

Alone in a crowd ... transcribing together

A crowd-sourcing project by Cambridge Digital Library
undertaken during the University of Cambridge's closure
period due to the Coronavirus outbreak in 2020

[Project Home](#)

1.1 About

"Alone in a crowd ... transcribing together" is a crowdsourcing campaign by [Cambridge Digital Library](#) with the aim of transcribing digitised material that does not have any existing research project to do so.

As much of the University of Cambridge is currently physically closed due to the coronavirus outbreak, we have been turning our attention to the possibilities of what can be achieved remotely.

At its core, Cambridge Digital Library values access, collaboration, engagement, innovation and inspiration. In these times of physical isolation, digital platforms can provide a focal point for those with a variety of interests and skills to come together as a community.

Our digital objects are not simply facsimiles. While fantastic images of unique and distinctive collections provide virtual access, transcription unlocks their potential to be something more than the original item.

The benefits include:

- Aiding discovery – with so much digital content, it can be tricky to explore
- Enriching our understanding and aiding accessibility
- The application of further digital research methods eg automated analysis

What do we do with your transcriptions?

Once a transcription file is complete, we incorporate it into the metadata file of the item. As well as being indexed by our search engine, the transcription of each page will then display adjacent to the digitised image when a user selects to view the transcription tab. The metadata files in Cambridge Digital Library are also downloadable from the site under a [CC BY-NC](#) creative commons licence.

1.2 What are we transcribing?

We have so much amazing content that it's difficult to know where to start. To get this going, we'd like to focus on a collection of nearly 400 small notebooks used by botanist [Oliver Rackham](#). A Fellow at Corpus Christi College, Cambridge, Rackham would record his observations of nature in these notebooks.

Now part of the archive collections at Corpus Christi, they contain a wealth of data about the natural world that's just shouting out to be unlocked by being transcribed!

If you want to know more about Rackham and his work, check out the [Oliver Rackham Collection](#) page on Cambridge Digital Library, where you can also explore the notebooks.

2.0 Step-by-Step Guide

2.1 Get set-up

2.1.1 Step 1: Create Account

First things first, you're going to need a GitHub account.

Go to: [GitHub](#)

Create a username, add your email address and create a password. After authenticating that you're a human it will ask you to select a plan - choose the free option. It will then give you the option to let GitHub get to know you a bit better with some questions. You can just scroll to the bottom of the page and click "Skip this step".

Check your email – you should have an email from GitHub asking you to verify your email address. Go ahead and do this, which will then open a browser tab that should signal you have authenticated your email and ask you to create your first "repository".

Unless you want to learn a bit more about GitHub and do anything fancier for your own work, you won't really need an initial repository, but go ahead and create one:

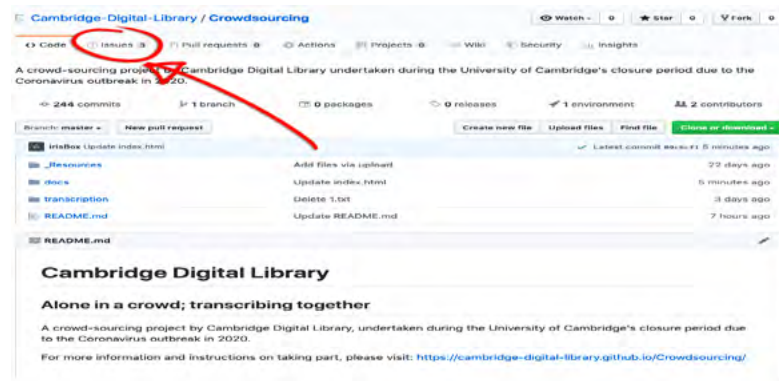
- Give your repository a name (you can just call it "test" or "my-repository")
- Select "Private"
- Click "Create repository" at the bottom

This will open a page up that shows a whole load of stuff you won't need to worry about. If you wanted to learn a bit more about Git and GitHub, there should be a prompt to "Read the guide", but we'll cover everything you should need to know here – you don't need to overwhelm yourself with all the coding stuff in order to help with our Cambridge Digital Library transcription project.

