
# Healthcare.gov Security Analysis – Congressional Hearing November 19, 2013

Version 1.0

TrustedSec, LLC
E: info@trustedsec.com
11565 Pearl Road
Suite 301
Strongsville, Ohio 44136
1.877.550.4728

Disclosure statement: Information contained in this report was obtained through passive analysis of readily available information. Under no circumstance did TrustedSec conduct any type of "hacking" efforts or attempt to exploit any weaknesses in the healthcare.gov website.

To Whom It May Concern, November 15, 2013

TrustedSec performed an open-source analysis of the security around the healthcare.gov website. This report contains information regarding the concerns for the security around the website and the ability to keep United States citizen information protected to an adequate level. TrustedSec did not perform analysis through "hacking" techniques, as our organization was not authorized to perform offensive activities against the site.

Instead, TrustedSec utilized information readily available on the Internet as well as analysis of information presented back from the website to perform the assessment. What this analysis shows us is that as an attacker, there are known exposures in the healthcare.gov website today that could lead to significant compromise of the website and information. Additionally, the website is integrated into multiple agencies including some of the largest collections of United States citizen data – this includes the Internal Revenue Service (IRS) and other federal agencies.

Based on our evaluation of the website, we have serious concerns over the security of the website and the ability to protect information. This document will explain our approach, what was identified, and the future roadmap to ensuring that the website and its integration into multiple agencies can be successful and secure.

We appreciate the opportunity to present this information to government officials and look forward to our testimony on November 19, 2013.

Sincerely,

David Kennedy
CEO, Founder - TrustedSec
11565 Pearl Rd. Suite 301
Strongsville, OH 44136
E: INFO@TrustedSec.com

# Table of Contents

# 1.0 Executive Summary

The Affordable Health Care Act was a sweeping change to the availability and affordability of health insurance for much of the United States population. The act provided a conduit for integration into multiple state-exchanges, as well as navigates citizens of the United States to different and competitive pricing. In order to support the integration process, the healthcare.gov website was created to provide a centralized approach and easy navigation to the general public. In order to meet the deadlines of the website, contractors were brought in to build and develop a customized solution to the website interface. Based on our research and the exposures identified, the healthcare.gov website is at critical risk for unauthorized access.


In traditional development lifecycles, websites are created formally with two major components (depending on methodologies i.e. waterflow, agile, etc.). A formal development process takes into consideration multiple teams and groups and merges them into one cohesive development process that integrates several areas. When a group of developers or several hundred developers work on something new, typically a framework is utilized (often referred to as a content-management system or CMS). This framework is used to ensure consistency while the logic on the background is developed in order to make the website work. In the case of the healthcare.gov website, two frameworks are utilized for content generation and the underlying framework. These are called Jekyll (https://github.com/mojombo/jekyll) and Bootstrap (https://github.com/twbs/bootstrap).

Frameworks provide a continuous way to have consistency and make the "look and feel" the same, however it does not actually create the functionality behind the website. This requires a formal development team to produce code in order to integrate into multiple federal and state departments as well as provide results to the end user based on the information provided. The website cost an estimated $624 million and consists of over 500 million lines of code. With the number of lines of code, this is one of the most complex applications ever written in the history of applications. To put this in comparison, the Microsoft Windows 8 operating system, which is the latest, has an estimated 50 to 80 million lines of code and has over 25 years of development and maturity. It should be noted that with 80 million lines of code, the Windows operating system has had a significant amount of "exploits" that have hit their product line since it's early existence. Additionally, the Linux Kernel which runs the most popular open-source distributions such as Ubuntu, Debian, Fedora, Redhat, etc. has roughly 15 million lines of code.

