**Ethical Submission: Data collection on a genomic privacy online evaluation tool (GENOPRI)**

|  |
| --- |
| **Timestamp** |
| 29/08/2019 |
| **Study Title:** |
| Data collection on the awareness level of web developers with regards to Subresource Integrity (SRI) as well as if SRI is well understood and used according to its specification |
| **Researcher(s):** |
| Bertil Chapuis, Kévin Huguenin, Mauro Cherubini, Olamide Omolola |
| **Email address(es) of researcher(s):** |
| bertil.chapuis@unil.ch, kevin.huguenin@unil.ch. mauro.cherubini@unil.ch, olamide.omolola@unil.ch |
| **Responsible Professor:** |
| Kévin Huguenin |
| **Responsible Professor is:** |
| Organizer |
| **Research question(s) to be answered by the study:** |
| 1. To what extent are developers aware of SRI? 2. To what extent do developers using SRI in their web applications/sites? 3. To what extent do developers understand the functioning of SRI (general and specific behavior in the specification/user-agent implementation)? 4. What are the current practices of developers when using SRI? (are these practice coherent with the specification) |
| **Study design and procedure:** |
| We propose to deploy a large-scale questionnaire to a random sample of web developers all over the world. We will deploy the questionnaire in two phases:   1. The first set of questionnaire will be disseminated via email. The sets of e-mails were obtained through following processes:    1. Domain owners. Domain owners can be contacted via e-mail at address “webmasters@domain” where domain is the considered domain name (e.g., videolan.org). The reasoning is based on the ICANN requirement where every domain has to provide a contact email and “webmaster@domain” redirects to the contact email provided by the domain owner. The domains will be chosen as follow from the Common Crawl dataset (CC): *N* random domains that do use SRI; *N* random domains that do not use SRI. Only domains with at least *k* distinct URLs in CC are considered. Typical values are *N ~ 2’000, k ~ 5*.    2. Wordpress plugin/theme authors. The “readme” file of the wordpress plugin/theme usually contains the e-mail of the authors. The authors of the plugin/themes are web developers because wordpress is a web content management system (CMS). *N ~ 20’000*    3. Node JS package authors. *N ~ 50’000*    4. Google web developer contacts. To be explored. 2. The next phase occurs if the response to the questionnaires are not sufficient. The possibilities for this phase are presented below:    1. Web developers on MTurk. Paid distribution of questionnaires with services like MTurk targeting category “web developers”. *N ~ 500.*    2. HEIGVD alumni working as web developer / webmaster. To be explored. *N ~ 50.* + phone/in-person interview. |
| **Type of compensation offered to the participants (monetary, other, none):** |
| None at first. There is no recruitment, we just collect responses to our questionnaire. However if the response is not sufficient, we will use commercial survey platforms e.g., MTurk |
| **Does the experimental procedure involve deception of participants?** |
| No |
| **If yes, why is deception necessary?** |
|  |
| **Are there any reasonably foreseeable risks or discomforts to participants and or groups/ communities?** |
| No |
| **If yes, indicate probability, magnitude, and duration of each (note that risks may be physical, psychological, social, legal and/or economic) and outline provision to minimize risk** |
|  |
| **Describe the types of research data to be collected, treated and analyzed and their origin(s)?** |
| * Questionnaire data |
| **During the experiment, how will subject-subject anonymity be assured?** |
| We have the emails of the respondents which is not open to the public. These emails are only used for distribution of the questionnaires and their responses cannot be linked to their emails. |
| **Will the anonymity of the participants be reasonably guaranteed in the final data-set?** |
| Yes |
| **If the anonymity of the participants will not be guaranteed, please explain why** |
|  |
| **Briefly describe how (where and on what media) the data will be stored, backed-up, managed and curated in the short to medium term (during the lifetime of the project). It is highly recommended to use SWITCHDrive. If you plan NOT to use SWITCHDrive, explain why and include a description of the protection mechanisms that will be put in place (e.g., cryptography, access control).** |
| Secure server at UNIL + SWITCHDrive **(Bertil, please check this)**. |
| **For how long will the data kept after the end of the project?** |
| Data is collected and stored for the duration of the research project on SRI |
| **If the data-set is intended to be shared with other institutions, under which conditions will it be shared and what data protection mechanisms do you intend to further apply (if any).** |
| It will only be shared with the project partners from HES-SO, CHUV and SDSC**(Bertil, please check this, I think Swiss army is now part)**. |
| **If yes, briefly state why it is suitable and under what conditions:** |
|  |
| **How many participants will you (roughly) need for the study?** |
| N/A |
| **Do you need a minimum number of participants per session in order to be able to run a session?** |
| No |
| **If your minimal session size is larger than 13 participants, please explain why it is not possible to run smaller sessions:** |
|  |
| **When do you plan to run the study?** |
| Start in September 2019 |
| **Do you plan to use the ORSEE subject-pool for the recruitment of participants?** |
| No **(Bertil, please check this)**. |
| **If no, how do you plan to recruit the participants?** |
| No recruitment needed but we mass mail participants **(Bertil, please check this)**. |
| **For SNSF application: Do you need an approval letter from the HEC Ethics Committee?** |
| N/A |
| **Do you have any further questions or comments? Please note them here or contact us via hec-labex@unil.ch** |
|  |
| **Accepted by Ethics Committee** |