protocols.io

🌐 Annotating PDF documents for the GeoArchive V.2

Sky Bristol[1]

[1]USGS

Version 2 ▼

Sep 01, 2021

In Development   *Share*

This protocol is published without a DOI.

GeoArchive

Sky Bristol

DISCLAIMER

ABSTRACT

This protocol provides steps to catalog NI 43-101 Technical Reports within the Zotero reference management system and use PDF highlighting and annotation to identify specific citation metadata and other descriptive information from the texts. It is part of a body of work being developed to manage a GeoArchive digital library of important reference materials in minerals assessments, energy assessments, and other work of the US Geological Survey.

PROTOCOL CITATION

WHAT'S NEW

I refined the protocol to point toward using the Zotero group library if desired and changed annotation details to indicate areas where we need several specific classifications of the same type of information (e.g., dates and institutions).

LICENSE

CREATED

Sep 01, 2021

LAST MODIFIED

Sep 01, 2021

PROTOCOL INTEGER ID

52882

MATERIALS TEXT

PDF files for the NI 43-101 reports referenced in this protocol or other types of materials may be obtained from web sites or specialized portals as is the case with the financial disclosure reports. It is assumed that you have these materials available to start this protocol.

DISCLAIMER:

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to protocols.io is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with protocols.io, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

BEFORE STARTING

Determine what PDF annotation tool you are going to use. You likely have something built in by default for your operating system to open and read PDFs. On Mac this may be the Preview app or something else. On Windows, this could be a free version of Adobe Acrobat Reader. Some operating system configurations may be set to open PDF files in a web browser, but you will want to open the file in the actual PDF tool itself as you may not be able to save the PDF from a web browser view.

Setup Collection in Zotero

1   Setup a local collection for the reports/documents you want to start annotating in Zotero.

We have created a group library in Zotero that can be used for this work, or you may elect to work in a separate collection in your own library ("My Library" in the Zotero client). If you would like to use the GeoArchive Group Library, you must create a Zotero account and be granted access to the group. This will let you accomplish the rest of the steps from the group space, making your report items, files, and annotations immediately available to other members of the group.

1.1   Create collection
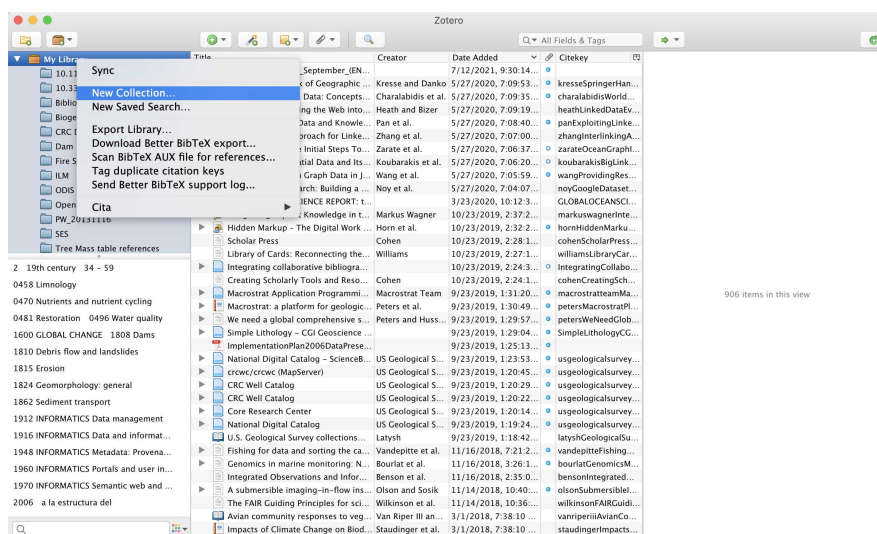
1.2
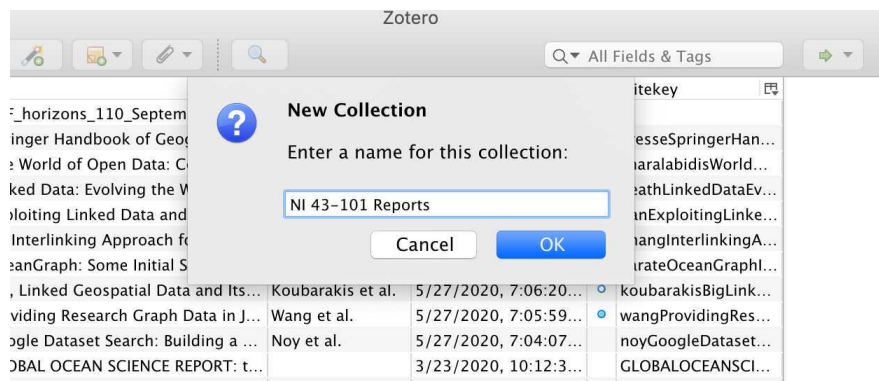
You might call this NI 43-101 Reports or some other name of your choosing. This is currently your local working space for these files, so it doesn't particularly matter what you call it.

