

Fifteenth Annual

Bio·IT World
CONFERENCE & EXPO '16



Enabling Technology. Leveraging Data. Transforming Medicine.

April 5-7, 2016

Seaport World Trade Center • Boston, MA

Securing Personal Genomic Data: The Potential of The Blockchain

Richard Shute
Curlew Research
April 6, 2016

“Land was the raw material of the agricultural age.
Iron was the raw material of the industrial age.
Data is the raw material of the information age.”

Alec Ross, “The Industries of the Future”, www.alecross.com



What this presentation is – and what it is not!



- It is NOT...
 - ... a sales pitch!
 - ... going to go into all the maths, the logistics and the technology of bitcoin and blockchain
 - So nothing on Byzantine Generals, Merkle trees, nonces, the pros and cons of a “hard fork”, Proof-of-Work, hashing...
 - Here’s a good place to go for that:
 - » <http://blockstrap.com/en/a-complete-beginners-guide-to-blockchain-technology/>
 - ... going to give you “the solution”
 - ... going to give you all the answers



This presentation IS ...

... going to try and be informative.

... looking to stimulate a wider debate.

... aiming to be a bit provocative!

... looking to extrapolate "disruption" within one industry (FinTech) and apply it elsewhere (BioTech).

... going to talk about POTENTIAL not "here-today" fact.

... and:-

It starts with three women ...

Three important women



1. Anne Wojcicki

- CEO & Co-Founder of 23andMe



2. Jennifer Lawrence

- Film star and Academy Award Winner



3. Henrietta Lacks

- Died from cervical cancer on Oct. 4th 1951.



Here's a clue ...



- These companies are all linked to the explosive growth of online personal data and content



iCloud

- Documents
- Photos
- Social Media Interjections and Content
- Music, Videos
- Shopping and Money
- Etc.



The connection?

“It’s all about YOU!”



- Anne Wojcicki
 - CEO of company whose purpose is to enable YOU to access and securely control YOUR genomic data.
- Jennifer Lawrence
 - Was one of a number of celebrities whose personal data – in this case photos (images of HER) – were hacked/stolen from an online repository (iCloud - allegedly).
- Henrietta Lacks (HeLa)
 - Her ultimate YOU, in the form of adenocarcinoma cells from her tumour, were appropriated, immortalised, cloned, and ultimately sequenced without her or her family’s consent.

→ A critical need for modern times ...

Secure, controlled access to YOUR data

- “Just a third of organisations routinely encrypt customer data – potentially putting it at risk of being breached if it did fall into the wrong hands – while only **one in five encrypt health data, which is arguably some of the most sensitive data about an individual.**”



- <http://www.zdnet.com/article/encryption-still-a-low-priority-for-too-many-cloud-users/> , Feb. 29th 2016

Especially when we now have ...

The \$999 Genome!

Via:

<https://twitter.com/EricTopol/status/706491998165807104>

For \$999, Veritas Genetics Will Put Your Genome on a Smartphone App

Getting your entire genome decoded is now more affordable than ever. Will consumers buy it?

by Antonio Regalado March 4, 2016

MIT
Technology
Review

Common Medical Test/Scan/Med	Cost ⁺
Head CT scan	\$1,200
Abdominal CT scan	\$1,420
MRI scan	\$2,611
Echocardiogram	\$1,300
10 "Most Important" Lab Tests*	\$ 319
Crestor 1 year supply from Costco	\$3,222
Xarelto 1 year supply from Costco	\$4,444
3 Pharmacogenetic tests Quest [^]	\$ 975
Whole Genome Sequence	\$ 999

Price of sequencing your genome falls to \$999

By Gary Robbins | 2:17 p.m. March 4, 2016



Announced at Future
of Genomic Medicine
Conference
4 March 2016

UT
San Diego

Curlew
Research

* A "package" significantly discounted via DirectLabs.com

+ vary widely, these are average US charges from centers w/ data available on the web

[^] 3 SNPs which account for 0.00000000001% of the genome

Data security is a particular challenge for genomic data



... Notwithstanding the volumes of data (file size) and the data complexity

- YOUR genome is the most personal, fundamental and defining expression of who and what you are.
 - It's the ultimate YOU!
- OUR genomic and healthcare information is increasingly being electronically captured, stored, made available and commoditized
- And patients themselves are driving this revolution (if Accenture are to be believed)...



