

Synthetic Healthcare Database for Research (SyH-DR)

A Synthetic Nationally Representative All-Payer Claims Database

Data Request Guide

AHRQ Publication No. 24-0019-2-EF
April 2024



Purpose

The Synthetic Healthcare Database for Research (SyH-DR) Data Request Guide is intended for researchers who want to apply for access to SyH-DR. This step-by-step guide includes:

1. Instructions to complete and submit the application package,
2. An overview of the AHRQ application review and communication process, and
3. The system requirements for accessing SyH-DR.

The appendixes to this document include the required SyH-DR application and data use agreement forms.

Please note: Your application must be approved by AHRQ before you receive access to SyH-DR. The information about accessing SyH-DR is provided so that you may prepare for the data download at your convenience.

More information about the data is available in the SyH-DR Codebook and Sampling, Weighting, and Synthetization Methodologies at <https://www.ahrq.gov/data/syh-dr.html>.

Email any questions about SyH-DR or the application process to DataInnovations@ahrq.hhs.gov.

Table of Contents

Purpose	i
I. How To Apply for Access to SyH-DR.....	1
II. AHRQ Application Review and Communication Process.....	3
Overview	3
Confirmation of Receipt	4
Information Request	4
Application Under Review	4
Application Disapproved	4
Application Approval	4
Instructions for Download.....	4
Download Confirmation	4
III. System Requirements for Accessing SyH-DR	5
General System Requirements.....	5
Appendix A: SyH-DR Request Form	7
Appendix B: SyH-DR Data Use Agreement (DUA).....	12
Appendix C: A Brief Introduction to Using SyH-DR.....	16
Instructions for Microsoft Windows Users: Generating the RSA Key Pair Using WinSCP	17
Instructions for MacOS Users: Generating the RSA Key Pair Using Terminal.....	21

I. How To Apply for Access to SyH-DR

1. Verify that SyH-DR is an appropriate database to use for your research purposes by reviewing all information provided on the SyH-DR web page, including this Data Request Guide.

Figure 1. AHRQ SyH-DR Web Page (<https://www.ahrq.gov/data/syh-dr.html>)



2. Fill out the SyH-DR Request Form found in Appendix A of this document. If possible, do not submit handwritten request forms. [Jump to Appendix A.](#)
3. Review and sign the SyH-DR data use agreement (DUA) found in Appendix B of this document. DUAs may be handwritten. [Jump to Appendix B.](#)
4. Obtain a signed DUA from each person on your research team who will be a SyH-DR data user. Team members do not need to complete a separate SyH-DR Request Form.
5. Submit your application by emailing DataInnovations@ahrq.hhs.gov. Create a new email with the subject line "Application Package – Applicant Last Name." If possible, do not submit image files.
6. Optional: Approved data recipients are required to generate an RSA Key Pair to download the SyH-DR data files. You may want to submit the RSA Public Key as part of your application package to streamline the data download process after approval. Otherwise, you will be required to send the RSA Public Key once your application is approved. [Jump to Appendix C: Instructions for Creating an RSA Key Pair.](#)
7. Use the template below as the body of the email. Be sure to complete each of the bolded and bracketed items. Regardless of who sends the Application Package email, the applicant

information in the body of the email must be the same as the information on the SyH-DR Request Form.

Figure 2. Template Application Package Email

The attached SyH-DR Application Package is from:

[Applicant Name]

[Applicant Institutional Affiliation]

[Applicant Email Address]

[Applicant Contact Phone Number]

The following materials are included in this Application Package:

- 1 SyH-DR Request Form
- **[# of DUAs attached]** SyH-DR DUA(s)
- **[RSA Public Key (delete if not submitting)]**

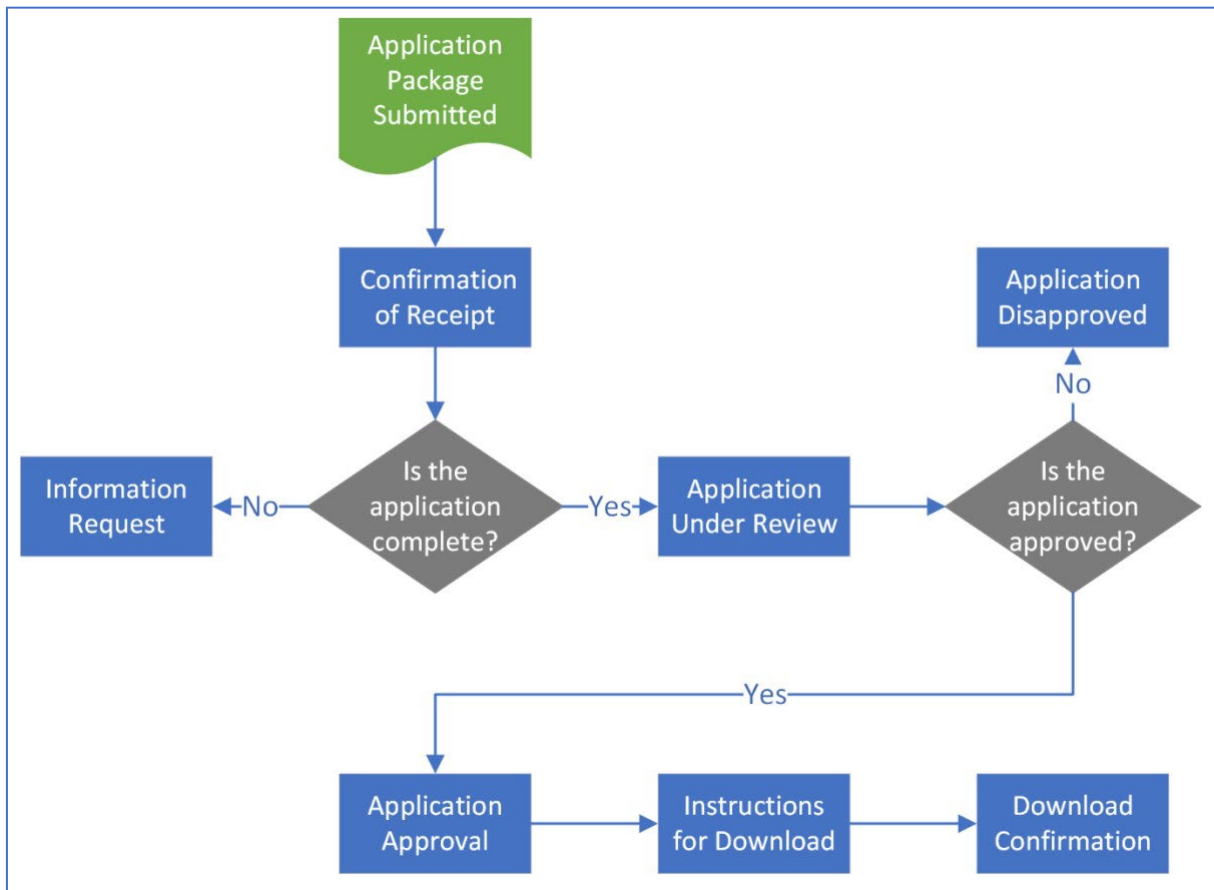
8. Send the email. You will receive a confirmation of receipt within 1-2 business days. If you do not receive a confirmation, resend the email with the attachments.