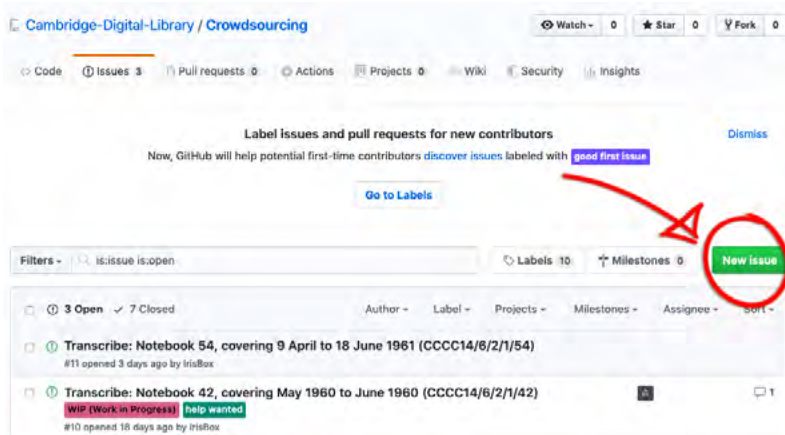
2.1.2 Step 2: Join our project

Now you have a basic account set up, you can find and join the project. Go to the project homepage: [Cambridge Digital Library - Crowdsourcing](#)

Click on the "Issues" tab of the repository



Next, click the green button "New Issue"



Type "Add me" in the "Title" field and then click the green button "Submit new issue".

This will let us know that you want to join in. We can then add you to the project by sending you an invitation to join. Once we do this, you should receive an email notification.

Go ahead and click the green button "View Invitation" in the email, which should open a browser tab with a green button to accept the invitation.

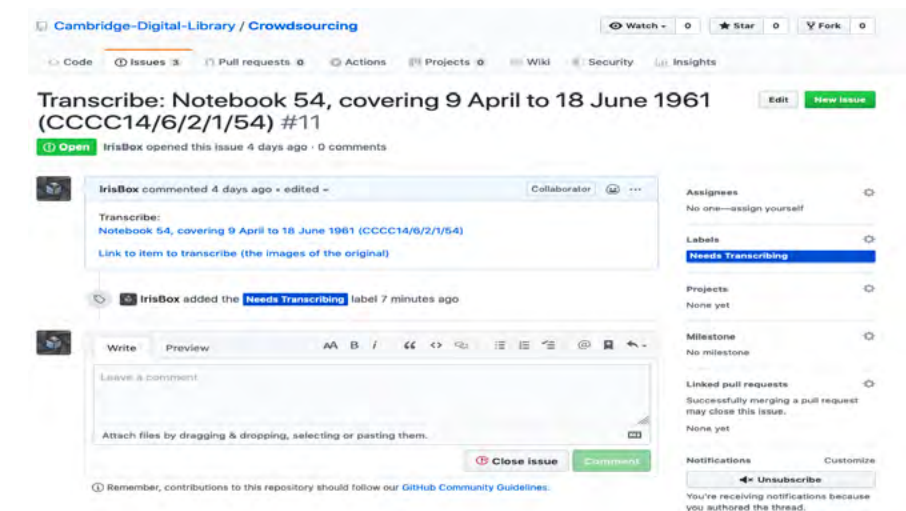
Welcome aboard! :-)

2.2 Starting your first transcription

2.2.1 Step 1: Find an item you want to transcribe

Find an item to transcribe by going to our [GitHub "Issues"](#) list.

You will see a list of "Issues" - these are like tasks that need doing. They have labels so that we can keep track of what needs doing, what is complete and what stage each file is at. Find one that has a blue "Needs Transcribing" label and click on the issue title. You should now see further details of the task, including the title and url of the original item, and a link to the text file that you will need to transcribe into.



