

Web Packaging - A non-profit publisher perspective

Submitted by Dan York

7 June 2019

Non-profit organizations could greatly benefit from an open standard for packaging websites that creates better user experiences, particularly in mobile environments. Done well, an open web packaging format could enable non-profit organizations to reach larger audiences that can help the organizations execute on their missions.

Given the typically smaller budgets and web staff of non-profit organizations, there is an opportunity to level the playing field and make room for content from nonprofits and other small organizations. Many of these organizations are unable to fully participate in some of the current proprietary packaging systems.

However, there is a danger that if not done as an open standard that will be widely deployed, a common web packaging format could solidify the ongoing consolidation within the Internet's content layer.

We need to ensure that whatever common web packaging format emerges does not further enshrine "gatekeepers" of content, but rather promotes an open Internet where all voices can participate.

Background

The Internet Society (ISOC) is a global non-profit organization that directly operates a collection of 20+ websites. These sites primarily use WordPress as a content management system (CMS). Most of the sites are currently hosted with a managed hosting provider whose service is based on Amazon Web Services (AWS) servers. Most sites make use of Amazon's Cloudfront content delivery network (CDN).

Additionally, the Internet Society has over 130 Chapters and Special Interest Groups (SIGs). These are separate legal organizations who independently operate their own individual websites. They do share many of the same objectives and use cases of the global organization, and could potentially realize many of the same benefits of an open standard for web packaging.

Note: The author is an employee of the Internet Society, but this paper does not represent an Internet Society position. This background is provided for context.

Use Cases

With a vision of "The Internet is for everyone", part of our mandate within the Internet Society Strategic

Communications team is to:

- Ensure our content is *available* to everyone, everywhere.
- Ensure our content is *discoverable* to everyone, everywhere.

In doing so, we strive to make our sites as fast as possible, ensure they meet accessibility standards, ensure they work well in mobile environments, and ensure they embrace and promote usage of Internet standards. There are several use cases where some form of packaging web sites would greatly benefit us.

Availability - Mobile caching / pre-fetching

A primary use case is in providing a better user experience for mobile users by accelerating the performance of mobile web pages through caching and pre-fetching systems. We can see in our analytics the continued rise in mobile users. I have seen through my own user experience the speed increase through visiting sites using Google's Accelerated Mobile Pages (AMP), Facebook Instant Articles, or Apple News. The sites appear to be almost immediately visible, versus waiting for a regular website to load.

Availability - Low bandwidth / high latency environments

A good portion of our work involves reaching people who live in parts of the world where there is limited connectivity to the Internet. In some places, people have to bring their smartphones or laptops to a local place with WiFi and download content over that connection. It would be ideal to be able to provide them a package of appropriate pages that they could download on their device while at the WiFi hotspot, and then be able to read and use later while offline.

We also have people in parts of the world where electricity goes on and off at somewhat random times, and Internet access with that. Again, the ability to download a package of content could help them learn and read more in the absence of actual connectivity.

The content could be:

- information about grant programs that could help increase connectivity
- information about how to make their local systems more secure
- a multi-page tutorial or implementation guide on routing security
- a set of recent blog posts to help the person stay informed about what the organization is doing
- documentation about how to start a Chapter or become a member
- or many other sets of content.

We do currently use a CDN to optimize the delivery of our content to locations such as this. But we can see benefit in some form of packaging that can be delivered by the CDN on our behalf.

It is important to note that there must be caution in terms of not making the “package” so large that it takes a long time to load in a low-bandwidth, high-latency environment. That concern would need to be considered in thinking how to construct the package.

Low bandwidth environments are a place where I see the "signed origin" work proposed as part of the current Web Packaging proposal could be of use. I could see a case where someone might download a package of content from a WiFi hotspot to a device - and then later share that package to someone else over another local network connection (for example, a Bluetooth connection between devices). It would be very useful for the second recipient to be able to know for certain that the package originated from our sites in its current form.

Availability - An extreme example

To be a bit more specific, let's consider a more extreme edge case. One of our sites, the Internet Society Foundation, will provide grants for various projects, including for increasing connectivity in various regions of the world. We currently are funding several projects creating “community networks” in different parts of the world.

For someone to apply, they need to visit the site, read the information, perhaps see examples of what we have funded, learn about the application criteria, and then complete an online form. This content exists on multiple pages on the site.