Import Files to Zotero

2 Copying the files into Zotero means that Zotero will place them into its own directory structure. This facilitates Zotero working with the files, linking to them with internal references, and other actions. For our purposes, this simply makes sure that we have all of the files in one place along with any metadata we can start developing for them so that we can move it all to a shared, group library.

2.1 Drag and drop files
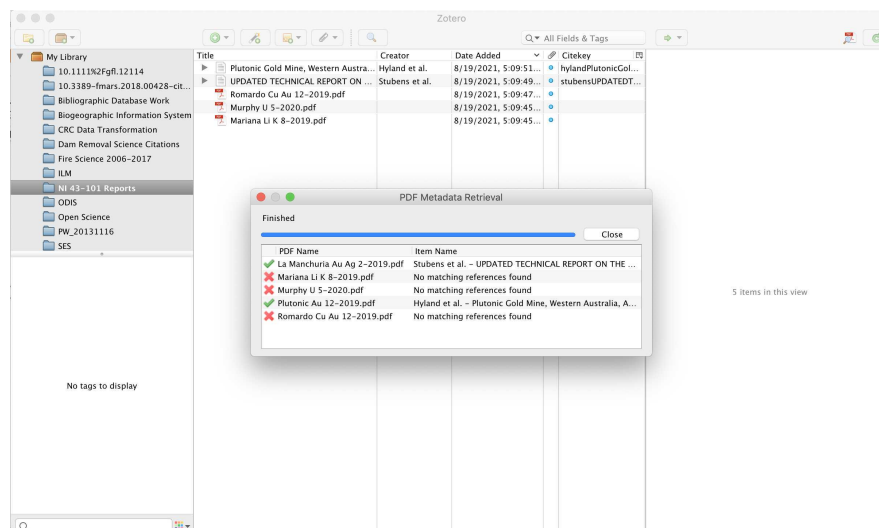
Dragging and dropping PDF files to a Zotero collection will automatically run a process to try and extract information from the PDFs to fill out citation metadata. This is hit and miss for many PDF files like the NI 43-101 reports. Some of them are well structured with at least some metadata that Zotero will interpret.

Click Close once this process is complete.

Establish Proper Item Types

3 After dragging and dropping a batch of files into Zotero, you may see that some have been brought in as journal article type items while others are simply files that do not yet have a Zotero item created for them. Zotero defaults to "thinking" that everything is a journal article as this is the most common use case. We need to establish items for everything and set everything to the appropriate type. This section covers that work.

