

# A Simple and Efficient Method for the Management of Multiple Electronic Health Record-Driven Phenotype Projects

Cosmin A. Bejan, PhD<sup>1</sup>, Joshua C. Denny, MD, MS<sup>1,2</sup>

<sup>1</sup>Department of Biomedical Informatics, <sup>2</sup>Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

## Abstract

Based on our experience of working with the Electronic Health Record (EHR) at Vanderbilt University Medical Center (VUMC), we developed a simple method of good scientific practice for the management of multiple phenotype projects. Our method relies on the Google Sheets application (<https://docs.google.com/spreadsheets/>) that allows for cooperative editing of documents by multiple investigators at the same time. Our framework proved to be efficient when working with multiple teams of investigators from both inside and outside VUMC.

## Background

Researchers in biomedical informatics are often involved in multiple EHR-based phenotype projects that may happen simultaneously or may be in different phases of their development life cycle. Each of such phenotype projects involves one or multiple investigators, requires the design and implementation of one or multiple phenotype algorithms, and usually results in executing dozens of experiments with various configuration parameters.

## Method

The office suite made freely available by Google as both web and mobile applications has quickly become popular for its simplicity. These applications, including Google Docs, Sheets, and Slides, allow for an efficient collaboration where users can simultaneously edit documents, can see character-by-character document changes by other users in real time, can track the past edits, and are able to revert the documents through a revision history mechanism. Out of these applications, we adapted Google Sheets to create a practical framework for the management of our phenotype projects. In general, for each project, we created a spreadsheet that we shared with all the investigators participating in the project. A standard document contains one main sheet with all the experiments performed for the corresponding project and multiple auxiliary sheets with information used in these experiments. Figure 1 depicts one of our projects that involves sensitive time searches for the discovery of rare adverse events associated with various medications.[1] For instance, to identify patients who experienced severe allergic reactions such as Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis (SJS/TEN) caused by phenytoin, an investigator added the medication in 'MED lists' and SJS/TEN billing codes in 'ICD lists'. As shown in Figure 1, time constraints for filtering patients with adverse drug reactions occurring after a medication are specified in the main sheet. The configuration of each experiment is automatically processed, the corresponding phenotype algorithm is executed, and the extracted patient cohort is returned to the investigator for further analysis. Additional information sheets integrated in other projects include: 1) annotation guidelines and relevant examples for textual information extraction projects; 2) the evaluation of phenotype algorithms; and 3) SQL queries for cohort size estimations.

	Search index	Search type	List name	Time interval	Restriction	Count	Date
355	Search 64	[CASES]	Investigator name			165	March 2, 2017
356	a	MED	med01	[ t - 10, t + 60]			
357		ICD	icd01	[ t - 60, t + 10]	count >=2		
358		intersect with					
359	b	MED	med01	[ t - 10, t + 60]			
360		ICD	icd03	[ t - 60, t + 10]	count >=3		
361		intersect with					
362	c	MED	med01	[ t - 10, t + 60]			
363		LAB	lab01	[ t - 48, t + 48]			
364							
365	Search 65	[CONTROLS]	Investigator name			725	March 7, 2017
366		MED	med05		>= 12 weeks		
367		no ICD	icd19				

Figure 1 Screenshot of a phenotype project in Google Sheets.

## Conclusion

We presented a simple and easy to implement method that proved to be efficient for the management of intensive, collaborative phenotype projects. The type of phenotype projects tested with this method ranges from applications that combine structural clinical information to projects mainly based on natural language processing technologies.

## References

- 1 Garon S, Shade L, Derrick MI, et al. Development of Specific Electronic Phenotypes for Severe Cutaneous Adverse Drug Reactions Facilitates Genetic Discovery. *J Allergy Clin Immunol* 2017.