Daniel Marley 09/10/19 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 and more effective each year as we continue to create and give away more and more 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. Yet, we sign off on our data when presented with a long document of “Terms and Conditions” without reading the specifics: do any of us know with certainty what of our exact information is out there about us and who its being sold to?

Enter our project of “Data Guardian”, a service to track the usage of our data on a decentralized blockchain, acting as an immutable leger of our data usage in conjunction with the companies who collect it. In practice, before an individual creates an online presence, they will create an account with our service. We will 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, and would report anytime they use that user’s data to the Data Guardian, for us to store in our blockchain record for the individual to view.

Include transitionsThere are several technical challenges within this project, such as the design of the blockchain and the method of consensus for the blockchain – obviously we may not sustain such a project with the common method of data mining. We may, however, use a delegated algorithm that would function off of a group of trusted internal nodes who would find a means of voting to verify transactions on the blockchain. 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, and we personally will not store any of the data beyond what the basic policy tables are and any reported uses. 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. There is no mention of seeking funding

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