Ulysses: Introducing a workflow R package for assisting in the development of OHDSI studies

Martin Lavallee\(^1\), Asieh Golozar\(^2\)
Odysseus Data Services\(^1\), Northeastern University\(^2\)

Background
Running an OHDSI study contains lots of organizational complexity in terms of organizing code and proper documentation to communicate the code. While examples for constructing OHDSI studies have been presented, for example the SOS challenge, there is no clear workflow towards developing an OHDSI study as a piece of software available in a github repository. The OHDSI community would benefit from a workflow tool that will help standardize the development of network studies and improve its organization. This gap led to the development of a new R package called *Ulysses* (Useful Learning Yielded Structuring and Setting Epidemiology Studies) dedicated towards assisting in the development of an OHDSI study. *Ulysses* is planned to be submitted to the HADES suite of OHDSI R packages.

*Ulysses* draws inspiration from the *usethis* R package\(^1\), which is a workflow tool for the development of R packages. Similar to OHDSI studies, there are several administrative steps and procedures required to develop a stable and transparent R package. *Usethis* helps R programmers navigate R package development by supplying functions that automate simple tasks or other useful steps needed to meet this goal. *Ulysses* can provide a similar solution for OHDSI studies, improving the organization, communication and development of OHDSI studies by supplying simple functions to guide developers towards a study that is transparent, robust and reproducible. The OHDSI community would benefit from a tool that enforces standards and organization in OHDSI studies because it makes it easier for study nodes to execute network studies from a recognizable structure and provide guidance to new researchers seeking to build an OHDSI study if they follow a common workflow. In this software demo, we will showcase an example of how *Ulysses* can be used to start a new OHDSI study and help initiate necessary tasks for organizing and communicating the study to the OHDSI data network.

Methods
*Ulysses* assists with the development of an OHDSI study by automating repetitive tasks and offering organization to the study. This package intends to host a series of functions intended to assist in both the development (tools for the study host) and execution (tools for the study node) of studies.

The first major feature introduced by *Ulysses* is a standard directory structure to OHDSI studies, as seen in figure 1. This structure is initiated upon calling a function to start a new OHDSI study as a R project.
The directory structure is meant to provide an intuitive and easy to follow means of organization for study files both for developers and study nodes alike. The analysis folder contains files dedicated to the analysis. The cohortsToCreate folder stores the cohort definitions required for the study. The documentation folder stores the study documents for the study such as the study protocol, how to run and the contribution guidelines. The extras folder contains all files that are ancillary to the study. The logs folder stores the study logs from the execution and the results folder stores the results of the execution. In addition to these folders, Ulysses contains a _study.yml file that stores all the meta information for the study useful for templating documents. By providing a standard directory for OHDSI studies, it will make it easier for study nodes to recognize how to execute studies and for developers to create studies that can be successfully executed.

The second major feature introduced by Ulysses is templates for documents that are vital for communicating scientific and technical information about the study. Like usethis, Ulysses leverages logicless templating using the Mustache implementation for R called whisker. Some templates provided by Ulysses include: README.md, NEWS.md, HowToRun.md, StudyProtocol.Rmd, ContributionGuidelines.md, AnalysisScript.R, and StudyRepoRequestEmail.txt. Providing templates makes it quicker to start-up important documentation and suggest best practices for developing a strong OHDSI study.

Results

This software demo of Ulysses walks through tasks required for developing an OHDSI study, using examples from the Book of OHDSI. Topics covered in this demo include:

1) Initiating an OHDSI study: Ulysses::newOhdsiStudy(…)
2) Reviewing the directory structure for an OHDSI study (as a piece of software)
3) Automating the initiation of essential documentation: Ulysses::makeReadMe, Ulysses::makeStudyProtocol
4) Development roadmap
Conclusion

In this software demo, we hope to introduce Ulysses to the OHDSI community. Ulysses intends to streamline OHDSI studies to conform to a known structure and standard making it easier to communicate, execute and develop studies. We aim to develop this package alongside strategus and continue to incorporate best practices into the R package to help OHDSI study developers and nodes alike. Ulysses is an open collaboration for all OHDSI developers intended to source best practices for building network studies across the community.

References