PATIENTS WANT A HEAVY DOSE OF DIGITAL

Healthcare consumers in the United States want a digitally enabled care experience, and they are initiating it with greater use of digital tools and electronic health records.

High performance. Delivered.



Patients should own their medical data



- “Healthcare consumers are taking control of their data.”
 - Accenture, 2016
- “Unpatients—why patients should own their medical data.”
 - Kish & Topol, *Nature Biotechnology* 33, 921–924 (2015).
- Even President Obama has said [recently](#): “I would like to think that if somebody does a test on me or my genes, that that’s mine.”
 - NY Times, February 26th 2015

- *So how can patients take control of THEIR own genomic data securely and confidently?*
 - And not have a “Jennifer Lawrence” done to THEIR most personal data

Furthermore, how do we balance ...

- People's increasing desire for more, personal digital healthcare information

... **AND**

- The ever-reducing cost to get that information

... **WITH**

- The need for better security

... **BUT WITHOUT**

- Sacrificing accessibility and control?



Tough huh ...!



Blockchain technology

A potential solution to secure,
controlled access to YOUR data

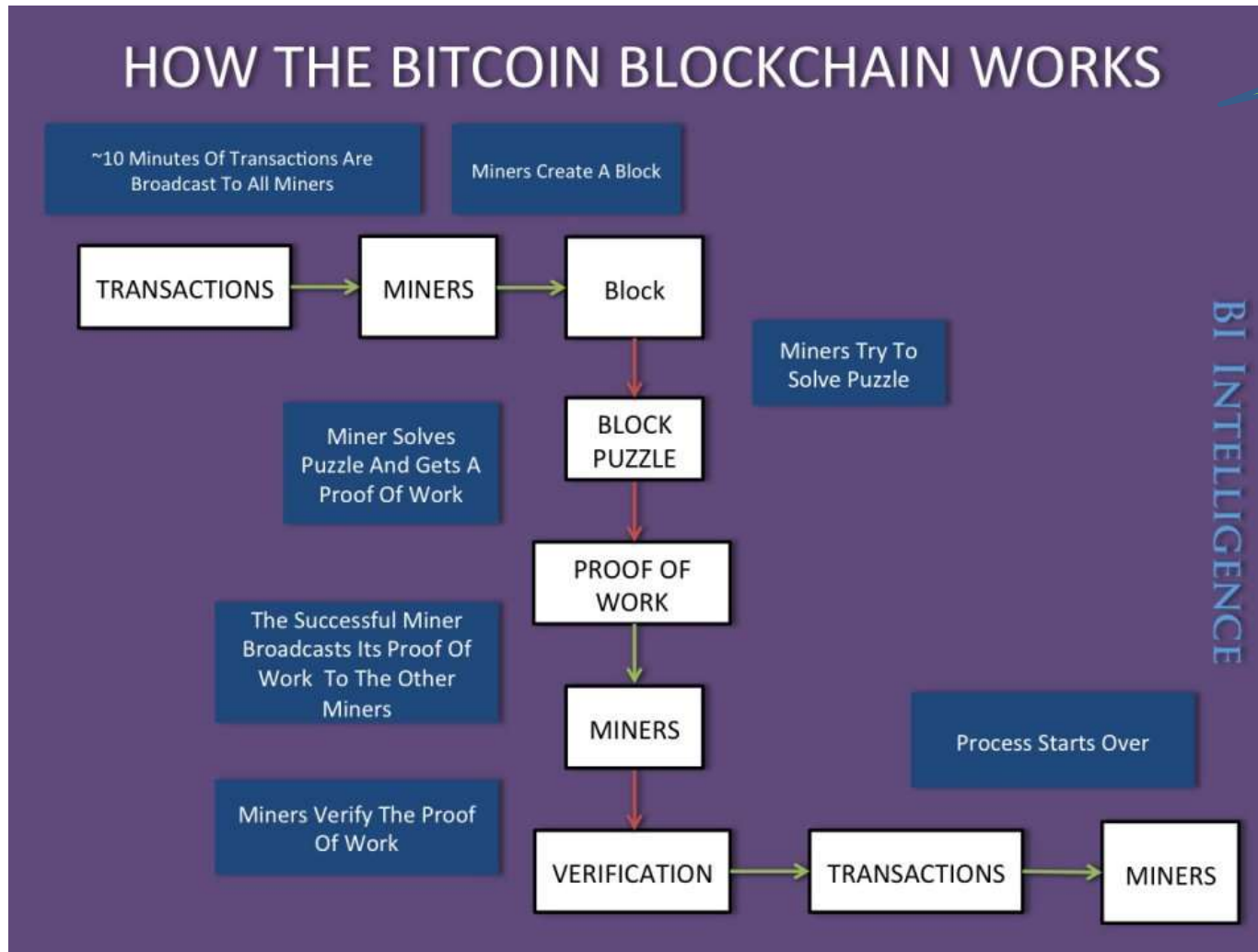
So, what is “The Blockchain”