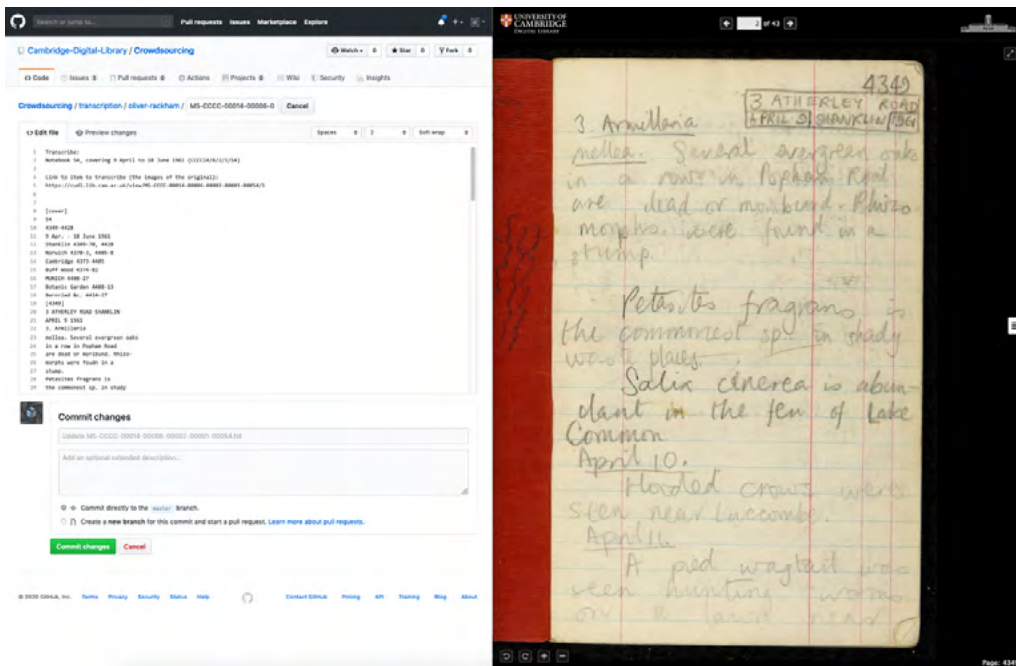
You can view the original item using the link and have a look through a few pages to get an idea of what it will be like to transcribe.

If you like the look of it, head back to the issue and look for the "Assignees" section of the task (usually to the right of the screen or try scrolling down).

This should read "No one - assign yourself". You can click on the "assign yourself" text to let us know you have chosen to work on that item.

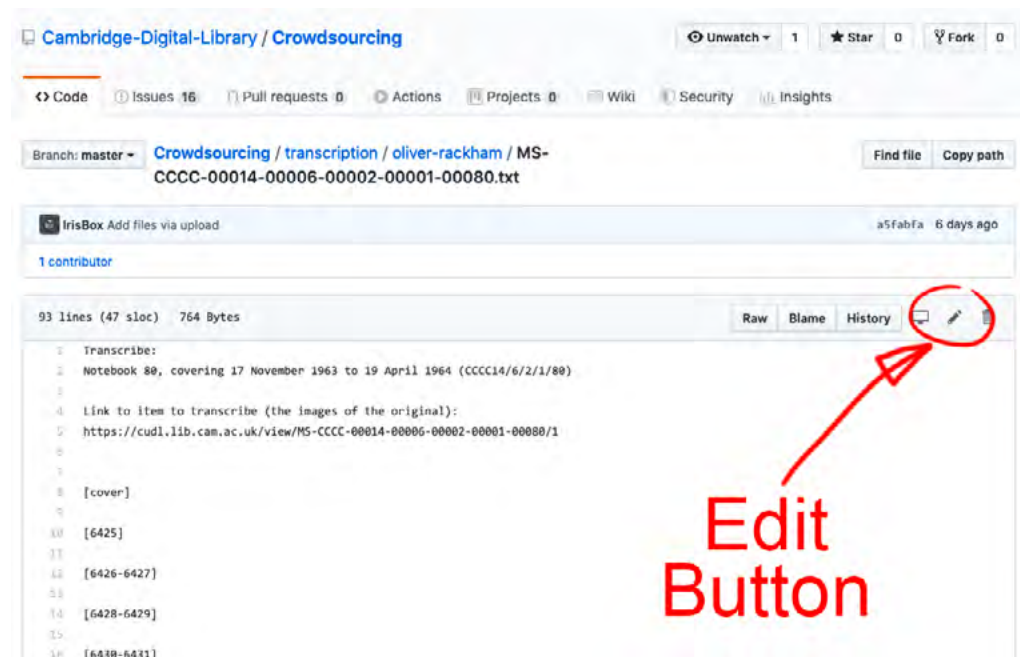
The "labels" section of the task, below "Assignees", will allow you to change the label from "Needs Transcribing" to "Work in Progress" by clicking on the word "Labels".