3.1 Select files that were not processed to create items, right click, and select Create Parent Items from the list of options.
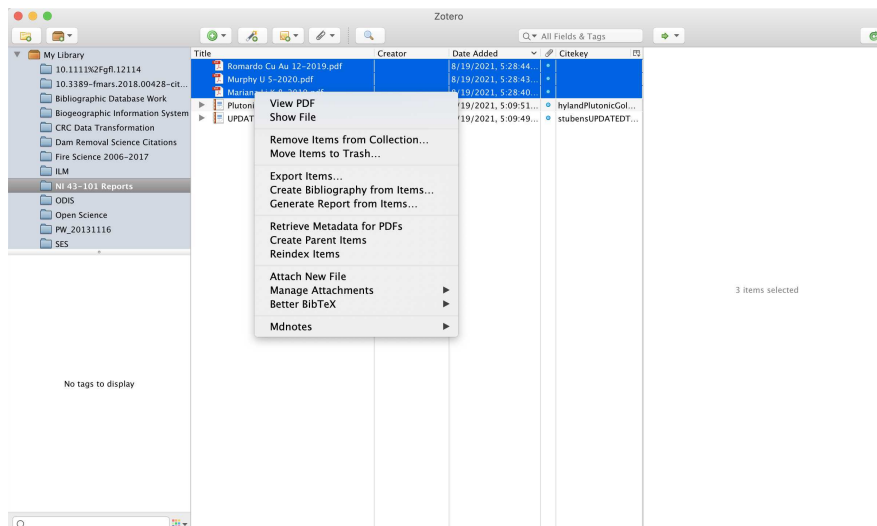
3.2 Change all item types to "Report"
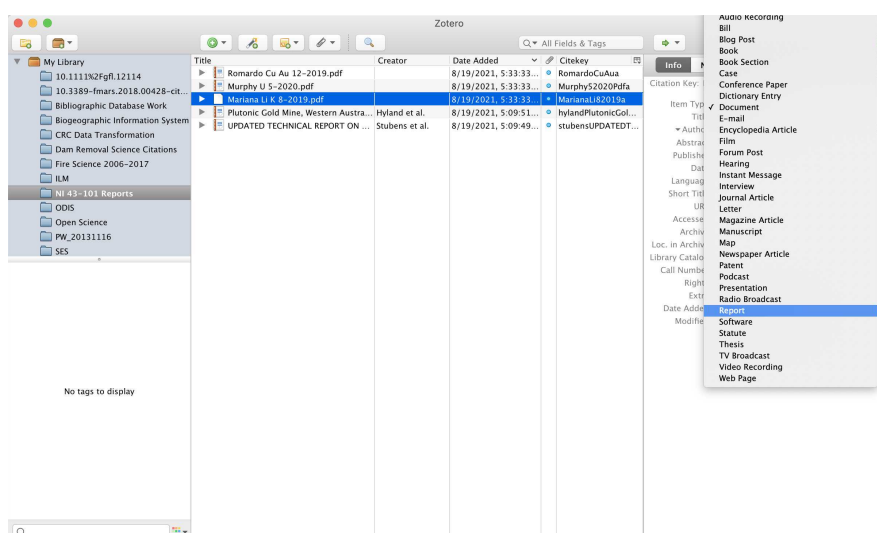
Select an item from the list to display that item's metadata form in the right-hand pane, click the current Item Type (typically either Journal Article or Document) to display the list of selections, select the "Report" type

4   We can now start annotating the PDFs documents stored with Zotero. Part of this will be developing a few fields of basic citation metadata. If the import process was able to successfully find things like title and authors, it is not necessary to annotate these elements separately. We already have them and can move that information into a shared library if we continue using Zotero in that capacity or export the information to include in another system. The following steps outline the approach to annotation that we can refine over time.

We will be evolving the annotation protocol over time to include more types of information content. If you have thoughts about the types of annotation that should be included as you go through this part of the protocol, please login with a protocols.io account and post your comments directly to the protocol.

4.1   Open a PDF for annotation. Zotero should be set up to use your system default PDF reader. This will vary between operating systems and circumstances. You can change this setting in Zotero preferences if desired.

You can usually either double-click on the item in the Library or open the "twisty" next to the item to view its file and double click that to open.