In words

“A blockchain is a digital, distributed transaction ledger, with identical copies maintained on multiple computer systems controlled by different entities. Anyone participating in a blockchain can review the entries in it; users can update the blockchain only by consensus of a majority of participants. Once entered into a blockchain, information can never be erased; ideally, a blockchain contains an accurate and verifiable record of every transaction ever made.”

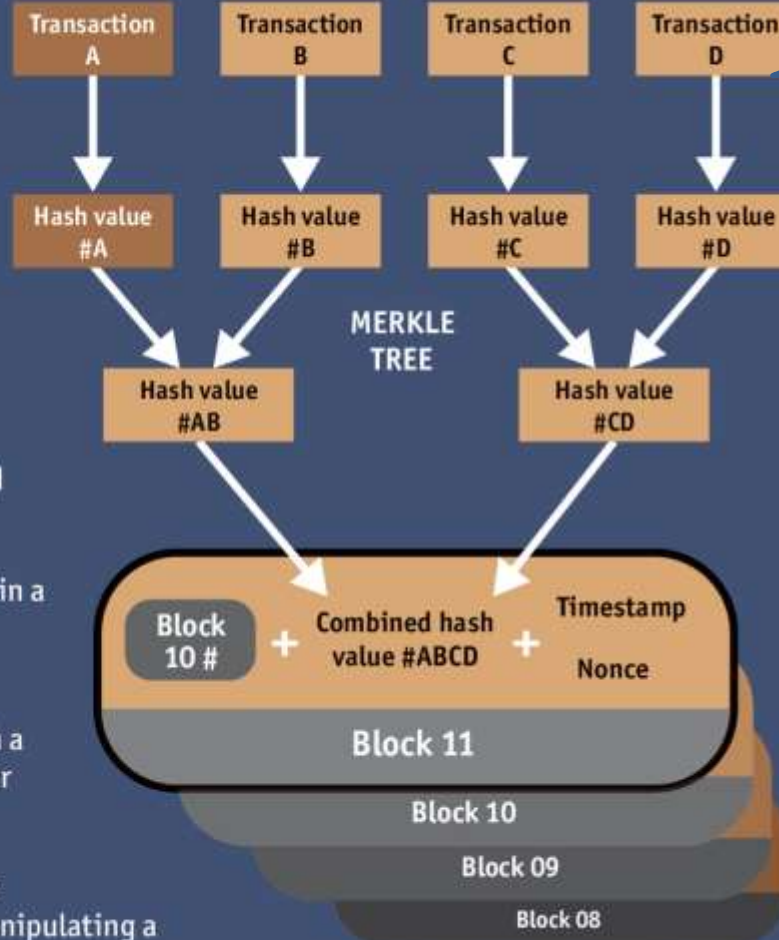
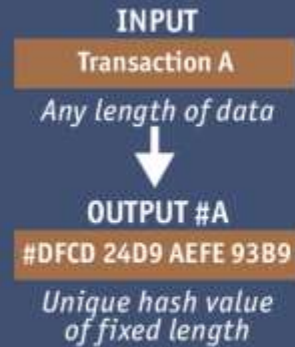
What is the blockchain?

