2021

Assessment of the Privacy and Security Practices of the Indian Academic Websites

Chanlang Ki Bareh
North-Eastern Hill University, shanlang88@gmail.com

Assessment of the Privacy and Security Practices of the Indian Academic Websites.

Abstract

The study presents the comprehensive assessment of the Indian Institutions of National Importance (IoNI) website's privacy and security practices. These 130 IoNI websites were selected because of the status conferred to premier institutions by the Government of India. This empirical study was based on webometric methods that assessed the quantitative aspects of information technologies structures on the web. The study also employed content analysis for analyzing the sites privacy policies. The assessment reveals numerous security vulnerabilities like the adoption of HTTPS may not always provide better protection to their users because 39 websites do not have TLS encrypted connection; still 62 websites use the older encryption TLS 1.2 version. Though Google tracking is moderately prevalent in 53 websites, only 4 websites enabled the privacy protection parameters. Also, only 26 percent of the 130 websites have privacy policies, yet their readability score is very poor. It was also found that none of the websites provides a cookie consent form or opt-in/opt-out option on their landing page. This may indicate that institutions are not concerned with the current privacy and security standard. Finally, based on the analysis, recommendations are made like increasing users' awareness, implementing simple privacy policy, reducing unnecessary tracking through informed consent, and promoting the use of privacy-enhancing technologies. Web tracking is common, yet few studies demonstrate its extent and consequences from academic websites context. It is hoped that these valuable insights into the current state of privacy and security may incite the need for updated information security.

Keywords: Privacy concern, Web tracking, HTTPS, Google analytics, Privacy policy, Cookies, Indian Institutions of National Importance.

1. Introduction

The exact number of websites keeps growing every minute, and there are over 1 billion sites on the world wide web i.e., 1,213,277,377 according to Netcraft's June 2021 Web Server Survey **(Netcraft, 2021).** This growth of websites and web services brought people and the Internet closer to each other and makes life and work a little easier with the readily available information at their fingertips. This online interaction often leads to the ever-growing collection and disclosure of users' personal information that service providers collect, store, and track users for various purposes. The majority of Internet users are unaware of these tracking activities, which rendered them lose control over their personal data. In addition, online preferences for entertainment, browsing logs, medical information, geolocation provides valuable insight for marketers. To this concern, every online activity over the web or social media post is recorded and stored to create helpful insight for personalized service or generate revenue by selling users' personal data to data aggregators.

Collecting personal data for tracking purposes is done via filling up online forms or it can be collected without users' knowledge and consent by analyzing the IP headers, HTTP requests, queries made in search engines, JavaScript and Flash programs embedded in web pages. Unfortunately, the collection of personal data does not end here. Sometimes, webmail services reputation exceeds to scanning and processing user's e-mail, even if they are received from a user who did not consent to message inspection (**Bujlow, Carela-Espanol, Sole-Pareta, & Barlet-Ros, 2017**). Firms not just collect and processes personal information of their site visitor but also from other parties. The merging between firms active in the third-party tracking industry raises unique privacy and fundamental rights challenges that are often missed in regulatory decisions and academic discussions of data and market concentration (**Binns & Bietti, 2020).**

With the ever-growing digital technologies, users' data became more beneficial to marketers for target advertisements. This causes Internet users and privacy advocates to raise concern about potential privacy infringement to the ever-growing big data (**Bauer, Bergstrom, & Madsen, 2021).** Users' privacy is endangered by using different sniffing and spying tools that allow government and private entities to passively track and monitor users' activities on the Internet (**Al-Shehari & Zhioua, 2018**). Criminals and terror groups also exploit this technology for nefarious purposes like scamming people, identity theft and various economic mal-practices against individuals or businesses **(Potoglou, Dunkerley, Patil, & Robinson, 2017)**. According to Barracuda Networks, over 1000 educational institutions cyber-attacks take place in India during June-September 2020. This is because most schools and colleges started conducting online classes using various teaching-learning software during the 2020 pandemic creating waves of spear-phishing attacks (**Ahaskar, 2021).**

Previous studies in web tracking and web privacy have generally focused on users' attitude towards privacy and web tracking, tools/technologies to protect tracking and privacy, and mechanisms that trigger web tracking. To the author's knowledge, an assessment of the current privacy and security mechanism of the Indian Institute of higher education website does not exist to date. As pointed that the measurement studies of web tracking are critical in order to ensure transparency to Internet users', technologists, policy makers, trackers and help users' understand how their data is collected and used to facilitate informed decision about privacy (**Lerner, Simpson, Kohno, & Roesner, 2016; O'Brien, Young, Arlitsch, & Benedict, 2018)**. This may help increase site owner/admin awareness on their current practices of privacy and security. Thus empower users' knowledge on personal information collection, which in turn may help guard against harmful trackers. As web traffic grows in size, protecting data is now becoming a necessity than ever before **(Yessine Borchani**, **2020).** The privacy and security practices in these Institutions of National Importance (IoNI) websites need to be assessed as a right of each student using these sites. Overall, with the absence of the Privacy Act in India, this study's contributions will enrich the privacy literature and may also assist the regulatory authority in understanding the prevalent privacy and security practices of IoNI websites. In order to grasp the nature and extent of privacy and web tracking that occurs on IoNI academic websites, the author proposed to address the following questions:

1.1. Do IoNI sites implement HTTPS with proper redirect practices?

1.2. Do IoNI sites have SSL/TLS Certification?
1.3. Do IoNI sites use Google Analytics and implement privacy protection parameters?
1.4. Do IoNI sites have robust privacy policies in place?
1.5. Do IoNI sites collect cookies and provide consent form?

The study presented the 130 academic websites of India's Institutions of National Importance (IoNI) basic privacy and security mechanisms employed in these academic sites. The study presents the following theoretical background on web tracking and privacy concerns on web tracking. The methodology then followed it. The author further laid out the results and discussion section and finally placed recommendations followed by limitations and avenues for future research.

2. Theoretical background

2.1. Web Tracking

The practice of tracking people's online activities has existed ever since companies started monetizing or advertising their products. It is certainly not a risky action in their missions to maximize incomes. However, it causes users' privacy and security concerns in many recent cases since outsider elements could utilize the accumulated information for malicious activities **(Malandrino & Scarano, 2013). Rossi (2021)** stated that the development of tracking technology like cookies and fingerprinting bypass the non-identified nature of IP addresses. These techniques are now commonly used for personalized advertisement by industry. The tracking types include analytics, usability tracking, tracking for personalized advertising, cross-origin requests, local storage for session data, ad-block detection, location data, and fingerprinting **(Pilton, Faily, & Henriksen-Bulmer, 2021).** It is also accomplished by different techniques, including tracking cookies, pixel tags, beacons, and other sophisticated mechanisms **(Cook, Nithyanand, & Shafiq, 2020).** Such tracking is not comfortable, as exemplified by one respondent, who worried that collecting personal information may cause her harm by the third-party site (**Melicher et al., 2016).**

An example by **Khormali et al. (2021)** shown that if a user typed www.example.com into a web browser, a server would map that name to an IP address, e.g., 1.2.3.4, corresponding to that domain and other Internet activities, e.g., web browsing, transferring files, rely on DNS to quickly provide the information necessary to connect users to remote hosts. Another practice is the deployment of unauthorized or unauthenticated Internet of Things (IoT) to inject false data into the system, enabling access to data and data breaches occurrence at different levels of the system (**Ari et al., 2020**). The tracking mechanism is also done in the form of GET/POST requests; these are requests made to websites or web services that usually send a response. Once a user is identified, a GET/POST request could be made to retrieve that user's profile picture but is more frequently used for requesting data about that user from third parties, including their browsing history. Another technique to track unknown devices and match cross-device is the user-device pair management module, which relies on two modules - matching cross-device user tracking and device finger printing modules to create a corresponding file generated for the device. This form of tracking the unknown device is processed to generate the corresponding user device pair by user-device pairing,

therefore, identifying the user with the unknown device (**Liu & Zhang, 2021**). In many cases, what concerns many users is the re-purpose of the same tracking data for multiple services and data ostensible collected for analytics might later be used/sell for targeted advertising and security **(Binns & Bietti, 2020; Javed, Salehin, & Shehab, 2020)**. Also, since users share similar behaviour across devices, user tracking gained more popularity for advertisers (**Liu & Zhang, 2021)**.

However, tracking is commonly done via Storage-based tracking mechanisms, which rely on data stores on PC/Devices. The most popular form of storage-based tracking mechanism is a cookie. Cookies and local storage were obtained by requesting the storage objects which could be sent to the extension interface (Cook, Nithyanand, & Shafiq, **2020)**. Tracking is also extended to other organizations, as **Narayanan and Reisman (2017)** remarked the extent of NSA or other surveillance agencies that uses cookies ability to track a person browsing traffic**.** According to **Rodriguez, Torres, Flores, and Benavides (2019)** cookie is a simple text format that are not virus or automatic codes. Cookies on the users' side are structured as key-value pairs that include identifiers that uniquely identify a user (**Cook, Nithyanand, & Shafiq, 2020**). While some did not mind if cookies were used to restore the browsing sessions, others complained that it clogged their computer **(Pilton, Faily, & Henriksen-Bulmer, 2021).** The theft of cookies to track or steal identity also extends to advanced techniques such as ever-cookies, cookie syncing, fingerprinting of browser type, webRTC, audio context etc., (**Samarasinghe & Mannan, 2019**). Others also view user profiling as a result of third party cookies (**McCarthy & Yates, 2010**), since the use of cookies to modify app behaviour is considered profiling (**Benjumea et al., 2020**). This third-party code is typically embedded on multiple sites or apps, making it easier to track user behaviours (**Binns & Bietti, 2020**). According to **Pilton, Faily, and Henriksen-Bulmer (2021)**, local storage is a different type of storage-based tracking mechanism which allows for larger data to be stored in a similar way. On the other hand, Fingerprinting is a method of identifying a device by creating a unique key. It uses different technologies to create a unique identifier for a device based on various factors, such as operating system, browser version, and screen size.