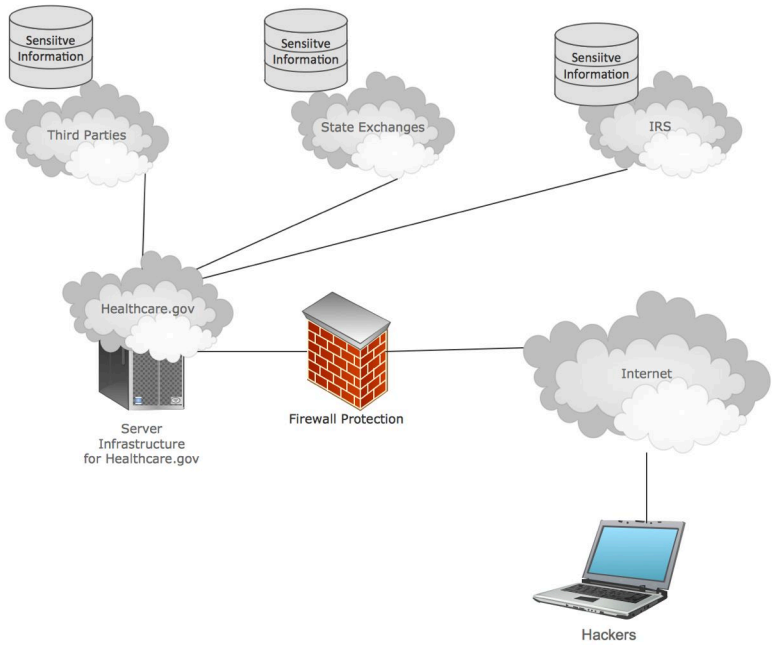
Microsoft has one of the largest and most sophisticated security development, protection, and remediation processes today. This process has taken years to mature and places security at the forefront. With a website that is over 6 times more complex than the Microsoft operating system and developed in an extremely short period of time, there is and was no foreseeable way to build security into the website.  This is apparent based on our research and what exposures we as well

as other security researchers have uncovered since the website's initial release. Based on our findings, we are confident that the security around the application was not appropriately tested prior to release, that the safeguards to protect sensitive information are not in place, and that there are and will continue to be for a significant amount of time serious security concerns with the website unless direct action is taken to address these concerns.

Again, TrustedSec has not performed direct "hacking" on the website, however based on the information contained within this document and issues that you will see walking through the report, there are clear indicators that even basic security was not built into the healthcare.gov website. TrustedSec is confident based on the exposures identified that the website has critical risks associated with it and security concerns should be remediated immediately.

While TrustedSec may not have the full picture of the underlying technologies, based on the research identified and public information available about how the system integrates into other federal and state departments, there is serious cause for concern with the website.

In it's simplest form a website is the programming and the "logic" behind how a user interfaces with a website and how it behaves. Behind the scenes are databases, supporting infrastructure such as routers, switches, and other technology devices to make things work. The claim thus far on the healthcare.gov website is that there is no actual sensitive information stored on the actual webserver itself. This may be accurate however in order for the website to pull the information needed, it requires tight integration into multiple state and federal sites as well as third parties. In order for this to work, integration through other databases or web services is required. Following is an example of how this may work within the healthcare.gov infrastructure.



In the previous depiction, an attacker would circumvent the website and gain control of trusted connections between the healthcare.gov website, its databases, and ultimately the integration into all of the other areas. This is one of the most likely scenarios and major concerns for the current healthcare.gov website. If a vulnerability or exposure is identified on the website, it can directly impact the federal and state governments.

Also note that TrustedSec identified multiple severely critical exposures that it is not publishing publicly until they have been addressed.



# 2.0 Public Information Analysis

This section covers areas of public information that were disclosed through other researchers, or through information that has been made public since the launch of healthcare.gov.

## 2.1 Healthcare.gov Targeted About 16 times

Reference link: http://investigations.nbcnews.com/_news/2013/11/13/21440068-healthcaregov-targeted-about-16-times-by-cyberattacks-dhs-official-says

This is one the most alarming statistics released publicly. It's highly inaccurate and not probable for a publicly facing website with such high profile to only experience "about 16" attacks. What this statement shows is the lack of a formal detection and prevention capability within the website and it's infrastructure. This should be one of the most alarming pieces of information released to date as it shows that there is little to no visibility into what actual attacks are occurring on the website. This means that in the event that the website is hacked (or already has been), the attacks would go largely un-noticed and the website would remain compromised for a long period of time. On average, while working for an international Fortune 1000 company, our main website was attacked over 230 (averaged 232 attacks a day for the year of 2012) times a day with a much smaller footprint and profile, and less publicity than the healthcare.gov website.