II. AHRQ Application Review and Communication Process

Overview

After an Application Package is submitted, AHRQ will send email notifications as the application moves through the review process. AHRQ strives to provide access to SyH-DR as quickly as possible, but the exact timing of each notification can vary depending on the content of the Application Package, timeliness of applicant response, and number of applications. The diagram below provides an overview of AHRQ communications through the application review process. Each rectangular box represents an email notification.

Figure 3. Overview of the AHRQ Application Review and Communication Process



Confirmation of Receipt

AHRQ will send a Confirmation of Receipt within 1-2 business days after an Application Package email is submitted.

Information Request

If an Application Package is incomplete, AHRQ will send an Information Request that specifies what component of the application is missing. The applicant should reply to the Information Request email and include the missing information or document.

Application Under Review

Upon verification that an Application Package is complete, AHRQ will send an Application Under Review notification.

Application Disapproved

If an application is not approved, AHRQ will send an Application Disapproved notice to the applicant. AHRQ created SyH-DR based on agreements with entities that provided the original sources of data. In some cases, the application may be disapproved because it is not consistent with these agreements.

Application Approval

Once an application is approved, AHRQ will send an Application Approval notification. The Application Approval will include a list of any additional steps that the data recipient must complete so that AHRQ can provide download access, such as generating the RSA Key Pair.

Instructions for Download

Once download access has been established for the data recipient, AHRQ will send the Instructions for Download email. The data recipient should reply to that email with any questions related to downloading the data.

Download Confirmation

AHRQ will track downloads through the secure file transfer protocol (sFTP); however, AHRQ will send a Download Confirmation email to verify that the data download was successful. The data recipient should reply to that email either confirming successful download or stating the nature of the download issue. If AHRQ does not receive a response, download access will be revoked by default after 5 business days.

III. System Requirements for Accessing SyH-DR

General System Requirements

The following general system requirements provide baseline guidance on the hardware and software needed to download and access the SyH-DR data files.

Hardware Requirements

Hard drive space: At least 50 GB of available hard drive space is required to download the SyH-DR data files in a single format (CSV, SAS, or Stata). More available hard drive space is required to download the data in more than one format.

RAM: A minimum 1 GB of available RAM is required to access the SyH-DR data. Depending on the statistical analysis software used, some SyH-DR data files may require 16 GB of available RAM. If the system has less than 16 GB of available RAM, we strongly advise subsetting the data using the read-in programs available for download with the data files. In addition, closing unnecessary programs before accessing the data may increase the amount of available RAM.

File transfer protocol client

The file transfer protocol (FTP) client enables secure handling of file uploads and downloads. The FTP client serves two important functions:

1. First, you will use the FTP client to generate the RSA Public/Private Key Pair. (Instructions for both Microsoft Windows and MacOS users are included below.)
2. Then, you will use the FTP client to connect to the SyH-DR sFTP and download the data files. (Instructions for Download are emailed after Application Approval.)

For Microsoft Windows users: Use WinSCP, an open source FTP application available for Microsoft Windows. [Download WinSCP](#).

For MacOS users: Use the Terminal application that automatically comes with MacOS. You should be able to find Terminal by searching your Applications folder or in Launchpad. (You should not need to install this application.)

Email questions about the FTP client requirement to DataInnovations@ahrq.hhs.gov.

RSA public/private key pair

Before you can receive access to download the SyH-DR files, you must generate an RSA Public/Private Key Pair using your FTP client. The RSA Key Pair is required to securely transmit the SyH-DR files. The public key is a unique encryption key that you must send to AHRQ. The private key is a unique decryption key that must be saved to your computer. *Do not share the private key.*

[**Jump to Instructions for Microsoft Windows.**](#)

[**Jump to Instructions for MacOS.**](#)

Note: You may send the RSA Public Key with your Application Package or after approval.

Email questions about the RSA Key Pair requirement to DataInnovations@ahrq.hhs.gov.

Third-party zip utility

The SyH-DR data files are compressed for download efficiency. After you download the SyH-DR data files, you will need to use a third-party zip utility to decompress the files. You may use any zip utility of your choosing. If you do not already have a zip utility installed on your computer, 7-Zip is a popular open source zip utility.

[Download 7-Zip for Microsoft Windows.](#)

[Download 7-Zip for MacOS.](#)

Statistical analysis software

The SyH-DR data are currently available as CSV, SAS, and Stata files to give you the flexibility to conduct your analysis in the statistical analysis software of your choosing, such as SAS, Stata, or R. Also provided are SAS and Stata read-in programs to load the data files of interest in either software program.

Appendix A: SyH-DR Request Form

Agency for Healthcare Research and Quality (AHRQ) Request for Synthetic Healthcare Database for Research (SyH-DR)

INSTRUCTIONS FOR COMPLETING THE SYH-DR REQUEST FORM AND DATA USE AGREEMENT

*Note: SyH-DR is only authorized for use within the United States and its territories.
International applicants are ineligible.*

Please reference the below terms and definitions as you complete the SyH-DR request form and DUA:

- **Collaborating Organization:** Organization that works with the requesting organization. If the data custodian does not want to assume responsibility for the data security of a collaborating organization, then the collaborating organization should request the data separately.
- **Data Custodian:** Individual who will be responsible for observance of all conditions of use on behalf of the requesting organization, including the establishment and maintenance of security arrangements to prevent unauthorized use. Please note that if the data custodian does not want to assume responsibility for the data security of a collaborating organization, then the collaborating organization should request the data separately.
- **Data Requester:** Individual responsible for submitting the complete SyH-DR application on behalf of the requesting organization and who will act as the primary point of contact between AHRQ and the requesting organization. Please note there can only be **one** Requester per SyH-DR application.
- **Data User:** Individual(s) who will have direct access (on site or VPN) to raw data and analytic files.
- **Title of Study/Project:** A name assigned by the Requesting Organization to the intended research for which the SyH-DR data will be used. The Data Requester should enter this name on the SyH-DR request form, and the Data Custodian and Data Users should include this name within the signature blocks on the DUA.
- **Requesting Organization:** Primary organization requesting the data.

Please submit any questions, the completed request form, and signed DUA(s) to DataInnovations@ahrq.hhs.gov.

Agency for Healthcare Research and Quality (AHRQ) Request for Synthetic Healthcare Database for Research (SyH-DR)

1. Title of Study/Project: _____

2. Data Requester Information

First Name: _____ Middle Initial: _____

Last Name: _____

Title / Position: _____

Requesting Organization/Affiliation: _____

Type of Organization: Not for Profit For Profit Other _____

Street Address Line 1: _____

Street Address Line 2: _____

City: _____ State: _____ ZIP: _____

Email Address: _____

Phone Number: _____

3. Data Custodian Information *(Please leave blank if the Data Requester will serve as the Data Custodian.)*

First Name: _____ Middle Initial: _____

Last Name: _____

Title / Position: _____

Requesting Organization/Affiliation: _____

Type of Organization: Not for Profit For Profit Other _____

Street Address Line 1: _____

Street Address Line 2: _____

City: _____ State: _____ ZIP: _____

Email Address: _____

Phone Number: _____

4. Please provide the names and affiliations (requesting organization or collaborating organizations) of all individuals on your research team who may access the requested data as part of this application.