## 4.2   Highlight and annotate title (if necessary)
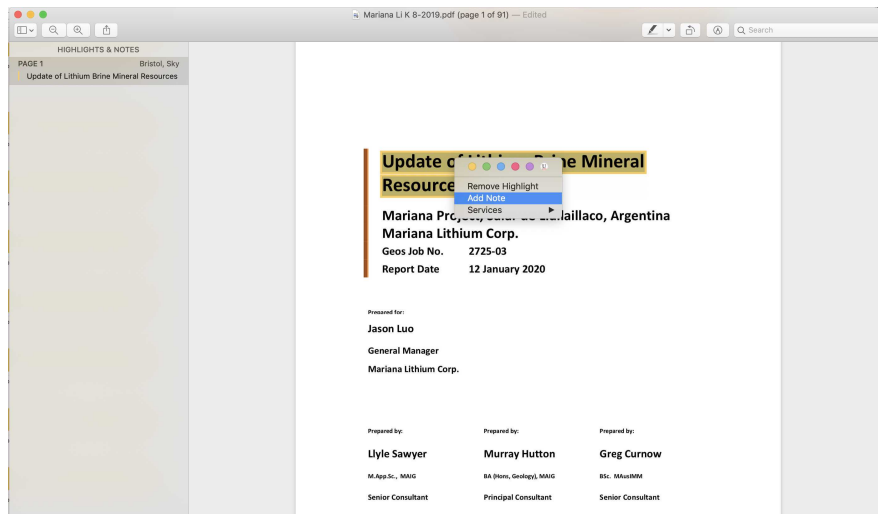
Select and highlight text for the title of the report that would be appropriate to list in a bibliographic reference. This will vary based on the report and can be found on a formatted title page or elsewhere in the document.

PDF reader software will vary on how to highlight a section of text, but this generally involves selecting text with a mouse and clicking some type of button or highlight feature. You can then add a note to that highlighted section. This will also vary and may involve right clicking the highlighted text or some other action to add a note.

Use **title** as the simple annotation for this element.
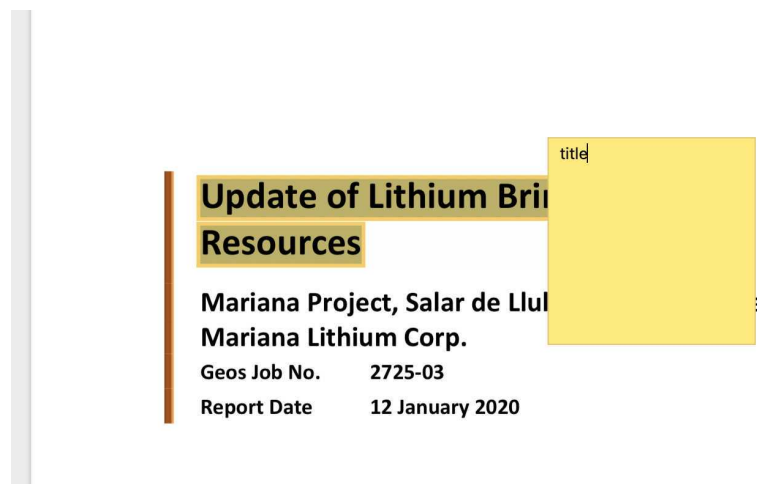
## 4.3   Highlight and annotate dates for the publication
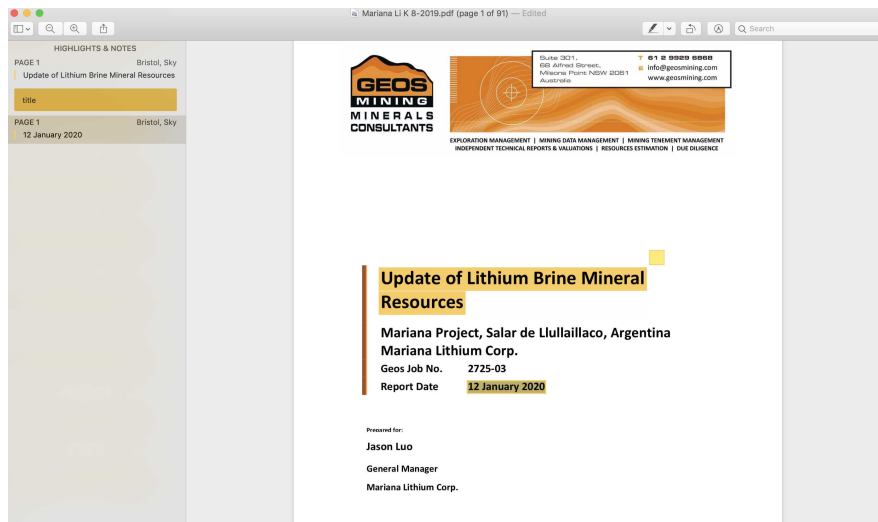
The most important date for the NI 43-101 Reports should be listed as "Effective Date." There may be other dates listed as well such as "Report Date." We need to identify at least one date to use as the publication date for the citation metadata and will revise our process to use the most appropriate annotated date for this purpose.