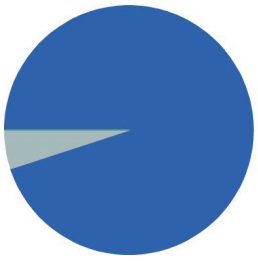
Additionally, basic reconnaissance was performed on the healthcare.gov website, and it appears that there are little to no preventative measures in place to stop attackers from hitting the website continuously, nor detect attackers. The only precaution that appears to have been taken is the website does not allow browsing from The Onion Router (TOR) which masks traffic and locations over the Internet (privacy related).

Analysis on Attacks: TrustedSec has an open-source project called Artillery (https://github.com/trustedsec/artillery and https://www.trustedsec.com/downloads/artillery/), which actively monitors attack vectors geographically from all over the world. Even including websites that are not well known, the breakdown shows that websites are attacked roughly 32 times per day on average.

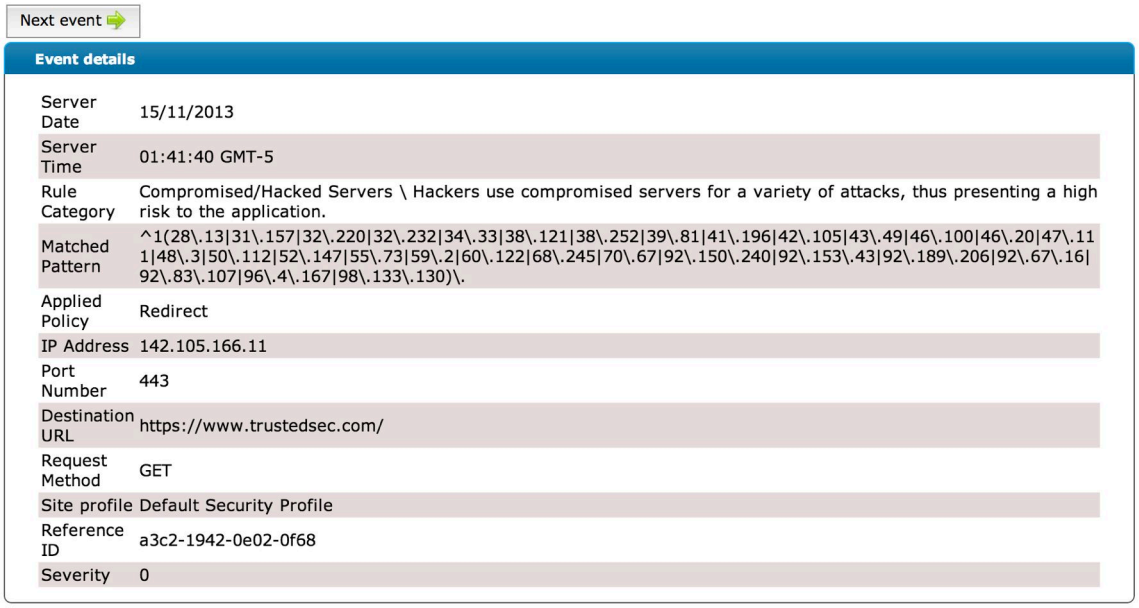
As an example, https://www.trustedsec.com received 46,689 known attacks in a one month timeframe:

**Attack Count For All Sites**

| Site Name | Attack Count | Percentage |
|---|---|---|
| www.trustedsec.com:443 | 44350 | 94.99 % |
| www.trustedsec.com | 2335 | 5.00 % |
| trustedsec.com | 4 | 0.01 % |
| **Total count** | **46689** | |



