Where will your website be in 20 years?
September 23, 2023

Intro
Where will your website be in 20 years?
Today I'm going to talk about how to how to save WordPress websites for long periods of time,
        Focus a few options, particularly some new technology called Arweave.

First I want to discuss what makes websites disappear.

Link Rot
Link Rot. Content Drift. Content Rot.
As I thought about it Two aspects
        1. Links that are broken – when someone reorganizes their permalinks in WordPress
                Doesn't redirect the old ones
                The content is still there, but all inbound links to it are broken
        2. Harder one - Web Content itself is gone
                Pages get changed or deleted
                Didn't pay monthly hosting bill
                In this case the links work, but can't get to the content

At first glance, who cares?
        So my website has a few 404s – so does everyone in this room!
        And there's the odd link from Wikipedia that doesn't work any more

Always on the move
But there's something beneath the surface here
Broken and rotten links seem like a minor annoyance
        But underneath, they chip away at the assumption that the internet is forever
        Instead – the internet like a huge glacier.
                It looks frozen in time, seems like it will last forever,
                But it's actually always moving - big chunks calve into the ocean all the time.

Link Rot Spread 1
When we launch a site, all the links are healthy. Client finds all the bad ones, we fix 'em
Then after 1 year, over 20% of cited links may be dead or otherwise inaccessible

Link Rot Spread 2
        Then in five years, a great decline. 50% of links can be affected
        Maybe the site has been refreshed, or gone offline.
        As time goes on, "link rot is inevitable."
        A progression that continues – and rarely reverses.

How far does it go?

Link Rot Stats
Hard to get Canadian stats on this. But currently:
> 20% of all science, medicine, technology article links
> 50% of cited links in US Supreme Court opinions
> 70% of all links in legal journals – maybe that's why AI hallucinates legal rulings
> The big one: 72-98% of everything on the internet disappears in 20 years.
>> All the work we do, the sites we build - gone.
>> Where will your website be in 2043?
>> Almost everything on every site we've built will be gone.

Good Riddance
Some of it needs to go. I don't miss Netscape. World will survive less cat memes.
> But some of it is a shame to lose.

WordPress
Share a bit of my origin story with WordPress.
About 15 years ago, working in Toronto for a mid-size non-profit – about 20 employees.
> We hired a high-end agency to create our website.
>> Agency had worked with the Toronto Raptors
>> They *charged* like they worked with the Toronto Raptors
>> Six figure website for a mid-size charity
> After about 9 months all we got back from them was a shell
>> A little gradient background
>> Our logo in the corner
>> NO CONTENT at all
>> Just a few blank pages in the navigation.
>> Honestly felt like 90 minutes of work.
>> "Now it's ready for you to put all the content in!"
>> Where in the 80 page contract did it say
>>> you actually wouldn't put any content on the website?

> This was back in the day when the Board would fly out to see a website launch
>> Almost inconceivable now
>> Here is the board flying in – and I have a blank website
>> Panic mode – working days and nights with this horrid proprietary CMS
>>> I've never built a website! Wut?
>> Got something crappy put together, and at least Board doesn't fire me
>>> A hot mess

At the end of this I thought, "there's got to be a better way to build websites for charities"

> And I then I found WordPress.
> Then I started Vibrant Content.
>> Now we're trying to be that "better way" for mid-size charities
>>> And we've had fun clients from rural Zimbabwe to urban Mongolia
>> Our main clients, the clients I love, are research institutions

Life cycle of a research website
Here's the life cycle of a research website:
> This is an actual one – part of the 98% that has disappeared.

Here's how it works
> Professor has an idea, gets colleague onboard
> Applies for grant funding from somewhere, usually the government
> After a long wait – huzzah! Approved!
> Quickly hire some grad students
> Gives us a small pile of $ to make a website
>> And generally hands-off with a lot of creative latitude
>> Amazing! We get to have fun.
> Pay for us to host and maintain the site for a few years

> But three years later, we send a hosting bill. It doesn't get paid.
>> The grant funding has totally dried up
>> Not even $75 a month to keep the lights on for their site
> So we shut it down and all the research the Prof and grad students did disappears
> The Professor is on to the next grant

> So pointless! All this money, all this work for research that only lasted 1000 days online.
>> Link rot at its worst.

