

A DISTRIBUTED CONTENT ADDRESSABLE NETWORK FOR OPEN EDUCATIONAL RESOURCES

Stephen Downes
*National Research Council Canada,
Ottawa, Ontario, Canada*

ABSTRACT

We introduce Content Addressable Resources for Education (CARE) as a method for addressing issues of scale, access, management and distribution that currently exist for open educational resources (OER) as they are currently developed in higher education. CARE is based on the concept of the distributed web (dweb) and, using (for example) the Interplanetary File System (IPFS) provides a means to distributed OER in such a way that they cannot be blocked or paywalled, can be associated with each other (for example, as links in a single site, or as newer versions of existing resources) creating what is essentially an open resource graph (ORG), and when accessed through applications such as Beaker Browser, can be cloned and edited by any user to create and share new resources.

KEYWORDS

Open Educational Resources, Distributed Web, Dat, Interplanetary File System, Open Resource Graph, Content-based Addressing

1. INTRODUCTION

The distributed web potentially solves issues proponents of decentralization have long sought to address. One challenge is traffic, which overloads a single server. A second issue is latency, or the lag created by accessing resources half a world away. Additionally, some resources may be subject to national policies creating the need to differentiate access. And finally, if the centralized source is unavailable for some reason, then access for the entire world is disrupted.

For the World Wide Web, many of these challenges are addressed with Content Distribution Networks (CDN). (Benghozi & Simon, 2016) In essence, a CDN creates a local version of a website in different geographical regions. When a person in that region requests a resource, they are served a copy of the resource from the local server, rather than the original from a server much further away. This reduces traffic on the home server and makes access faster for the end user. Companies such as Cloudflare and Akamai now serve as much as half the content traffic on the internet (yet they are almost invisible to end-users).

The solution proposed by advocates of the distributed web is in many respects very similar. Content is stored on multiple servers. And when a web user requests that content, it is served from the nearest server. In the distributed web, however, rather than belonging to a company such as Akamai, these servers are individual users' computers. This model, called the peer-to-peer web, has a history of continuous development including services such as Napster, Gnutella, Tor and BitTorrent. (Troncoso, Isaakidis, Danezis & Halpin, Harry, 2017) These are called 'peers' and the system as a whole is called a 'peer-to-peer' (P2P) network.

More recently, a set of proposals called "Web3" (corresponding to a JavaScript library called Web3.js) (Stark, 2018) applies methods of chaining encrypted data structures to create what may be characterized as a "stateful" distributed web. "The ability to easily and efficiently transfer value P2P is at the heart of finance and efficient markets. If you can't hold state in the Internet, you can't transfer value without centralized institutions acting as clearing entities." (Voshmgir, 2018). Beyond obvious applications such as distributed token networks such as Bitcoin or Ethereum, Web3 may offer a response to the issues of centralization and commercialization afflicting OER.

2. CONTENT ADDRESSABLE RESOURCES FOR EDUCATION

2.1 Open Web and Open Educational Resources

The open internet began as email lists and Usenet groups. It grew through blogs and personal websites, from humble personal pages to sprawling websites such as Wikipedia. The open internet thrived in the age of social networks, online classrooms, and massive open online courses. When we think of the internet, we usually think of the internet as ‘open’, and this openness is most often found in the form of open resources.

The philosophy of ‘open’ that characterized the early internet was also reflected in the concept of open education. “Open education is a philosophy about the way people should produce, share, and build on knowledge. Proponents of open education believe everyone in the world should have access to high-quality educational experiences and resources, and they work to eliminate barriers to this goal.” (opensource.com; Colpaert, 2018) This access is often supported by means of Open Educational Resources.

Open Educational Resources (OER) are teaching, learning and research materials that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions. (UNESCO, 2002) There is a large base of literature and practice associated with OER. (Weller, 2016) Numerous repositories containing OER have been developed. (Atenas & Havemann, 2013).

2.2 Challenges to Open Educational Resources

In today’s internet, however, we see companies and institutions pushing back against openness. Proprietors of copyright content such as music, videos, articles and research publications have demanded that internet services block access to free copies of this content, and that they require payment for access to these resources. (Aversa, Hervas-Drane & Evenou, 2019) In addition, content owners and vendors began making money through advertisements. Both subscription-based and advertising-based models encouraged the growth of technology that herded users into content silos and that tracked and analyzed their behaviour. (Papadopoulos, Snyder & Livshits, 2019).