You're now ready to begin transcribing (Yay!) - to do so, find and click the "Transcription text file to edit:" link in the issue. If you right-click the link and select "Open in new window" then you will be able to see the text file in the new window and then open the link to the images in the first window so that you can see both. You might want to position the two browser windows next to each other so that you can see both the original image that you are transcribing from, and the text file in which you are editing the transcription, like in the image below:



2.2.2 Step 2: Start editing the text file

So, in one of your windows you should now see the text file that you are going to edit. To begin editing, find and click the “pencil” icon - this should be to the top-right hand side of the file viewing window. This is indicated in the image below:



Now, following our transcription guidelines below you can begin transcribing!

2.2.3 Step 3: Save your work

If you want to take a break from transcribing the item, you will find a green button at bottom of the text file in Github that says “Commit changes”. Go ahead and click this and your work will be saved.

2.2.4 Step 4: Find your work to carry on editing

Log back into your [GitHub account](#).

When you are logged into GitHub there are some site navigation links along the very top of the web page. One of these is “Issues” - clicking on this will take you to a list of all the tasks that you

have assigned yourself to and you should be able to find the links to both the images of the original item and your transcription text file in your issue.

You can also go straight to our entire [list of tasks](#), log in from there (top right) and then find the task you’ve been working on or a new one to start.

2.2.5 Step 5: Finished transcribing an Item?

Once you think you have finished transcribing an item, find your task in your list of “Issues” and click on the task title. Click on the “Labels” section and go ahead and change the label to “Transcription Finished”.

Next you can go ahead and assign yourself to a new item task to transcribe!

2.3 What do I do if I get stuck and need some help?

Go to the task that you need help with. You can add a “Help Wanted” label in the Labels section of the task. You can also add comments to the task where it says “Leave a comment”, in this section, you can describe what you need help with, and if you include “@Cambridge-Digital-Library I need some help with this one!” then we will get a notification and be able to head over and help you as soon as we can.

If you have friends or colleagues that are also transcribing with us, you can also use this feature to ask each other questions, discuss things between yourself or tell each other if you find something exciting in the item you are transcribing and want to share!

3.0 Transcription Guidelines

In this guide, we will describe how we would like you to transcribe for the purposes of this project. We have tried to keep it as simple as possible but please do get in touch if you are unsure about anything or have any questions. Don’t worry if you come across anything you’re not sure of, simply carry on around it and ask us about it later.

We will also describe a basic form of “mark-up” that uses square brackets around certain indicators. These will allow us to identify specific elements in the document later on.

3.1 A quick guide to our simple transcription mark-up

This is a quick guide to the basic “mark-up” indicators we would like you to use in your transcriptions. Detailed instructions on using this mark-up follow below.

You might want to copy and paste these into a simple text file on your computer and keep them close at hand to copy and paste directly into your transcriptions to save you repeatedly typing them out.

- [????] = Illegible or unclear text
- [!!!!] = Deleted or obscured text
- [SYMBOL] = Symbols and non-standard characters
- [DIAG] = Presence and location of any graphical material (sketch, diagrams, photo etc.)
- [TABLE] = Use at beginning and end of any tabulated text or data

3.2 Spelling, grammar & punctuation

This is an opportunity to show your rebellious side and throw out the rule book!

Why? Because we need you to replicate the written text exactly as it is in the original document.

This is harder than you'd think, since our brains normally try to autocorrect things, but you need to preserve all of the original spelling, punctuation and grammar. This also extends to the word order and things like abbreviations – we need you to simply copy what is written, as it is written.

Why is this important? Different people at different times in history have used spelling and punctuation for different reasons. Notebooks for example are often personal documents and therefore contain many quirks. We do not want to put a subjective contemporary slant on the text, but to replicate a computer readable objective copy of it.

NB: Depending on your web browsers settings, they may try to autocorrect as you type, so keep an eye on it! If you encounter this and can't find how to switch it off, please contact us and we'll do our best to help.

3.3 Line-breaks

Please try and preserve line-breaks as this will make it easier to review your transcription. Since GitHub is geared towards software development and computer code, its editing interface will display the text per line (rather than wrapping it to fit it all in the window). This means each line of text is automatically numbered. Once you come to the end of transcribing a line in the original document image, simply hit "return" or "enter" and this should begin a new line as you would normally expect in Word for example.

