

Thesis Portfolio

Automated Analysis of General Data Protection Regulation Violations in WordPress Plugins
(Technical Report)

Preserving Digital Privacy in the Light of Big Data
(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Patrick Thomas
Spring, 2021

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Automated Analysis of General Data Protection Regulation Violations in WordPress Plugins

Preserving Digital Privacy in the Light of Big Data

Thesis Prospectus

Sociotechnical Synthesis

Over the past several decades, Internet usage has greatly grown and now allows for people to access a myriad of online services, such as banking, shopping, and social media. Much of the Internet nowadays is fueled by online advertising and marketing, which in turn relies on the bulk collection and analysis of personal information. This runs against digital privacy or the idea that users should have the choice to maintain control of who accesses and collects their data.

One way to further digital privacy is to support initiatives like the General Data Protection Regulation (GDPR). In some situations where a web application's source code is publicly available, that code can be analyzed for a subset of GDPR violations. One such situation is WordPress plugins, which extend the functionality of WordPress beyond blogging. The technical project aims to build an analysis tool capable of evaluating WordPress plugins for GDPR violations. This is challenging in that violations can take on many different forms, and that analyzing source code that contains multiple languages is hard to model. These violations can range from unwarranted personal data collection to unauthorized personal data transmission to third parties.

As opposed to simply identifying digital privacy violations in a specific set of web application plugins according to GDPR, the STS research aims to answer what we as a society should do towards stopping datafication and preserving digital privacy. First, some ways that personal data is collected are explored, as it gives useful context to the solutions. Then, the importance of digital privacy is explained as well as challenges in users' understanding of privacy. Several existing, known methods are considered and evaluated as to how they support digital privacy and fundamental ideas behind digital privacy, like autonomy and self-ownership. These methods include ad blockers, DNS privacy, government regulation, advocacy, and

education. Ad blockers and DNS-over-HTTPS, while realizing some benefits for the user, only tangentially support digital privacy but still help to achieve it. On the other hand, a combination of regulation, advocacy, and earlier privacy education in undergraduate curricula more directly support the ideals of digital privacy. Those three combined help change the course of datafication rather than help fix the resulting problems, helping us reach a more private future.

Overall, the work on the aforementioned projects was somewhat successful. Through the STS research, I learned a great deal about digital privacy and its values as well as what exists beyond ad blocking to remedy datafication. In retrospect, I would have chosen a more narrow topic though, as the scope of the project was very hard to contain and work with. The technical project similarly taught me a great deal about new technologies to me, like graph databases. Stronger project management would have served us well, and better use of GitHub, including features like GitHub issues, would have made the project a lot more organized.