the EndNote WOS database have the following meaning: *Neptunium 4*, *Neptunium IV*, *Np 4*, and *Np IV* are EndNote groups of references organized by those respective search terms. The EndNote group *Np Pass 1* contains articles that the researcher believed to be potential solubility studies, using the keyword search engine within the EndNote program. The EndNote group *Np FINAL* lists the articles that manually passed criteria G2-G10 from [1], and required further human investigation.

Information Only

2.6.2 Curium

The search terms used to obtain citations from OSTI and WOS were *Curium III*, *Curium 3*, *Cm III*, *Cm 3*, and *Cm (III)*. Because the quantity of citations on the OSTI website was too large to manually download, a script was used to download the citations using the OSTI API [3]. The EndNote Group *CmPass 1* in the OSTI.enl database contains citations containing *Curium III*, *Curium 3*, *Cm III*, *Cm 3*, and *Cm (III)* search terms that the researcher believed to be potential solubility studies, using the keyword search engine within the EndNote program. The *Cm Final* EndNote group were citations that appeared to possibly pass criteria G2-G10 from [1], and required further human investigation.

The EndNote groups in the EndNote WOS database mean the following: *Curium III*, *Curium 3*, *Cm III*, are groups of citations organized by those respective search terms. The EndNote group *Cm Pass 1* contains citations that the researcher believed to be potential solubility studies. The EndNote group *Cm FINAL* contains citations that appeared to possibly pass criteria G2-G10 from [1].

3 Conclusions

Automating the literature search for determining the actinide uncertainty distributions remains a work in progress. Different researchers had different skill levels with automated tools. This combined with time pressures to accomplish the task led to different levels of success. Regardless, we believe that this methodology will lead to a more transparent and hopefully more complete literature search.

Both WOS and OSTI now allow for some degree of record keeping on searches. In theory those online searches should return the same set of references if they are repeated. This hasn't been tested, however. Such a logging option in future literature searches could remove the need for the Python scripting described herein.

4 References

1. Domski, P. and M.B. Nemer, *Revised Literature Screening Criteria for the WIPP Actinide Solubility Uncertainty Analysis for CRA-2024, Revision 1*. 2023, Sandia National Laboratories: Carlsbad, NM. SAND2023-01264O. ERMS 579911.
2. Clarivate. *Web of Science API Expanded*. 2023 [cited 2023; Available from: https://developer.clarivate.com/apis/wos.
3. U.S. Department of Energy. *API Documentation*. 2023 [cited 2023; Available from: https://www.osti.gov/api/v1/docs.
4. Walmsley, P., *Definitive XML Schema*. 2nd ed. 2013, Upper Saddle River, New Jersey: Prentice Hall.
5. Friedl, J.E.F., *Mastering Regular Expressions*. 2006, USA: O'Reilly.
6. JSON. *Introducing JSON*. 2023 [cited 2023 12/10/2023]; Available from: https://www.json.org/json-en.html.

Information Only

7. Clarivate. *EndNote: Definition and notes about groups*. 2023 [cited 2023; Available from: https://support.clarivate.com/Endnote/s/article/EndNote-Definition-and-notes-about-groups?language=en_US.

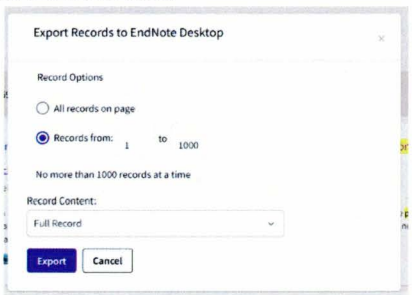
Information Only

Appendix A. Instructions Given to CPG Geochemistry Researchers

Below are the instructions that were given to each researcher performing the literature search. In the future, better guidance should be developed for researchers to annotate how they manually screened in/out references using the G criteria.