Data User	Affiliation

5. In the space provided below, please describe your intended use of the SyH-DR data (250 words or less).

Your response should include the following:

- A clear statement of the research questions you plan to address.
- The overall purpose and goals of your research.
- An explanation of how you and/or your organization will use the output generated from your SyH-DR analyses.
- The expected final products and anticipated audiences (e.g., client reports, peer-reviewed manuscripts). If your audience includes clients, please describe the type of clients you serve.

6. How did you hear about SyH-DR? (Please check one box.)

- AHRQ/SyH-DR Website Conference Publication Colleague
 AHRQ Outreach Email Other: _____

7. Confirmation and Signature

- I have read all information on the SyH-DR web page, including the data use agreement (DUA), and understand that the use of this database is restricted to the individuals named on this application.
- A signed DUA for all individuals named on this application is included with this request.
- I understand that my complete application must be approved by AHRQ before I receive access to SyH-DR. AHRQ shall have sole discretion with respect to the determination of SyH-DR access approval.

Data Requester Signature: _____ Date: _____

Appendix B: SyH-DR Data Use Agreement (DUA)



Synthetic Healthcare Database for Research Data Use Agreement

**Agency for Healthcare Research and Quality (AHRQ)
U.S. Department of Health and Human Services**

WARNING: ANY EFFORT TO DETERMINE THE IDENTITY OF INDIVIDUALS OR ESTABLISHMENTS IS PROHIBITED BY LAW AND SUBJECT TO FEDERAL PENALTY.

This Data Use Agreement (“Agreement”) governs the disclosure and use of data in the Synthetic Healthcare Database for Research (SyH-DR), which is maintained by the Agency for Healthcare Research and Quality (AHRQ). Section 944(c) of the Public Health Service Act (42 U.S.C. §299c-3(c)) (“the AHRQ Confidentiality Statute”) requires that data collected by AHRQ that identify individuals or establishments be used only for the purpose for which they were supplied. SyH-DR may only be used for research, analysis, and aggregate statistical reporting projects. AHRQ does not authorize the use of SyH-DR for commercial or competitive purposes affecting establishments; to determine the rights, benefits, or privileges of individuals or establishments; for criminal and civil litigation, including expert witness testimony; for law enforcement activities; or for any other purpose incompatible with the AHRQ Confidentiality Statute.

The undersigned data recipients provide the following assurances concerning SyH-DR:

Protection of Individuals

- I will not release or disclose any information that directly or indirectly identifies persons. This includes attempts to identify individuals through the use of vulnerability analysis or penetration testing.

Protection of Establishments

- I will not publish or report, through any medium, data that could identify individual establishments directly or by inference.
- When the identities of establishments are not provided in the datasets, I will not attempt to use the dataset to learn the identity of any establishment.

Limitations on Data Use, Sharing, and Disclosure

- I will not use or disclose the dataset, or any part thereof, except for research, analysis, and aggregate statistical reporting, and only as permitted by this Agreement.
- I will not use the dataset for unauthorized purposes. AHRQ does not authorize the use of SyH-DR for commercial or competitive purposes affecting establishments; to determine the rights, benefits, or privileges of individuals or establishments; for criminal and civil litigation, including expert witness testimony; for law enforcement activities; or for any other purpose incompatible with the AHRQ Confidentiality Statute.
- I will not store, use, or access the SyH-DR dataset outside of the United States and its territories.

- I will not redistribute SyH-DR by posting on any website or publishing in any other publicly accessible online repository. If a journal or publication requests access to data or analytic files, I will cite restrictions on data sharing in this Agreement.

Safeguards

- I will ensure that the data are kept in a secured environment and that only authorized users (individuals who have signed the Agreement) will have access to the data.

Responsibility

- I acknowledge and affirm that I am personally responsible for compliance with the terms of this Agreement, to the exclusion of any other party, regardless of such party's role in sponsoring or funding the research that is the subject of this Agreement.
- I acknowledge and affirm that interpretations, conclusions, and/or opinions that I reach as a result of my analyses of the datasets are my interpretations, conclusions, and/or opinions, and do not constitute the findings, policies, or recommendations of the U.S. Government, the U.S. Department of Health and Human Services, or AHRQ.
- I will acknowledge in all reports based on these data that the source of the data is the "Synthetic Healthcare Database for Research (SyH-DR), Agency for Healthcare Research and Quality."
- I will indemnify, defend, and hold harmless AHRQ and the data organizations that provide data to AHRQ for SyH-DR from any or all claims and losses accruing to any person, organizations, or other legal entity as a result of violation of this Agreement. This provision applies only to the extent permitted by Federal and State law.
- I agree to report the violation or apparent violation of any term of this Agreement to AHRQ without unreasonable delay and in no case later than 30 calendar days of becoming aware of the violation or apparent violation.

Terms, Breach, and Compliance

Any violation of the terms of this Agreement shall be grounds for immediate termination of this Agreement. AHRQ shall determine whether a data recipient has violated any term of the Agreement. AHRQ shall determine what actions, if any, are necessary to remedy a violation of this Agreement, and the data recipients shall comply with pertinent instructions from AHRQ. Actions taken by AHRQ may include but not be limited to providing notice of the termination or violation to affected parties and prohibiting data recipients from accessing SyH-DR in the future.

In the event AHRQ terminates this Agreement due to a violation or finds the data recipients to be in violation of this Agreement, AHRQ may direct that the undersigned data recipients immediately return all copies of the SyH-DR to AHRQ or its designee without refund of purchase fees.

Acknowledgment *(To be completed by the Data Users and Data Custodian.)*

I understand that this Agreement is requested by the United States Agency for Healthcare Research and Quality to ensure compliance with the AHRQ Confidentiality Statute.

I understand that a violation of the AHRQ Confidentiality Statute may be subject to a civil penalty of up to \$10,000 under 42 U.S.C. §299c-3(d) and that deliberately making a false statement about this or any matter within the jurisdiction of any department or agency of the Federal Government violates 18 U.S.C. § 1001 and is punishable by a fine, up to 5 years in prison, or both.

My signature indicates that I understand the terms of this Agreement and that I agree to comply with its terms.

Signed: _____

Print or Type Name: _____

Title of Study/Project: _____

Affiliation: _____

Date: _____

Email: _____

Data Custodian Acknowledgment *(Only to be completed by the Data Custodian.)*

I am designated as Data Custodian and shall oversee and comply with the observance of all conditions of use and the establishment and maintenance of security arrangements as specified in this Agreement to prevent unauthorized use.

My signature indicates that I understand the terms of this Agreement and that I agree to comply with its terms.

