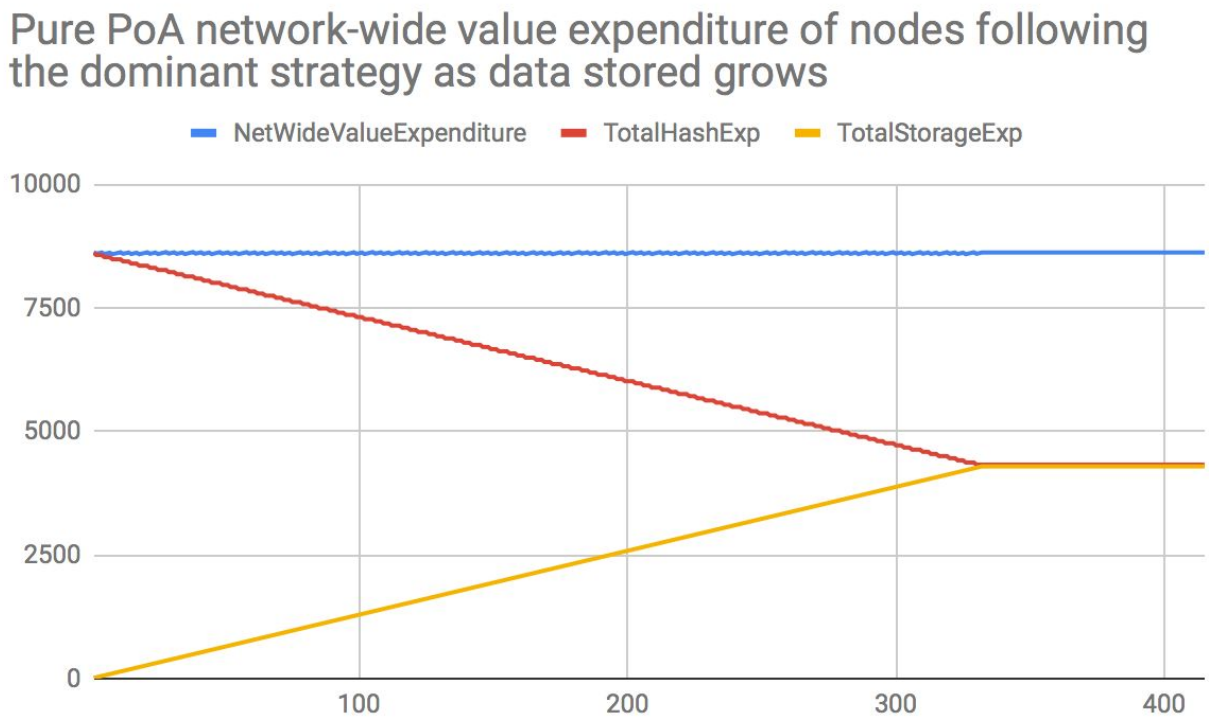
# The Arweave position on knowledge permanence and pricing.

## Version: DRAFT-31-May-2019

**Basic mechanism:**

Proof of work cryptocurrencies expend large quantities of value in the process of mining. In the Arweave network, currently £676.80 worth of value can be expended network-wide on mining every hour, while still leaving 10% profit for the nodes. This provides a large value pool that, with careful mechanism design, can be reassigned from wasted effort (guessing random numbers) to a socially useful purpose (preserving knowledge and history).

By forcing miners to compete over storage as well as hashing (see proof of access mechanism in the lightpaper), assuming that miners do not invest more money in storage hardware (if their returns are near their `expenditure + opportunity cost`), the best way for miners to optimise their profits while the weave grows is to reassign some of their mining value expenditure from hashing to storage. Eventually, storage and hashing expenditure in the network reach an equilibrium.



