



## Toward a containerized pipeline for longitudinal analysis of open-source software projects

Allan Miller  
amiller17@luc.edu

George Thiruvathukal  
*Loyola University Chicago*

Konstantin Laufer  
*Loyola University Chicago*

Emmanuel Amobi

Sean Higgins

*See next page for additional authors*

Follow this and additional works at: <https://ecommons.luc.edu/grs>

---

Miller, Allan; Thiruvathukal, George; Laufer, Konstantin; Amobi, Emmanuel; Higgins, Sean; Maliakal, Linette; Meister, Emily; Putter, Jean-Luc; Rose, Alex; Synovic, Nicholas; Von Hatten, Sophie; Warkentin, Jonathan; and Zugschwert, Martin, "Toward a containerized pipeline for longitudinal analysis of open-source software projects" (2020). *Graduate Research Symposium*. 4.  
<https://ecommons.luc.edu/grs/2020/posters/4>

This Open Access is brought to you for free and open access by the Conferences at Loyola eCommons. It has been accepted for inclusion in Graduate Research Symposium by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).

---

### **Presenter Information**

Allan Miller, George Thiruvathukal, Konstantin Laufer, Emmanuel Amobi, Sean Higgins, Linette Maliakal, Emily Meister, Jean-Luc Putter, Alex Rose, Nicholas Synovic, Sophie Von Hatten, Jonathan Warkentin, and Martin Zugschwert



Preparing people to lead extraordinary lives

# Toward a containerized pipeline for longitudinal analysis of open-source software projects

Allan Miller, Emmanuel Amobi, Sean Higgins, Linette Maliakal, Emily Meister, Jean-Luc Putter, Alex Rose, Nicholas Synovic, Sophie Von Hatten, Jonathan Warkentin, Martin Zugschwert, Konstantin Läufer and George K. Thiruvathukal  
Loyola University Chicago, Chicago, Illinois, USA



## Research Goal

Develop a research pipeline for investigating software metrics of GitHub projects known to indicate project health and code quality. Our design approach aims to be:

- Language Agnostic
- Modular
- Parallel/Distributed (Efficient)
- Visually-Appealing

## Language Agnostic

GitHub Projects contain many file types for different source languages. Instead of analyzing each language separately, our pipeline is focused on process metrics:

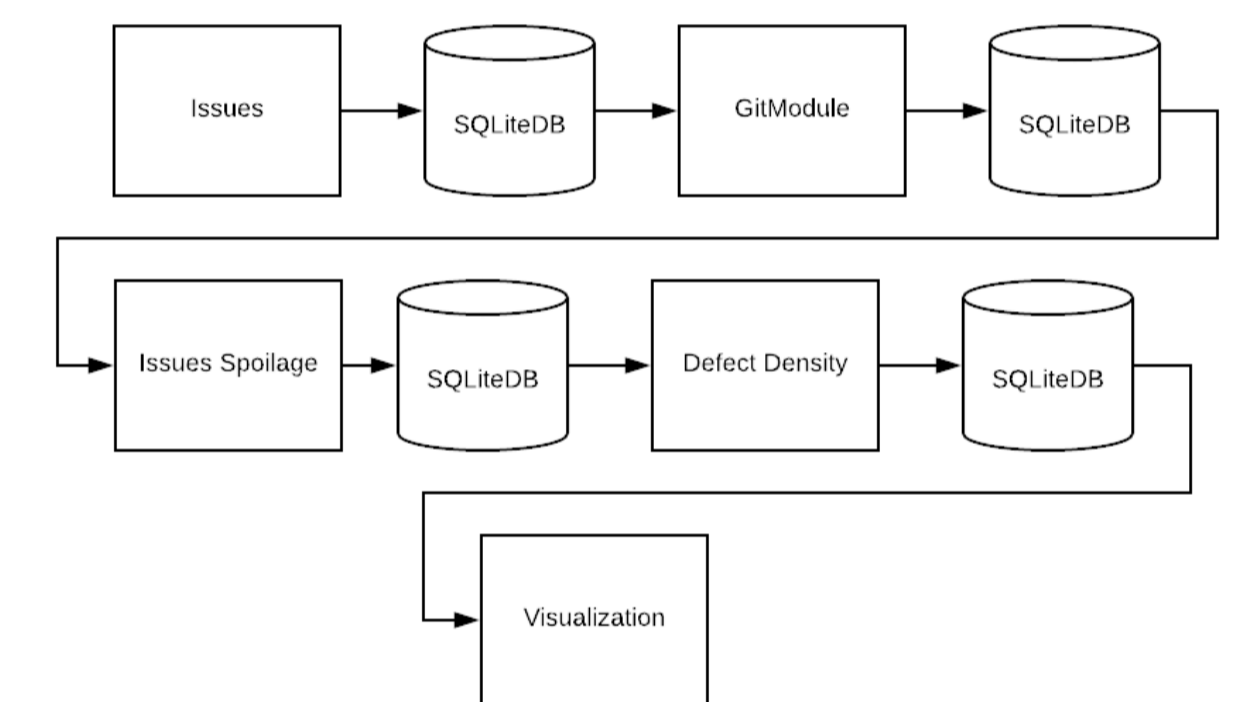
- Defect Density
- Issue Spoilage

## Example Visualization from Our Pipeline



## Dockerized Modular Pipeline

### Modular Sequence:



## Acknowledgments

- Thanks to the Software Systems Laboratory: <https://ssl.cs.luc.edu/>
- Previous work: <https://arxiv.org/abs/1804.02053>