The challenges faced by the open web are reflected in the challenges faced by OER. For example, providers of massive open online courses (MOOCs) have begun to create barriers, charging first for certification and then for access to content itself. (Shah, 2017) In the world of OER the same thing happened with open textbook publisher Flat World Knowledge started charging for access. (Lederman, 2012) The temptation to monetize OER is always present for centralized services like Open Stax, Alison, Top Hat and Lumen Learning. (Aspesi, et.al., 2019)

By dint of subscription fees, value-added services, or advertising and surveillance, these services must contemplate one business model after another based on enclosing open content and requiring some form of authentication to access. And as David Bollier says, the enclosure of open content is one of the greatest threats to the internet. “Enclosure is about dispossession. It privatizes and commodifies resources that belong to a community or to everyone, and dismantles a commons-based culture.” (Bollier, 2011).

The practical application of OER in education today faces numerous challenges, a number of which were described by Sukaina Walji and Cheryl Hodgkinson-Williams (Walju & Hodgkinson-Williams, 2018).

- While OER are being created, we are seeing limited re-use, and almost no adaptation to create new or localized resources
- Licensing remains a mystery to many people, and there isn’t clarity about what license to use, how to license, or even whether certain licenses are actually OER
- It is not easy to create and upload OER to repositories, nor is it easy to use OER in the context of a course or the creation of course materials.
- Models for support and sustainability of OER remain elusive, and projects continue to depend on uncertain sources such as institutional funding, foundations and national or international bodies.

Additional problems exist, for example:

- OER remain hard to discover; there isn’t a good way to search for OER, and learning object metadata (LOM) is difficult to use, and didn’t actually facilitate discovery
- Individual OER often lacked educational support materials such as quizzes, assignment banks, or other materials

- There is no mechanism for ensuring the quality of OER or the appropriateness of OER in a given educational context.

As Tim Berners-Lee wrote recently, “for all the good we’ve achieved, the web has evolved into an engine of inequity and division; swayed by powerful forces who use it for their own agendas.” (Berners-Lee, 2018) His own project, Solid, is a tentative first step toward re-decentralizing that new web.

2.3 Content Addressing

One major difference between the traditional World Wide Web and Web3 lies in how these resources are addressed. On the traditional web and in CDNs, we use the location of a resource. The URL corresponds to an IP address (for example, <http://www.downes.ca> corresponds to 167.99.39.236) and to retrieve a resource, the browser sends a request to that address. However, in the distributed web we use content-based addressing. In essence, we search for resources based on what it is rather than where it is.

The content of a resource (whether it’s text, a web page, an image, whatever) is used as input to a hash algorithm that produces a scrambled string of characters - the hash - of the resource. Depending on the algorithm and the length of the hash produced, each hash is an essentially unique identifier for that resource. So instead of using a URL to request a resource, we use this unique identifier. (Sicilia, Sánchez-Alonso & Barriocanal, 2016). A peer sends a request to the closest peer, which either sends us the resource, or passes the request along to more peers. A content receiver can verify fidelity by using the hash algorithm to ensure the hash of the content received is the same as the hash of the content that was requested.

One significant current project implementing such a protocol is called Dweb (for ‘distributed web’ or ‘decentralized web’). (Ayala, 2018) It’s being called the next big step for the World Wide Web. The Dweb is based on the dat protocol, (<https://www.datprotocol.com/>) which is essentially a mechanism for finding and distributing content-addressable resources by their hash. We may see more and more resources with addresses like this in the future:

`dat://502bdf152d00a35f9785f78d107b9037b5eca9354bcf593e7b4995f9be97a614/`

This address is in fact the dat:// address for the first Content Addressable Resource for Education (CARE). If you access this resource using a peer-to-peer Dweb application you will find a set of pages containing the National Research Council’s Vision and Principles statement (in both official languages, set to photos I took myself). CARE, along with the associated concepts of CARE Packages and CARENet, is a new type of Open Educational Resource.

2.4 Peer Applications

In order to participate in the distributed web, it is necessary to have a peer application. This is an application that runs on your computer and communicates with other nodes in a P2P network to share resources. One such application is the Beaker Browser. The browser allows users to explore Dweb resources, ‘clone’ those resources locally, and create or edit new resources. Beaker manages Dweb functionality like creating hashes and chaining resources together. (Robinson, Hand, Madsen, Buus & McKelvey, 2018).