Additionally, you can see the direct attacks as they occur and the heavy attack volume for a website that is purely dedicated to a specific industry:



Next event

**Event details**

| | |
|---|---|
| Server Date | 15/11/2013 |
| Server Time | 01:41:40 GMT-5 |
| Rule Category | Compromised/Hacked Servers \ Hackers use compromised servers for a variety of attacks, thus presenting a high risk to the application. |
| Matched Pattern | ^1(28\.13\|31\.157\|32\.220\|32\.232\|34\.33\|38\.121\|38\.252\|39\.81\|41\.196\|42\.105\|43\.49\|46\.100\|46\.20\|47\.111\|48\.3\|50\.112\|52\.147\|55\.73\|59\.2\|60\.122\|68\.245\|70\.67\|92\.150\.240\|92\.153\.43\|92\.189\.206\|92\.67\.16\|92\.83\.107\|96\.4\.167\|98\.133\.130)\. |
| Applied Policy | Redirect |
| IP Address | 142.105.166.11 |
| Port Number | 443 |
| Destination URL | https://www.trustedsec.com/ |
| Request Method | GET |
| Site profile | Default Security Profile |
| Reference ID | a3c2-1942-0e02-0f68 |
| Severity | 0 |

Based on Internet statistics, it is evident that the website would be attacked significantly more based on pure Internet volume and not including targeted attacks.

## 2.2 Security Warnings Ignored

Reference Link: http://www.pcworld.com/article/2063220/lawmakers-healthcaregov-security-warnings-came-before-launch.html

One of the more alarming trends is that the actual security testing of the website was deferred due to project delays. The website was launched without formal testing and with known risks around the security of the applications. Even further, there was little to no security built into the website or through the development. With the complexity of the website, this would indicate that the website will suffer from significant security concerns for a long period of time unless significant action is taken to address the issues and flaws within it.

## 2.3 Personal Information Disclosure

Reference Link: http://arstechnica.com/information-technology/2013/10/healthcare-gov-deferred-final-security-check-could-leak-personal-data/

Recently, an exposure identified shared personal information with third party groups such as rum-collector.pingdom.net and doubleclick.net (statistical information).



## 2.4 Email Enumeration Exposures

Reference Link: http://swampland.time.com/2013/10/28/exclusive-password-reset-security-glitch-fixed-on-healthcare-gov/

In the referenced link, an email disclosure vulnerability was identified that would allow an attacker to enumerate email accounts for individuals. While this may seem minor, the ability to identify who has registered on the healthcare.gov website makes it significantly easier to target individual accounts and utilize social-engineering techniques to compromise the system. As an example, TrustedSec's CEO was on the Katie Couric show recently and showed how easy it was in under ten minutes to compromise someone online once the email address was exposed: https://vimeo.com/77102165.

## 2.5 Multiple Exposures Identified

Reference Link: http://blog.isthereaproblemhere.com/2013/10/appalled.html

The mentioned link shows multiple exposures identified including the ability to brute force user accounts through the error messages, reveal password reset codes without access to the actual account or email address, reveal email addresses, and reveal the security questions. These issues are prone to multiple areas of attack and show a lack of formal security practices around the website.

## 2.6 Other User Information Exposed

Reference Link: http://blog.heritage.org/2013/11/02/exclusive-healthcare-gov-users-warn-of-security-risk-breach-of-privacy/

User logged into the healthcare.gov website and saw information from a completely different persons profile (PDF document).

## 2.7 Additional Reference

http://www.popularmechanics.com/technology/how-to/computer-security/can-healthcare-gov-keep-your-data-safe-16119563
http://apnews.myway.com/article/20131022/DA9JEPK81.html
http://www.forbes.com/sites/theapothecary/2013/10/01/healthcare-gov-crashes-during-first-day-why-massachusetts-never-had-this-problem/
http://dailycaller.com/2013/11/13/hacking-tool-destroy-obamacare-poses-new-threat-to-health-care-website/
http://fedscoop.com/decoding-healthcare-gov-security/
http://blogs.wsj.com/digits/2013/11/11/chart-a-car-has-more-lines-of-code-than-vista/

# 3.0 Actual Analysis

TrustedSec conducted analysis of the website and identified a number of exposures that could expose United States citizen's sensitive information or direct exposures that could actually lead to the compromise of the website. Note that several exposures were not posted publicly because they expose extremely sensitive information.

