The Need for “Data Guardian” In 2019 digital ad revenue topped $100 billion for the first time1, the reason being their efficacy. Targeted ads become more effective each year as we continue to create and give away data, revealing more about who we are individually. As of last year, 2.5 quintillion bytes of data are being generated every day2. This data can range from basic public information we input ourselves – our name, our age, location – to usage data we may not realize we are signing away – such as shopping habits, frequently visited sites, travel habits, job fields, etc. Companies know the value of this, as American companies alone spent $19 billion this year acquiring it3. This is the cause for our project, “Data Guardian”: a means for a user to track the data they share with companies.

Now, we achieve this goal as “Data Guardian” as a service upkeeps a ledger of company reported data usage. In practice, before an individual creates an online presence, they will create an account with our service. We will then allow them to create a custom data policy, mandating what data companies have access to, specifying a lease for how long the company may use said data, and requiring companies to report when they use that data. Then, when the individual creates an account on another service, such as Facebook, they link their new Facebook account to their “Data Guardian” account. Facebook would agree to the Data Guardian policy terms in exchange for access to the users data; as part of that agreement, Facebook then report to the Data Guardian any instance in which they use that user’s data. The Data Guardian then reviews this report and adds it to our ledger, for the individual to view.

Shifting from this idealized view, there are several technical challenges within this project, such as the design ledger and the means of maintaining its security and preventing false reporting. Another technical challenge is how to build such a system with the ability for scale, as the amount of internet users is only set to grow; Blockchain is built for growth however, as we will not store any of the data beyond the basic policy tables per user and reported data-usage from companus. Lastly, there is the challenge of how to account for company participation; companies, however, are looking to re-establish trust after several public exposures of unknown data use – such as Cambridge Analytica in 20164. This would be a means for them to operate, still with a profitable data centered model, but without heavy regulations that are sure to come if the industry may not regulate itself. Lastly, the project would be funded from the government or an NGO, dedicated to watchdog organizations.

1. “How Much Data Do We Create Every Day?”, https://www.forbes.com/sites/bernardmarr/2018

/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should- read/#70f458160ba9

2. “Digital ad revenue in the US surpassed $100 billion for the first time in 2018”,

https://www.cnbc.com/2019/05/07/digital-ad-revenue-in-the-us-topped-100-billion-for-the-first- time.html

3. **“**The WIRED Guide to Your Personal Data (and Who Is Using It)”, https://www.wired.com/story/wired-guide-personal-data-collection/

4. “Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens”,

https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html