Beaker also helps users with a dat name service. Hash addresses (like the one above) are long and difficult to remember. A name service allows us to associate a simple string with a hash address (in exactly the same way the Domain Name Service (DNS) associates URLs with IP addresses). So an address in Beaker might look like this: `dat://enoki.site/` For more Dweb resources open a Beaker browser to this website: `dat://taravancil.com/explore-the-p2p-web.md`

The dat:// protocol is only one of a number of current projects based on creating a content-addressable distributed web. One of the other major initiatives is called the blockchain. In the case of the blockchain, the resources in question are entries in financial ledgers. Another initiative, Git (with services based on the protocol like GitHub and GitLab), chains resources in different versions or branches of a software development project. An ambitious project to bring all these under a single umbrella is called the Interplanetary File System (IPFS) along with the associated project, Inter Planetary Linked Data (IPLD).

2.5 Content Addressable Resources for Education

Content Addressable Resources for Education (CARE) is proposed as a new medium for free and open learning resources, essentially replacing OER as it exists today. The differences will be as follows:

- Because CARE are content-addressable, they are stored and access in the web as a whole, rather than in a specific location, and hence cannot be blocked or paywalled
- As part of the distributed web, CARE are also associated with each other (for example, as links in a single site, or as newer versions of existing resources) creating what is essentially an Open Resource Graph (ORG).
- Accessed through applications such as Beaker Browser, CARE can be cloned and edited by any user to create and share new resources.

While we have seen more traditional contents, such as books, media and music, being distributed through IPFS and Dweb, it is important to underline that CARE consist not only of educational content, but interactive applications and service interfaces as well.

As demonstrated in Figure 1, a high-level overview of CARE, resources are uploaded into IPFS, where they receive a content-based address. This address is stored on an Ethereum blockchain. In order to upload, retrieve, view and edit OER, an application similar to Beaker is employed, using IPFS and Web3 Javascript libraries (this application will be demonstrated at the conference).

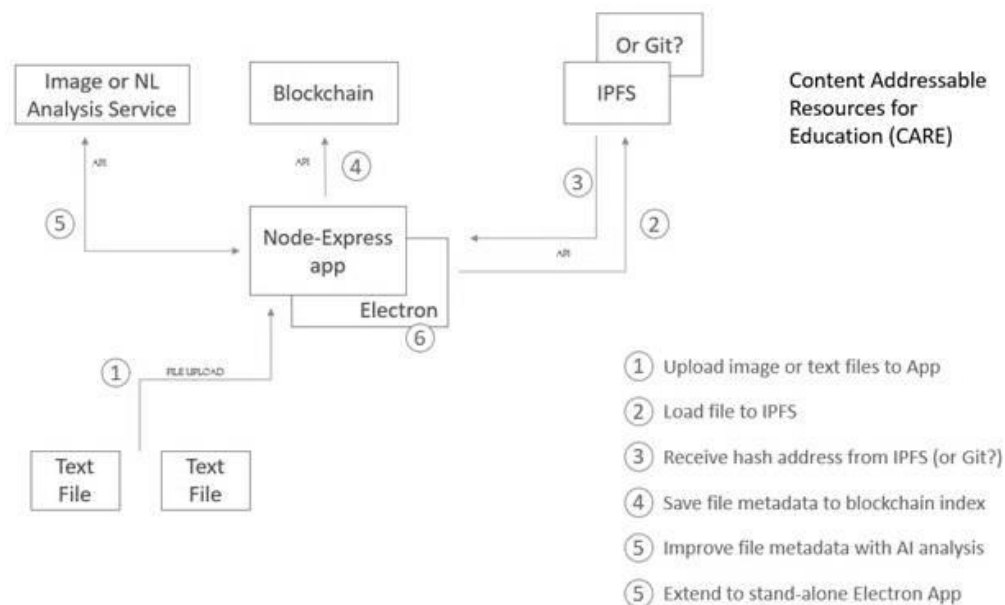


Figure 1. Content Addressable Resources for Education network diagram

3. CONCLUSION

In our work thus far, we have found the distributed web is very much in flux and that practical applications will depend on the resolution of some significant issues. Among them are:

- Speed - though the distributed web can be very fast, in practice, it often isn't, partially because of the time it takes to locate individual content-addressed content, and partially because upload speeds can be very slow for average users. In response, many people look to the cloud to host Dweb or IPFS nodes.
- Ease of Use - while it may seem that creating and sharing a web resource using Beaker or IPFS should be easy, in practice (as E-Learning 3.0 participants experienced first-hand) it can be daunting, especially since applications don't always work and guides are minimal.