Signed: _____

Print or Type Name: _____

Title of Study/Project: _____

Requesting Organization/Affiliation: _____

Date: _____

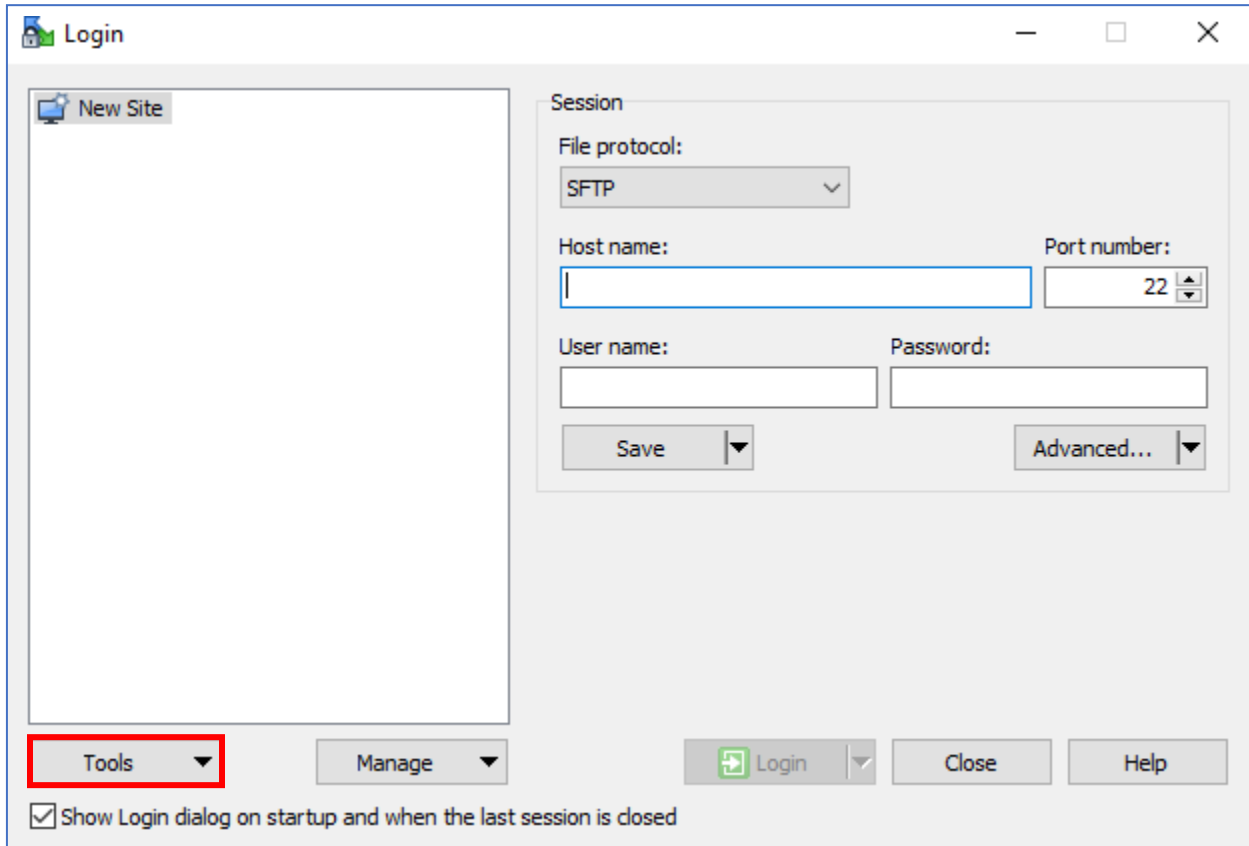
Email: _____

Appendix C: Instructions for Getting an RSA Key Pair

Instructions for Microsoft Windows Users: Generating the RSA Key Pair Using WinSCP

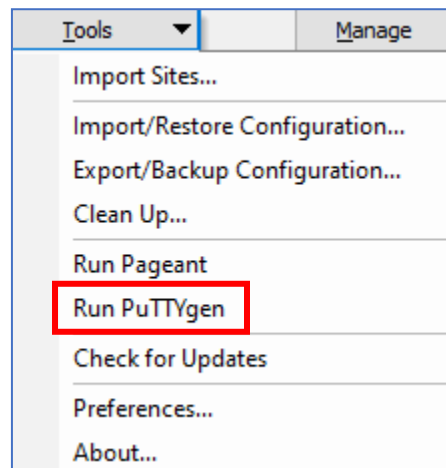
1. Launch WinSCP. The Login window should open automatically.
2. From the Login window, click the **Tools** dropdown button highlighted below.

Figure 4. WinSCP Login Window



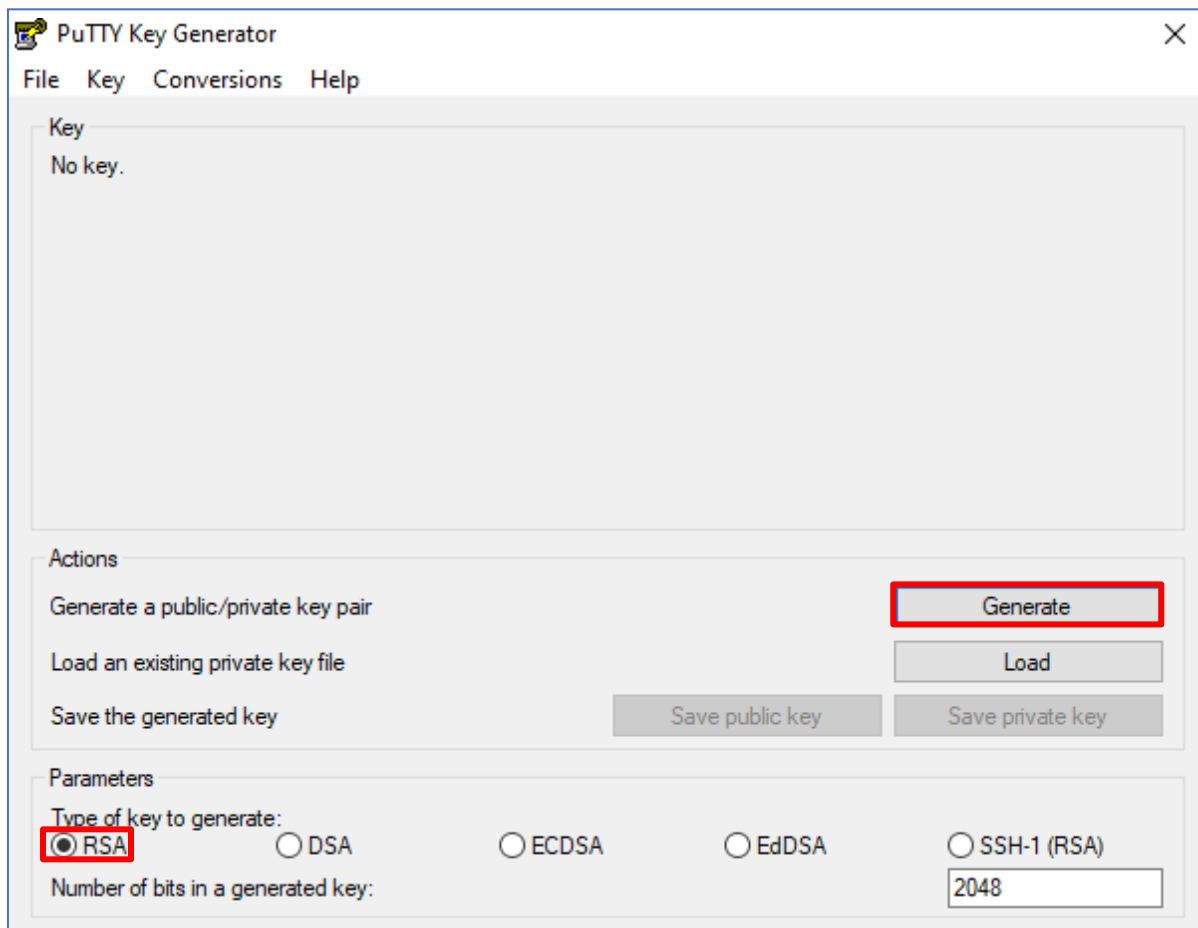
3. In the Tools dropdown list, select **Run PuTTYgen**.