In a simple diagram



What is the blockchain?

Making a hash of it



In a more complicated diagram!

Each transaction in the set that makes up a block is fed through a program that creates an encrypted code known as the hash value.

Hash values are further combined in a system known as a Merkle Tree.

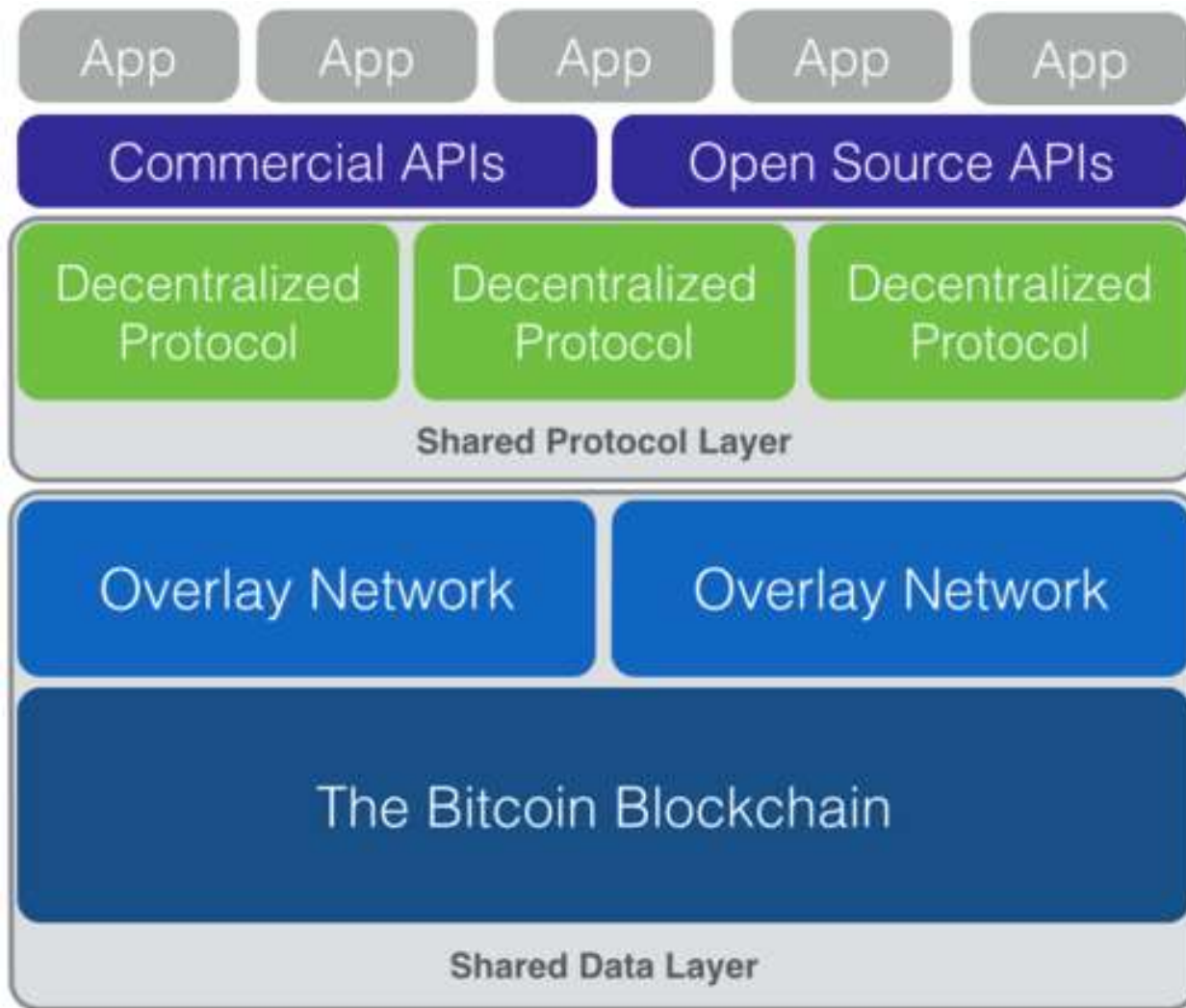
The result of all this hashing goes into the block's header, along with a hash of the previous block's header and a timestamp.