Web analytics which was identified as early as 1996 in the dataset has emerged as one of the most essential activities in e-commerce because of its ability to study the behaviour of customers (**Lerner, Simpson, Kohno, & Roesner, 2016)**. According to **Katuu (2018**), web analytics is defined as "the measurement, collection, analysis, and reporting of Internet data for the purposes of understanding and optimizing web usage". Many web analytics services includes, e.g., Clicky, Piwik, Google Analytics, StateCounter or LuckyOrange, used by thousands of websites serve millions of users (**Leiva & Huang, 2015)** are widely used for tracking and analyzing commercial activity (**Garcia, Garcia-Nieto, & Aldana-Montes, 2016; Katuu, 2018**). In addition, web analytics also allows one to assess technical specs such as incorrect links; outbound links that lead to non-existent domain; error 404; reloading site pages, and indexing by search robots. In short, it helps to identify users' search behaviour (**Redkina, 2018).**

With online advertising revenue raging in billion, it is no surprise that Google search engine monetization turned into one of the world most recognized brands (**Quintel & Wilson, 2020**). Though personalized advertisement poses privacy threats (**Cha, 2011**), Internet users still find it beneficial since it provides personalized information **(Strycharz, Smit, Helberger, & Noort, 2021; Liu & Zhang, 2021).** In some cases, users find that targeted advertisements could save them time and money (**Melicher et al., 2016)**. The use of web

analytics like Google tracking services (Tag Manager and Google analytics) specialized in users behaviour data (**Farney, 2016**) are widespread in users' tracking (**Katuu, 2018)**. For example, **O'Brien, Young, Arlitsch, and Benedict (2018)** survey of 279 libraries found that 88% of the sample studied implemented Google Analytics or Google Tag Manager. It was also found that 9 out of 20 predominant third-party domains belong to trackers and also confirm the extensive tracking capacity of Google (**Schelter & Kunegis, 2016**).

With today's big data analytic capacity, access to users' personal data added more advantages (**Chang, Wong, Libaque-Saenz, & lee, 2018**). Industries like Telecom, Banks, Credit card companies are using big data to build marketing tools based on customer preferences **(Jun, Yoo, & Choi, 2018)** and, this intensified with the presence of Information externalities which made it possible to draw inferences about users from already existing data shared by the users (**Choi, Jeon & Kim, 2019**). These externalities complicate the digital-data industry, which is commonly referred to as third-party tracking, in which firms not just collect personal data of its own but instead use data of other first-party services. This third-party tracking industry raises unique privacy challenges that are often not included in regulatory decisions (**Binns & Bietti, 2020)**. The presence of these third-party trackers is extensive can be extensive in the form of advertisers or simple content embedded on first-party sites like conversion tracking, acceleration of content loading or provision of widgets. Therefore, the presence of third-party tracking is relatively higher (**Samarasinghe & Mannnan, 2019**).

2.2. Privacy concerns over web tracking

The increased usage of the web leads to the collection of personal information which can be used to identify a specific individual. This online interaction usually leads to a massive amount of information gathered and stored by service providers **(Ali, Zaaba, Singh, & Hussain, 2020).** According to **Jamin et al. (2019)**, this person's information is used to identify from details like name or other description associated with the person. The main concerns according to **Cozza et al. (2020)**, regarding information collection is not just for personal information (e.g., name, date of birth, address, email, etc.), but also sensitive information (e.g., health data, political views, biometric data, etc.). Surprisingly, these practices of collecting personal data are unaware to a larger extent, and this undermined the privacy of people associated with online tracking (**Bashir, Arshad, & Robertson, 2016**). While **Choi, Joen, and Kim (2019)** is of the opinion that even if the collection of personal information requires users' permission and people are aware of the privacy risks, the economy is still dictated by an excessive collection of personal information, which resulted in the loss of privacy.

The collection of online personal data boosts users' experiences but poses concerns like privacy infringement (**Cha, 2011**). Users' experience can be enhanced by personalizing the pricing based on personal behaviour on a particular website. This is often collected via first-party cookies or across websites through aggregators, like Google Analytics, which uses third-party cookies **(Schmidt, Bornschein, & Maier, 2020).** For example, by recording users' preferences of product choices, buying patterns or uses cookies to alert repeat visitors to special offerings **(Stearn, 1998).** This benefit of users' tracking for personalized service is also reported by 69% of participants who are willing to provide personal information about their tastes, interests, and preferences for online advertising purposes if the website compensated them for the information they give away (**Frik & Mittone, 2019)**. This implied

that the collection of personal information is found to be quite beneficial for the users. Hence, the consolidation of web tracking could be both positive and negative for privacy **(Binns & Bietti, 2020).** In addition to site users' visited and products users' buy, the tracking is also done for price discrimination, personalization of services and government surveillance (**Bujlow, Carela-español, Solé-Pareta, & Barlet-Ros, 2017**).

Various studies have shown that users are concerns about their privacy associated with web tracking. **Melicher et al. (2016)** respondents felt that invasion of privacy occurs when their personal information is in the hand of the trackers. In fact, **Frik and Mittone (2019)** reported that 43% of their participants have personally fallen victim to the invasion of privacy. An example would be the DoubleClick Company who has been sued multiple times due to privacy violations relating to cookie tracking practices (**Hormozi, 2005**). **Brown, Ghani, Hoque and Rehman (2012)** findings also revealed that none of the websites gave list of affiliates, and this implies that people do not know which websites their personal information or non-personal information is shared with. This fact is concerning to many of the faithful Internet users. Similarly, many websites failed to address in their privacy policy the types of data shared with unknown third parties and whether this personal information is under the control of a host website or not. In some cases, websites take advantage to the absence of proper regulations **(Cha, 2011)**.

Numerous studies indicate that failure of the websites to protect users' personal information also affected their trust towards the organization. It's been observed that some websites exploit the trust of individuals by selling, sharing, or analyzing their data **(Pilton, Faily, & Henriksen-Bulmer, 2021). In Melicher et al. (2016)** study, over 51% of the respondents felt that tracking search activities on Google or Yahoo impacted their feeling. **Martin (2020)** results also show that the secondary uses of personal data, like sharing it with a third party and changing the use of information have the same impact on consumer trust. Moreover, about third of **Melicher et al. (2016**) respondents, 37% reported affecting their trust in the tracking party. In line with the privacy calculus of **Dinev and Hart (2006)**, privacy risks include collection and types of information, handling errors, unauthorized secondary use (e.g. third party access and sharing) and improper access to personal information.

There is a need to improve website security and privacy practices. As pointed out by **Chang, Wong, Libaque-Saenz, and lee (2018)**, intrusion into private data, disclosure of sensitive information and exposure of individuals data to third parties could be avoided if proper and transparent measures are displayed on these sites. Consider that the request for personal information by sites may create privacy concerns but the imposition of strict passwords and secure authentication in compliance with the law and possessed security certificates may mitigate privacy concerns (**Frik & Mittone, 2019).** Meanwhile, users may also resort to taking matters into their own hands to block online trackers since stringent privacy regulations and the ability to block trackers may limit retailers to track individuals explicitly **(Kakatkar & Spann, 2019).** Certain mechanisms currently exist to help educate users in context. While understanding and being aware regarding organizational practices might contribute partial control over personal information disclosure **(Frik & Mittone, 2019);** other privacy protection techniques designed may also help mitigate potential privacy violation **(Huang & Bashir, 2018).** After all, when there is a lack of clear communication and protection around web tracking, it is easy to compromise users' privacy.

3. Methodology

Webometric as a method was applied as it covers the quantitative aspects of web studies such as web link analysis, web usage log, e.g., browsing log and web technology analysis **(Björneborn & Ingwersen, 2001).** The webometric also covers analysis on network communication using infometric and other quantitative analysis **(Jeyshankar & Babu, 2009). Lorentzen (2014)** used webometric query for web mining and even remarked to computer science concepts usage in web metric network. Nevertheless, Webometric methodology is one of the empirical methodologies popularly employed in the field of library and information science research **(Pal, Kar, & Sardar, 2020; Verma & Brahma, 2017; Jeyshankar & Babu, 2009; Jalal, Biswas, & Mukhopadhyay, 2009).** Nevertheless, **Muruganandham (2019)** believes that this technique is broader still and, based on this methodology web content analysis of site privacy was undertaken by **(Thompson, Mullins & Chongsutakawewong, 2020; O'Brien, Young, Arlitsch & Benedict, 2018)** to determine sites privacy policies. These methods provided a rigorous and thorough evaluation to assess India's Institutions of National Importance (IoNI) website privacy and security practices. The following section details the data sample and design methodology.

3.1. Sample

The sample for the current study includes all the 130 academic websites of the Indian Institutions of National Importance (IoNI) under the Department of Higher Education, Government of India **(Ministry of Education, 2021).** This is a status conferred by the parliament of India to premier institutions under public undertaken. According to **Nangia (2020)**, these are institutions that "serve as a pivotal player in developing highly skilled personnel within the specified region of the country/state." Many of these institutions are highly ranked under the National Institutional Ranking Framework (NIRF), and few have found a place in global rankings.

3.2. Data Collection

Under the Webometric method, the data collection period spanned over May – July 2021 and all the 130 IoNI academic websites were assessed manually using a fresh installed Mozilla browser. The data collection method followed for the study is a covert observation in which the researcher's professional identity and academic intentions are partially or fully hidden from those involved in the study **(Lugosi, 2008).** When this data collection technique was chosen, it was not intended to create risks since the observed subjects for the present study are information structures publicly hosted on machines (**O'Brien, Young, Arlitsch & Benedict, 2018).**

3.3. Procedures

The following steps are unique to each research questions below:

3.3.1. RQ1: Do IoNI sites implement HTTPS with proper redirect practices?

Step 1. The author takes the Institution keyword from the Ministry of Education website and pastes the same in the browser.

Step 2. The author assesses the landing page and inspects for the presence of a secure connection (HTTPS) between the users' browser and the site URL.

Step 3. Assessment of the server redirects non-secure requests to secure connections of the page requested or vice-versa using the following examples:

- For example: If a non-secure URL is requested (e.g. http://www.nits.ac.in/), does the web server respond with HTTP 301 redirecting users to secure URL (e.g. https://www.nits.ac.in/)

And'

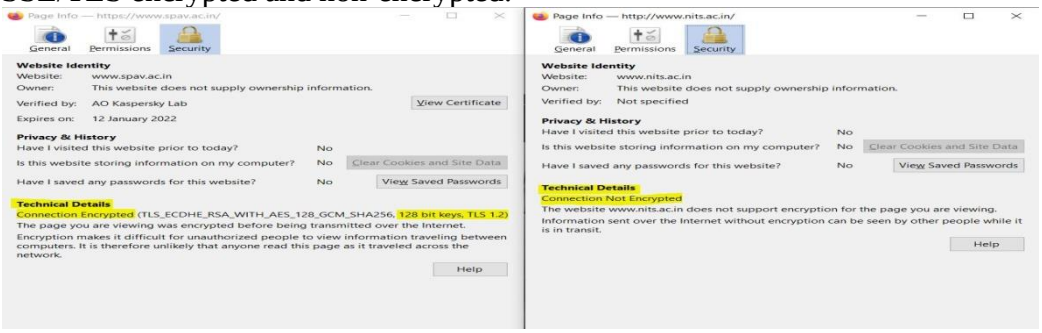
- For example: When a secure URL is requested (e.g. https://www.nits.ac.in/) does the web server redirect the users to a non-secure URL (http://www.nits.ac.in/)

3.3.2. RQ2: Do IoNI sites have SSL/TLS Certification?

To check if the connection is SSL/TLC secured or verified. The page info is accessed by pressing the shortcuts key Ctrl + I or one can use the browser setting options.

