University of Louisville
Journal of Respiratory Infections

ORIGINAL RESEARCH

A Software Tool for Automated Upload of Large Clinical Datasets Using REDCap and the CAPO Database

**William A. Mattingly[1], Christopher Sinclair[1], Danna Williams[1], Matthew Grassman[1], Stephen Furmanek[1], Kimberley Buckner[1], Mohammad Tahboub[1]*

Abstract

Introduction: Obtaining clinical data from healthcare sources is necessary for conducting clinical research. New technologies now allow for connecting a research database to Electronic Medical Records remotely, allowing the automatic import of clinical research data. In this paper we design and evaluate a REDCap extension to import clinical records from an external health database.
Methods: Many hospital EHRs are designed to use secure file transfer protocol (SFTP) repositories for data communication. We develop a REDCap plugin to connect to an external SFTP file repository for the import of clinical record data. We use the CAPO instance of REDCap and a sample set of clinical pneumonia variables for the connection.
Results: The plugin allows the input of record data in a much shorter time than traditional data entry in addition to being less error prone. However, the formatting of the data in the SFTP file repository must be exact in order for the import to be successful. This can require setup time on the part of EHR IT staff.
Conclusion: Developing a direct connection from EHR to research database can be an effective way to lower the overhead for conducting clinical research. We demonstrate a means to do this using REDCap and SFTP.

DOI: 10.18297/jri/vol2/iss1/7
Received Date: February 12, 2018
Accepted Date: February 27, 2018
Website: https://www.louisville.edu/jri
Affiliations:
[1]University of Louisville Division of Infectious Diseases, Louisville, KY 40202

Introduction

The adoption of new technologies in clinical research has been improving over time and continues to improve and change the way clinical research is conducted [1]. Accessible and affordable software like REDCap [2] and R [3] lower the barrier to entry for independent investigators while maintaining the high accuracy and reliability needed for large studies. As the number of studies grows so does the amount of data needed to answer tomorrow's research questions and data collection remains an integral part of the clinical research enterprise [4].

The importance of accurate data collection to clinical research has pushed the development of software features to support this need. REDCap has features that allow it to connect to other systems for data import, export, and statistical analysis. Software has been developed to support these types of connections with legacy data collection solutions [5]. Recent work has also helped to bridge the gap between electronic health record systems (EHRs) and data warehouses and surveillance systems by providing a standard interface to connect to multiple different types of EHR database [6]. Further development of this work could one day allow for connection to all major types of EHR database, but currently only a few are supported.

Despite many different EHR providers and platforms, most support a standard protocol for the secure transfer of files to external parties. This protocol is called the secure file transfer protocol (SFTP) and is designed to establish an encrypted data connection between two consenting parties. In the United States the Health Insurance Portability and Accountability Act (HIPAA) Security Rule requires that reasonable and appropriate measures be taken to secure patient data that is in transmission, and this includes encryption protocols like SFTP.

EHR systems allow for the fast export of a large number of records into an SFTP repository, which can then have access granted to external investigators. This is especially useful in situations where longitudinal data is being collected or patient data may change, as the SFTP repository will have the most recent changes to patient data as soon as they are made in the EHR. It is also useful for studies involving the biosurveillance of patient populations since new patients matching study criteria will continually be added to the system over a prolonged period.

Performing clinical research studies depends on patient information, and getting data to the point where it can be analyzed, such as in an electronic data capture system, is a major bottleneck for studies. Data collection will often involve the time consuming process of research staff navigating through the medical record and recording study data points in an electronic format. In this paper we discuss the design and implementation

*Correspondence To: William A Mattingly, PhD
501 E Broadway, Suite 120B
Louisville, KY 40202
bill.mattingly@louisville.edu

of SFTP-2-REDCap, a REDCap plugin to support the fast upload of data into a REDCap clinical research project.

Methods

The international Community Acquired Pneumonia Organization (CAPO) database is a database consisting of over 9000 cases of clinical pneumonia from over 130 clinical practices across the world [7]. The CAPO database uses the REDCap software, an electronic data capture software used by over 2700 research and educational institutions across the world, as its platform. REDCap is designed to support data capture for research studies, providing an intuitive interface for validated data entry, audit trails for tracking data manipulation and export procedures and automated export procedures for seamless data downloads to common statistical packages. REDCap also supports extensibility through special plugins, small programs that work with REDCap to provide additional features.

We design the plugin SFTP-2-REDCap which imports data fields from an SFTP server into a REDCap project. A diagram of the process is shown in **Figure 1**. REDCap uses the popular PHP programming language [8] to provide its features, and its plugins are written using the same language. Providing the appropriate security requires SSL encryption, the standard for secure internet communication [9]. We use phpseclib [10], a well-tested and secure extension to PHP that provides seamless SFTP connection and secure transfer of patient information.

