Web Annotation Architecture and Scope

@JAKEHARTNELL
I love imagining the future.
“Post modern as f@#$ and genuinely cool.” ~@RansomSilver

“The book really captures the confusion and social complexities of growing up online.” ~D

“A full-blown 3D Technicolor dystopia.” ~L.S.E.

“It’s so great to read when you’re high!” ~I School Student

“A work of pure genius.” ~My sister (slightly biased)

“I liked it.” ~Professor Ray Larson

23rd Century Romance

The future of sex and relationships.
Let’s imagine a future with a truly readwriteweb!
Let’s make private annotation

In the browser of 2018!
A few things have changed.

#understatement #peoplestillusehashtags
Web Annotation is successfully implemented in browsers and in reading applications. We can now address things not just on the granularity of a page, but of anything on that page.
In January 2014, documents provided by Edward Snowden showed that a Canadian spy agency used a unique identifier to follow thousands of Canadians as they moved about the country. The tracking all originated from an unnamed airport.

It got us thinking: how hard would it be to replicate this little experiment, with small? Could I use one of my own online identifiers as a way to track my own movements through time and space?

The answer, perhaps unsurprisingly, is: yes. It’s easy to do, and it’s revealing about what I do, when I do it, and where I go.

Like many other websites, Ars Technica employs a system of voluntary user logins. These logins allow you to do things like leave comments at the bottom of every story and engage in our user forums. Each time you log in to Ars, we record the date, time, and IP address that you logged in from. This is a common practice: nearly every website maintains similar records. Typically though, Ars only keeps one record per user of the last date, time, and IP address used. We do not keep any historical records of login data.

However, Ars lead developer Lee Aylward was kind enough to make an exception—for me. For 11 days in February 2014, Ars tracked all of my logins. The working theory was that since I’m telling Ars who I am (my login name is the frequently used and obvious “canvast”), and loading the site multiple times per day, my logins would actually give Ars a clear idea of my actions and movements.

In turn, I sent this 11-day log along to Nicholas Weaver, a computer security researcher at the International Computer Science Institute in Berkeley, California. It took Weaver just a short amount of time to write a Python script that converted the raw CSV data file (including Unix time notation). It would start with a line like this:

```
1392058430,335607,
```

And Weaver’s creation could turn it into something much more human-readable, like this:


Learning Dvorak, Part 1
My QWERTY typing speed, my first typing lesson, and progress by the end of part 1.

Stay in the know with

Latest News
- FASTER, HIGHER, STRONGER
  More Wi-Fi is better: FCC expands use of 5GHz spectrum
- ECHOS AND PAIN
  Amid CEO fallout, Mozilla affirms support for “LGBT marriage equality”
- BRAIN TRANSPLANT
  Maryland puts its health exchange site before death panel
- THE NOT-SO-PIRATE DAY
  Pirate Bay Prez to ‘choose a little'
In January 2014, documents provided by Edward Snowden showed that a Canadian spy agency used a unique identifier to follow thousands of Canadians as they moved about the country. The tracking all originated from an unnamed airport.

It got us thinking: how hard would it be to replicate this little experiment, writ small? Could I use one of my own online identifiers as a way to track my own movements through time and space?

The answer, perhaps unsurprisingly, is yes. It’s easy to do, and it’s revealing about what I do, when I do it, and where I go.

Like many other websites, Ars Technica employs a system of voluntary user logging. These logs allow you to do things like leave comments at the bottom of every story and engage in our user forums. Each time you log in to Ars, we record the date, time, and IP address that you logged in from. This is a common practice: nearly every website maintains similar records. Typically though, Ars only keeps one record per user of the last date, time, and IP address used. We do not keep any historical records of login data.

However, Ars lead developer Lee Aylward was kind enough to make an exception—me. For 11 days in February 2014, Ars tracked all of my logins. The working theory was that since I’m telling Ars who I am (my login name is the frequently used and obvious “cfnarvar”) and loading the site multiple times per day, my logins would actually give Ars a clear idea of my actions and movements.

In turn, I sent this 11-day log along to Nicholas Weaver, a computer security researcher at the International Computer Science Institute based in Berkeley, California. It took Weaver just a short amount of time to write a Python script that converted the raw CSV data file (including Unix time notation). It would start with a line like this:

f392058430,335007,[IP REDACTED],wp-admin/post.php?post=410003&action=edit&message=10

And Weaver’s creation could turn it into something much more human-readable, like this:

Between Fri Feb 14 07:03:51 2014 and Fri Feb 14 10:58:38 2014 at SecuredServers.com

**FURTHER READING**

**WHAT THE NSA CAN DO WITH “BIG DATA”**

The NSA can’t capture everything that crosses the Internet—but it doesn’t need to.

**Learning Dvorak, Part 1**

My QWERTY typing speed, my first typing lesson, and progress by the end of part 1.

