



REDLeTr: Workflow and tools to support the migration of legacy clinical data capture systems to REDCap



William D. Dunn Jr.^{a,b}, Jake Cobb^c, Allan I. Levey^a, David A. Gutman^{a,b,*}

^a Department of Neurology, Emory University, Atlanta, GA, USA

^b Department of Biomedical Informatics, Emory University, Atlanta, GA, USA

^c College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

ARTICLE INFO

Article history:

Received 24 November 2015

Received in revised form 21 June 2016

Accepted 26 June 2016

Keywords:

Electronic health record

Biomedical informatics

ETL

REDCap

Clinical informatics

Clinical research informatics

ABSTRACT

Objective: A memory clinic at an academic medical center has relied on several *ad hoc* data capture systems including Microsoft Access and Excel for cognitive assessments over the last several years. However these solutions are challenging to maintain and limit the potential of hypothesis-driven or longitudinal research. REDCap, a secure web application based on PHP and MySQL, is a practical solution for improving data capture and organization. Here, we present a workflow and toolset to facilitate legacy data migration and real-time clinical research data collection into REDCap as well as challenges encountered.

Materials and methods: Legacy data consisted of neuropsychological tests stored in over 4000 Excel workbooks. Functions for data extraction, norm scoring, converting to REDCap-compatible formats, accessing the REDCap API, and clinical report generation were developed and executed in Python.

Results: Over 400 unique data points for each workbook were migrated and integrated into our REDCap database. Moving forward, our REDCap-based system replaces the Excel-based data collection method as well as eases the integration into the standard clinical research workflow and Electronic Health Record.

Conclusion: In the age of growing data, efficient organization and storage of clinical and research data is critical for advancing research and providing efficient patient care. We believe that the workflow and tools described in this work to promote legacy data integration as well as real time data collection into REDCap ultimately facilitate these goals.

Published by Elsevier Ireland Ltd.

1. Introduction

1.1. Background and significance

Properly integrating clinical research data from a variety of systems can be a daunting challenge, especially given the specific demands of an academic medical environment. Data collection instruments are often developed to fulfill specific needs, often with little thought placed on future changes or interoperability with other instruments or central management. In our clinical research environment, this has led to a number of stopgap solutions ranging from old and unmaintainable Microsoft Access databases and Visual Basic applications, paper forms, and Microsoft Excel spreadsheets. The relative inaccessibility to and lack of communication between these data sources present challenges not only from a

clinical research perspective, but also in providing optimal clinical care.

While numerous studies agree that improvements in electronic data capture (EDC) systems can lead to improved efficiency [3,4], accuracy [5,6], cost savings [7], and ultimately improvements in health care delivery, the implementation of these systems is often challenged by common barriers such as user motivation, regulatory requirements, economic and personnel resources, and graphical user interface interaction [8,9]. Various measures can be taken to lower these barriers to transition. For example, adopting online EDCs based on a Software as a Service (SaaS) models or offering no-cost licensing arrangements can save significant investment in information technology, form editing functions allow systems to stay practical and integrated in a rapidly changing clinical research environment, and precise quantification of costs and benefits can inspire institutions to finally adopt EDCs [10,11].

There exists a wide range of both proprietary and open source solutions in the EDC domain to support organized, flexible data capture and analysis in healthcare or research settings (see reviews [6,12]). Oracle Clinical is an example of a commercial integrated clinical management system that allows remote data capture [13].

* Corresponding Author at: Department of Neurology, Emory University School of Medicine, 1639 Pierce Drive, Atlanta, GA 30322, USA.

E-mail address: dgutman@emory.edu (D.A. Gutman).

While its benefits include enforcing standards and consistency through internal global libraries and allowing a flexible study design that can take into account even the most complex trial scenario, the system is limited by common constraints observed in commercial EDCs, namely prohibitive costs for implementation and management [14]. An example of an open source EDC for management of clinical research studies is OpenClinica [15], which has been approved for every stage of clinical research and has a global community spanning over 100 countries. Some key features are the easily-designed flexible studies through an intuitive GUI (Graphical User Interface), automated visit scheduling, direct access to data, and the ability to integrate with other systems such as EHR and PACS.

Seeking to better organize and store our clinical research data, we have adopted REDCap (Research Electronic Data Capture), another open source, web-based EDC backed by a MySQL database engine for data storage and manipulation [16]. REDCap includes a wide range of special features that have accelerated its initial development in 2004 to a multi-national consortium today comprised of users from over 1650 institutions [17]. Key features include collaboration access across institutions, role-based security restrictions, quality assurance mechanisms, centralized data storage and backup, and data export options for common statistics packages [16,18]. The REDCap user interface allows less-technical users to access and view the project data. In addition, the integrated GUI form designer greatly simplifies form design for clinicians and staff as it does not necessitate a software developer or database administrator. Advanced features and plugins contributed internally and from the external community continue to add to the appeal of REDCap. Available plugins include capabilities to visualize geographical data and generate advanced plots [19], “hooks” to insert custom code and features within standard REDCap environment [20], and pairing with Open Data Kit (ODK, opendatakit.org) software to allow data collection and upload to REDCap directly from mobile devices [21].