By calculating the GB-hour cost of storage:

```
HDD Cost / (HDD capacity * Mean Time Between Failure (in hours))
```

and the value emitted by the network per hour:

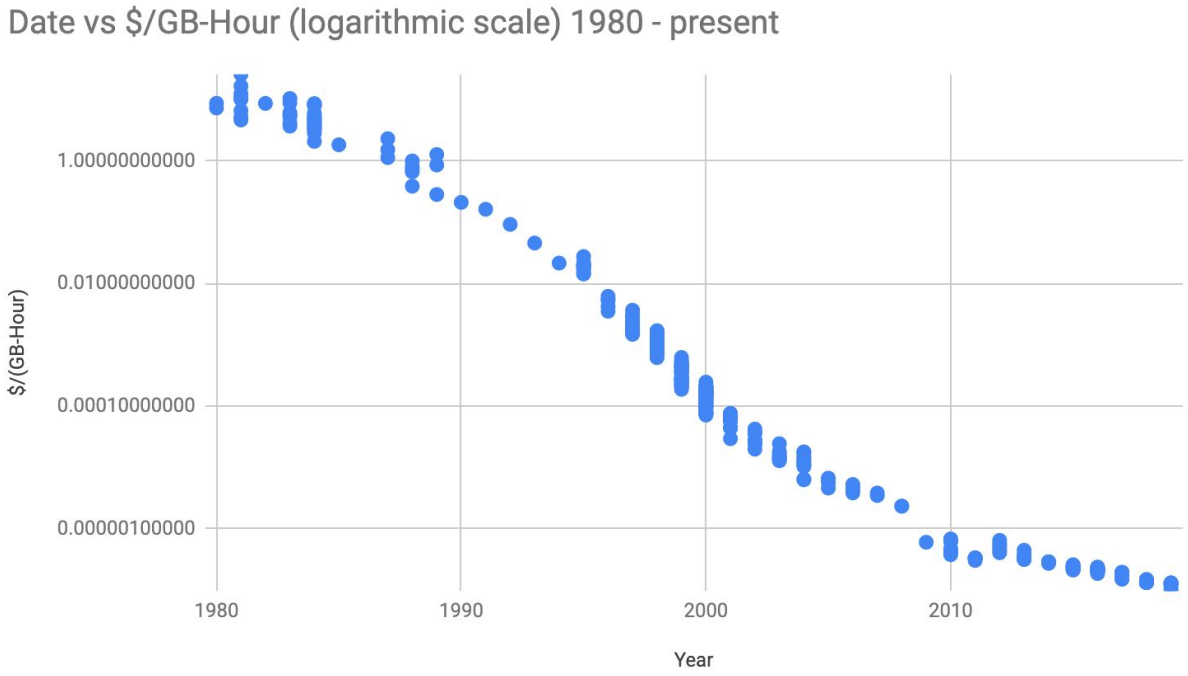
```
(tokens emitted by inflation + tokens emitted from the TX reward pool) * fiat cost of one token
```

we can then calculate the maximum size that the blockweave can support before triggering over-bound protections (see section below):

```
fiat emission rate / GB-hour cost / number of replications / 2
```

The maximum supportable size is halved inline with the equilibrium of the dominant strategy of value (re-)assignment in proof of access (see figure above).

In the current network, the maximum supportable weave size is ~333.06 PB. This can be considered the 'capacity' of our append-only hard-drive. For comparison, in 2018 The Internet Archive's Wayback Machine held ~25 PB of data. For further context, consider that Bitcoin currently emits $651k every hour, and miners spend close to this value (Bitcoin mining has a notoriously low profit margin). If 50% of this value (as is the eventual equilibrium in proof of access) was expended on storage rather than hashing, even through the comparatively inefficient route (in GB-hour terms) of Amazon S3, the Bitcoin network could pay for the storage of 17.49 exabytes of data. Relative to hashing, storing large quantities of data is extremely cheap.



Over the last 50 years, data density of commercially available drives has increased at an annual average rate of 30% (normally in the range of 15% and 50% for any given year). Data density and hard drive capacity/pricing are highly correlated. Please see the 'Data density and pricing expectations' section below for an overview of our reasoning and analysis regarding data density progress over time.