- Finding resources - there isn't yet a good Dweb search engine. Additionally, resources can disappear when a host goes offline. This has led to the development of semi-centralized intermediaries such as Hashbase (<https://hashbase.io/>) (which make money by offering always-open nodes).
- Acceptance - many institutions officially disapprove of peer-to-peer services and block .torrent and other P2P traffic; additionally, many P2P sites are associated with blockchain and may therefore also be blocked by institutional internet services
- Appropriation for questionable and possible illegal content and services. With no central point of origin, there is no means to control these types of content, which raises questions about both their legality and their vulnerability.

Future work will be focused on addressing speed issues with a set of known CARE repositories functioning as IPFS nodes (known as CARE Net) and the development of multi-part CARE resources (known as CARE Packages). The emphasis will be to facilitate not only discovery but to also develop mechanisms for content creation through remixing and reusing existing resources.

REFERENCES

- Aspesi, C., et al, 2019. SPARC*Landscape Analysis: The Changing Academic Publishing Industry – Implications for Academic Institutions. In *Copyright, FairUse, Scholarly Communication, etc.*, 99.
- Atenas, J. and Havemann, L., 2013. Quality assurance in the open: an evaluation of OER repositories. *INNOQUAL: The International Journal for Innovation and Quality in Learning*, 1(2) pp. 22–34
- Aversa, P., Hervas-Drane, A. and Evenou, M., 2019. Business model responses to digital piracy. *California Management Review*, 61(2), pp. 30-58.
- Ayala, D., (2018. Introducing the Dweb. *Mozilla Hacks* (weblog), July 31, 2018.
- Benghozi, P. and Simon, J., 2016. Out of the Blue: The Rise of CDN Networks. *Communications & Strategies*, 1, issue 101, p. 107-128.
- Berners-Lee, T., 2019. One Small Step for the Web... *Inrupt* (web site). October 23, 2018.
- Bollier, D., 2011. The Commons, Short and Sweet. Weblog post, July 15, 2011.
- Colpaert, J., 2018. Exploration of affordances of Open Data for Language Learning and Teaching. *Journal of Technology and Chinese Language Teaching*, 9(1), 1–14.
- Lederman, D., 2012 Fleeing from 'Free'. *Inside Higher Ed*, November 5, 2012.
- Papadopoulos, P; Snyder, P. and Livshits; B., 2019. Another Brick in the Paywall: The Popularity and Privacy Implications of Paywalls. *arXiv:1903.01406*
- Robinson, D. C., et al, 2018. The Dat Project, an open and decentralized research data tool. *Scientific Data* volume 5, Article number: 180221 (2018).
- Shah, D., 2017. Coursera Experiments with A Single Subscription Price for the Entire Catalog. *Class Central*, November 8th, 2017.
- Sicilia, M. Á., Sánchez-Alonso, S. and Barriocanal, E., 2016. Sharing Linked Open Data over Peer-to-Peer Distributed File Systems: The Case of IPFS. In *Metadata and Semantics Research: 10th International Conference*, MTSR 2016, Göttingen, Germany, November 22-25, 2016, Proceedings, pp.3-14.
- Stark, J., 2018. Making Sense of Web 3. L4 Blog, *Medium*. June 6, 2018.
- Troncoso, C., et al, 2017. Systematizing Decentralization and Privacy: Lessons from 15 Years of Research and Deployments. *Proceedings on Privacy Enhancing Technologies* 2017 (4):404–426. De Gruyter.
- UNESCO, 2002. UNESCO Promotes New Initiative for Free Educational Resources on the Internet. Web site.
- Voshmgir, S., 2018. Token Economy: How Blockchains and Smart Contracts Revolutionize the Economy. *BlockchainHub Berlin Edition* 27 Jun 2019.
- Walju, S. and Hodgkinson-Williams, C., 2018. Conversation with Sukaina Walji and Cheryl Hodgkinson-Williams. Video. Week 5 of *E-Learning 3.0*, November 21, 2018.
- Weller, M., 2016. Different Aspects of the Emerging OER Discipline. *Revista Educacao e Cultura Contemporanea*, 13(31), 404-418.