Apart from the features described above, our choice of REDCap for this project was driven by two practical considerations. Our umbrella institution has an enterprise REDCap installation that is maintained and supported by the information technology group for a modest yearly fee (<\$250 a year/project). Given the typical complexities of installation, backup and maintenance of a server that houses protected health information, this nominal fee obviates many of the administrative challenges we would have faced using a separate database system. In addition, REDCap provides a robust API that allows the import, export, and modification of data that can significantly aid in automating the data migration process. While not an absolute dependency, the existence of a python based wrapper (pyCAP) that simplified interaction with the REST-based API was also a key advantage.

1.2. Objective

In this work, we describe the workflow, challenges, and tools we have developed (REDLeTr, or REDCap Load, Extract, Translate, Revise) to facilitate the migration of legacy clinical research data and real-time data collection into a REDCap environment with added features to facilitate integration in standard clinical workflow. This will increase the ability to take advantage of previous data as well as to facilitate future research. The case study we use to describe this process is the migration of neuropsychological testing data from the NIH-funded Alzheimer’s Disease Research Center (ADRC) and affiliated memory clinics at Emory University. In addition, as the applicability of tools to transfer valuable legacy clinical data to centrally organized and flexible research databases is widespread, we have provided a resource containing code and explanations on an accompanying site (<https://github.com/dgutman/RedLETR>) and believe that it will be of use to other organizations seeking to standardize data capture and improve efficiency.

[com/dgutman/RedLETR](https://github.com/dgutman/RedLETR)) and believe that it will be of use to other organizations seeking to standardize data capture and improve efficiency.

2. Materials and methods

2.1. Data description

Legacy data consisted of approximately 4000 Excel spreadsheets corresponding to patients seen in the ADRC memory clinics at Emory University between 2011 and 2014. This research-oriented clinic is staffed by approximately 15 nurses and physicians in addition to social workers, researchers, and other support staff and sees approximately 1500 patients per year through primary patient encounters, follow up visits, and research studies. As part of each standard visit, patients are administered a comprehensive neuropsychological battery containing tests that measure a wide variety of functions (Supplemental Table 1). These data are later stored in the institutional electronic medical record system (EMR, Cerner Millennium, www.cerner.com). Each spreadsheet consists of at least nine individual forms/tests evaluating the patient’s cognitive function across a number of domains and a summary sheet containing the overall raw scores as well as corresponding normative scores adjusted for a patient’s age, race, education, and/or gender.

2.2. Software

Functions for extracting data, converting to formats compatible with REDCap integration, and normative scoring were developed in Python (Anaconda Python distribution [22], Python version 2.72). The main prerequisite for this project was a working REDCap server with the ability to upload data using the API. The PyCAP toolkit was used to communicate between Python and the REDCap API [23]. We have included a set of package dependencies for the application itself in the requirements.txt file on our GitHub page, along with instructions on how to setup a Python “virtualenv” for this work. Due to the iterative nature of this process, we found iPython Notebook [24], which provides a web-based GUI for Python development, especially useful. With the exception of the Excel password removal using a 64-bit Windows 7 machine due to its use of a Windows DLL, all of our code was executed on a Linux system running Ubuntu 12.04 and 14.04 64-bit as well as MAC OS-X 10.1.

2.3. Data migration of legacy neurology data to REDCap

Our general workflow for migrating legacy data to the REDCap environment is similar to a standard Extract-Translate-Load (ETL) [25] process, and we describe our procedure in this framework. Because this data migration to an improved system was part of quality improvement and regular operations, Institutional Review Board (IRB) approval was not required for this project.

2.3.1. Pre-extraction

In order to leverage existing Python tools, we first used the Microsoft Office Migration Manager [26] to convert any XLS (a legacy Excel file format) files to XLSX (the current Excel format; see Code Snippet 1 in the supplemental section). In addition, to allow our tools access to password-protected sheets, we used the win32com.client Python library to unlock/unprotect and re-save each XLSX file (see Code Snippet 2 in the Supplemental section).

2.3.2. Extraction: extracting data from Excel worksheets

We leveraged the openpyxl Python library [27] to extract content from individual cells by location from each XLSX spreadsheet. In most cases, such as the score on the “Trails” exercise from the

Download English Version:

<https://daneshyari.com/en/article/516061>

Download Persian Version:

<https://daneshyari.com/article/516061>

[Daneshyari.com](https://daneshyari.com)