Figure 5. WinSCP Tools Dropdown List



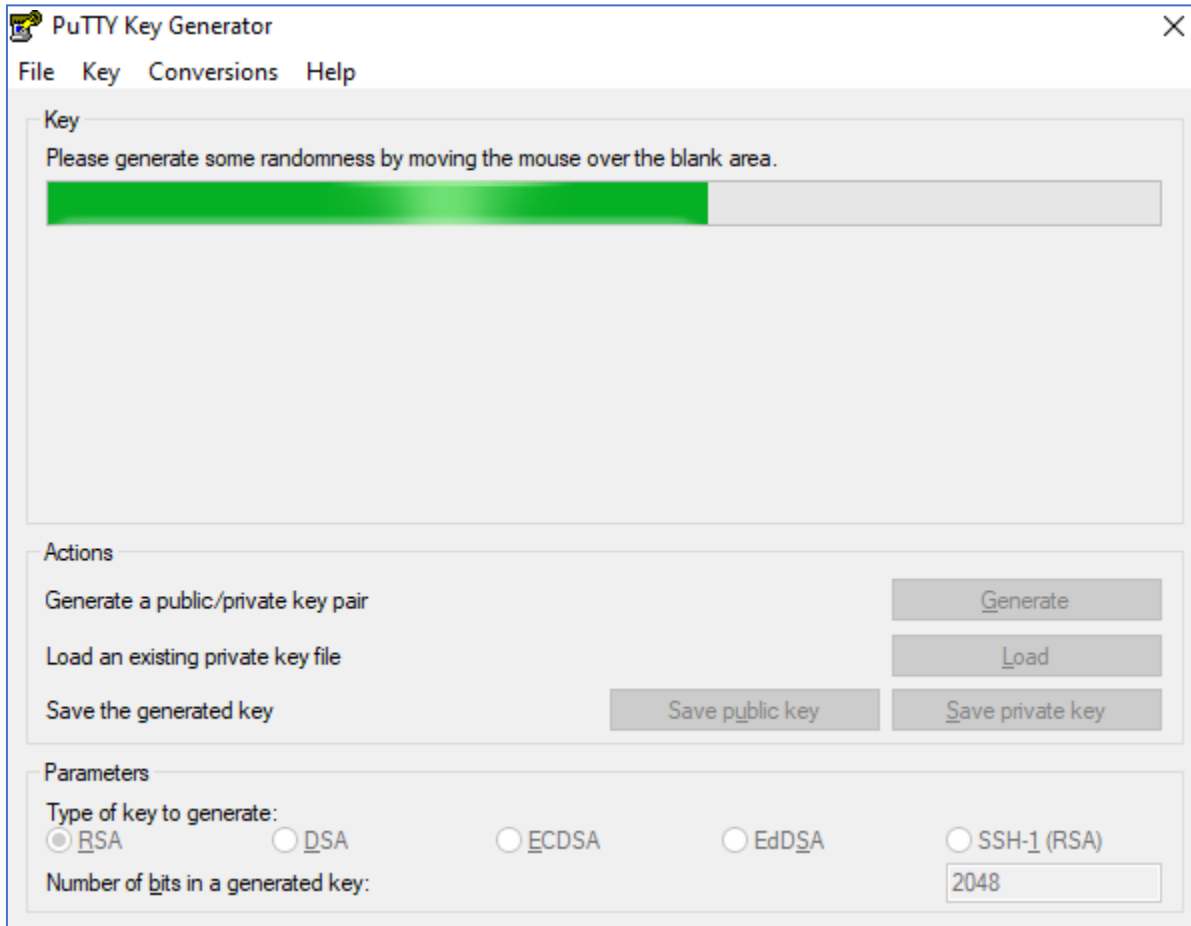
4. The PuTTY Key Generator window will appear.
 - a. In the Parameters section of the window, verify that the **RSA** radio button is selected under “Type of key to generate.”
 - b. Then select the **Generate** button.

Figure 6. WinSCP PuTTY Key Generator Window



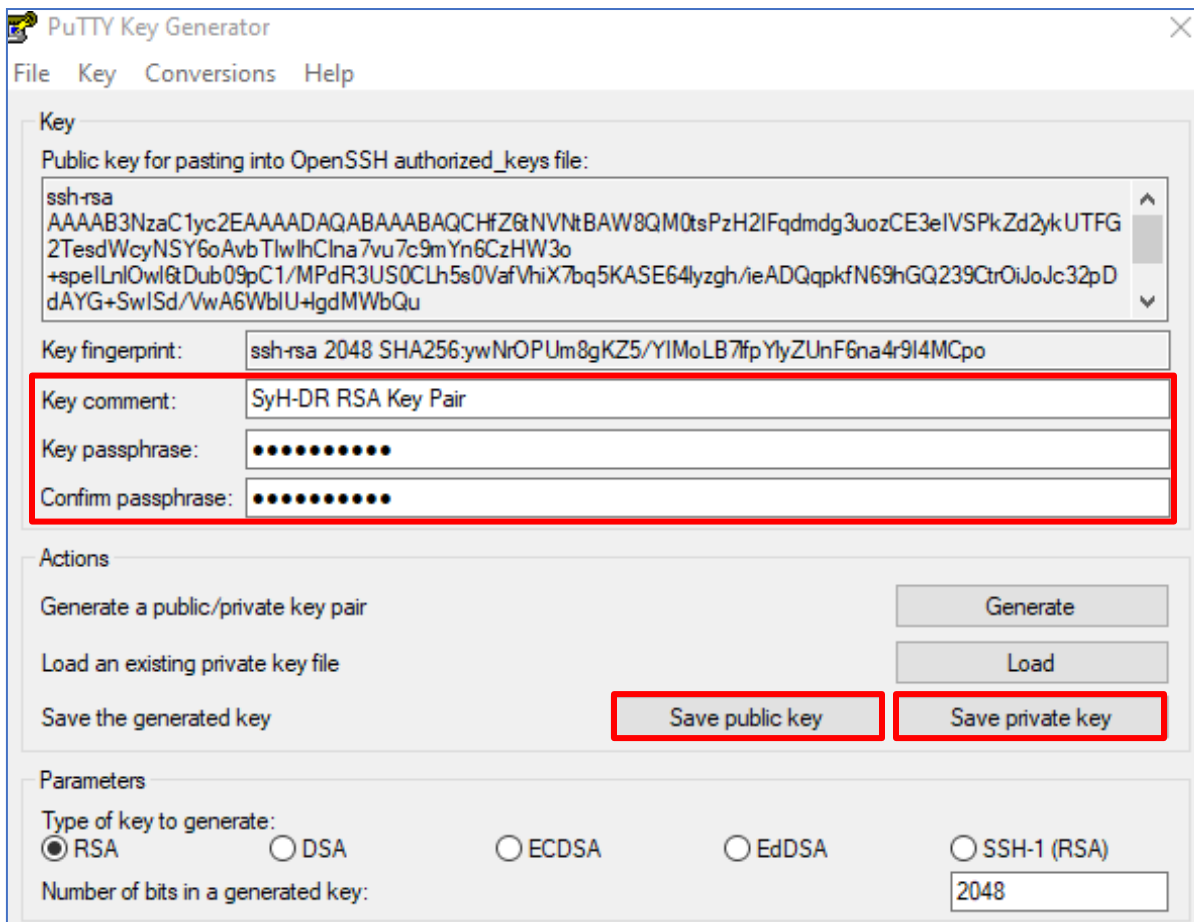
- The RSA Key Pair generation will begin. Make random movements with your mouse over the blank area of the PuTTY Key Generator window to generate the key.