- For example, upon accessing this browser page, press the shortcuts key Ctrl+I on the landing page. Page info will immediately display as depicted in Figure 1. The security tab highlights the technical details of connection that are encrypted with TLS between the users' browser and the IoNI websites.

Fig 1. SSL/TLS encrypted and non-encrypted.



3.3.3. RQ3: Do IoNI sites use Google Analytics and implement privacy protection parameters?

Step 1. To determine whether the IoNI sites implement Google Analytics? Each IoNI webpage is manually checks from the HTML code by pressing the shortcut Ctrl+U or' right-click the mouse and select "View page source," or used the browser setting to access the webpage HTML code.

Step 2. Each page is then checked for the presence of global site tag (gtag.js) **(Google Analytics, 2021a)** or Google analytics tag (analytics.js) (**Google Analytics, 2021b)** and, Google tag manager (gtm.js) **(Google Tag Manager, 2021).**

Step 3. If one of the tracking code from step 2 is detected. In that case, the author then proceeds to find out if the site enabled privacy protection parameters like - disable Google Analytics (**Google Analytics, 2021e);** disable advertising features **(Google Analytics, 2021f);** disable pageview measurement **(Google Analytics, 2021c);** anonymize IP addresses **(Google Analytics, 2021d)**.

3.3.4. RQ4: Do IoNI sites have robust privacy policies in place?

To address this question, the privacy policy content will be analyzed by going through each (Privacy link, Privacy Policy, Privacy statements, etc.) usually located in the footer note/section.

Step 1. The site is check for the availability of privacy policy.

Step 2. Site with privacy policy is then checked for word count using the basic text analyzer generated by (https://www.online-utility.org/).

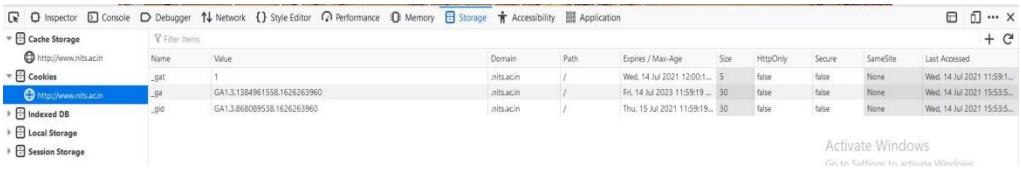
Step 3. The policy is then assessed based on the statement readability using the Flesch-Kincaid grade level formula, which is used to evaluate privacy policies (**Javed, Salehin, Shehab, 2020**). This tool is publicly available online at (https://www.online-utility.org/).

Step 4. The privacy statement is then assessed for the presence or absence of crucial information regarding data collection and the purpose of data collection.

3.3.5. RQ5: Do IoNI sites collect cookies and provide consent form?

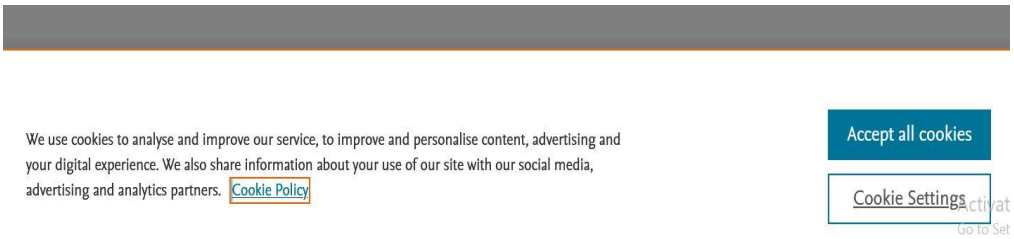
Step 1. Each IoNI site is manually check for the presence of both persistent and non-persistent cookies by pressing the shortcut key Ctrl+Shift+I or' used the browser privacy and security setting option to manually search for each URL browsed to locate the number of cookies stored in the browser. The Figure 2 below depict the description and number of cookies stored in the browser.

Fig 2. Description/No. of cookies using the shortcut key Ctrl+Shift+I



Step 2. The observations is also made for the presence of protection features such as cookie-consent form, privacy notice or opt-out/opt-in feature upon landing on the site for the first time. Figure 3 below depict the example of consent form and opt-out/opt-in feature.

Fig 3. Consent form and opt-out/opt-in.



4. Data Analysis and Interpretation

4.1. RQ1: Do IoNI sites implement HTTPS with proper redirect practices?

Fig 4. Https implementation

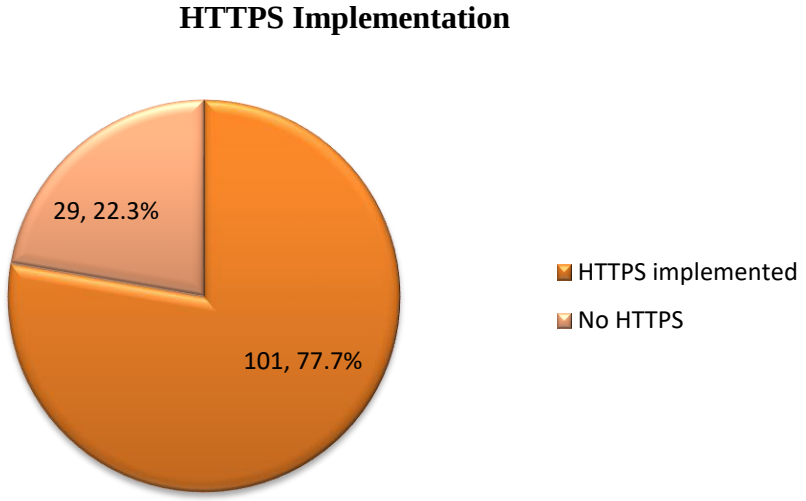


Fig 5. Http/Https redirect practices

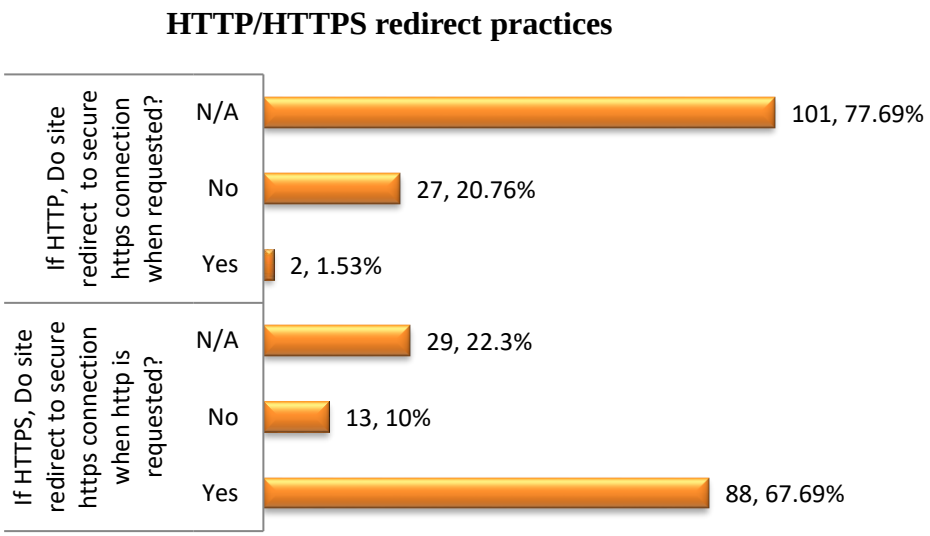
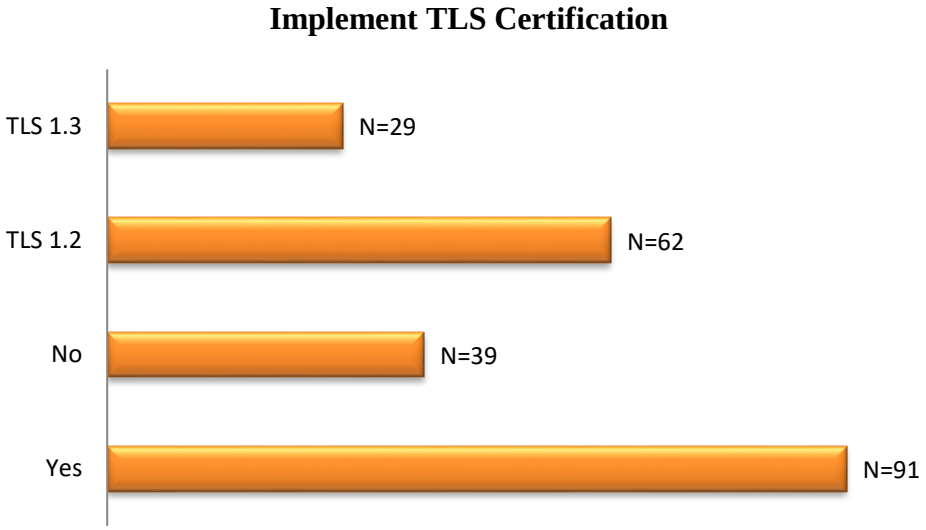


Figure 4 indicates that HTTPS implementation on IoNI websites is widespread (N = 101, 77.7%), with many websites offer secure connections while non-secure connections without HTTPS occupy only 22.3 percent. This absence of secure connection (https://) still undermines users' privacy and may be susceptible to privacy violation. The redirect practices in Figure 5, from non-secure URL manual requests (i.e. http://) to secure (i.e. https://) is negligible with only 1.53 percent reported to support secure redirect practice. Also, to see if secure connection (https://) remain secure when http is requested, revealed significant number (N = 88, 67.69%). Thus, offering secure redirect practices to users and protect their privacy.

Earlier literature has given way for discussion about user privacy and the concerns with websites that implement HTTPS. **O'Brien, Young, Arlitsch and Benedict (2018**) found that majority of the websites implement basic encryption technology via HTTPS which is also reflected in the current finding of 77.7 percent. Of the finding, the redirect practice to a secure connection from non-secure connection is less at 1.53 percent, and this trend is also seen in O'Brien etal study. This is not true in China where web services are at risks due to the use of HTTP **(Huang, Zhang, Li, & Xin, 2019**). Similarly, this is also seen in the audit of 800 pages across 40 websites by **Thompson, Mullins, Chongsutakawewong, (2020)** on the use of HTTPS encryption when only half of Australian and one-third of Thai sites use this technology.

