Universal Access to All Knowledge

Decentralization at the Internet Archive

Arkadiy Kukarkin
Internet Archive
Quick Caveats

- not "official" views of the Archive, my own exégesis
- 30min 7min 25min!
- does not contain Ethereum
Why decentralize the Archive? (in spirit)
“Universal Access to All Knowledge”
So we need to be able to preserve our digital history in
Why decentralize the Archive? (in practice)
Physical Location Risk

SAN FRANCISCO BAY REGION EARTHQUAKE PROBABILITY

62% probability for one or more magnitude 6.7 or greater earthquakes from 2003 to 2032. This result incorporates 14% odds of quakes not on shown faults.

9% Probability of magnitude 6.7 or greater quakes before 2032 on the indicated fault

Increasing probability along fault segments

Expanding urban areas
The Economist Intelligence Unit’s 2020 Democracy Index
167 countries scored on a scale of 0 to 10 based on 60 indicators

- Full democracies
- Flawed democracies
- Hybrid regimes
- Authoritarian regimes

Source: The Economist Intelligence Unit
### Network Bottlenecks

#### New Accounts Created per Month

<table>
<thead>
<tr>
<th>Month</th>
<th>Accounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 1998</td>
<td>600000</td>
</tr>
<tr>
<td>Apr 1998</td>
<td>400000</td>
</tr>
<tr>
<td>Jul 1998</td>
<td>200000</td>
</tr>
</tbody>
</table>

#### Users added per month (archive.org):

<table>
<thead>
<tr>
<th>Month</th>
<th>Users Added</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul 1996</td>
<td>16</td>
</tr>
<tr>
<td>Jul 2002</td>
<td>2,154</td>
</tr>
</tbody>
</table>
Network Bottlenecks
our baseline: ~$650/TB-“forever” (1)
S3 Infrequent Tier: ~$160/TB-year
Filecoin: ~$0.0020/TiB-year (*) (2)
Storj: ~$48/TB-year (**) (3)
arweave: ~$2900/TB-“forever” (4)
Others: $? (***)

(1) https://economicmodel.dshr.org/
(2) https://file.app/
(3) https://www.storj.io/pricing
(4) https://ar-fees.arweave.dev/

(*) or “negative” for DataCap
(**) however see https://www.storj.io/node
(***) talk to me, Swarm people!
(and... positive things)

- content addressing
- transparent link preservation
- web3 forward compatibility
- etc etc
2 Demos and some Musings
Demo I: Streaming from Storj

you might need this:

```javascript
javascript:javascript:location.href='https://link.us1.storjshare.io/s/jxcipzkkgperv3nvngya6l64hjq/archive.org/download/' + location.pathname.replace(/\/$/, '').split('/').pop()
```
What do we need?

- actual UI (/assign @parkan)
- browser support (hello Brave/Opera/???)
- Storj to decentralize metadata (satellite)
Demo II: You might need to know about WARC

(W)ARC Headers

- **Format:** warc
- **WARC-Type:** response
- **WARC-Target-URI:** http://www.bbc.co.uk/
- **WARC-Date:** 2013-07-25T13:13:26Z
- **WARC-Payload-Digest:** sh1:35UTF11PHQZrAVWJTRzJ8BD9O1E
- **WARC-IP-Address:** 213.58.244.69
- **WARC-Record-ID:**
- **Content-Type:** application/http
- **Content-Length:** 101075

HTTP Headers

- **Status Code:** 200 OK
- **Server:** Apache
- **X-Cache-Action:** HIT
- **X-Cache-Hits:** 22
- **X-CDN-X-Requested-With:**
- **X-Cache-Age:** 50
- **Cache-Control:** private, max-age=0, must-revalidate
- **Content-Type:** text/html
- **Expires:** Thu, 25 Jul 2013 13:13:37 GMT
- **ETag:** "f4bc79f1ca2944dc0268f23038749:9"
- **X-LB-NoCache:** true
- **Connection:** close
- **Content-Length:** 100695

Content

Open Content
What do we need?

- encryption/ACLs
- indexing/metadata
- scale!
Thank you!

Arkadiy Kukarkin
Decentralization Lead
arkadiy@archive.org + https://getdweb.net/
@parkan