- - - - - - - - -

Stop the 404
Shift this talk from the problem to the solution

Start with the basics. A 404. The error when a webpage can't be found.
> How can we fix broken pages on our sites
> Some of you know all about WordPress and 404s

How can you fix 404s *on your own sites*

Detect 404s
The cool thing about 404s is they show what pages are users looking for and can't find
> So boost user experience by getting rid of them.

The sad truth is that 404 detection becomes an exercise in
> "look at all these PHP files malicious bots looking for and can't find"

But it's still worth doing.
Helpful tools to detect 404s
> Redirection plugin – classic, free (doesn't work with Gutenburg blocks)
> AIOSEO – a premium one
> Yoast SEO Premium will do this automatically for you…if you pay

Fix 404s
Once you have a sense of your 404s
      Use the same plugins mentioned above
            Redirection is cheap and cheerful – free redirects
            Yoast Premium does this automatically when you change a WordPress permalink
      I'd recommend Server Redirects for 404s – if your host offers it
            Faster – because inbound traffic doesn't have to hit a plugin before redirection
                Gets redirected right at the server level
            A bit longer to setup, but worth it

But what about external 404s?
Next level
What if you really don't want users to hit a 404 – when they click *away* from your site?
      Generally less control over
      In one sense, you probably don't care – not your problem
      But sometimes (e.g., Supreme Court citations) it really matters

Perma.cc
      Project of Harvard Library
      Easy to use
            Makes a HTML copy of single page and a screenshot
            Kind of like a faster Internet Archive
      Downside - Just 10 free links then you pay for more

archiveforever.xyz –
      Developed on top of Arweave (which we'll discuss)
      Fast
      Totally Free (for now)

Live demo of archiveforever.xyz
      <>Let's do a quick demo of Vibrant Content home page<>
      Heads up
            No delete button – once it's up, it's up
            Long, long, long URLs

But what about whole websites?
Oooo – up another level.
Now we are getting to the heart of it.
How do you save whole websites for 20 or even hundreds of years…for a single price?
      How can you stop link rot from eroding almost the entire web every 20 years?

Options
Options thin out

Wayback Machine – many of you are familiar with
        Amazing resource – 840 billion saved web pages
                Hands up if you've lost a page, rebuilt it by looking it up on Wayback Machine
                   …lost a whole site, and rebuilt it by looking it up on Wayback
                But limited scope, often just 1 snapshot
                And. It. Is. Slow.
And…WordPress itself wants in on the permanence game
        Interesting.
        Pay them for 100 years of domain + hosting for an eye-watering $38k

Is there a better way?

Enter Arweave
New protocol to store data for very long periods of time
   World's first permanent data storage platform

Intriguing
        My whole business model based on recurring revenue
                …because I went to WordCamps that said I should do that!
                And they were right!
        But if I could pay once and store a site for 100s of years – game changer
        Started exploring Arweave as an arbitrage on my own business
                Eventually got to know some of the early teams working on Arweave
                And one ended up as a client – Ar*Drive* – a kind of Dropbox for Arweave

How does Arweave work?

Blockchain based
So like other blockchain projects -
        decentralized – there's no one corporation at the wheel
        self-sovereign - the data is provably your own
        immutably time stamped
But whereas Bitcoin's blockchain just marches a long in a straight line –
        one block after another forming a *chain of blocks*.

Blockweave
        Arweave uses a "block weave" - multi-directional, multi-stranded 'weave' of data
                optimized for data storage and very high redundancy.

Arweave 2

So as I understand it, upload a file to Arweave
        Within a matter of hours around 800 replicas around the world
                As of now, ~150 TB of data has been uploaded to the Weave
                        but including all replicas,
                        network size is something like 92 PB stored globally*
        And the integrity of your file checked every 15 seconds
                Claims to have the "highest level of data validation currently available"
        All of this makes it very very hard to delete

Arweave says they'll store your file for a minimum of 200 years for a single price
        But how do those who store your data get paid for 200 years?