## 3.1 Undisclosed Exposures

Reference Link: Not disclosed.

TrustedSec has identified critical exposures for the healthcare.gov website as well as sub-sites which it cannot disclose at this time due to responsible disclosure principles and the possible impact of sensitive information disclosure.

## 3.2 Open URL Redirection

Reference Link (provided by independent security researcher Gillis Jones):
http://finder.healthcare.gov/cms/sites/all/modules/ckeditor_link/proxy.php?url=http://example.com

When clicking on the above link, users could visit the website thinking they were going to the legitimate healthcare.gov website but instead be redirected to a malicious website that would completely hack their computer.

## 3.3 Vulnerability Query String XML Output

Reference Link: https://spa.healthcare.gov/search-server/search?test='test'

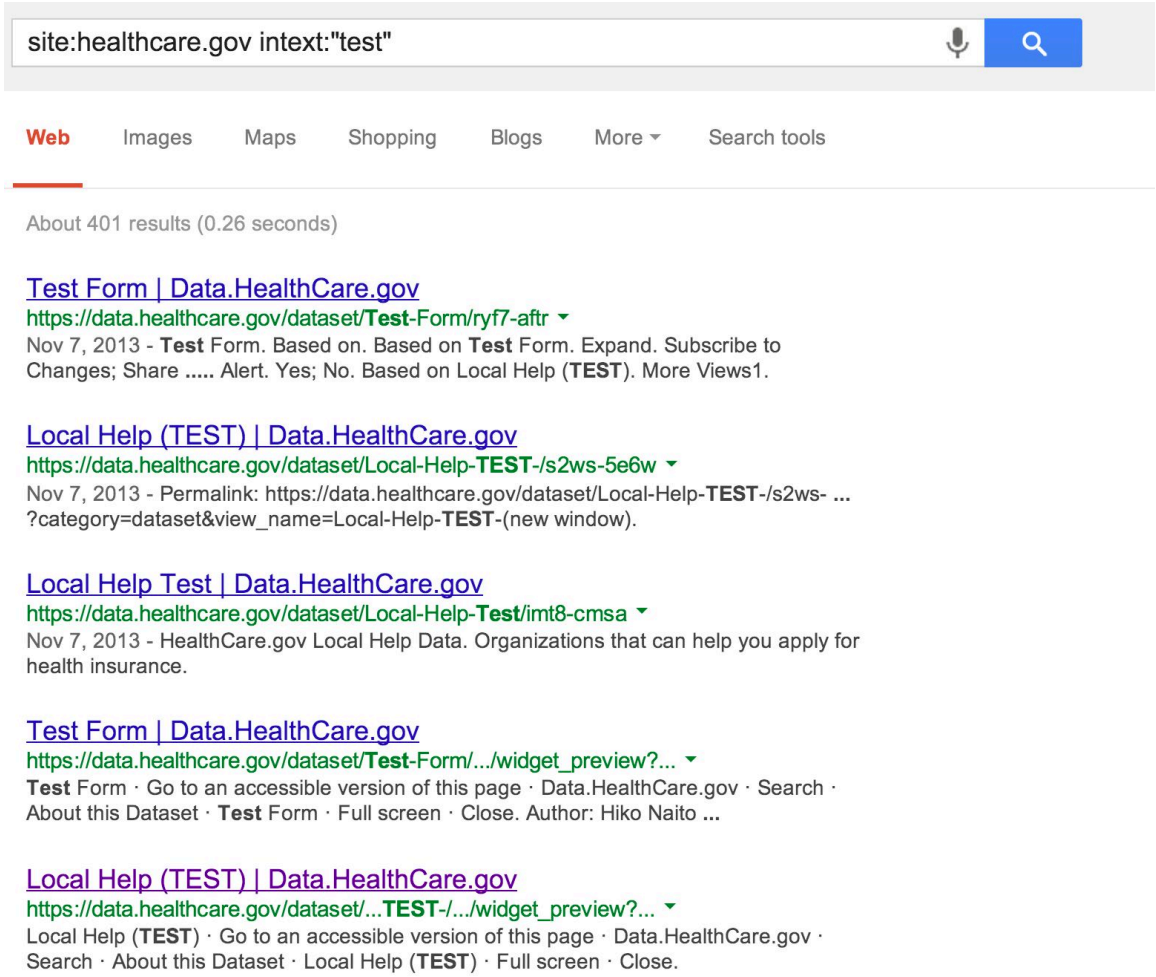
Within spa.healthcare.gov you have the ability to manipulate the response data to whatever you want by changing the query string parameter "test" to whatever you want.