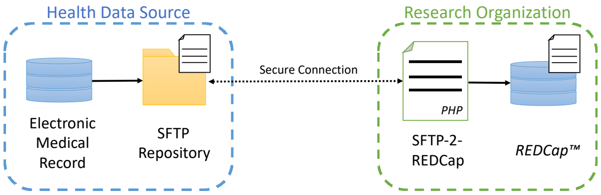
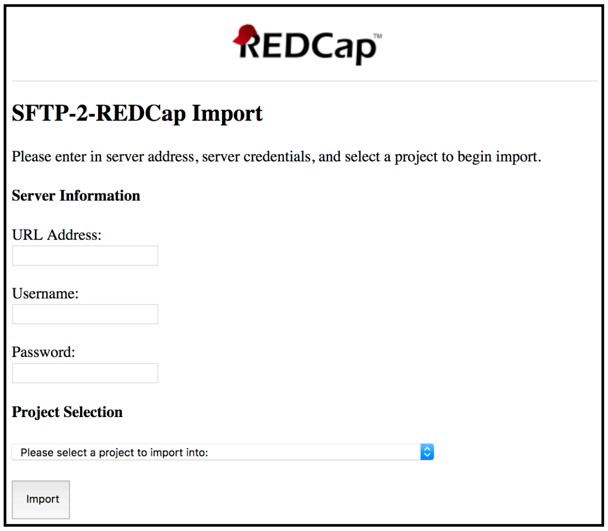


Figure 1 Communication is shown from a Health Data Source, such as an EHR, to REDCap. An EHR deposits patient information into a secure FTP repository, which can then be accessed via SFTP-2-REDCap using credentials provided by the source. A secure SSL connection completes the transfer, and saves the record(s) into REDCap's database.

Our plugin can be integrated into REDCap by copying its files into the plugins directory that is part of every REDCap installation. The plugin can then be accessed by navigating to the base url of your redcap installation followed by /plugins/sftp-redcap.php. SFTP-2-REDCap requires REDCap version 6. The user interface for SFTP-2-REDCap is shown in **Figure 2**.

A user can enter the server credentials for an external SFTP server, including the URL address of the server, the username and password. This is the minimum information that would be provided by the party granting access. A user can then choose into which project they would like records to be imported from the active projects on that REDCap installation.

Once the project has been selected, the user can finalize import by clicking the import button. If there are any errors, regarding project selection or SFTP credentials, they will be displayed and must be corrected before continuing. If upload is successful, the total number of variables successfully imported will be displayed.



SFTP-2-REDCap Import

Please enter in server address, server credentials, and select a project to begin import.

Server Information

URL Address:

Username:

Password:

Project Selection

Please select a project to import into:

Figure 2 Connection screen for SFTP-2-REDCap.

To test the data import process into the CAPO database, we chose a selection of data points related to the Pneumonia Severity Index [11] or PSI which is used to classify patients into risk classes based on demographics, medical history, lab values, and presence of pleural effusion. The list of data points and their identifiers are shown in **Table 1**.

Table 1 List of imported data points related to the Pneumonia Severity Index and their identifiers.

Record Variable Description	Variable Name
Record ID	record_id
Age	dem_age
Sex	dem_sex
Nursing Home Resident	risk_nursinghome
Neoplastic disease	hx_neoplastic
Liver Disease	hx_liver
CHF History	hx_chf
Cerebrovascular Disease	hx_cva
Renal Disease	hx_renal
Altered Mental Status	exam_mental
Respiratory Rate	exam_rr
Systolic blood pressure (mmHg)	exam_sbp
Temperature (Degrees Celsius)	exam_temp
Heart rate (pulse)	exam_hr
Arterial pH	lab_abgph
Blood Urea Nitrogen (BUN)	lab_bun
Serum Sodium	lab_na
Serum Glucose (mg/dL)	lab_glucose
Hematocrit (%)	lab_hematocrit
Partial pressure of Oxygen (mmHg)	lab_abgpao2
O2 Saturation	exam_o2satvalue
Pleural effusion: X-ray	cx_pe
Pleural effusion: CT scan	ct_pe

We simulate three sets of data having 10, 100 and 1000 cases with values for each of these variables. We assume there is no missing data so our tests will provide an upper bound for measuring the time taken to import the entire data set.