As a consequence of the annualised growth rate (AGR), the capacity of our 'append-only hard drive' naturally grows between 15% and 50% per year, without any increase in value expenditure in the network.

Subsequently, as long as the network does not grow more than 15-50% of the maximum supportable weave size within a given year, we can support the growth of the network without adding a cent to the overall cost of maintaining it.

In practical terms, this means that we can store between 49.9 and 166.5 PB in the next 12 months without affecting long-term sustainability or pricing. This is very similar in principle to the concept of 'living off the interest' of a large sum of money -- as with university endowments, etc.

The average transaction in the Arweave network is very small -- just 99.3 KB. Thus, we could store 502 billion average sized data transactions (assuming a conservative 15% data density AGR) into the permaweb in the next year without affecting the sustainability or future pricing of storage in the network.

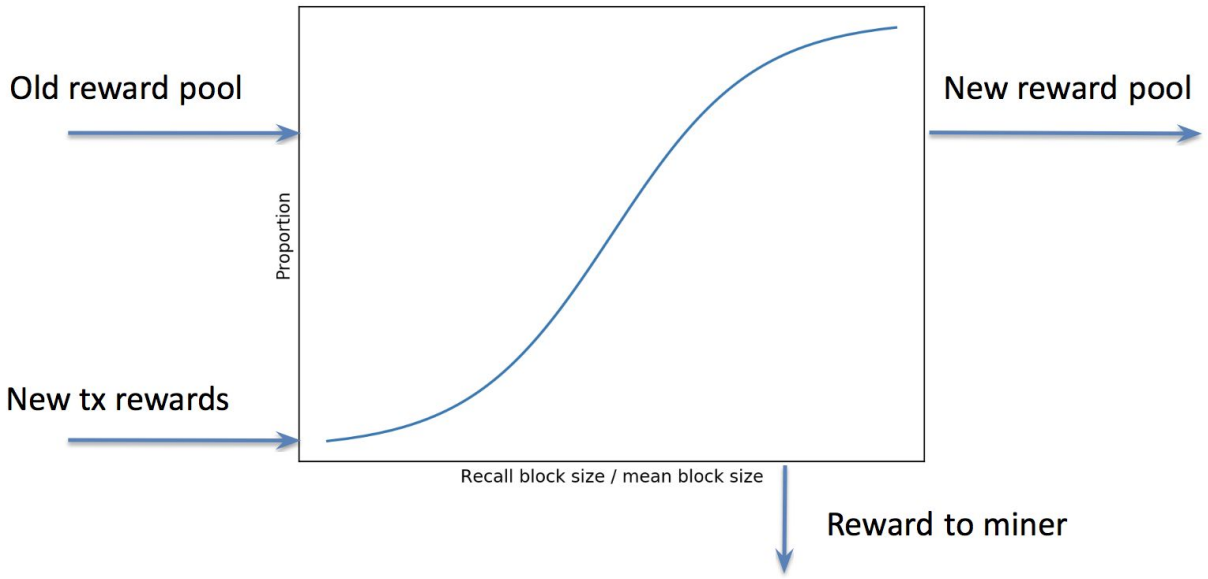
In 2018, The Internet Archive stored 344 billion pages. It is our expectation that the market for permanent data storage is unlikely to be higher than that of the centralised web. Consequently, it is our position that this mechanism alone should be sufficient for keeping the cost of weave maintenance from increasing along with the size of the permaweb.

Notably, this model does not take into account any change in the value of the Arweave token, despite the enormous growth in token demand that would be required to reach 15% of the capacity of the ‘append-only hard drive’. As can be seen by the maximum weave size calculations above, positive increase in the cost of the Arweave token would also correlate with increases in the maximum supportable weave size.