4.2. RQ2: Do IoNI sites have SSL/TLS Certification?

Fig 6. Implementation of TLS Certification

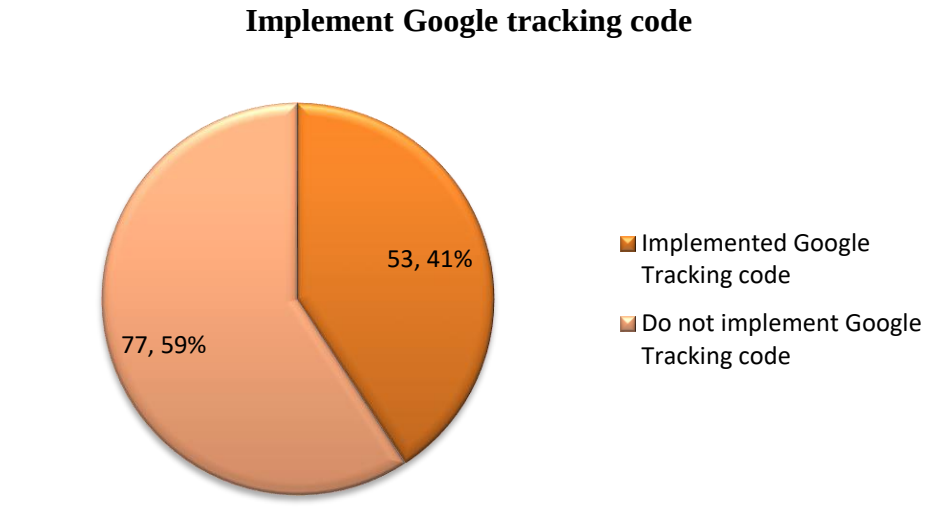


In Figure 6, the IoNI websites were also inspected for the presence of SSL/TLS (Secure socket layer/Transport layer security) enabled web application that provides public-key certification for authentication and establishes secure session between web servers and the Internet users. It was found that majority of the websites (N = 91, 70%) supported protection for the data sent via (HTTP) between users and web servers which is known as HTTP over TLS (HTTPS). This digital certificate ensures safe and secure transaction of data for the Internet users and prevents threats like man-in-the-middle attack (MITM) that is commonly known for attacking SSL/TLS enabled web applications (**Das & Samdaria, 2014**).

The 91 websites were then evaluated based on the version they used for Transport Layer Security (TLS) protocol. It was observed that majority of the IoNI websites (N = 62) still uses the older protocol TLS 1.2 whereas only (N = 29) uses the latest version TLS 1.3 which was standardized in August 2018 by the Internet Engineering Task Force (IETF). TLS 1.3 offer better security while TLS 1.2 suffer numerous attacks **(Arfaoui et al., 2019; Alqattaa & Aßmuth, 2019).**

4.3. RQ3: Do IoNI sites use Google Analytics and implement privacy protection parameters?

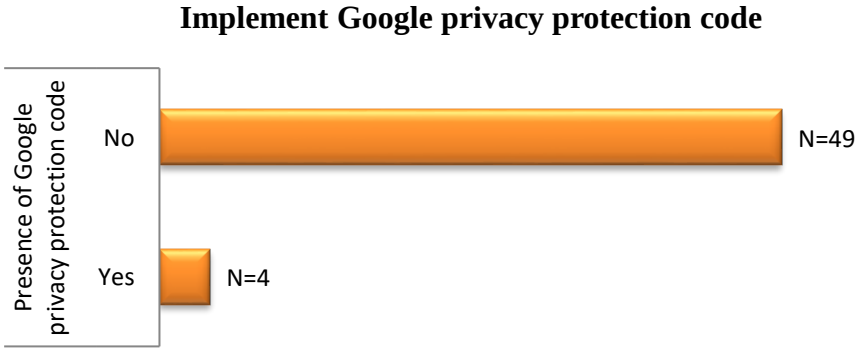
Fig 7. Implementation of Google tracking code



The study evaluates the implementation of Google analytics/Global site tag and/or Google tag manager in Figure 7. The results demonstrate 41 percent out of the 130 websites have implemented Google analytics/Global site tag and/or Google tag manager. Still a majority of 59 percent do not practice Google analytics for web tracking.

Early literature has covered a wide array of user privacy and the concerns with websites that implement Google Analytics (**Quintel & Wilson, 2020**). While Installing Google Analytics is useful as a search tool (**Farney, 2016**); proves beneficial for decisions making (**Fisher, 2018),** and is a helpful utility to improve campaign planning, strategic communication and message design **(Kent, Carr, Husted, & Pop, 2011).** However, it is not free from privacy violation due to the users' data being subjected to re-used for target advertising **(Binns & Bietti, 2020; Akiyama et al., 2017);** track IP addresses **(Kent, Carr, Husted, & Pop, 2011)** In addition to regular tracking practices, web tracking is extended when Google tracking services are included in Third-party scripts, and advertisers can then set global tracking identifiers which are possible to share data/track users uniquely across multiple websites (**Merzdovnik et al., 2017**; **O'Brien, Young, Arlitsch, & Benedict, 2018**).

Fig 8. Implementation of Google privacy protection parameters



From the 41 percent (N = 53) number of websites who have implemented Google analytics/Global site tag and/or Google tag manager; the results in Figure 8, demonstrate that only (N = 4) implement Google privacy protection code whereas majority of the websites (N = 49) do not enabled privacy protection code.

In the face of these results, it is clear that these 49 IoNI websites that practice web tracking must ensure the privacy protection of their users by not collecting information for personalization, ad targeting, collection of IP addresses and other personally identifiable information. **Obrien, Young, Arlitsch, & Benedict, (2018**) is also of the opinion to websites that implemented Google Analytics to have activate the available privacy-protection feature.

4.4. RQ4: Do IoNI sites have robust privacy policies in place?

Fig 9. Availability of privacy policy

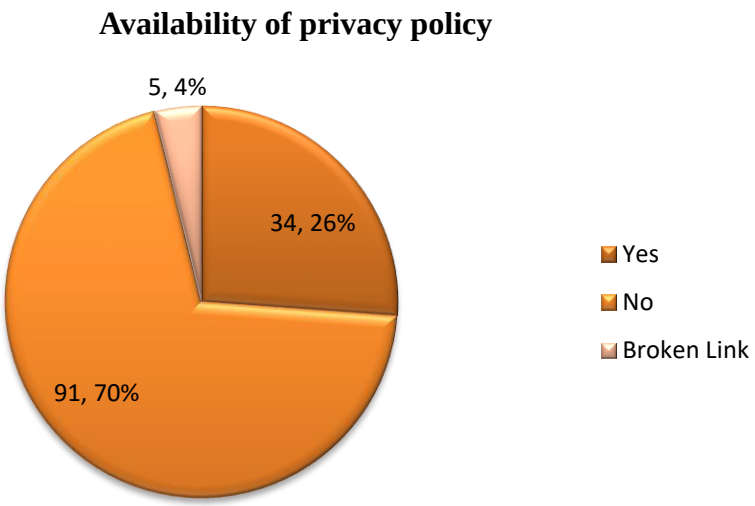
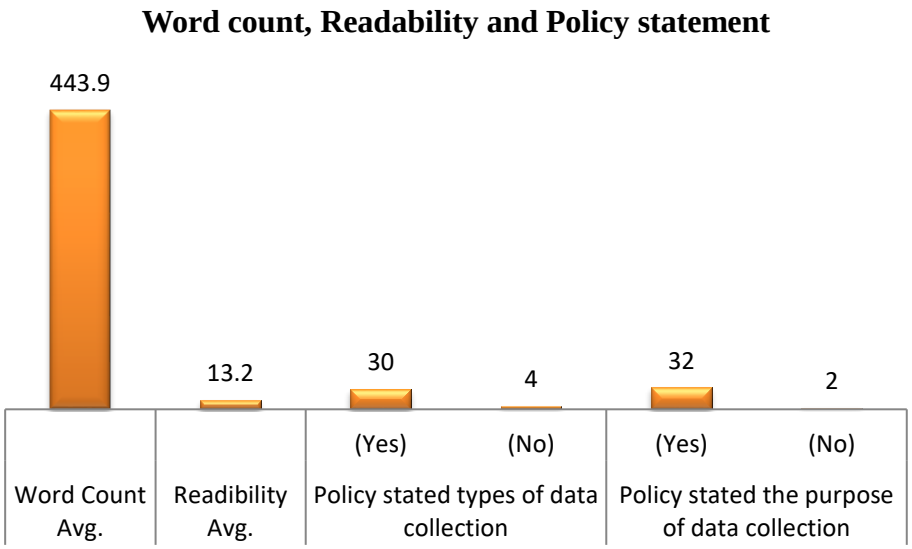


Figure 9 showed that only 26 percent (N = 34) have privacy policies while the majority of these websites, 70 percent do not have privacy policies in place with 4 percent of broken/inactive policy links. The lack of privacy policy in IoNI websites deprived the users' of Informed consent, thereby rendering users' unaware of how their personal information is collected, stored and processed.

The availability of privacy policy is crucial for a website because it informs users about the current practices relating to data collection, usage of data, control of data, protection and the usage of technologies such as website beacon and cookies **(Ali, Zaaba, Singh, & Hussain, 2020).** Present research showed that the lack of privacy policy in the website hinder a role in relaying important information between users and websites (**Brown, Ghani, Hoque, & Rehman, 2012)**. The important of privacy policy is laid down by

numerous literatures in the past studies; it's important is not just to detailed organisational practices regarding personal information **(Earp, Anton, Smith, & Stufflebeam, 2005**), but to instil trust in the website (**Chang, Wong, Libaque-Saenz, & lee, 2018**) and inform users regarding their choice towards privacy preferences (**Kaur et al., 2018).**

Fig 10. Word count, Readability and Policy statement

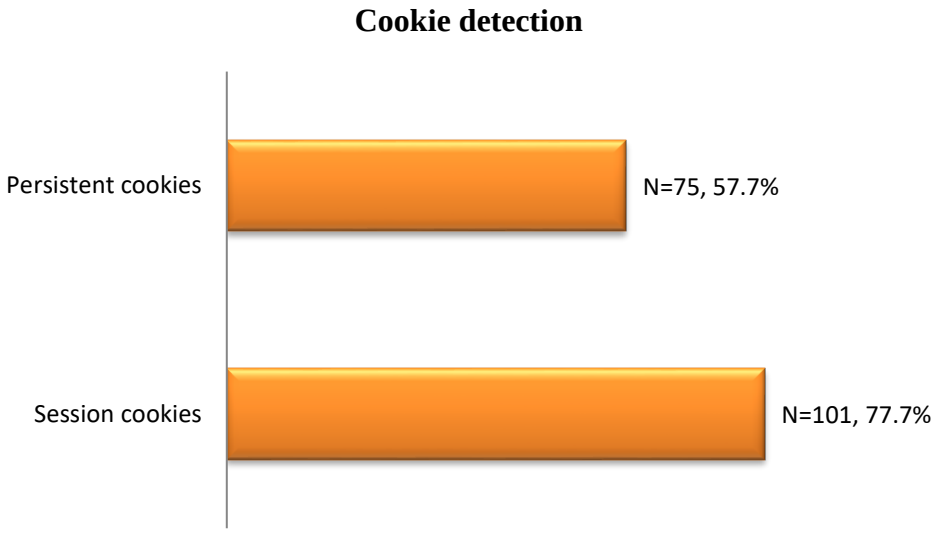


In Figure 10, further evaluation were undertaken to understand the policy statement of (N = 34, 26%). All statements are in English, and the average word is 443.9 of these 34 websites. The average readability score of these websites is considerably high, with a score of 13.2. According to the Flesch-Kincaid grade formula, a score from 12+ to 15 refers to texts that are 'very difficult' to read and are meant for college graduates and above (**Flesch, 1979**). For the presence of the term 'types of data collected,' in the IoNI policy statements, the Figure 10, showed that the majority (N = 30) mentioned the types of identifiable or non-identifiable collected for data processing. Also, the majority of the website (32 out of the 34 policy statements) includes the purpose of data collection.