## 3.4 Test Domains Exposed on the Internet

Reference Link: https://test.healthcare.gov

Test domains are exposed to the Internet, which is often an area for focus of attack. Additionally, there is a significant amount of test data already indexed all over the Internet.






## 3.5 Exposed Profiles

Reference Link: Google -> site:healthcare.gov inurl:profile

It appears that individual user accounts and names are indexed via Google and can expose profile information of individuals that sign up on data.healthcare.gov.

[redacted] | Data.HealthCare.gov
https://data.healthcare.gov/profile/[redacted]..✓
[redacted] Joined on June 17, 2013 Last logged in November 11, 2013.
Datasets. 1. Forms. 1. [redacted]

[redacted] Data.HealthCare.gov
https://data.healthcare.gov/profile/[redacted]
smith_ca. Joined on October 01, 2013 Last logged in October 01, 2013. [redacted]
Followers (0). Following (1). [redacted] Datasets ...

[redacted] | Data.HealthCare.gov
https://data.healthcare.gov/profile/[redacted]
Nov 7, 2013 - [redacted] DataSlate Developer Intern, Socrata Washington, District of Columbia, United States. Joined on June 03, 2013 Last logged in ...

[redacted] Data.HealthCare.gov
https://data.healthcare.gov/profile/[redacted]
10+ items - Skip to main content Skip to footer links. Hello, Unknown User ...

2 RY2011 MLR Dataset 20121206 134,882 views.
3 RR Submission Version Policy 116,306 views.

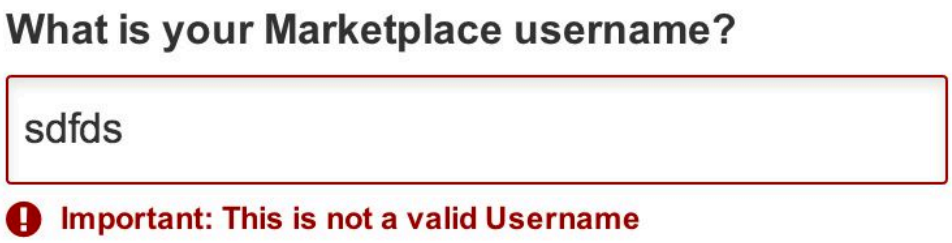
[redacted] | Data.HealthCare.gov
https://data.healthcare.gov/profile[redacted]
New Mexico QHP Individual Market Dental Landscape 10-7-13. For instructions on how to read and use this data, please view the documentation available ...