Finally, this begs the question: 'Won’t the permaweb’s growth rate exceed that of the data density AGR?' We think that it will. However, because the Arweave’s current data storage capacity is so many orders of magnitude higher than the amount of data currently stored, there is plenty of capacity for this early stage growth. Even if we experienced Facebook’s growth rate from 2004-2012, perpetually followed by two times the growth rate of the current web (as estimated by the average growth in yearly Internet Archive crawl pages since 2013), the permaweb would not near its capacity for at least 100 years.

**TX Reward Pooling**

Instead of giving transaction rewards directly to the miner that finds the block that the data was added to, Arweave instead puts the tokens into a pool that is passed from block to block. Miners then take a portion from this pool relative to the size of the recall block that was required to mine the new block, providing a 'smoothing' effect on the incentive to select blocks for storage based on their size.




The main motivating factor for the reward pooling mechanic is to lessen the effect of 'hit and run' style situations in which a permaweb app becomes extremely successful for a short period, but then wanes in popularity after. The TX reward pool essentially collects tokens during these periods of fast-growing data usage, in order for them to be released in periods of slower-growing usage.

Transactions are currently priced such that they are ~385x more expensive than the base cost of perpetual storage, assuming just a 1.25% AGR in data density/pricing. The perpetual cost of storage is calculable as the infinite sum of GB-hours, observing the decaying rate of GB-hour costs. We intend to make pricing less conservative and the TX reward pool more aggressive in storing value in an upcoming protocol upgrade.

**Over-bounds protection:**

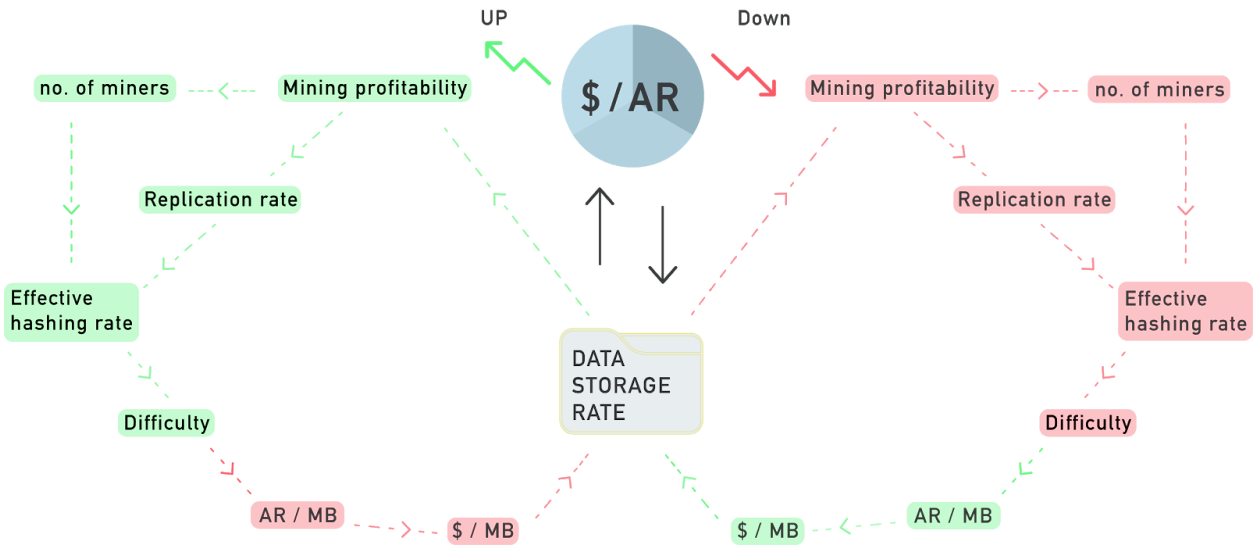
Given the reasoning above, we find it exceptionally unlikely that the weave will grow at such a rate that over-bounds protection should *ever* be required.