Figure 7. RSA Key Pair Generation in Progress



6. Once the RSA Key Pair generation is complete, the Public Key and Key fingerprint will appear.
 - a. In the Key comment field, type a comment that will help you identify this key in the future. This comment will be displayed whenever WinSCP asks for the Key passphrase.
 - b. Then enter a secret passphrase of your choosing in the Key passphrase field and the Confirm passphrase field. *Be sure to securely save this passphrase for future use when downloading the SyH-DR files. We cannot look up or reset this passphrase on your behalf. If you do not have the passphrase, you will need to recreate the RSA Key Pair.*
7. Select the **Save public key** button and save the public key to your computer. **This is the file that you will send to AHRQ as an email attachment.**
8. Select the **Save private key** button and save the private key to a private drive on your computer.

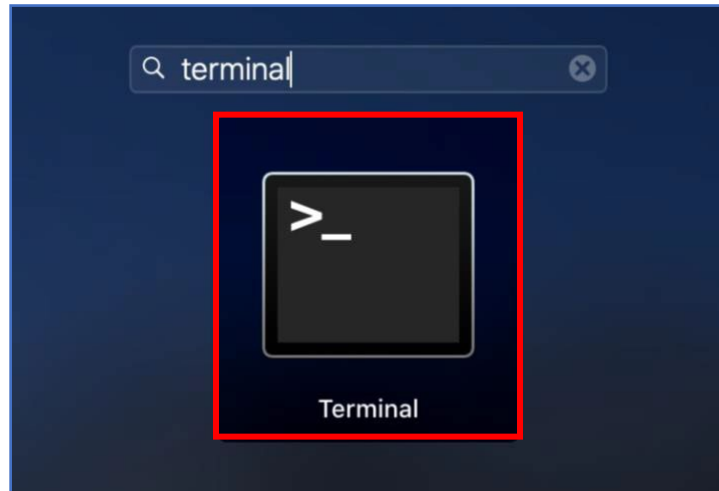
Figure 8. Generated RSA Key Pair



Instructions for MacOS Users: Generating the RSA Key Pair Using Terminal

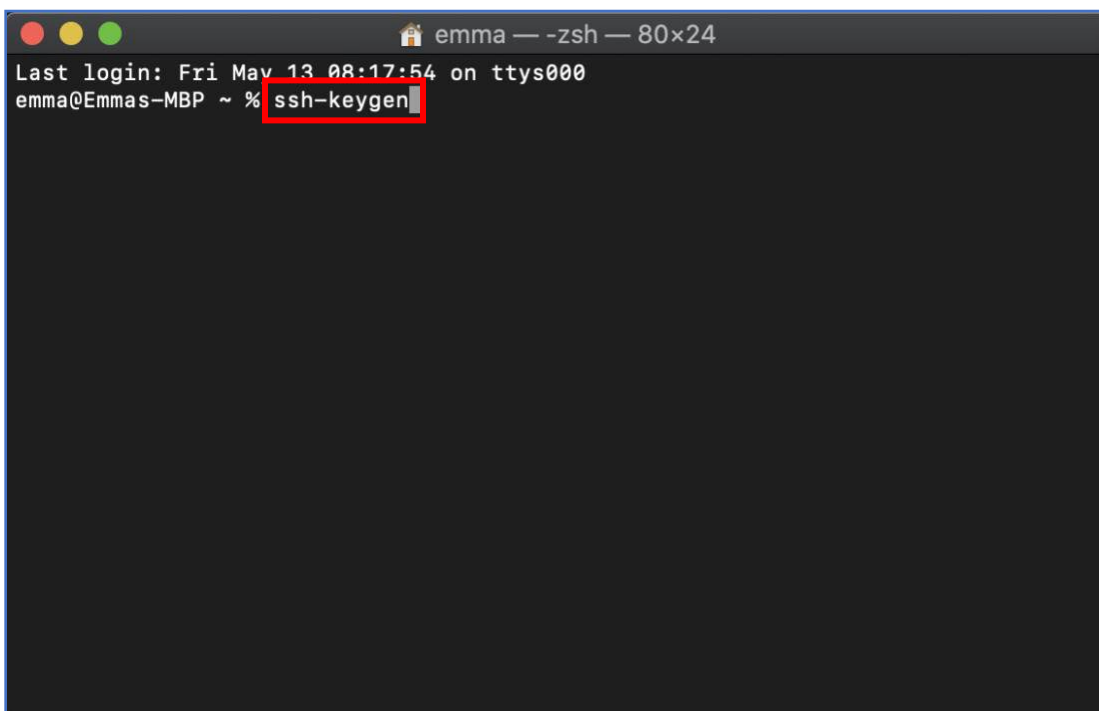
1. From Launchpad, type “Terminal” in the search bar. Then, select the Terminal icon to launch the application.