### 3.6 Username Enumeration

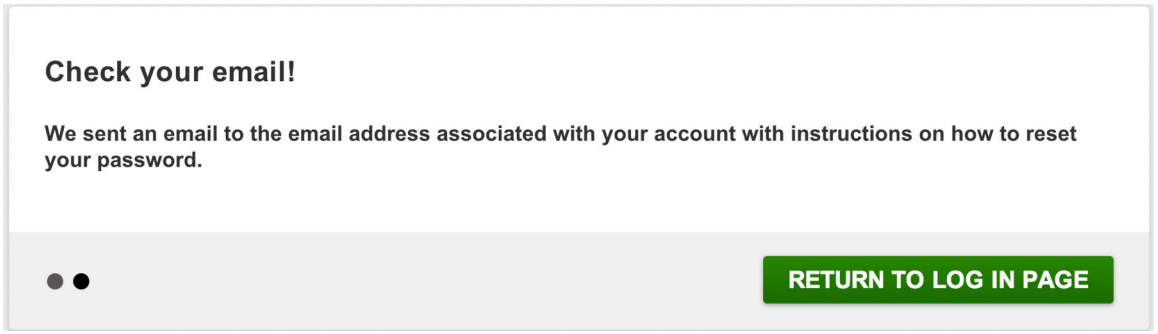
Reference Link: https://www.healthcare.gov

When logging into the website, the website will let you know when an invalid username is specified and when an invalid password is specified. This will allow an attacker to enumerate userID's used in the website.

Invalid user:




Valid user:



## 3.7 Privacy Sign off

Reference Link: https://www.healthcare.gov/individual-privacy-act-statement/

Information is shared with multiple third parties and other government agencies:

> In order to verify and process applications, determine eligibility, and operate the Marketplace, we will need to share selected information that we receive outside of CMS, including to:
>
> 1. Other federal agencies, (such as the Internal Revenue Service, Social Security Administration and Department of Homeland Security), state agencies (such as Medicaid or CHIP) or local government agencies. We may use the information you provide in computer matching programs with any of these groups to make eligibility determinations, to verify continued eligibility for enrollment in a qualified health plan or Federal benefit programs, or to process appeals of eligibility determinations. Information provided by applicants won't be used for immigration enforcement purposes;
> 2. Other verification sources including consumer reporting agencies;
> 3. Employers identified on applications for eligibility determinations;
> 4. Applicants/enrollees, and authorized representatives of applicants/enrollees;
> 5. Agents, Brokers, and issuers of Qualified Health Plans, as applicable, who are certified by CMS who assist applicants/enrollees;
> 6. CMS contractors engaged to perform a function for the Marketplace; and
> 7. Anyone else as required by law or allowed under the Privacy Act System of Records Notice associated with this collection (CMS Health Insurance Exchanges System (HIX), CMS System No. 09-70-0560, as amended, 78 Federal Register, 8538, March 6, 2013, and 78 Federal Register, 32256, May 29, 2013).

## 3.8 Experian Third Party Verification

Reference Link: http://krebsonsecurity.com/2013/10/experian-sold-consumer-data-to-id-theft-service/

Verification information shared with Experian recently was identified in selling consumer information to ID theft services.

## 3.9 jQuery File Upload exposed

Reference Link: https://www.healthcare.gov/marketplace/global/en_US/js/jquery.fileupload.js

Upload forms are often an area for an attacker to upload malicious content and attempt to execute it or use it in social-engineering campaigns.



### 3.10 HTML5 Cross-Origin Sharing

Reference Link: https://www.healthcare.gov

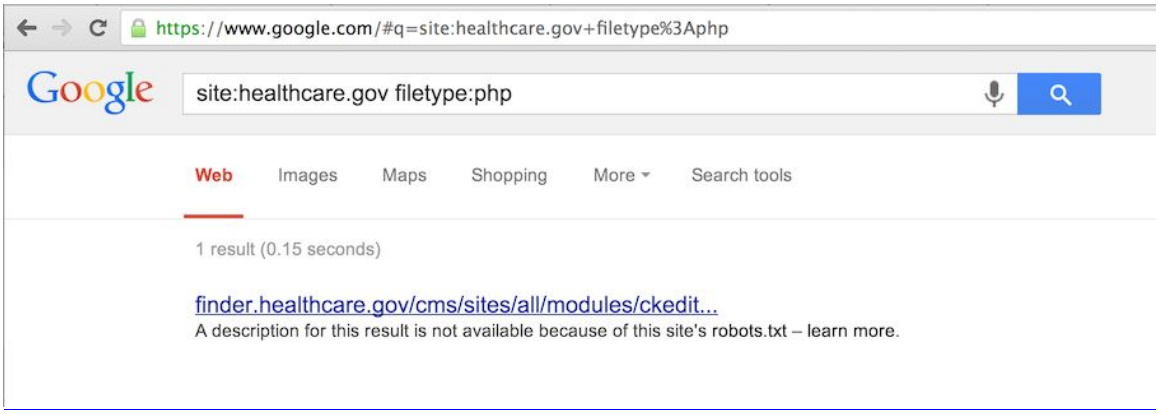
For a detailed list of cross-origin sharing, refer to this link:
https://code.google.com/p/html5security/wiki/CrossOriginRequestSecurity