Now, imagine you are in a remote Inuit village in the Arctic that perhaps only has Internet access through one of the various satellite systems that provide periodic connectivity. Your Internet access is limited to the time when one of the satellites is passing overhead. During one pass, you might learn about the Foundation grants or somehow wind up on the Foundation's site. Maybe you can read the first page or so... but then the satellite goes out of range. Now you have to wait until the next satellite pass to get the next page(s) of information. And maybe you get all the info in the next pass... or maybe you don't and have to wait for the next pass... or maybe you give up.

It seems to me that a much better approach would be to have some way to package up a set of pages so that if someone visits, say, the page on “Medium and Large Grants”, they actually get a package of all the various pages needed to apply for a grant. The person in the remote community could pull down this whole package during one satellite pass. They could then read all the various info they need, without waiting for the next satellite pass.

Note that it could be the person downloading the package to their own browser / computer / device - or it could be the ground station downloading the package into a local caching system and making the pages available to serve people in the community.

While this may be an extreme example, there are many remote communities throughout the world with these types of connectivity challenges.

Discoverability - Mobile search results

An additional use case is placement in mobile search results within the platforms. Search engines are how a large percentage of people learn about and get connected to our sites, services, and content.

What we see within the current platforms is that priority is given in mobile search results to content that uses that platform's packaging format. Google's mobile search favors sites offering AMP content. Facebook results favor sites offering Instant Articles. Apple News will only show sites using its format.

This favorable placement provides an extremely strong motivating factor for an organization to consider using as many of these different systems as possible.

However, given the economic requirements and the typical resource challenges of nonprofits and other small businesses or organizations, there is a danger that if the current proprietary packaging systems continue, we may lose these smaller voices.

Ideally, the open standard for web packaging would be accepted by the platform providers so that sites packaged in that format would receive favorable placement in search results on par with the platform's proprietary format (if the platform continues to use that format).

Choosing Consolidation - or Complexity

The challenge with adopting the existing systems for website packaging is that each platform uses its own proprietary format and has its own separate packaging requirements.

Like many nonprofits, we are not resourced to support custom development work. Instead, we rely on "plugins" built within the larger WordPress developer ecosystem. Right now, to support the different packaging formats/systems, we must install a different plugin for each platform. This adds to the complexity of our site. As you install more plugins, you:

- increase the potential for conflicts between the plugins;
- decrease the security of the site, as you are adding in more code from external vendors;
- increase the administrative overhead in terms of updates;
- increase the overall size of the site for backups;
- increase the system load, which can lead to decreased performance; and
- increase the challenges in troubleshooting any issues on the website.

For those reasons, we seek to avoid adding too many additional plugins to the site. The level of complexity to support all the current packaging systems is more than we wish to add to our sites.

However, if we choose to support only one mobile caching / pre-fetching system, we are then locking ourselves

into that one platform. We are also feeding into the ongoing consolidation that I see within the Internet content space.

Seeking A Standard

In my ideal world, there would be one form of web packaging that we could generate from our sites that could then be consumed by multiple different caching / pre-fetching systems. This would allow us, for instance, to only install a single plugin and generate a single feed.

Questions I would like to see discussed

In the ESCAPE workshop, beyond the discussion of how a standardized web packaging specification can meet the use cases outlined above, some of my additional questions include:

- If a Web Packaging specification could be standardized through the IETF or other standards bodies, what commitment can we see from major platforms to adopt and promote this standard?
- What are the conditions under which a standardized Web Packaging specification could help ensure content on the Internet remains open and not increasingly under the control of large platforms?
- What are the conditions under which such a standard could lead to centralization/consolidation?
- How could we enter into the standardization process in a way that would avoid the conditions leading to centralization?
- How can web packaging lead to scenarios that help decentralization?
- How can web packaging be constructed in a way that ensures smaller voices of nonprofit organizations, individuals, and other smaller organizations can also be viewed on the Web?

Personal Background

My current role with the Internet Society is as the Director of Web Strategy, managing a small team responsible for the publishing of content across our various sites, as well as the development of many aspects of the sites. My role also includes search engine optimization (SEO), search engine marketing (SEM), and all aspects of analytics across our sites. I have worked with the Internet since the mid-1980s and with the Web since its early days in the beginning of the 1990s. I have built websites since that time using a wide range of systems, frameworks, and CMSs, with a focus on WordPress since 2005. My responsibilities have included teams managing websites for both commercial/for-profit and non-profit employers, and have involved both strategic/policy/business and tactical/execution aspects.

Within the IETF, I have been participating in the WPACK mailing list and face-to-face meetings since the initial side meeting at IETF 101.

As noted previously, this paper is a personal submission and does not represent an Internet Society position.