**STAY IN THE KNOW WITH**

**FASTER, HIGHER, STRONGER**

More Wi-Fi is better: FCC expands use of 5GHz spectrum

**ECHOS AND PAINS**

Amid CEO fallout, Mozilla affirms support for “LGBT marriage equality”

**BRAIN TRANSPLANT**

Maryland puts its health exchange site before death panel

**THE NOT-SO-PIRATE BAY**

Spirits Bay, Brad-Aggh way out of line
In January 2014, documents provided by Edward Snowden showed that a Canadian spy agency used a unique identifier to follow thousands of Canadians as they moved about the country. The tracking all originated from an unnamed airport.

It got us thinking: how hard would it be to replicate this little experiment, with small? Could I use one of my own online identifiers as a way to track my own movements through time and space?

The answer, perhaps unsurprisingly, is: yes. It's easy to do, and it's revealing about what I do, when I do it, and where I go.

Like many other websites, Ars Technica employs a system of voluntary user logins. These logins allow you to do things like leave comments at the bottom of every story and engage in our user forums. Each time you log in to Ars, we record the date, time, and IP address that you logged in from. This is a common practice: nearly every website maintains similar records. Typically though, Ars only keeps one record per user of the last date, time, and IP address used. We do not keep any historical records of login data.

However, Ars lead developer Lee Ayliward was kind enough to make an exception—me. For 11 days in February 2014, Ars tracked all of my logins. The working theory was that since I'm telling Ars who I am (my login name is the frequently used and obvious “cflavin”) and loading the site multiple times per day, my logins would actually give Ars a clear idea of my actions and movements.

In turn, I sent this 11-day log along to Nicholas Weaver, a computer security researcher at the International Computer Science Institute based in Berkeley, California. It took Weaver just a short amount of time to write a Python script that converted the raw CSV data file (including Unix time notation). It would start with a line like this:

```
1392056430,3356077,[IP REDACTED]/wp-admin/post.php?post=410003&action=edit&message=10
```

And Weaver’s creation could turn it into something much more human-readable, like this:

```
Between Fri Feb 14 07:53:51 2014 and Fri Feb 14 10:58:58 2014 at SecuredServers.com
```

Here we can write a new HTML document which is associated with this page.

It contains selections, like the one we first highlighted:

> “Like many other websites, Ars Technica employs a system of voluntary user logins.”

In fact, we can make multiple selections and associate multiple documents with this page.
What happens when you make an annotation?

- The browser captures selection information which it passes to a new html document.
- The selection appears inside this new html document as an annotation object. Perhaps an `<annotation>` element that contains all the necessary information for selection.
- The document must be stored somewhere.
“Channels”

● A.K.A annotation document stores.
● Each channel can have a variety of settings associated with it and offer features like groups or custom editors.
● *Users* control which annotation channels they subscribe to.
Settings

Sign in
Signed in as Jake.Hartnell@gmail.com. Manage your synced data on [Google Dashboard](https://google.com).

Disconnect your Google Account...  Advanced sync settings...

Annotations
Channels
You are currently fetching relevant annotations from these channels, click on a channel to remove or change settings.

Manage Private annotation data  Advanced settings...

Search
Set which search engine is used when searching from the omnibox.

Google  Manage search engines...
What is reading annotations like?
The browser loads a page
It then queries all channels the user is listening on.

https://annotations.Hypothes.is
https://annotations.rapgenius.com
etc.
Annotation documents are loaded into a sidebar (or some other space) where they can be seen.
We can filter by channel, search annotations, and apply additional filters and sorts to find relevant documents and information.
OK, you've got 20 million songs in your pocket. Now what? With endless choice comes a serious challenge: cutting through all that noise to find the next song that will change your life. In this era, the real rock star are the curators—the people, tools, and algorithms that bring you the music you’ll love. With the help of special guest DJ Questlove, we’ve assembled the ultimate guide to discovering the ultimate tracks (including Questo's personal favorites). Your perfect playlist awaits.
OK, you’ve got 20 million songs in your pocket. Now what? With endless choice comes a serious challenge: cutting through all that noise to find the next song that will change your life. In this era, the real rock star are the curators—the people, tools, and algorithms that bring you the music you’ll love. With the help of special guest DJ Questlove, we’ve assembled the ultimate guide to discovering the ultimate tracks (including Questo’s personal favorites). Your perfect playlist awaits.
But how do these documents connect to the page?
There is <annotation> element that gets placed into the page around selections that are referenced in documents loaded from channels. It is a type of link, its specifics are out of the scope of this talk, but it is *spatial* and *contextual*, based on where the browser is querying for annotation documents.
More goodies...

We then can have different methods of visualizing, discovering, and interacting with the selections and annotation documents.

Other extensions will be able to provide additional functionality.

It will even be possible to search for additional documents that reference this page.
In summary

Three matters the W3C should address.
1: Annotation as ‘advanced linking’

We should be able to link to any selection on a page. The browser should handle this linking/selection as well as attachment of annotations.
2: User control

Give people the greatest amount of control in managing what they see. This is incredibly important for combating noise.
3: A space

The browser should allow a space for these attached documents to live and be viewed, and it should be a store for personal documents/notes.
We need people to be able to link their thoughts to things.
The End

@JakeHartnell
jakehartnell@hypothes.is