Highlight and annotate one or more of the following dates using these terms:
- **effective date**
- report date
- publication date

- bolded term is required

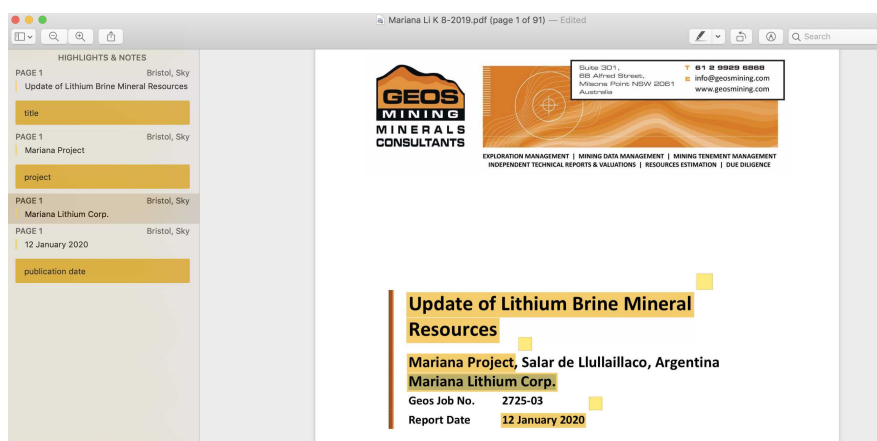## 4.4 Highlight and annotate the project name
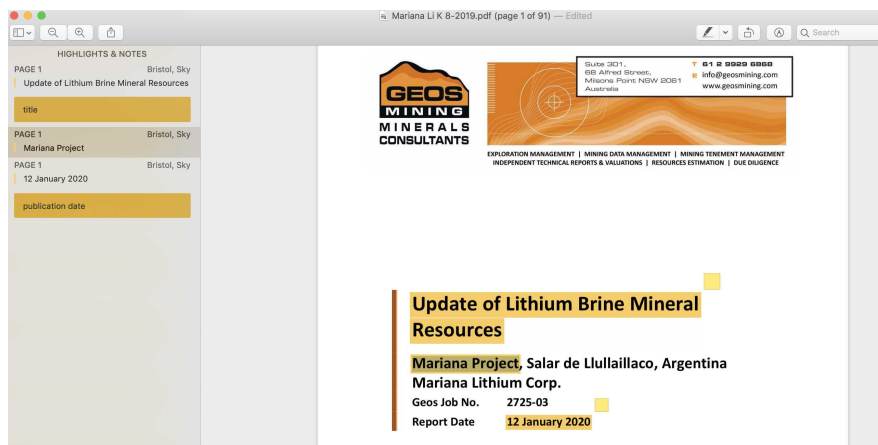
The project name may be found in different parts of the documents and may or may not include the word, project, in the text. We are looking for some type of simple shorthand reference to the particular mineral development project that can be used to identify sometimes multiple reports associated with the same project or can be found in other references. In some cases this may be the name of a mine or mining property that could be "well known" in other reference sources.

Use **project** as the simple annotation for this element

4.5   Highlight and annotate companies (or institutions) operating the mineral development project and/or producing the report

The name of the company operating the mine or exploration project as well as the name of the company responsible for generating the report may be listed in various places in the document.