Steps Involved in Performing the Actinide Uncertainty Analysis:

1. Literature Search
    a. EndNote will be used to assemble citations into a database file that is machine readable, and then again for a final human screening. The first step in this process is to obtain a copy of EndNote through SAMS.
    b. From Web of Science and OSTI databases, gather citation information related to the actinides and their oxidation state analog of interest (i.e., either 3+ or 4+). For the initial on-line screening, the search terms should be as general as possible:
        i. **For An(III), search for**: Americium, Neodymium, Curium, and Plutonium and Am(III), Nd(III), Cm(III), and Pu(III). We could also try Am(*), Nd(*), etc. if either WOS or OSTI allows for wildcards.
        ii. **For An(IV) search for**: Thorium, Neptunium, Uranium, and Plutonium, and similarly to i. above, perhaps Th(IV), and or Th(*), etc.
    c. The citation information will need to be downloaded as a machine-readable file that can be imported into EndNote.
        i. Citation information here means the basic publication information (author, title, etc) including the abstract, and keywords. The Python script mentioned below will screen by searching all of these fields.
        ii. To download a list of references from Web of Science, choose Export near the top of the screen -> EndNote Desktop -> Records from X to Y and Record Content: Full Record



You will have to manually set the numbers in the records from fields, to retrieve all of the records that you have obtained in your web search. Web of Science will only allow you to download 1000 references at a time.

iii. To download a list of references from OSTI, choose Save Results near the top of the screen then RIS. This will also download no more than 1000 references and this site will not give you the option to choose a range of

Information Only

record numbers to download. Hence you can limit records to <= 1000 by choosing a publication date range.

iv. If the number of references is in the 1M+ range for a given element, we can use the WOS and OSTI APIs to download this citation information over the course of several hours to days even, instead of manually.
    1. https://www.osti.gov/xml-services
    2. https://developer.clarivate.com/apis/wos

d. Import all of the citation information collected into an EndNote database. If separate people are working on the literature search it is useful for each person to search for a particular subset of the literature and have their database contain only that subset. This isn't required.

e. Note that 2-4 Gb is about the limit of an EndNote Database, given that they use SQLite. We may have to break up databases if they're larger than this, but we haven't seen a database bigger than ≈ 100 Mb.

f. Once you have produced a roughly screened database of citations in EndNote, go to File -> Export -> Save As Type -> .XML . This will produce an XML database with all of the citation information you have collected.

2. Initial screening using scripting

a. Use the Python script to perform an initial automated screening of your database. The Python script will produce a new XML file that must be imported into a new EndNote database. While this is admittedly cumbersome, it allows the researcher to quickly repeat different screenings, as desired, using the local database file rather than waiting for the web to update your search. The Python script also allows for a far more complicated set of search criteria via a nested set of search terms using regular expressions.

b. The first step in using the Python script is to set up the filenames for the input file, output file, and a path. This occurs via the class constructor `parsetlit(path,inputfile,outputfile)`, where
`path` = path where the input and output XML files can be found and written respectively
`inputfile` = name of the EndNote generated XML inputfile
`outputfile` = name of the XML outputfile

c. Next the search expressions must be defined as Regular expressions: https://learn.microsoft.com/en-us/dotnet/standard/base-types/regular-expression-language-quick-reference

d. In the Python scrip, the variable `patt1` is a Regular Expression that comprises a set of search terms that would screen in a reference.

e. The variable `patt2` is a Regular Expression of search terms to screen out a reference, applied on the references that were screened in in step c.

f. The variable `patt3` is a Regular Expression that will screen in references that were screened in in step c and not screened out in step e; `patt3` contains a final check that the reference contains the desired element, is specific to the element being searched for, which is they only reason why it's a separate search.