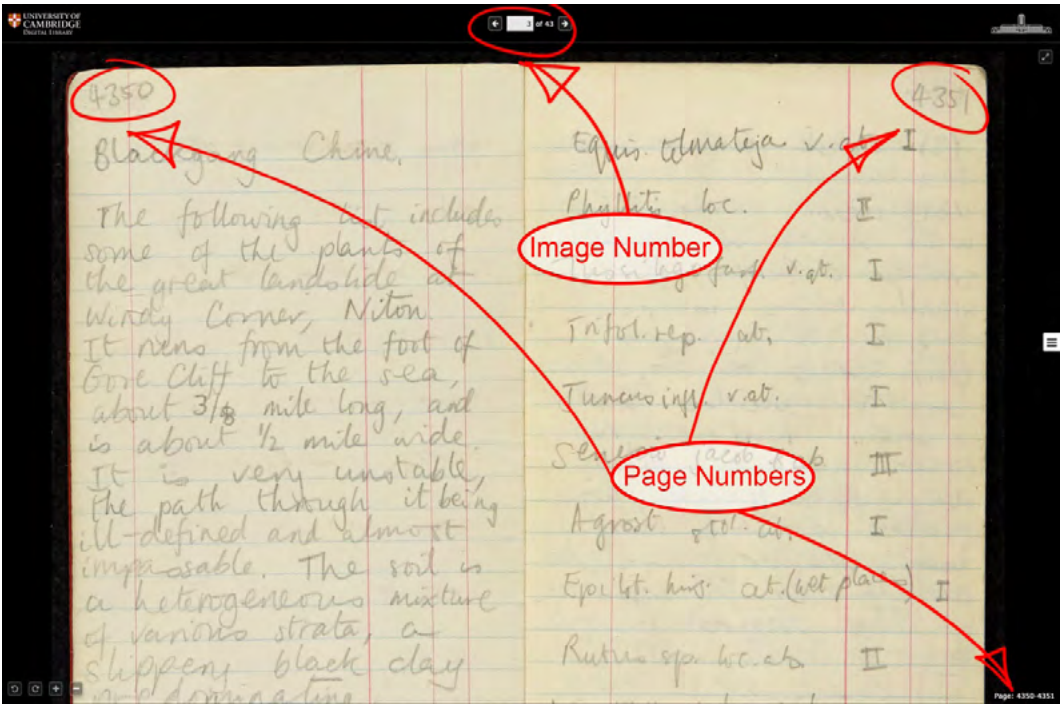
3.4 Pages

The text file you will edit will have all of the pages in the notebook laid out in a list in square brackets.

For example:

- [5128-5129]
- [5130-5131]
- [5132-5133]

These page references are the same as those found in the bottom right-hand corner of each image on the Digital Library.



To transcribe text on a page, simply start typing underneath the relevant page reference.

For example:

- [5128-5129]
- Type what is written on the page ...
- [5130-5131]
- [5132-5133]

If there is no text on a page, simply leave it blank and move onto the next page.

3.5 Font styling (bold, underline, italic, superscript, indentation etc.)

In order to make this process as simple as possible, we do not want you to worry about any font styling such as bold, underline, superscript or indentation etc.