Highlight and annotate one or more of the following company names/institutions using these terms:
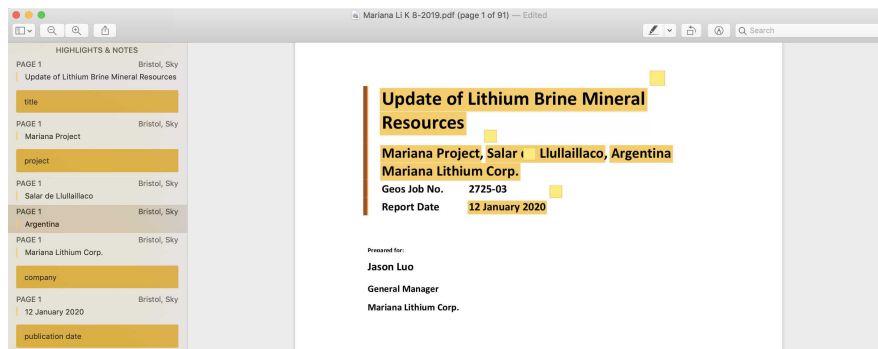- **report producer**
- project operator

- bolded term is required

The name of the company that produced the report is required in order to build a proper citation string for the report. The company operating the mine or development project should also be listed and will further aid in future work with the information.

4.6   Highlight and annotate significant place names

Basic geographic context for the projects can be found in various places in these reports. What we are looking for are a few simple things like country names, provinces/states, counties, and other administrative units that can be used to generally locate the area where a mining property exists. These have been used to create a folder structure in the past, but we can pull them out as place name keywords as an alternative.

Use **place** as the simple annotation for this element, highlighting and annotating as many places as are significant and useful in roughly geolocating the development project and associated geoscientific information.

4.7 Highlight and annotate report authors
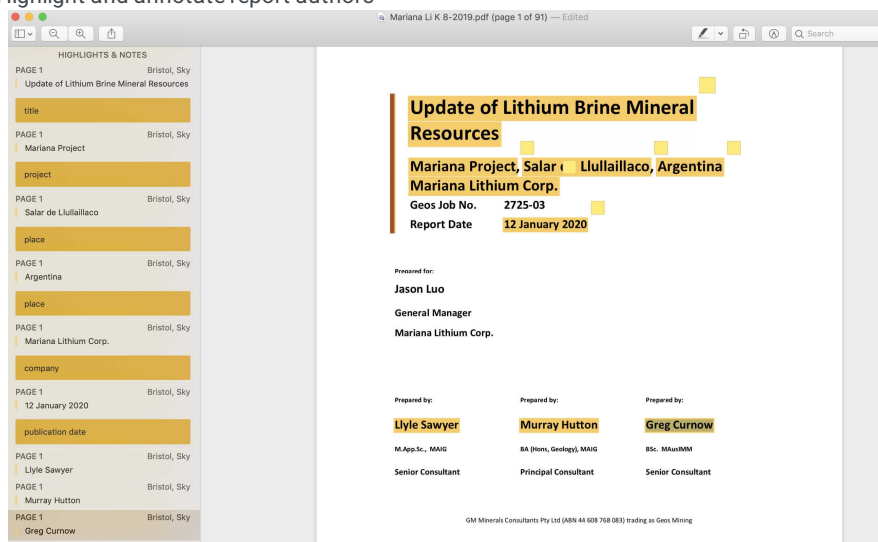
Author names can be listed in various ways in the NI 43-101 Reports. We are looking for the best reasonable text to use in identifying report authors for citation metadata.

Use author as the simple annotation for this element

Prepared by:

Llyle Sawyer

M.App.Sc., MAIG

Senior Consultant

Prepared by:

Murray Hutton

BA (Hons, Geology), MAIG

Principal Consultant

author

GM Minerals Consultants Pty Ltd (ABN 44 608 768 083) trading as Geos Mining

When extracting these annotations for use, multiple highlights using the same annotation will be pulled out in an ordered list based on where they appear in the document, giving us the ability to produce an ordered list of authors.

4.8 Highlight and annotate latitude and longitude coordinates if provided

We ultimately want to identify point coordinates or more sophisticated geospatial footprints for as many of the mining properties as we can. These may be listed in the document in various places, and if we can identify patterns in how these are documented, we can potentially develop extraction algorithms. In the near term, highlight whatever form the coordinates are in, and we can extract and process those further.