Nonetheless, in the extremely unlikely event that it is required, here's how it would work:

As the number of replications of the average piece of data in the network reaches a critical level (as detected by changes in the *effective* hashing rate on each block), the network begins to increase the cost of storage. This triggers one of two things to happen.

1. Users still believe that the value of the storage is higher than the cost, so they continue to store data. As the price of storing data has increased and the amount of data stored is the same, the profitability of mining would increase. As mining profitability increases, the number of miners increases. New miners bring new hard drives, increasing the replication rate, thus creating a self-adjusting feedback loop for storage pricing.

2. Users believe that the price of storage is too high, so they reduce their storage rate, slowing the rate of growth in the network. Miners continue to generate profits from the tokens added to the TX reward pool in prior better market periods.




**Fiat price stabilisation**

While cryptocurrencies and decentralised systems gain adoption, we expect that the volatility that we have seen in the markets over the last decade will continue. Pricing volatility presents problems for the adoption of decentralised utility networks like Arweave. Subsequently, through the same feedback loop cycle outlined in the ‘Over-bounds protection’ section, Arweave employees a form of heuristic (not oracle) based fiat price stabilisation.

As the price of AR relative to USD (or any other fairly stable store of value) increases, the profitability of mining also increases. Higher mining profits brings more miners, raising the replication rate, the effective hashing rate, and the difficulty of mining. Higher difficulty lowers the AR per MB cost of storage, thus rebalancing the fiat price of storage. The same cycles works in the inverse direction, raising AR per MB cost when appropriate, too.

**Data density and pricing expectations:**

Data density increasing and data storage costs decreasing is important to the arguments made in the above sections. While it is clear when plotted on a graph that this pattern has been sustained and consistent for >50 years, it is very reasonable to ask 'why would you assume that this will continue in the future? What would happen if it stops?'.

As well as the pattern of previous data density AGR over prior years, there are three main reasons to believe that this trend will continue:

1. Unlike in the CPU realm, where computation speeds are approaching the point at which theoretical physical limits are being reached and Moore's Law is slowing, this is *far* from the case with data density:

Maximum data density currently in production: 1.66 x 10^12 bits / cm3

Maximum data density achieved in research: 2.5 x 10^25 bits / cm^3

Theoretical maximum data density: 1.53 x 10^67 bits / cm^3

From our current position, at a data density AGR of 30% it will take 434 years to reach the maximum theoretical limit (20% AGR: 697 years, 10% AGR: 1,329 years).

2. Even if advances in data density slow down, data reliability (Mean Time Between Failures in our GB-hour cost formula) continues to increase and and arguably has an even brighter future. The core metric of our argument (GB-hour costs) responds equally to changes in data density and data reliability.

3. Because of the rate of growth of humanity's data, the incentive to develop storage mechanisms with a cheaper GB-Hour cost in the future is enormous. This means the likelihood that high data density/data reliability will remain a possibility, but not an actuality, is extremely low.

**End-game**

All technologies exist in cycles which eventually fade. We don't expect the Arweave will be any different. At some point in the future (maybe 50, maybe 100 years) the final Arweave node will be turned off as humanity migrates to a new immutable data store. It is our expectation that at this point, the cost of storing the data in the Arweave will be minuscule. Just as a copy of Gopherspace can now fit on a $1.89 USB stick (and will soon be uploaded to the permaweb), we expect that copies of the Arweave will end up in whatever the next storage medium that humanity shares will be.

Further, because of Arweave's incentive mechanisms, the network encourages humans to make copies, which they almost never remember to delete if they turn off their nodes. We have spoken to numerous miners that have disconnected a node (typically because competition became too high), and we have yet to find an instance where the node operator actually deleted the data afterwards. At this point, given the level of churn in crypto networks, it is likely that there are more cryptographically verifiable copies of the data offline, than there are online. This 'viral' effect to the data stored in the network will further increase the number of copies of the data in the permaweb that are persisted into humanity’s next storage mechanism, and the one after this, etc.