Figure 9. Search for the Terminal Application



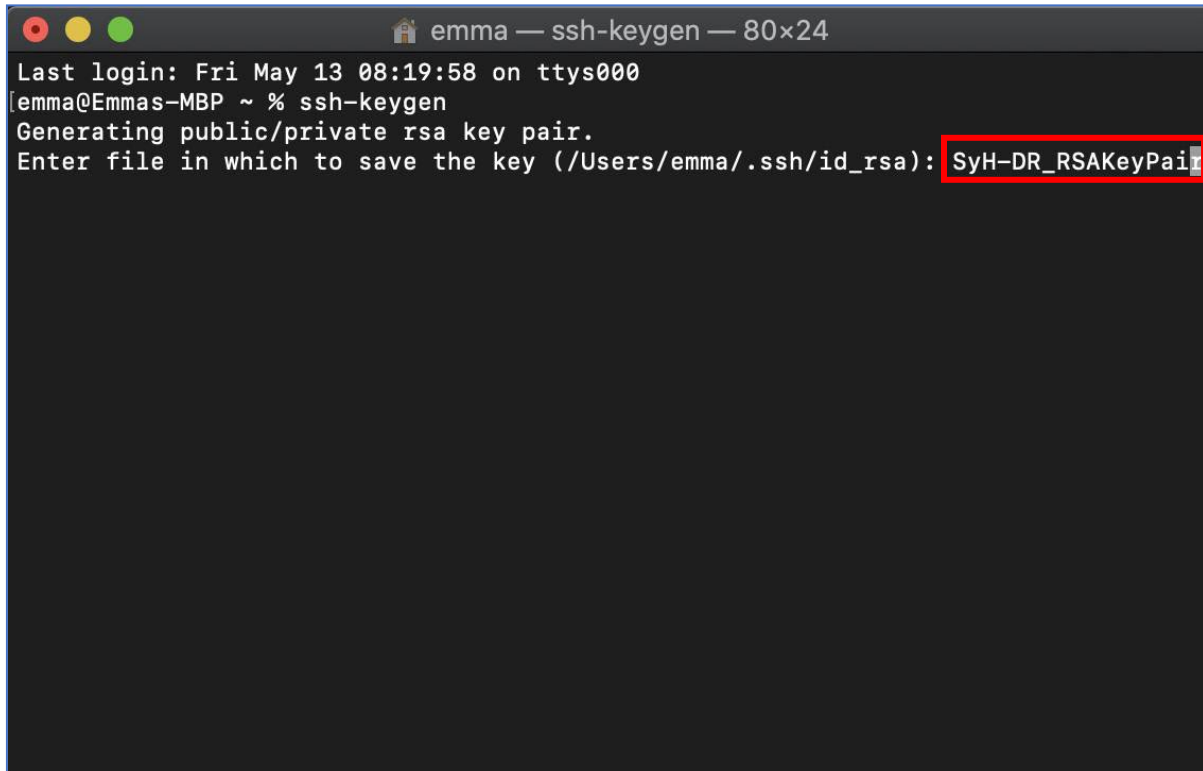
2. A new Terminal session window will appear. *Note: Depending on your system settings, this window may appear with a white or black background.*
3. In the Terminal session window, type “ssh-keygen.”
 - a. Press the **Enter** key on your keyboard.

Figure 10. Terminal Session Window



4. The Terminal session window should now say “Generating public/private rsa key pair.”
 - a. Type a descriptive file name that will help you identify this key in the future, as shown below.
 - b. Press the **Enter** key on your keyboard.

Figure 11. Terminal Generating Key Pair



```
emma — ssh-keygen — 80x24
Last login: Fri May 13 08:19:58 on ttys000
[emma@Emmas-MBP ~ % ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/Users/emma/.ssh/id_rsa): SyH-DR_RSAPKeyPair
```


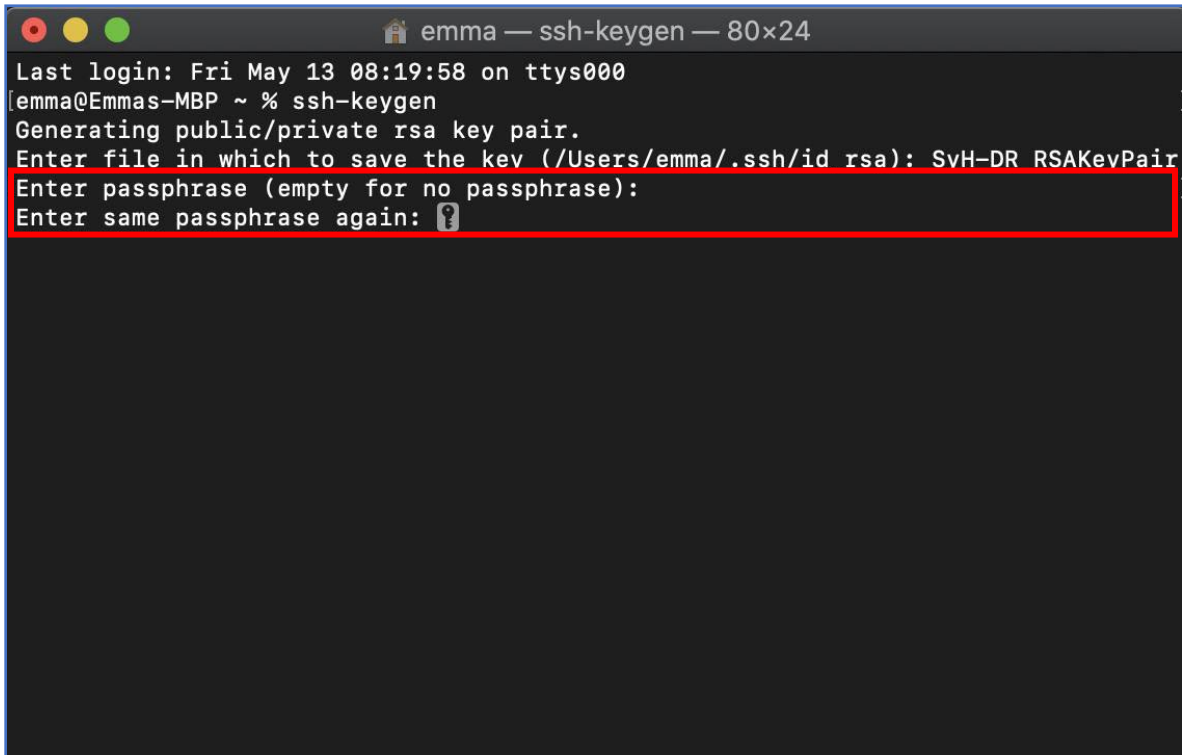

5. The Terminal session window will prompt you to enter a passphrase.
 - a. Enter a secret passphrase of your choosing. *Note: The typed characters will NOT appear on the screen. You will only see a Key icon.* 
 - b. Press the **Enter** key on your keyboard, and then enter the same passphrase again. *Be sure to securely save this passphrase for future use when downloading the SyH-DR files. We cannot look up or reset this passphrase on your behalf. If you do not have the passphrase, you will need to recreate the RSA Key Pair.*