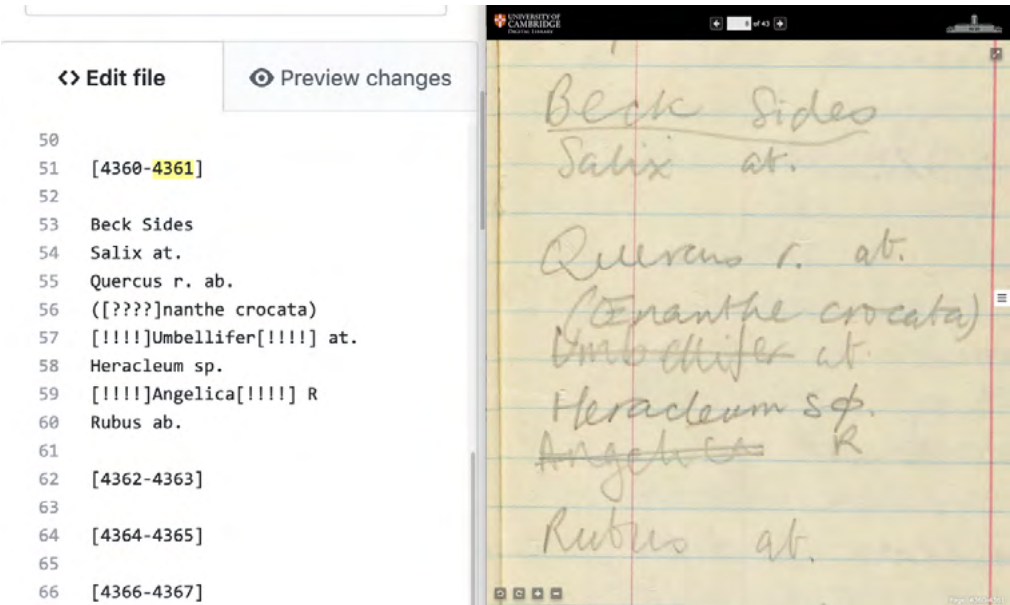
3.6 Illegible or unclear text

Some styles of handwriting, or even the author's own style and haste, mean it is not always easy to read what has been written in the original document. Don't worry about this at all, simply use a pair of square-brackets around four question marks ("[????]") to indicate what you can't make out.

If some of the letters are clear, you can enter what you can make out and use the above to indicate any relevant section of a word that is unclear. For example, using the phrase "Cambridge Digital Library", if you can only make out some of the letters then your transcription might look a bit like this: "Ca[????]b[????]ge Dig[????] Li[????]y"

Why is this important? As we said above, our aim is to re-create an objective copy of the text in the original document rather than a subjective interpretation of it. Therefore it is safer to be explicitly clear that something isn't clear rather than introduce any potentially incorrect text. Often it might also be the case that a fresh pair of eyes might see or read something differently to you. This method makes it easy to find any unclear text and perhaps go back to it later on.

Here is a screengrab showing examples of both illegible text and deletions:



3.7 Deletions

If someone has obscured any text in the original document, it would still be useful for us to know that text has been deleted or indeed, if you can make any of it out clearly, what that text was. For example, sometimes an author might put a "cross" over or put a line through a word or chunk of text, or they might have obscured it completely by scribbling/blotting over the top of it for example.

In these situations, please indicate the deleted text by using a similar method as you did for any

illegible text but this time use exclamation points (“[!!!!]”) instead of question marks (“[????]”). You can see an example of this in the screengrab image above.

This time though, we would need to know where the deletion begins and ends. Therefore if you can make out some or all of the original text, please insert “[!!!!]” at the beginning and end of the word, or section of words, that has been deleted or obscured.

So if the word “scribble” in the phrase “I don’t like the scribble word” has been completely obscured, your transcription would read: “I don’t like the [!!!!] word”

If it had been crossed out or had a line put through it, your transcription would read: “I don’t like the [!!!!]scribble[!!!!] word”

3.8 Symbols and non-standard characters

If you encounter any symbols or non-standard characters, by which we mean anything you can’t enter quickly and easily using a modern computer keyboard, then please indicate these using a pair of square brackets around the capitalised word “symbol” (“[SYMBOL]”). This will mean we can go back over your transcriptions and easily identify and insert any symbols or odd characters at a later date.

3.9 Graphical material (Sketches, diagrams, photographs etc)

As with any unusual symbols, it might help us in the future to be able to identify any graphical material that exists in the original text. By this we mean any sketches, diagrams or even if a photo has been inserted. Don’t worry about transcribing any text that is located within any sketches/diagrams etc, simply insert a pair of square brackets around the capitalised abbreviation “diag” (“[DIAG]”) and move on with transcribing the rest of the text. We’ll hopefully be able to go back through your transcription later on and handle these.

3.10 Tables and tabulated data

Some documents also contain tables of text or data and it would be useful for us if you could transcribe these. Formatting tabulated data in plain text can be complex, so we just want you to transcribe these as simply as possible.

Please indicate the beginning and end of tabulated data by inserting a pair of square brackets around the capitalised word “table” (“[TABLE]”) on a new line. Between these two indicators you can use regular spaces and line-breaks to transcribe as normal. Please don’t add any additional characters that are not present in the text such as commas, slashes or pipe symbols to indicate columns for example. This way it will be easy to review the transcribed words without causing any formatting problems later on or including any characters that the original author didn’t use.

If you are not sure about transcribing tabulated data, you can just insert the “[TABLE]” indicator on a new line and move on to the next piece of regular text. We can then go back over these and review them later.