Endowment

This is the second piece of the equation.
When you upload a file, a portion of the price goes into an endowment.
Similar to the way big universities have endowments and use them year after year to pay profs.
        Arweave's endowment used to pay data storers over the long haul

Growth

Since 2018 the platform has seen big growth

Use Cases

OK – so who is using it? Why haven't I heard of it?
        Early users are mostly developers
        Permanent code repositories, gateway software – a lot of infrastructure projects

Popular in China
        Seems to elude the Great Firewall, no 'one server' where files can be deleted

Digital scrapbooking
        Save family photos for future generations

Some Business Use Cases
        Accounting firms – keep tax records

Still early days
        If you want a deep dive into all this, look up Sam Williams

Objections

Economic objection – Arweave is a crypto project, dependent on the AR token,
        the native token of the Arweave ecosystem
                How reliable is that?
        Suppose the answer is: *Time will tell*
        But FWIW, Arweave is one of the larger projects in crypto
                Endowment designed to compensate for fluctuations in the value of AR

Still, I have a bit of Betamax syndrome
	For those who weren't alive then
	Betamax was a competitor to VHS – better tech from Sony
	But it lost the popularity contest to VHS
	Why at the back of every thrift store in Canada you find VHS not Betamax
	The best tech doesn't always win the day
	Could this be the same?
	Time will tell

Interesting one: data permanence vs permanent access
	Arweave promises to *store* your data permanently for a one time price
	They do *not* promise to give you access to it permanently
	Someone explained it this way:
		Suppose you are moving out of your apartment and going away on a long trip
		You go to your friend who has a garage and ask
			"will you store my stuff for a while, I'll give you $100"
		Friend says "sure."
		You move your stuff in
		Later, after your trip, you go back to your friend's house.
		"Hey friend! Had a great trip. By the way, I need to get some of my stuff."
		"Woah, woah! You only paid me to store your stuff, not to actually see it again."

	Yikes! Permanent access means you can always actually get to your data.
	Or less extreme – you are back from your trip, and your friend is rarely at home
		Can't get to your stuff cause they are never home
	Developers – have a work around, good open source gateways you can run yourself
		In essence: build your own key to the garage with your stuff
	But non-developers – it's a concern

	Access is subsidized now, but that could be a problem later.
	I think the Arweave ecosystem knows this – working on a forever free option
		and if you are a developer, good open-source tools to run your own gateway.

Emerging Competitors
	In 2023, Arweave is not the only permanent storage provider
	At least one other player - Lighthouse.Storage out of India
		also uses an endowment model but on FileCoin
	And in the digital scrapbooking category there's forever.com
	But both are for file storage.
		Neither for hosting websites permenantly

Helpful
Here's where I land:
	Objections are real, but Arweave also shows promise
	Really the only to have a serious claim to permanent data
	One of the most helpful tools to deter link rot stop 404s

- - - - - - - -

10 Steps to Permanence
　　　　ArDrive manifest demo
　　　　10 Steps – not really available online, our own method through trial-and-error

Demo Permasite
Goal is to create a permanent website – using my own.
You can have a look at the permanent site and compare it with vibrantcontent.ca
Arweave charged me $0.24 CDN for this

What Arweave tool will we use?
Mentioned that Arweave is technically complex
　　　　In 2023, starting to see some consumer facing apps
　　　　　　　Like Archive Forever, ArDrive, Akord, Dragon Deploy
　　　　In this example, I'll work with ArDrive
　　　　　　　Because I know them best
　　　　　　　Think they are the easiest and most robust

1. Build Static Site
Does not preserve the entire backend of WordPress, contact forms
Will grab JS – so some elements may literally move
But no backend access (at least for now!)
Need a static copy of your site
　　　　　　　Simply Static – not as awesome at JS, CSS preservation
　　　　　　　Offline Pages Pro – oldie but goodie, but only preserved for you
　　　　　　　Others – use CLI, whole process on CLI