The privacy policy is a statement of a website that mentioned how the personal information of site visitors is gathered, processed and disclosed to other parties for further uses. Even with the presence of privacy policies, the strict execution of policies is not guaranteed, so the 34 websites were analyzed further to understand their transparency and readability measurement from the text content perspective. The finding showed the privacy text to be 'very difficult,' to read and, as remarked by **Ali, Zaaba, Singh, and Hussain (2020)** privacy policy should be written in a clear and straightforward language.

4.5. RQ5: Do IoNI sites collect cookies and provide consent form?

Fig 11. Cookie detection



The IoNI website were then assessed for the acceptance of cookies. In Figure 11, session cookies on the users' site were detected from the Mozilla browser (N = 101, 77.7%). Persistent cookies from the first party/third party were also detected (N = 75, 57.7%). It was also found that none of the IoNI websites provides a cookie consent form or opt-in/opt-out option on their landing page. The absence of this feature violates the informed consent and users' ability to trust the website with their personal information.

Session cookie ends or gets deleted when the user exits the web browser whereas persistent cookie, also known as the transient cookie has an expiration date and stayed in the users' hard disk until it gets removed or reached the end of its expiration date. This evaluation is done to understand the prevalence of persistent cookies, which revealed that 57.7 percent of IoNI websites store persistent cookies. This practice is common for most websites (**Kaur et al., 2018**). While using cookies can be helpful and harmful simultaneously, designing effective cookie management schemes to balance between helpful and harmful cookies is quite challenging (**Yue, Xie, & Wang, 2010).** Also, receiving both technical and legal know-how empowers users to protect themselves against unnecessary cookies **(Strycharz, Smit, Helberger, & Noort, 2021).**

5. Recommendations

5.1. Implement secure HTTPS

When one visits a site via HTTPS, the URL looks like this https://iit.ac.in and, when one visit site via plain (unencrypted) HTTP, it looks like this http://iit.ac.in. Whenever a user connect to a website via (HTTPS), attackers find it difficult and cannot intercept the network to alter the data transfer over the connection. This level of privacy and secure connection is vital for the IoNI websites. The significance of HTTPS is applying TLS encryption over the HTTP

protocol to protect data shared via the connection from interception by eavesdroppers. TLS Protocol, according to **Alqattaa and Aßmuth (2019**), ensures three basic properties: Authenticating using asymmetric cryptography, e.g. RSA; Confidentiality where encryption is used for all messages after a simple handshake, after which the data is visible to the endpoints (end-to-end encryption) and, Integrity which ensures reliability where no attacker can modify any part of the communication without being detected by the endpoints. In many ways, the evolution from SSL, TLS to the design of TLS 1.3 transform real-world authenticated key exchange by employing modern cryptographic mechanisms. The key schedule algorithm that expands a relatively short master key is much more complex than in previous versions, respecting the paradigm of separating keys used at different layers and for different purposes (**Arfaoui et al., 2019**).

Even though Google over chrome is now encrypted by HTTPS upto 90%+ across Google chrome. Still, Google said in its chromium blog that there's a lot more to do to make HTTPS the preferred protocol on the web (**Google Chrome, 2021**). Therefore, implementing HTTPS is a vital privacy-protecting mechanism to securely connect to servers via HTTPS (**Obrien, Young, Arlitsch, & Benedict, 2018).** HTTPS is typically used to protect the leak of sensitive information such as Credit cards; Sensitive cookies such as PHP session cookies; Passwords and Usernames; Identifiable information (Social Security number, State ID numbers, etc) and Confidential content. For privacy reasons, security experts also urged web traffic to be routed via HTTPS **(Drupal, 2021).** The implementation of HTTPS is also possible through Let's Encrypt, a free and open certificate authority (CA) provided by Internet Security Research Group (ISRG) who issues approximately 200 million certificates daily to secure web traffic via HTTPS (**Aas, 2021**).

5.2. Reduce unnecessary web tracking

Tracking types include analytics, usability tracking, personalised advertising tracking, and cookies tracking **(Pilton, Faily, & Henriksen-Bulmer, 2021).** This is accomplished by various techniques, including tracking cookies, pixel tags, beacons, and other sophisticated mechanisms (**Cook, Nithyanand, & Shafiq, 2020).** Other tracking methods called device fingerprinting use the browser's unique configurations and settings to track users' activity. Companies also use techniques to connect users' identities to the different devices use online and then cutomized ads across all devices (**Federal Trade Commission, 2021). The** use of web analytics and cookies allow web admin to record users browsing behaviour mainly for third party usage. Mergers between firms who practice third party tracking raises privacy concerns which is a challenge for regulators according to **Binns and Bietti (2020**), and this third party tracking is commonly widespread via advertisers, social media widgets in many websites or embedded directly in first party website **(Lerner, Simpson, Kohno, & Roesner, 2016).** Hence, it is recommended to reduce tracking by reducing association and integrating services with marketing agencies or third-party organizations.

For example, Social media track users via 'share' and 'like' buttons presented as widgets in many websites; Google Analytics is by far the commonly used analytics tool for websites, where up to 70% of the top 1 million sites globally, enabled Google to get an insight into the users behaviour over the internet (**Cookiebot, 2020**). On the other hand, Web analytics is designed to leak clients' personal information and its tracking capacity is enhanced when combined with other tools. For example, Google analytics is enhanced when combined with

Google AdSense, Google's popular cross-site advertising service. Tracking has severe implications on individuals' privacy but can be reduced by avoiding merging with third-party trackers. Employing techniques like anonymizing the IP addresses and other protection parameters also ensures privacy protection to Internet users (**O'brien, Young, Arlitsch, & Benedict, 2018).**

Tracking cookies are both beneficial and harmful (**Yue, Xie, & Wang, 2010**). Many cookies, marketing cookies notoriously track data about users, such as their IP addresses and their browsing activity. These, along with analytics cookies can be called tracking cookies (**Cookiebot, 2020**). Though it's been a challenge to design effective cookie management schemes (**Yue, Xie, & Wang, 2010).** It is recommended for the website owner/admin to help users make useful decision in accepting and rejecting cookie to reduce tracking. After all, regulation under the European Union through the General Data Protection Regulation (GDPR) has laid out tracking cookies regulations that allow cookies to be set with users consent only. This means that organizations can no longer track users without users' consent, i.e., without users acknowledging and accepting data collection and processing.

5.3. Simple and transparent privacy policy

Website's Privacy Policies aim to inform and educate users of a website's practices and processes personal data, its usage, how personal data is exchange between first and third party, how to control the flow and protection of data, and the use of application for data collection (e.g., website beacon and cookies) whenever users' visit the website **(Ali, Zaaba, Singh, & Hussain, 2020)**. A survey data of **Earp, Anton, Smith, and Stufflebeam (2005)** showed that individuals are more concerned with how their personal data is shared, lend, sold to other parties and wanted to have complete information in the form of a privacy notice stating how the website might use their data. Earp et al., also pointed out privacy policy as a signal that informs users about information that are important to them, thus informing and assuring security of the personal data collected and transferred over the web (**Earp, Anton, Smith, & Stufflebeam, 2005).**

Another suggestion to privacy statement by **Kaur et al. (2018)** is for the privacy policy to use fewer ambiguous words and focus on words like cookies that comply with the regional regulations. The words used in all policy statements should be consistent regardless of the domain and regulations, which in turn enhanced the readability of privacy policy statements when used with simple and clear text to read (**Javed, Salehin, & Shehab, 2020; Ali, Zaaba, Singh, & Hussain, 2020**). This will also instil trust in the website (**Chang, Wong, Libaque-Saenz, & lee, 2018**) and inform users regarding their choice of privacy preferences (**Kaur et al., 2018). Javed, Salehin, and Shehab (2020)** also give suggestions to improve the policy accessibility by embedding a link to privacy statements on the landing page; include protection to children data; privacy statements to be written with a reading grade level of 8 or less; statements in the native language, implement all privacy principles in compliance with regional regulations and provide contact information to address privacy grievances.

5.4. Awareness

With Internet services increasing access to personal data, raising user awareness about what privacy guarantees websites offer is crucial in today's world—finding revealed that users are less aware of privacy seals and cookies stored in their browser (**Vakeel, Das, Udo, & Baghi, 2017**) & (**Harding, Reed, & Gray, 2006**). Therefore, **Cai, Gantz, Schwartz, and Wang (2003**) conclude that parents and schools have to educate children to protect their privacy online and, once users become more aware of how service provider treats their data, they can demand that these services become more sensitive to online privacy (**Bergram et al., 2020)**. Some ways to create awareness are having users attend a course and read documents, follow good practices and know the 'Dos & Don'ts' on protecting their privacy (**Bhardwaj et al., 2020**).

Addressing privacy awareness is also the responsibility of the website owner or administrator. As pointed by **Miller and Wells (2007**), it is the management's responsibility to be aware and make amends appropriately. **Earp, Anton, Smith, and Stufflebeam (2005)** further suggests that organizations be aware and in control of the actual practices displayed in their privacy policy. Therefore, website owners and governments must be aware of their TLS objectives and the possible harm from unauthenticated blocking pages **(Alashwali, Szalachowski, & Martin, 2020)**. Enterprises need to be aware of the risk and be suitably concerned if they are not decrypting and inspecting SSL traffic from untrusted sources and IoT devices (**Omar Yaacoubi, 2019**). These trends demand greater awareness among website owners (**Javed, Salehin, & Shehab 2020**). These requirements might not only shape privacy-enhancing practices regarding a single use of data but rather make organizational changes that would effectively raise organizations' awareness and protection levels (**Harber, 2020)**.

Awareness of collecting, processing, sharing, and storing personal data strongly affects trusting the website **(Frik & Mittone, 2019)**. Though many organizations started awareness training on cyber-attack (**Bhardwaj et al., 2020)**, can still bypass the defences in many ways. To avoid the excessive invasion of personal privacy on the web might rest with individuals and organizations who are highly aware, well informed and adequately trained.