Use **latitude** and **longitude** as the annotations for these elements

1 Executive Summary

1.1 PROPERTY LOCATION

The Mariana property covers the Salar de [longitude] una plateau of western Salta Province, Argentina (latitude 24°48'30"S

1.2 OWNERSHIP

The mineral development and mining per held in the name of Litio Minera Argentina SA (LMA). LMA is the project o Lithium Corp. (MLC), a subsidiary of Jiangxi Ganfeng Lithium Co. Ltd. (Ganfeng).

9

4.9 Highlight and annotate commodity names
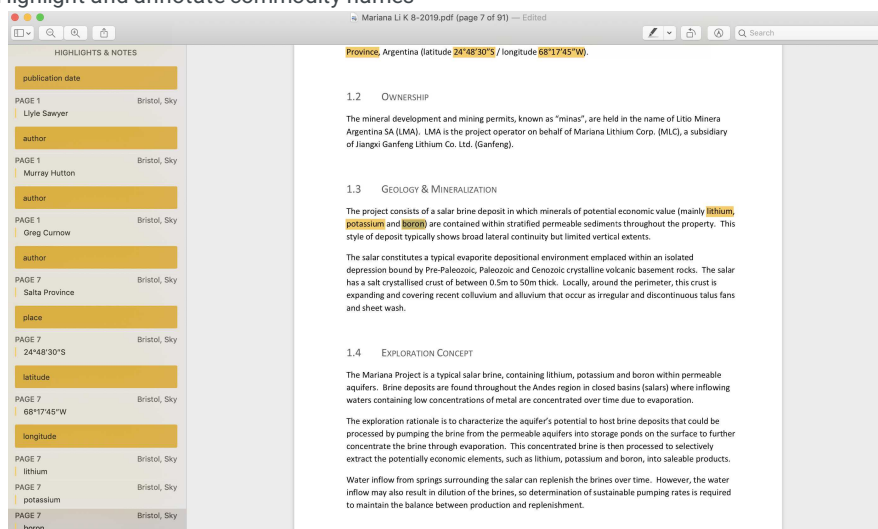
Commodity names may be listed and summarized in different places in the document but are generally mentioned by name in the Geology & Mineralization section. In the near term, we need a simple identification of mineral commodities by name to bring into tags for helping to navigate to reports applicable for a given assessment or analysis. Longer term, we will work to use named entity recognition algorithms to identify any and all minerals that may be of interest, using natural language processing to determine their relative "significance" in terms of what is being described in a given text.

Use **commodity** as the simple annotation for this element

4.10 Highlight and annotate tables of data and other important details for further use of the reports in assessments and other practices. We will monitor and discuss development of annotation conventions over time to develop a reasonable vocabulary for identifying additional important elements in these and other documents.

We will be employing a toolset for extracting tabular data and figures from these texts that we can improve on over time. It is not necessary to highlight an entire table or figure, but it will be helpful to highlight a title or caption for tables and figures of interest as well as providing a meaningful annotation.

Annotations can also include further meaningful text to aid in the use of the material. Simple keyword types of indicators like we are using for identifying citation elements are useful for running extraction processes to classify particular information for use, but we can experiment with a convention for delimiters to incorporate a classifier for software to use along with a longer annotation for human

consumption. For instance, something like the following might be a reasonable annotation on a particular table in a report. It tells a computer program that the table is of a particular type and then provides an opinion by the annotator for further investigation.

summary of mineral resources: This table shows a higher than expected concentration of potash than expected for this formation.

Saving and Processing

5   Save the PDF file at the end of or during annotation. Depending on the platform and software used, this may be done automatically. The highlights and annotations will be saved with the Zotero copy of the file.

At this stage, the most important thing is simply to get the annotations started within the files. Don't worry that the metadata you've highlighted is not a part of the citation information in Zotero yet. We'll get to that point shortly through a software process that will need to extract the annotations from the texts and place it back into Zotero.

If we do continue using the Zotero group library for a time, at least up to the limitation of 300MB of files per user, we can set up an automated process to extract the annotations and populate Report metadata content. This will follow the example outlined here but can be set up to run in the background.