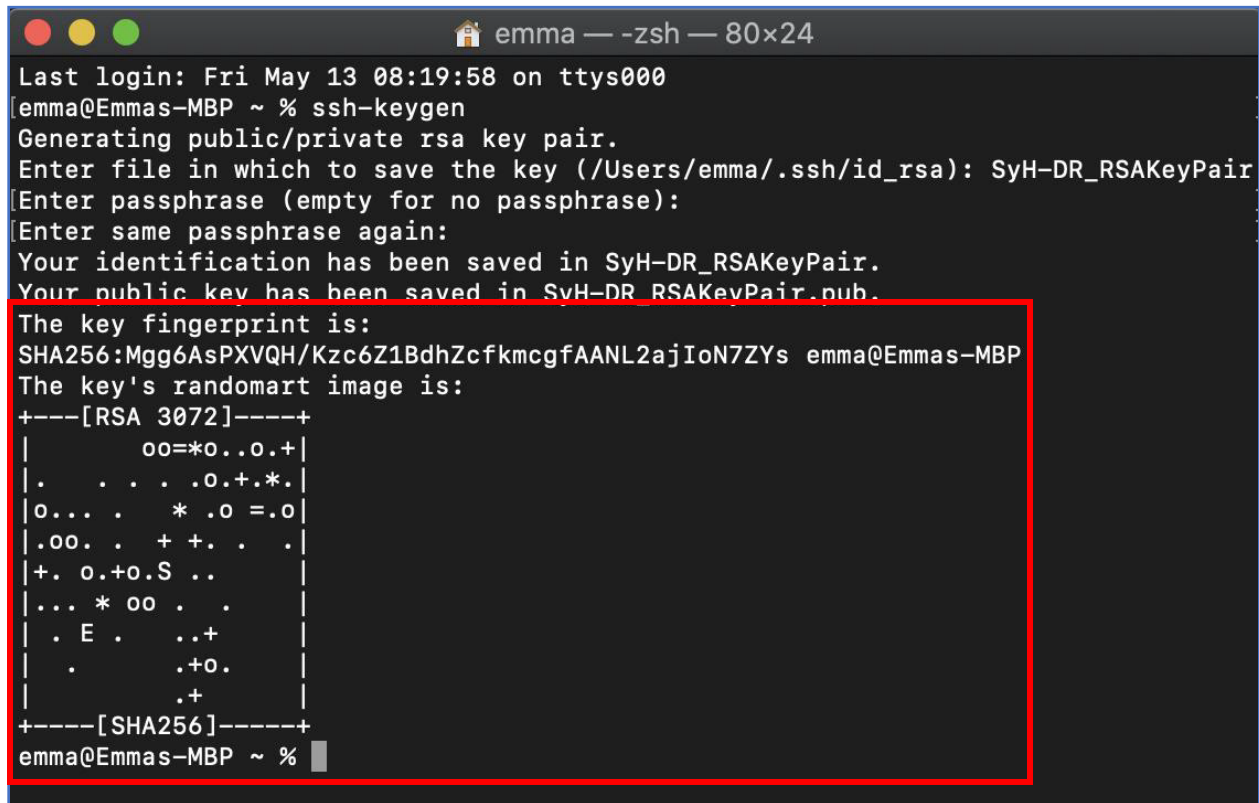
Figure 12. Enter Passphrase for the RSA Key



```
emma — ssh-keygen — 80x24
Last login: Fri May 13 08:19:58 on ttys000
emma@Emmas-MBP ~ % ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/Users/emma/.ssh/id_rsa): SvH-DR RSAKeyPair
Enter passphrase (empty for no passphrase):
Enter same passphrase again: 
```

6. Press the **Enter** key on your keyboard. The key fingerprint will appear in the Terminal session window.

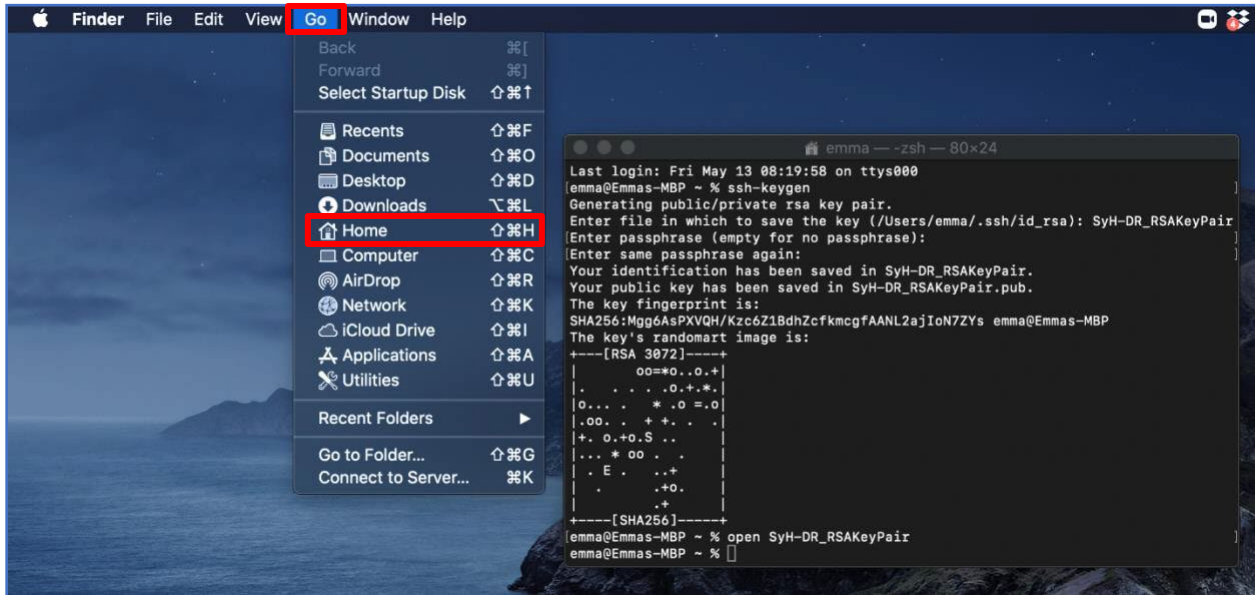
Figure 13. Key Fingerprint in the Terminal Window



```
emma — -zsh — 80x24
Last login: Fri May 13 08:19:58 on ttys000
[emma@Emmas-MBP ~ % ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/Users/emma/.ssh/id_rsa): SyH-DR_RSAPKeyPair
[Enter passphrase (empty for no passphrase):
[Enter same passphrase again:
Your identification has been saved in SyH-DR_RSAPKeyPair.
Your public key has been saved in SyH-DR_RSAPKeyPair.pub.
The key fingerprint is:
SHA256:Mgg6AsPXVQH/Kzc6Z1BdhZcfkmcgfAANL2ajIoN7ZYs emma@Emmas-MBP
The key's randomart image is:
+---[RSA 3072]-----+
|          oo=*o..o.+ |
|. . . . .o.+.*. |
|o... . * .o =.o |
|.oo. . + +. . . |
|+. o.+o.S .. |
|... * oo . . |
|. E . ..+ |
|. . .+o. |
|. . .+ |
+-----[SHA256]-----+
emma@Emmas-MBP ~ %
```

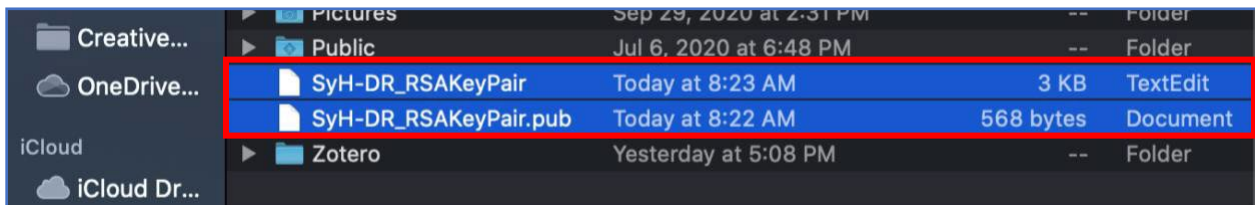
7. Next, verify that the files are saved to your computer.
 - a. Click the desktop area outside of the Terminal session window to show the **Finder** menu.
 - b. Select the **Go** dropdown menu and select **Home**.

Figure 14. Finding the Public and Private Key Files



8. Your computer’s file finder will open to your **Home** folder. Verify that there are two files with the file name you entered in Step 4.
 - a. The file with the “.pub” extension is the **Public key**. **This is the file that you will send to AHRQ as an email attachment.**
 - b. The file that does **not** have the .pub extension is the **Private key**. Make sure this file is saved to a private location on your computer.

Figure 15: Public and Private Key File





AHRQ Publication No. 24-0019-2-EF
April 2024
www.ahrq.gov