5.5. Informed consent

Every user generally has the right to privacy and intimacy, but more importantly, there shall be no encroachment into the private spaces of an individual without his/her consent (**Harber, 2020)**. To consent to something, one must be able to understand what one consent to and why? Sometimes, the website we entrust our personal data and give our trust is the first to manipulate and sell our data without our consent. Furthermore, according to **Ari et al., (2019)**, eavesdropping and monitoring without the observed person's consent and knowledge is another privacy invasion. After all, Informed consent is ones' right according to the European Union under the GDPR regulation. This right enables users' to remove consent, the right to be forgotten, and the right to remove users' data from the databases (**Slepchuk & Milne, 2020)**. Even though finding in a study like **Benjumea et al. (2020)** reported informing users about the right to give and remove consent, for example, consenting or declining the usage of specific cookies **(Schmidt, Bornschein, & Maier, 2020).** The present finding in all the IoNI website does not practice informed consent via cookie banners or opt-out notice which is why it is necessary to include information in a layered form, i.e. a link about the company's identity; purpose for which the cookies were stored; a classification of the

collected data, and information about the possibility to withdraw consent (**Bauer, Bergstrom, & Foss-Madsen, 2021)**.

Consent also comes at a cost due to market equilibrium characterized by excessive consent to collect data, resulting in excessive loss of consumer privacy (**Choi, Jeon, & Kim, 2019**). The excessive data collection in a different form (e.g., processed, analyzed, presented and shared in a system) makes it difficult for the Internet users to give their consent **(Ari et al., 2019).** Therefore, the use of transparent and user-friendly browsers or applications is highly recommended **(Rossi, 2021)**. Also, having a good contrast privacy policy in place serves as informed consent to users' visiting the site and enables users to choose which information to share with a particular site visited **(Schmidt, Bornschein, & Maier, 2020).** Thus, to efficiently preserve privacy in a system, informed consent is a significant requirement for collecting, storing and processing personal information.

5.6. Promote the use of Privacy protection tools/techniques

To provide better security and privacy improvements to the users is by suggesting the use of Privacy-enhancing technologies (PET) to secure connections between Internet users and the websites. The use of PET may guarantee safety in the online environment, and it is vital to cyber groom students by teaching essential protection technologies and techniques.

Whereby from the users' perspective, the use of the following tools and techniques comes highly recommend such as VPNs (**Kakatkar & Spann, 2019**); browser extension and plugins like Adblock, Ghostery, Adblocker, NoTrace, uBlock origin **(Melicher et al., 2016; Merzdovni et al., 2017; Mazel, Garnier, & Fukuda, 2019**; **Malandrino & Scarano, 2013);** HTTPS Everywhere plugin **(Electronic Frontier Foundation, 2021)**; check for privacy badger (**Mazel, Garnier, & Fukuda, 2019**); anonymous browsing (**Yang et al., 2019**) like Tor browser **(Al-Shehari & Zhioua, 2018).** Other studies promote the use of anonymous search engines, email encryption, or disconnecting social media tracking plugins, removing cookies, clear browsing history and block unwanted emails (**Huang & Bashir., 2018**); setting up browser preference to 'Do Not Track,' (**Malandrino & Scarano, 2013**; **Melicher et al., 2016).** Finally, opt-out of targeted advertising by considering an ad blocker, opt-out from websites that trade personal information, verifying one's identity and make prudent decisions before sharing information like credit, employment, insurance, housing and marketing products (**Federal Trade Commission, 2021).** According to cybercrime.gov, basic protection techniques includes the following example: learning to block and remove someone not comfortable; be selective about accepting friends on social media; remember to log out and use a strong password; be mindful of audio/video featuring self before sharing online; avoid taking/storing sensitive photos/video with mobile phones; disable cyberstalking by disable location service to social media and applications/websites or talk to friends and relative if one is a victim of cyber stalking and report visually disturbing images/videos like sexual abuse/child pornography to the concerned social media website (**Cybercrime, 2021).**

Privacy protection technologies have been designed to protect users from web tracking. Some of the technologies recommended from organizations perspective include encrypted processing data, obfuscation, anonymization, sticky policy, trusted platform module, data segmentation and trusted third party mediator (**Ari et al., 2019**); TLS encrypted connection **(Naylor et al., 2014);** cookie-consent notification and opt-out policy **(Obrien,**

Young, Arlitsch, & Benedict, 2018; McCarthy & Yates, 2010; Schmidt, Bornschein, & Maier, 2020); request domain to HTTPS-enforcement such as the HTTPS Everywhere **(Electronic Frontier Foundation, 2021).** Finally, the focus should be on Privacy and Security by design, reflecting a holistic approach to privacy at an organizational or enterprise level (**Harber & Tamo-Larrieux, 2020**).

5.7. Regulatory aspects

Even with the Indian Cyber Crime Coordination Centre (14C) scheme of the Ministry of Home Affairs, the Government of India focuses on strengthening the capacity of reporting, prevention, detection, investigation, training, research or the National Cyber Crime Reporting Portal (https://cybercrime.gov.in) to tackle cybercrime in India (**Ministry of home affairs, 2021**). The implementations of such initiatives are still lagging behind and needs to be popularized at the schools and college level to create extensive awareness and practical know-how.

India, via the Ministry of Electronic and Information Technology, intends to implement the Data Protection Bill 2018 **(Ministry of Electronics & Information Technology, 2021**), which the GDPR inspired. The bill covers principles such as data collection; data retention; collection notification; data sharing; sensitive data processing; user control; user access; security standards; quality control; grievance redressal, and accountability **(Javed, Salehin, & Shehab, 2020).** Even with the recent introduction in the parliament, the bill still lacks many necessary safeguards due to the dilute right to privacy and increased state power to surveillance which poses a significant concern **(Waris, 2021)**. This current bill is not likely to protect privacy adequately, which is why according to the **National Law Review (2021)**, the future revised bill is expected to witness changes in the regulations before the subsequent introduction in the assembly, and changes will be seen in industry-specific data policies; drone-related policies and new issues in cyber security and mandatory disclosure to the Government. From here, it is clear that the judiciary is much aware, and the recognition towards privacy right is more apparent than ever, but keeping in mind the ever-changing landscape of privacy and security issues is the one where the regulators will have to tackle ahead.

6. Limitation & Future work

Although the assessment and content analysis process is time-consuming, it is done to achieve the most accurate results from the study. In this work, the researcher only assessed the HTTPS ecosystem of the Indian Institute of National Importance (IoNI) under the Govt. of India and found that majority of these sites are partially secure but at the same time majority of these sites used TLS 1.2, which is an older version. The study only assessed the presence of Google Analytics tracking code in the landing page under the HTML code and found that the users' tracking is not extensive overall. The study only consider the Google tracking code and this limit the study because the presence of one or two third-party tracking scripts was also discovered in the study but was not taken into account due to the scope of the current study. Further, a different sample such as assessment of other public and private universities or comparative studies may yield different results, thereby limiting the present

study attempt to generalize the findings. To what extends users know and have the technical know-how to protect themselves from new ways that collect their personal information provides new context to further study. To better understand the privacy policies, the study analyzed the availability, frequency words, inclusion of data collection or types of data collected, and the policy readability score. To researcher understanding, this proves to be the most comprehensive analysis of privacy statements. However, further studies can look into in-depth keywords analysis, confidentiality agreements and other legal aspects of privacy statements. Finally, this study is just the beginning to understanding the security and privacy of academic websites in India and looks forward to researchers exploring other areas in related fields.

7. Conclusion

Academic institutions are secular and neutral ground with a long-held value for intellectual freedom; therefore, institutions utmost duty is to protect users' privacy on the web. Higher educational institutions are custodians of Intellectuals with maximum research outcomes, yet the actual privacy and security practices of most IoNI websites are not up to the mark. The empirical findings revealed that many IoNI websites are not secure enough. Even though tracking users is at its initial stage in these sites, the practice to enabled privacy protection features such as blocking Google tracking script, up-gradation to TLS 1.3, availability of privacy policy is comparatively low. The finding shows that IoNI privacy policies are not clear enough for the users to understand; they lack informed consent and do not provide opt-out control by giving users more freedom in sharing their personal information. Ultimately, showing the level of carelessness and lack of priority when it comes to protecting users privacy. In recent years, the knowledge of security flaws and how to exploit them has grown; finally, it is time to update their security and privacy features by following the study recommendation to reduce unnecessary tracking.

Reference

Aas, J. (2021). Preparing to Issue 200 Million Certificates in 24 Hours. Retrieved August 3, 2021, from https://letsencrypt.org/2021/02/10/200m-certs-24hrs.html

Ahaskar, A. (2021). Over 1000 Indian schools, colleges targeted in cyberattacks in Jun-Sep. Retrieved August 31, 2020, from https://www.livemint.com/technology/tech-news/over-1-000-indian-schools-colleges-targeted-in-cyberattacks-in-jun-sep-report-11604044942722.html

Akiyama, M., Yagi, T., Yada, T., Mori, T., & Kadobayashi, Y. (2017). Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots. *Computers and Security*, *69*, 155–173. https://doi.org/10.1016/j.cose.2017.01.003

Al-Shehari, T., & Zhioua, S. (2018). An empirical study of web browsers' resistance to traffic analysis and website fingerprinting attacks. *Cluster Computing*, *21*(4), 1917–1931. https://doi.org/10.1007/s10586-018-2817-4

Alashwali, E., Szalachowski, P., & Martin, A. (2020). Exploring HTTPS security inconsistencies: A cross-regional perspective. *Computers & Security*, *97*, 101975. https://doi.org/10.1016/j.cose.2020.101975

Ali, A. S., Zaaba, Z. F., Singh, M. M., & Hussain, A. (2020). Readability of websites security privacy policies: A survey on text content and readers. *International Journal of Advanced Science and Technology*, *29*(6), 1661–1672. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/9315

Alqattaa, A., & Aßmuth, A. (2019). Analysis of the Internet Security Protocol TLS Version 1.3. Retrieved from https://www.researchgate.net/publication/341680184_Analysis_of_the_Internet_Security_Protocol_TLS_Version_13

Arfaoui, G., Bultel, X., Fouque, P., Nedelcu, A., & Onete, C. (2019). The privacy of the TLS 1.3 protocol. *Proceedings on Privacy Enhancing Technologies*, *2019*(4), 190–210. https://doi.org/10.2478/popets-2019-0065

Ari, A. A. A., Ngangmo, O. K., Titouna, C., Thiare, O., Kolyang, Mohamadou, A., & Gueroui, A. M. (2020). Enabling privacy and security in Cloud of Things: Architecture, applications, security & privacy challenges. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2019.11.005

Bashir, M. A., Arshad, S., & Robertson, W. (2016). Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. In *Proceedings of the 25th USENIX Security Symposium* (pp. 481–496). https://doi.org/10.5555/3241094.3241132

Bauer, J. M., Bergstrom, R., & Foss-madsen, R. (2021). Are you sure, you want a cookie ? – The effects of choice architecture on users ' decisions about sharing private online data. *Computers in Human Behavior*, *120*, 106729. https://doi.org/10.1016/j.chb.2021.106729