Information Only

g. Note that there is no perfect set of `patt1 - patt3`, and we make no claims that any search (automated or human) is perfect. It's perfectly acceptable and expected that users may choose their own `patt1 - patt3`.
h. The output file (typically given the same name of the input file but with `_parsed` at the end, is then inputted into EndNote as a **new** EndNote library and the results are human inspected.
i. Each reference now has the results of the script screening as a JSON database, pasted into the Notes field of the reference. For example, the following reference was obtained from WOS during a search for thorium,

Andrews, L., Cho, H. G., Thanthiriwatte, K. S., & Dixon, D. A. (2017). Thorium and Uranium Hydride Phosphorus and Arsenic Bearing Molecules with Single and Double Actinide-Pnictogen and Bridged Agostic Hydrogen Bonds [Article]. *Inorganic Chemistry*, *56*(5), 2949-2957. https://doi.org/10.1021/acs.inorgchem.6b03055

The Python script produced the following JSON in the Notes field:


```
{
 "WIPP": {
  "Search_1": "failed to match search terms regex solub|dissol|formation constant|dissociation constant",
  "Relevance": {
   "Ionic Strength > 3M": "",
   "pH 8.3 - 11.2": "",
   "T IC < 1.4E-2": "",
   "TH(IV)": "",
   "Solubility Study": "",
   "Solvent Water": "",
   "Pressure/Temp Std Cond": "",
   "Solid Phase in WIPP database": "",
   "Aqueous sol of known comp": "",
   "Element and species in WIPP model": "",
   "original experiment": ""
  }
 }
}
```


Some of the above fields in the JSON need to be updated given the new uncertainty G criteria. Regardless, "Search_1": "failed... means that none of the search terms in `patt1` were found in any field of this reference. Next consider the reference:

Blyth, A. J., Hua, Q., Smith, A., Frisia, S., Borsato, A., & Hellstrom, J. (2017). Exploring the dating of "dirty" speleothems and cave sinters using radiocarbon dating of preserved

Information Only

organic matter [Article]. *Quaternary Geochronology, 39,* 92-98. https://doi.org/10.1016/j.quageo.2017.02.002

This reference produced the following JSON:


```
{
 "WIPP": {
  "Search_1_Terms": "include regex solub|dissol|formation constant|dissociation constant",
  "Search_1_Hit": "solub",
  "Search_2_Terms": "exclude regex molten|arctic|modular|reactor|serum|mantle|crust|cave|marine|weather|monaz|glacial|separation|recovery",
  "Search_2_Hit": "Cave",
  "Relevance": {
   "Ionic Strength > 3M": "",
   "pH 8.3 - 11.2": "",
   "T IC < 1.4E-2": "",
   "TH(IV)": "",
   "Solubility Study": "",
   "Solvent Water": "",
   "Pressure/Temp Std Cond": "",
   "Solid Phase in WIPP database": "",
   "Aqueous sol of known comp": "",
   "Element and species in WIPP model": "",
   "original experiment": ""
  }
 }
}
```


A hit was obtained, which added the key Search_1_Hit" with value "solub" to indicate which term was found in the reference. In the second part of the pattern search (`patt2`) a hit was found to the word "Cave", which was used to screen out this reference.

Finally consider the following reference:

Bonin, L., Aupiais, J., Kerbaa, M., Moisy, P., Topin, S., & Siberchicot, B. (2016). Revisiting actinide DTPA complexes in aqueous solution by CE-ICPMS and ab initio molecular dynamics [Article]. *Rsc Advances, 6*(67), 62729-62741. https://doi.org/10.1039/c6ra08121e