2. Create Account on ArDrive.
ArDrive is a file sync app similar to Dropbox, but with permanent data on Arweave
　　　　Heads Up – self-sovereign data, no corporation with your password
　　　　　　　So you don't want to lose it
　　　　Developers might want the CLI or to interact directly with Arweave

3. Create Drive – Make a Drive
The big red New button

4. Top Up – buy credits by credit card
This is new – used to only be able to pay in AR
　　　　Which you can get basically now where in Canada
　　　　Much easier now
　　　　Minimum spend of $5 US

5. Upload Static Files
If you use OPP – you get an index.html to folder of assets
　　　　With SS – you get them all in one folder
Either way – upload them

6. Wait for files to upload and mine
Keep browser tab open
Can take a minute – or up to 30 min
Basically wait for the file icons to go from yellow to green (or red if they fail)
            You may need to reload, re-enter password to check

7. Create Manifest
This is the special sauce
            Combines all the permafiles into one website
            What I did at the start of the talk
Step 7 after Step 6: Wait until mined before making manifest
            Crowded Folder (SS) - May want to give it a name that's easy to find in the folder

8. Wait again
Another minute, 10 min, 30 min
May depend how busy the network is
Permanence takes patience

9. Preview Manifest
Got a green light? Ready to go.
Click on the manifest file
Click on preview
<>Try manifest for the WordCamp site<>
https://fw2nachp2wsdei3m22n5uiwlgflypthw5yra3jvqeqivjqrrugpa.arweave.net/LbTQCO_VpDIjbNab
2iLLMVeHzPbuIg2msCQRVMIxoZ4/Content/2023.html

10. Huzzah
Saved forever
But a heads up that Arweave will crank you out a huge ugly URL
            You can minify it or redirect another URL to it
            …or even mask it.

Heads Up
Things to watch for:
            No delete button with Arweave. Once it's up, it's up.
                        So make sure this is a site you want future humanity to have
                        Can block content at a Gateway level but difficult.
            No password reset – don't lose it!
            Static site – forms won't work, eCommerce won't work
                        Not full WordPress…yet
            Non-English URLs are problematic
                        Difficulty putting a site with Chinese characters in the URL
            Permanence takes patience

Applications

Goodbye Sites

What happens when a website ends?
Or before you launch a new version?
Or say for an extra $x dollars, we'll create a permanent version
of your current site that you can always have for reference.

Client Thank You

Good relationship with a client
Offer for free as a "Thank you" to clients when they leave

Regular Perma-backups

I'm sure you have good daily backups on your server
Hopefully you have offsite backups
Why not make 800 backups for cheap?
We regularly upload our backups to the permaweb
Doesn't have to be a working site with manifest, just ZIP files

Developers: what if code you use regularly was permanently available

Not reliant on GitHub or PHP servers
What if you make apps or plugins where users get to choose version they want
Because they're all saved permanently
What if there's a WordPress plugin that lets you push static backups to Arweave
From within the WordPress dashboard
Or what if all of WordPress – even former versions could be hosted permanently
And work without having month-to-month subscriptions
Think we're just seeing the start of this.
And if any of that interests you – let's talk

The current internet connects people across vast geographic distances.
What if permanent data connects us across vast spans of time – to people we will never meet

I suppose, *time will tell*

REFERENCES

LINK ROT

https://www.theverge.com/2021/5/21/22447690/link-rot-research-new-york-times-domain-hijacking

https://www.cbc.ca/radio/spark/something-s-rotten-in-the-state-of-the-internet-and-archivists-are-worried-1.6047113

https://www.canadianlawyermag.com/news/general/beware-of-link-rot/272928

https://www.newyorker.com/magazine/2015/01/26/cobweb

ARWEAVE

Overview: https://help.ardrive.io/hc/en-us/articles/5300353421467-Arweave-Manifests

Deep Dive: https://docs.ardrive.io/docs/cli/using-the-cli.html#working-with-files

92 Petabytes: https://permaweb.news/arweave-ecosystem-weekly-report-88-89