### 3.11 CKEDITOR (HAS BEEN REPORTED AND REMOVED)

Reference Link: Google – site:healthcare.gov filetype:php (SINCE REMOVED)

CKEDITOR was installed on the finder.healthcare.gov website which contains multiple vulnerabilities. This has since been removed.

http://www.exploit-db.com/exploits/24530/
http://www.exploit-db.com/exploits/25493/



# 4.0 Website Recommendations

Complex websites such as this are bound to have exposures and "glitches," however it appears based on the sheer number of exposures and the lack of formal testing around security that there are systemic and serious concerns with the healthcare.gov website. Based on our experience, in large web applications such as this, there are a few options available in order to address the security concerns with the website.

## Option 1: Version 2.0 (Highly Recommended)

The website that is currently up is functioning in some capacity. The overly complex solution designed for the integration into state exchanges and other areas for real-time display of healthcare programs should be re-written from a code optimization standpoint. In something this complex, if design and code quality weren't created from the start, the fixes that we see now will only be small patches for a much larger problem. The first option would be to write a second healthcare.gov website in conjunction with what's currently up and running. This version "2.0" would be completely redesigned from the ground up with security and proper development processes established.

## Option 2: Shut Down and Fix

If the website is shut down for the time being in order to address the situation, this may allow a more rapid response to addressing security concerns with the website. A "penetration test" which is apparently in process on the website is not recommended at this point. A full source code review and dynamic logic testing with use cases on the application should be considered for a more in-depth review. This will alleviate some of the major security issues but based on the complexity and size, the remediation process will span seven to twelve months at a minimum.

## Option 3: Fix in Production

The term "production" refers to a site or application that is already up and running with normal user traffic. In this case, significant changes to a production environment need to undergo extensive testing before promotion from a QA/Dev/Test scenario. In a formal process, coding changes would occur, be tested in a formal setting in a non-production instance and then be promoted to production, or the "live site". This process definitely slows down the ability to introduce rapid fixes to the website as it could dramatically impact the end-user experience and functionality of the website.



## 5.0 Security Researchers

David Kennedy – Founder and CEO of TrustedSec (@HackingDave)

Scott White – Principal Security Consultant at TrustedSec (@s4squatch)

Alex Hamerstone – Practice Lead for Governance Risk and Compliance (@infosecdoc)

Gillis Jones – Independent Security Researcher (@Gillis57)



# 6.0 Risk Calculation Methodologies

During a technical review of an organization, basic criteria can be identified for the calculation of a risk that a specific vulnerability or exploit has to a company. TrustedSec utilizes the formula Risk = (Vulnerability + Threat) * (Impact – Countermeasres). There are several unknowns when calculating risk factors due to the likelihood of occurrence being a large uncertainty. TrustedSec cannot calculate likelihood due to many moving factors including discoverability, adversaries, timing, and opportunity.

TrustedSec can however calculate risk based on the vulnerability and how it could be utilized. Note that TrustedSec cannot calculate true impact due to not understanding the information available on all systems, the loss and damages, and the importance of the data to the company. TrustedSec can however calculate impact as it pertains to the impact it had towards the rest of the assessment and further compromising an organization.