which produces the following JSON:


```
{
 "WIPP": {
```


Information Only

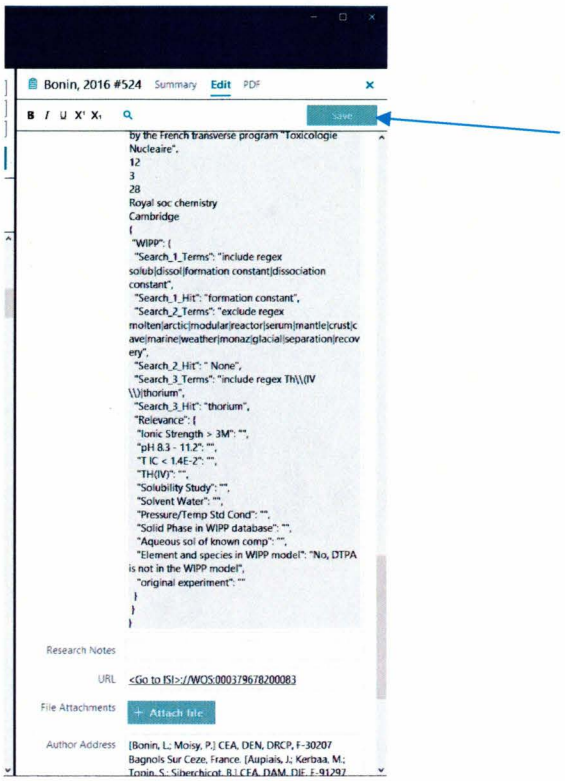
```
 "Search_1_Terms": "include regex solub|dissol|formation constant|dissociation constant",
 "Search_1_Hit": "formation constant",
 "Search_2_Terms": "exclude regex molten|arctic|modular|reactor|serum|mantle|crust|cave|marine|weather|monaz|glacial|separation|recovery",
 "Search_2_Hit": " None",
 "Search_3_Terms": "include regex Th\\(IV\\)|thorium",
 "Search_3_Hit": "thorium",
 "Relevance": {
  "Ionic Strength > 3M": "",
  "pH 8.3 - 11.2": "",
  "T IC < 1.4E-2": "",
  "TH(IV)": "",
  "Solubility Study": "",
  "Solvent Water": "",
  "Pressure/Temp Std Cond": "",
  "Solid Phase in WIPP database": "",
  "Aqueous sol of known comp": "",
  "Element and species in WIPP model": "",
  "original experiment": ""
 }
}
}
```


In this case, the reference passes all three searches from `patt1` – `patt3`, and hence this reference should be looked at by human eyes. We don't have DTPA in the WIPP geochemistry model, nor do we expect to have this in the repository so this reference would be screened out. The details of the specific WIPP screening criteria for references that need further human attention are given in 3. At this point the Relevance fields can be utilized to make notes of why this reference was human-screened out, by editing the JSON field Relevance in EndNote. E.g. in this case one might modify the field as follows:

"Element and species in WIPP model": "No, DTPA is not in the WIPP model"

When modifying the relevance fields, we should be consistent by stating yes or no first, then listing a reason why. This will make automatic reporting possible. Be sure to save your changes to the reference in EndNote by hitting the Save button at the top right-hand side of the screen.

Information Only



   a. Test your screening terms in d. - f. against papers that are known to be screened in. If it doesn't catch all of these, you need to adjust your screening terms.

3. Literature Screening
   a. Given a list of script screened in references, perform a human screening with the G criteria.
   b. High-Level Screening: May be performed by reading the title and abstract of the publication. Most publications identified in step 1 are screened out in this step.
      i. Ensure that the general criteria G2-G6 are met.
   c. Detailed Screening: Requires a more in depth read of the paper to determine its suitability for inclusion in the analysis.
      i. Ensure that the general criteria G7-G10 are met.
      ii. Meet with all team members and present the screening results to gain consensus.
      iii. Write up the rationale for including/excluding studies. This applies to studies that make it through the detailed screening level (Step 2.).
4. Prepare EQ3/6 input files for studies that are screened in.
   a. Copy solubility data from tables in the publication or digitize data from plots in the publication into Excel.
   b. Set up EQ3NR and EQ6 (when appropriate) input files to meet the experimental conditions of the studies.
   c. Run the code and process the output calculating the "D-values" where D = log S_m - logS_s where Sm = solubility measured and Ss = solubility simulated.

Information Only

d. Create empirical histograms and CDFs of the D values.

5. Reporting
    a. Include write ups from 3.b.iii.
    b. Prepare the data from Step 3 into plots and tables for the analysis report.

Information Only