Results

Upon entering the credentials for an external SFTP server and clicking the import button, the plugin successfully imports records into the CAPO REDCap project. Three different data files containing simulated data were tested and the time taken for each data import was recorded and shown in **Table 2**.

Table 2 Volume of data and time taken for import of three data files containing simulated data.

Number of Records	Total Imported Data Points	Time (seconds)
10	230	3.35
100	2,300	2.65
1000	23,000	8.30

For small numbers of records the majority of the time taken was for network transmission, and internet latency accounts for more time being taken for 10 records than 100. Even 1000 records and 23,000 data points takes less than 9 seconds to download from a remote SFTP server and import into a REDCap project.

To compare this to the time needed for standard data entry, we use a rough approximation of the average time needed to enter a single record being 10 seconds. This would include moving the mouse to the appropriate field, typing the data, and visually double checking the result. However, this would not include the time needed to navigate the different user interface screens of an electronic medical record, which could involve a substantial amount of overhead in some systems.

The time saved increases dramatically with respect to the number of records and data points entered, with 1000 records and 23,000 data points needing about 64 hours of data entry time using the above approximation.

Discussion

CAPO includes membership and data from over 130 different organizations and those who have shared data with CAPO repository have donated the time required for data entry into REDCap. There is a great potential for increased research by establishing electronic connections with some of the highest contributors in CAPO, and the same could be said of other institutions that wish to establish electronic connections to REDCap.

Arguably, the greatest potential lies in the transfer of lab values, as these are objective measurements that are usually stored in a structured format. Making this transfer automatic has the best potential for saving time and costly transcription errors. Measurements involving patient history are often found only in clinician notes and free text. Automatic transfer and recognition of this data is likely to require the oversight of a research clinician.

Since data in a remote SFTP server will fall under the management of the health data source, setup and selection of an appropriate data format has to be negotiated with their technology department, and setup will usually be prioritized to follow business critical applications. This can lead to a wait period for the initial setup of an SFTP data connection.

This plugin is meant to be a fast and simple way to load data from an external SFTP connection, and as such does not take advantage of REDCap's data import framework, which provides error checking and adjudication before overwriting old data stored in the database. We plan for future versions of SFTP-2-REDCap to take advantage of these REDCap features.

Since SFTP-2-REDCap uses SFTP technology, connections to international EHRs do not present any difficulties. However, languages having non-Latin characters and using Unicode for to represent variable names and values may cause consistency problems if REDCap is not setup to use the same character encoding.

The REDCap plugin SFTP-2-REDCap provides a convenient interface to import multiple data records from a remote SFTP server in a short amount of time. It is available to users of REDCap, and has the potential to streamline and automate the process of data collection for REDCap projects.

Acknowledgements: Thanks to the Community Acquired Pneumonia Organization for providing the data structure and REDCap installation for testing.
Conflict of Interest: None reported.
Funding Source: None reported.

References

1. Murphy SN, Dubey A, Embi PJ, Harris PA, Richter BG, Turisco F et al. Current state of information technologies for the clinical research enterprise across academic medical centers. Clin Transl Sci. 2012 Jun;5(3):281–4. https://doi.org/10.1111/j.1752-8062.2011.00387.x PMID:22686207
2. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009 Apr;42(2):377–81. https://doi.org/10.1016/j.jbi.2008.08.010 PMID:18929686
3. The R Foundation. The R Project for Statistical Computing. 2014. Available from: http://www.r-project.org/
4. Saczynski JS, McManus DD, Goldberg RJ. Commonly used data-collection approaches in clinical research. Am J Med. 2013 Nov;126(11):946–50. https://doi.org/10.1016/j.amjmed.2013.04.016 PMID:24050485
5. Dunn Jr WD, Cobb J, Levey AI, Gutman DA. REDLetr: Workflow and tools to support the migration of legacy clinical data capture systems to REDCap. International journal of medical informatics. 2016 Sep 1;93:103-10.
6. Campion Jr TR, Sholle ET, Davila Jr MA. Generalizable Middleware to Support Use of REDCap Dynamic Data Pull for Integrating Clinical and Research Data. AMIA Summits on Translational Science Proceedings. 2017;2017:76.
7. CAPO. Community-Acquired Pneumonia Organization. 2017. Available from: http://caposite.com/
8. The PHP Group. PHP: Hypertext Processor. 2017 [cited 2017]. Available from: https://secure.php.net/
9. Elgamal T, Hickman KE. inventors; Netscape Communications Corp, assignee. Secure socket layer application program apparatus and method patent

5,657,390. 1997.

10. http://phpseclib.sourceforge.net/. phpseclib. 2017. Available from: https://github.com/phpseclib

11. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003 May;58(5):377–82.