Benjumea, J., Ropero, J., Rivera-Romero, O., Dorronzoro-Zubiete, E., & Carrasco, A. (2020). Assessment of the fairness of privacy policies of mobile health apps: Scale development and evaluation in cancer apps. *JMIR MHealth and UHealth*, *8*(7), 1–20. https://doi.org/10.2196/17134

Bergram, K., Gjerlufsen, T., Maingot, P., Bezencon, V., & Holzer, A. (2020). Digital Nudges for Privacy Awareness: From consent to informed consent? In *Proceedings of the 28th European Conference on Information Systems (ECIS)* (pp. 1–16). Retrieved from https://aisel.aisnet.org/ecis2020_rp/64

Bhardwaj, A., Sapra, V., Kumar, A., Kumar, N., & Arthi, S. (2020). Why is phishing still successful? *Computer Fraud and Security*, (9), 15–19. https://doi.org/10.1016/S1361-3723(20)30098-1

Binns, R., & Bietti, E. (2020). Dissolving privacy, one merger at a time: Competition, data and third party tracking. *Computer Law and Security Review*, *36*, 1–19. https://doi.org/10.1016/j.clsr.2019.105369

Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, *50*(1), 65–82. https://doi.org/10.1023/A:1005642218907

Brown, J. D., Ghani, M. A., Hoque, M., & Rehman, U. A. (2012). *An Analysis of Web*

Privacy Policies Across Industries. *Worchester Polytechnic Institute*. Retrieved from https://digitalcommons.wpi.edu/iqp-all/3295

Bujlow, T., Carela-español, V., Solé-Pareta, J., & Barlet-Ros, P. (2017). A Survey on Web Tracking: Mechanisms, Implications, and Defenses. *Proceedings of the IEEE*, *105*(8), 1476–1510. https://doi.org/10.1109/JPROC.2016.2637878.

Cai, X., Gantz, W., Schwartz, N., & Wang, X. (2003). Children's website adherence to the FTC's online privacy protection rule. *Journal of Applied Communication Research*. https://doi.org/10.1080/1369681032000132591

Cha, J. (2011). Information privacy: a comprehensive analysis of information request and privacy policies of most-visited Web sites. *Asian Journal of Communication*, *21*(6), 613–631. https://doi.org/10.1080/01292986.2011.615942

Chang, Y., Wong, S. F., Libaqye-Saenz, C. F., & Lee, H. (2018). The role of privacy policy on consumers' perceived privacy. *Government Information Quarterly*, *35*(3), 445–459. https://doi.org/10.1016/j.giq.2018.04.002

Choi, J. P., Jeon, D.-S., & Kim, B.-C. (2019). Privacy and Personal Data Collection with Information Externalities. *Journal of Public Economics*, *173*, 113–124. https://doi.org/https://doi.org/10.1016/j.jpubeco.2019.02.001

Cook, J., Nithyanand, R., & Shafiq, Z. (2020). Inferring Tracker-Advertiser Relationships in the Online Advertising Ecosystem using Header Bidding. In *Proceedings on Privacy Enhancing Technologies* (pp. 65–82). https://doi.org/10.2478/popets-2020-0005

Cookiebot, 2021. (2020). Tracking cookies and the GDPR. Retrieved August 3, 2021, from https://www.cookiebot.com/en/tracking-cookies/

Cozza, F., Guarino, A., Isernia, F., Malandrino, D., Rapuano, A., Schiavone, R., & Zaccagnino, R. (2020). Hybrid and lightweight detection of third party tracking: Design, implementation, and evaluation. *Computer Networks*, *167*, 106993. https://doi.org/10.1016/j.comnet.2019.106993

Cybercrime. (2021). Online Safety Tips. Retrieved July 28, 2021, from https://www.cybercrime.gov.in/Webform/Crime_OnlineSafetyTips.aspx

Das, M. L., & Samdaria, N. (2014). On the security of SSL/TLS-enabled applications. *Applied Computing and Informatics*, *10*(1–2), 68–81. https://doi.org/10.1016/j.aci.2014.02.001

Dinev, & Hart. (2006). Privacy Concerns and Levels of Information Exchange: An Empirical Investigation of Intended e-Services Use. *E-Service Journal*, *4*(3), 25. https://doi.org/10.2979/esj.2006.4.3.25

Drupal. (2021). Enabling HTTP Secure (HTTPS). Retrieved August 3, 2021, from https://www.drupal.org/https-information

Earp, J. B., Antón, A. I., Aiman-Smith, L., & Stufflebeam, W. H. (2005). Examining Internet privacy policies within the context of user privacy values. *IEEE Transactions on Engineering Management*, *52*(2), 227–237. https://doi.org/10.1109/TEM.2005.844927

Electronic Frontier Foundation, 2021. (2021). HTTPS Everywhere. Retrieved August 3,

2021, from https://www.eff.org/https-everywhere

Farney, T. (2016). *Google Analytics and Google Tag Manager*. *Library Technology Reports: Expert Guides to Library Systems and Services*. Chicago. Retrieved from https://www.alastore.ala.org/content/google-analytics-and-google-tag-manager

Federal Trade Commission, 2021. (2021). How To Protect Your Privacy Online. Retrieved August 3, 2021, from https://www.consumer.ftc.gov/articles/how-protect-your-privacy-online

Fisher, J. (2018). Managing external links on the Leeds Beckett University Library website: strategies and approaches. *Collaborate: Libraries in Learning Innovation*, (3). Retrieved from https://ojs.leedsbeckett.ac.uk/index.php/COL/article/view/4556

Flesch, R. (1979). *How to Write Plain English: A Book for Lawyers & Consumers*. Newyork: Harper & Row.

Frik, A., & Mittone, L. (2019). Factors Influencing the Perception of Website Privacy Trustworthiness and Users' Purchasing Intentions: The Behavioral Economics Perspective. *Journal of Theoretical and Applied Electronic Commerce Research*, *14*(3), 89–125. https://doi.org/10.4067/s0718-18762019000300107

García, M. del M. R., García-Nieto, J., & Aldana-Montes, J. F. (2016). An ontology-based data integration approach for web analytics in e-commerce. *Expert Systems with Applications*, *63*, 20–34. https://doi.org/10.1016/j.eswa.2016.06.034

Google Analytics. (2021a). Add gtag.js to your site. Retrieved April 12, 2021, from https://developers.google.com/analytics/devguides/collection/gtagjs

Google Analytics. (2021b). The Google Analytics tag. Retrieved April 4, 2021, from https://developers.google.com/analytics/devguides/collection/analyticsjs

Google Analytics. (2021c). Measure pageviews. Retrieved April 10, 2021, from https://developers.google.com/analytics/devguides/collection/gtagjs/pages

Google Analytics. (2021d). IP anonymization with gtag. Retrieved April 10, 2021, from https://developers.google.com/analytics/devguides/collection/gtagjs/ip-anonymization

Google Analytics. (2021e). Disable Google Analytics measurement. Retrieved April 10, 2021, from https://developers.google.com/analytics/devguides/collection/gtagjs/user-opt-out

Google Analytics. (2021f). Disable Advertising Features. Retrieved April 10, 2021, from https://developers.google.com/analytics/devguides/collection/gtagjs/display-features

Google Chrome, 2021. (2021). Increasing HTTPS adoption. Retrieved August 3, 2021, from https://blog.chromium.org/2021/07/increasing-https-adoption.html

Google Tag Manager, 2021. (2021). Quick Start Guide. Retrieved April 5, 2021, from https://developers.google.com/tag-manager/quickstart

Haber, E., & Tamò-Larrieux, A. (2020). Privacy and security by design: Comparing the EU and Israeli approaches to embedding privacy and security. *Computer Law and Security Review*, *37*, 1411–1412. https://doi.org/10.1016/j.clsr.2020.105409

Harding, W. T., Reed, A. J., & Gray, R. L. (2006). Cookies and web bugs: What they are and how they work together. *Information Systems Management*, *18*(3), 17–24. https://doi.org/10.1201/1078/43196.18.3.20010601/31286.3

Hormozi, A. M. (2005). Cookies and privacy. *Information Systems Security*, *13*(6), 51–59. https://doi.org/10.1201/1086/44954.13.6.20050101/86221.8

Huang, H. Y., & Bashir, M. (2018). Surfing safely: Examining older adults' online privacy protection behaviors. *Proceedings of the Association for Information Science and Technology*, *55*(1), 188–197. https://doi.org/10.1002/pra2.2018.14505501021

Huang, J. K., Zhang, Z. X., Li, W. J., & Xin, Y. (2019). Assessment of the impacts of TLS vulnerabilities in the HTTPS ecosystem of China. In *2018 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI 2018* (Vol. 147, pp. 512–518). https://doi.org/10.1016/j.procs.2019.01.238

Jalal, S. K., Biswas, S. C., & Mukhopadhyay, P. (2009). Webometric analysis of Central Universities in India: A study. In *International Conference for Internet Technology and Secured Transactions, ICITST 2009* (pp. 1–9). https://doi.org/10.1109/icitst.2009.5402605

Jamin, J., Arifin, N. A. M., Mokhtar, S. A., Rosli, N. N. I. N., & Shukry, A. I. M. (2019). Privacy oncern of personal information in the ICT usage, Internet and Social Media perspective. *Malaysian E Commerce Journal (MECJ)*, *3*(2), 15–17. https://doi.org/doi.org/10.26480/mecj.02.2019.15.17

Javed, Y., Salehin, K. M., & Shehab, M. (2020). A Study of South Asian Websites on Privacy Compliance. *IEEE Access*, *8*, 156067–156083. https://doi.org/10.1109/ACCESS.2020.3019334

Jeyshankar, R., & B Ramesh, B. (2009). Websites of universities in Tamil Nadu: a webometric study. *Annals of Library and Information Studies (ALIS)*, *56*(2), 69–79. Retrieved from http://nopr.niscair.res.in/handle/123456789/5939

Jun, S. P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, *130*, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009

Kakatkar, C., & Spann, M. (2019). Marketing analytics using anonymized and fragmented tracking data. *International Journal of Research in Marketing*, *36*(1), 117–136. https://doi.org/10.1016/j.ijresmar.2018.10.001

Katuu, S. (2018). Using Web Analytics to Assess Traffic to the Mandela Portal: The Case of African Countries. *New Review of Information Networking*, *23*(1–2), 1–18. https://doi.org/10.1080/13614576.2018.1523741

Kaur, J., Dara, R. A., Obimbo, C., Song, F., & Menard, K. (2018). A comprehensive keyword analysis of online privacy policies. *Information Security Journal*, *27*(5–6), 260–275. https://doi.org/10.1080/19393555.2019.1606368

Kent, M. L., Carr, B. J., Husted, R. A., & Pop, R. A. (2011). Learning web analytics: A tool for strategic communication. *Public Relations Review*, *37*(5), 536–543.

https://doi.org/10.1016/j.pubrev.2011.09.011

Khormali, A., Park, J., Alasmary, H., Anwar, A., Saad, M., & Mohaisen, D. (2021). Domain name system security and privacy: A contemporary survey. *Computer Networks*. Elsevier B.V. https://doi.org/10.1016/j.comnet.2020.107699

Leiva, L. A., & Huang, J. (2015). Building a better mousetrap: Compressing mouse cursor activity for web analytics. *Information Processing and Management*, *51*(2), 114–129. https://doi.org/10.1016/j.ipm.2014.10.005

Lerner, A., Simpson, A. K., Kohno, T., & Roesner, F. (2016). Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study ofWeb Tracking from 1996 to 2016. In *Proceedings of the 25th USENIX Security Symposium* (pp. 997–1013). Retrieved from https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/lerner

Liu, J., & Zhang, T. (2021). Cross-device User Tracking via Hybrid Model. In *Procedia Computer Science- International Conference on Identification, Information and Knowledge in the internet of Things, 2020* (Vol. 187, pp. 83–88). Elsevier B.V. https://doi.org/10.1016/j.procs.2021.04.036

Lorentzen, D. G. (2014). Webometrics benefitting from web mining? An investigation of methods and applications of two research fields. *Scientometrics*, *99*(2), 409–445. https://doi.org/10.1007/s11192-013-1227-x

Lugosi, P. (2008). Covert research. In *Encyclopedia of Qualitative Research Methods* (Vol. 2, pp. 133–136). Retrieved from https://www.academia.edu/5786854/Covert_research

Malandrino, D., & Scarano, V. (2013). Privacy leakage on the Web: Diffusion and countermeasures. *Computer Networks*, *57*(14), 2833–2855. https://doi.org/10.1016/j.comnet.2013.06.013

Martin, K. (2020). Breaking the Privacy Paradox: The Value of Privacy and Associated Duty of Firms. *Business Ethics Quarterly*, *30*(1), 65–96. https://doi.org/10.1017/beq.2019.24

Mazel, J., Garnier, R., & Fukuda, K. (2019). A comparison of web privacy protection techniques. *Computer Communications*, *144*, 162–174. https://doi.org/10.1016/j.comcom.2019.04.005

McCarthy, L., & Yates, D. (2010). The use of cookies in Federal agency web sites: Privacy and recordkeeping issues. *Government Information Quarterly*, *27*(3), 231–237. https://doi.org/10.1016/j.giq.2010.02.005

Melicher, W., Sharif, M., Tan, J., Bauer, L., Christodorescu, M., & Leon, P. G. (2016). (Do Not) Track Me Sometimes: Users' Contextual Preferences for Web Tracking. In *Proceedings on Privacy Enhancing Technologies* (pp. 135–154). https://doi.org/10.1515/popets-2016-0009

Merzdovnik, G., Huber, M., Buhov, D., Nikiforakis, N., Neuner, S., Schmiedecker, M., & Weippl, E. (2017). Block Me if You Can: A Large-Scale Study of Tracker-Blocking Tools. In *Proceedings - 2nd IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 319–333). https://doi.org/10.1109/EuroSP.2017.26

Miller, C., & Stuart Wells, F. (2007). Balancing Security and Privacy in the Digital

Workplace. *Journal of Change Management*, *7*(3–4), 315–328. https://doi.org/10.1080/14697010701779181

Ministry of Education, 2021. (2021). Institutions of National Importance. Retrieved April 4, 2021, from https://www.education.gov.in/en/institutions-national-importance

Ministry of Electronics & Information Technology. (2021). Data Protection Framework. Retrieved July 20, 2021, from http://meity.gov.in/data-protection-framework

Ministry of Home Affairs, 2021. (2021). Details about Indian Cybercrime Coordination Centre (I4C) Scheme. Retrieved July 19, 2021, from https://www.mha.gov.in/division_of_mha/cyber-and-information-security-cis-division/Details-about-Indian-Cybercrime-Coordination-Centre-I4C-Scheme

Muruganandham, G. (2019). Webometrics research methods adopted in library and information science: An overview. *Library Philosophy and Practice*, (2869). Retrieved from https://digitalcommons.unl.edu/libphilprac/2869/

Nangia, V. (2020). Institutes of National Importance (INI), Institutions of Eminence (IoE) and National Education Policy (NEP) 2020. Retrieved July 14, 2021, from https://www.linkedin.com/pulse/institutes-national-importance-ini-institutions-eminence-nangia

Narayanan, A., & Reisman, D. (2017). The Princeton Web Transparency and Accountability Project. In T. Cerquitelli, D. Quercia, & F. Pasquale (Eds.), *Transparent Data Mining for Big and Small Data. Studies in Big Data* (pp. 45–67). Springer. https://doi.org/https://doi.org/10.1007/978-3-319-54024-5_3

National Law Review, T. (2021). Privacy and Data Protection – India Wrap 2020. Retrieved July 2, 2021, from https://www.natlawreview.com/article/privacy-and-data-protection-india-wrap-2020

Naylor, D., Finamorey, A., Leontiadisz, I., Grunenbergerz, Y., Mellia, M., Munafòy, M., … Steenkiste, P. (2014). The cost of the "S" in HTTPS. In *CoNEXT 2014 - Proceedings of the 2014 Conference on Emerging Networking Experiments and Technologies* (pp. 133–139). https://doi.org/10.1145/2674005.2674991

Netcraft. (2021). June 2021 Web Server Survey. Retrieved June 29, 2021, from https://news.netcraft.com/archives/2021/06/29/june-2021-web-server-survey.html

OBrien, P., W.H. Young, S., Arlitsch, K., & Benedict, K. (2018). Protecting privacy on the web: A study of HTTPS and Google Analytics implementation in academic library websites. *Online Information Review*, *42*(6), 734–751. https://doi.org/10.1108/OIR-02-2018-0056

Omar Yaacoubi, B. (2019). The rise of encrypted malware. *Network Security*. Elsevier Ltd. https://doi.org/10.1016/S1353-4858(19)30059-5

Pal, A., Kar, S., & Sardar, S. (2020). Webometric Analysis of ICSSR Sponsored Research Institutions in India. *Library Philosophy and Practice*, (3804). Retrieved from https://digitalcommons.unl.edu/libphilprac/3804/

Pilton, C., Faily, S., & Henriksen-Bulmer, J. (2021). Evaluating privacy - determining user privacy expectations on the web. *Computers and Security*, *105*, 102241.

https://doi.org/10.1016/j.cose.2021.102241

Potoglou, D., Dunkerley, F., Patil, S., & Robinson, N. (2017). Public preferences for internet surveillance, data retention and privacy enhancing services: Evidence from a pan-European study. *Computers in Human Behavior*, *75*, 811–825. https://doi.org/10.1016/j.chb.2017.06.007

Quintel, D. F., & Wilson, R. (2020). Analytics and Privacy. *Information Technology and Libraries*, *39*(3). https://doi.org/https://doi.org/10.6017/ital.v39i3.12219

Redkina, N. S. (2018). Library Sites as Seen through the Lens of Web Analytics. *Automatic Documentation and Mathematical Linguistics*, *52*(2), 91–96. https://doi.org/10.3103/s0005105518020073

Rodríguez, G. E., Torres, J. G., Flores, P., & Benavides, D. E. (2020). Cross-site scripting (XSS) attacks and mitigation: A survey. *Computer Networks*, *166*. https://doi.org/10.1016/j.comnet.2019.106960

Rossi, J. (2021). What rules the Internet? A study of the troubled relation between Web standards and legal instruments in the field of privacy. *Telecommunications Policy*, *45*(6), 102143. https://doi.org/10.1016/j.telpol.2021.102143

Samarasinghe, N., & Mannan, M. (2019). Towards a global perspective on web tracking. *Computers and Security*, *87*, 101569. https://doi.org/10.1016/j.cose.2019.101569

Schelter, S., & Kunegis, J. (2016). Tracking the Trackers: A Large-Scale Analysis of Embedded Web Trackers. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)* (pp. 679–982). Retrieved from https://www.semanticscholar.org/paper/Tracking-the-Trackers%3A-A-Large-Scale-Analysis-of-Schelter-Kunegis/6916e12f6233f9d7c46d7950a5196fed9783bdea

Schmidt, L., Bornschein, R., & Maier, E. (2020). The effect of privacy choice in cookie notices on consumers' perceived fairness of frequent price changes. *Psychology and Marketing*, *37*(9), 1263–1276. https://doi.org/10.1002/mar.21356

Slepchuk, A. N., & Milne, G. R. (2020). Informing the design of better privacy policies. *Current Opinion in Psychology*. Elsevier Ltd. https://doi.org/10.1016/j.copsyc.2019.08.007

Stearn, J. (1998). The 10 common myths of cookies. *Computer Fraud & Security*, (7), 13–15. https://doi.org/10.1016/s1361-3723(98)80006-7

Strycharz, J., Smit, E., Helberger, N., & Noort, G. van. (2021). No to cookies: Empowering impact of technical and legal knowledge on rejecting tracking cookies. *Computers in Human Behavior*, *120*, 106750. https://doi.org/10.1016/j.chb.2021.106750

Thompson, N., Mullins, A., & Chongsutakawewong, T. (2020). Does high e-government adoption assure stronger security? Results from a cross-country analysis of Australia and Thailand. *Government Information Quarterly*, *37*(1), 101408. https://doi.org/10.1016/j.giq.2019.101408

Vakeel, K. A., Das, S., Udo, G. J., & Bagchi, K. (2017). Do security and privacy policies in B2B and B2C e-commerce differ? A comparative study using content analysis. *Behaviour and Information Technology*, *36*(4), 390–403.

https://doi.org/10.1080/0144929X.2016.1236837

Verma, M. K., & Brahma, K. (2017). A webometric analysis of national libraries' websites in South Asia. *Annals of Library and Information Studies*, *64*(2), 116–124. Retrieved from http://nopr.niscair.res.in/handle/123456789/42441

Waris, S. (2021). India: Personal Data Protection Bill-Status Of The Legislation And Data Regulation Regime In India. Retrieved June 11, 2021, from https://www.mondaq.com/india/data-protection/1022956/personal-data-protection-bill-status-of-the-legislation-and-data-regulation-regime-in-india

Yang, X., Yi, X., Khalil, I., Cui, H., Yang, X., Nepal, S., … Zeng, Y. (2019). A New Privacy-Preserving Authentication Protocol for Anonymous Web Browsing. *Concurrency Computation*, *31*(21), 1–17. https://doi.org/10.1002/cpe.4706

Yessine Borchani, B. (2020). Advanced malicious beaconing detection through AI. *Network Security*. Elsevier Ltd. https://doi.org/10.1016/S1353-4858(20)30030-1

Yue, C., Xie, M., & Wang, H. (2010). An automatic HTTP cookie management system. *Computer Networks*, *54*(13), 2182–2198. https://doi.org/10.1016/j.comnet.2010.03.006