The header then becomes part of a cryptographic puzzle solved by manipulating a number called the nonce.

Once a solution is found the new block is added to the blockchain.

<http://www.economist.com/news/briefing/21677228-technology-behind-bitcoin-lets-people-who-do-not-know-or-trust-each-other-build-dependable>

What is the blockchain?



In a simple IT architecture diagram

Blockchain in quotations

“In the future all web applications are going to be following a blockchain structure.”

[Dan Larimer](#)

“The revolution will not be televised. It will be cryptographically time stamped on the block chain.”

[Dominic Frisby](#)

“Words like ‘blockchain’ and ‘mining’ and ‘dis-intermediation’ are as alien to us now as ‘browser’, ‘website’ or ‘URL’ were 20 years ago. But in a few years everyone will be using them.”

[Dominic Frisby](#)

“The blockchain is the new website. Fairly soon, every company will have a blockchain, or be on a blockchain, or several ones, just as organizations are involved in many websites today.”

[William Mougayar](#)

“We should think about the blockchain as another class of thing like the Internet; it is a new organizing paradigm for the discovery, valuation, and transfer of all quanta (discrete units) of anything.”

[Melanie Swan](#)



OK, but what is it in *simple* terms?

The blockchain explained ...
through cake and butlers!

In a 4 minute
video!



<https://youtu.be/wyfm92qqSh8>

In summary

“Blockchain technology is a big fancy word that describes the act of recording events in a database.” [Will Long](#)

But Blockchain Technology IS ...

- ... more than just about Bitcoin
- ... about secure, “trusted but trustless” transactions and the metadata associated with those transactions.
- ... about smart contracts and decentralization

And Blockchain Technology IS NOT ...

- ... just ONE blockchain – there are and can be many.
 - <http://www.ibm.com/blockchain/>
- ... just a place to manage digital cryptocurrencies.
 - *It's a lot more!*

→ Some examples of where blockchain technology is being used now ...



The blockchain, the ledger that underlies bitcoin, enables institutions to transfer and record assets other than the digital currency without an intermediary. Here's how it's being put to use.

<http://assets.bwbx.io/images/irGVsRUI8ODQ/v1/-1x-1.jpg>

Remittances

MeXBT in Mexico City provides a Web-based app that lets migrants send money via the blockchain to Mexico and withdraw cash from ATMs.



Diamonds

London startup Everledger uses 40 unique characteristics to 'fingerprint' large diamonds and record the stones on the blockchain to track them from the mine to someone's finger.



Property Titles/Land Records

Factom, based in Austin, Texas, is building a land title registry in Honduras on the blockchain so citizens can defend their property in court from unlawful seizures.



Interbank Payments

London-based Earthport and Ripple Labs, a San Francisco startup, have launched a new international payments network on a private blockchain.



Gold

Real Asset Co. in London permits gold bugs to record their bullion on the blockchain and will soon enable them to trade the metal.



Syndicated Loans

Digital Asset Holdings, Masters's firm, is creating a distributed ledger to handle the settlement of pooled corporate debt.



Private Shares

Nasdaq is developing a blockchain-based system for trading shares in closely held companies.



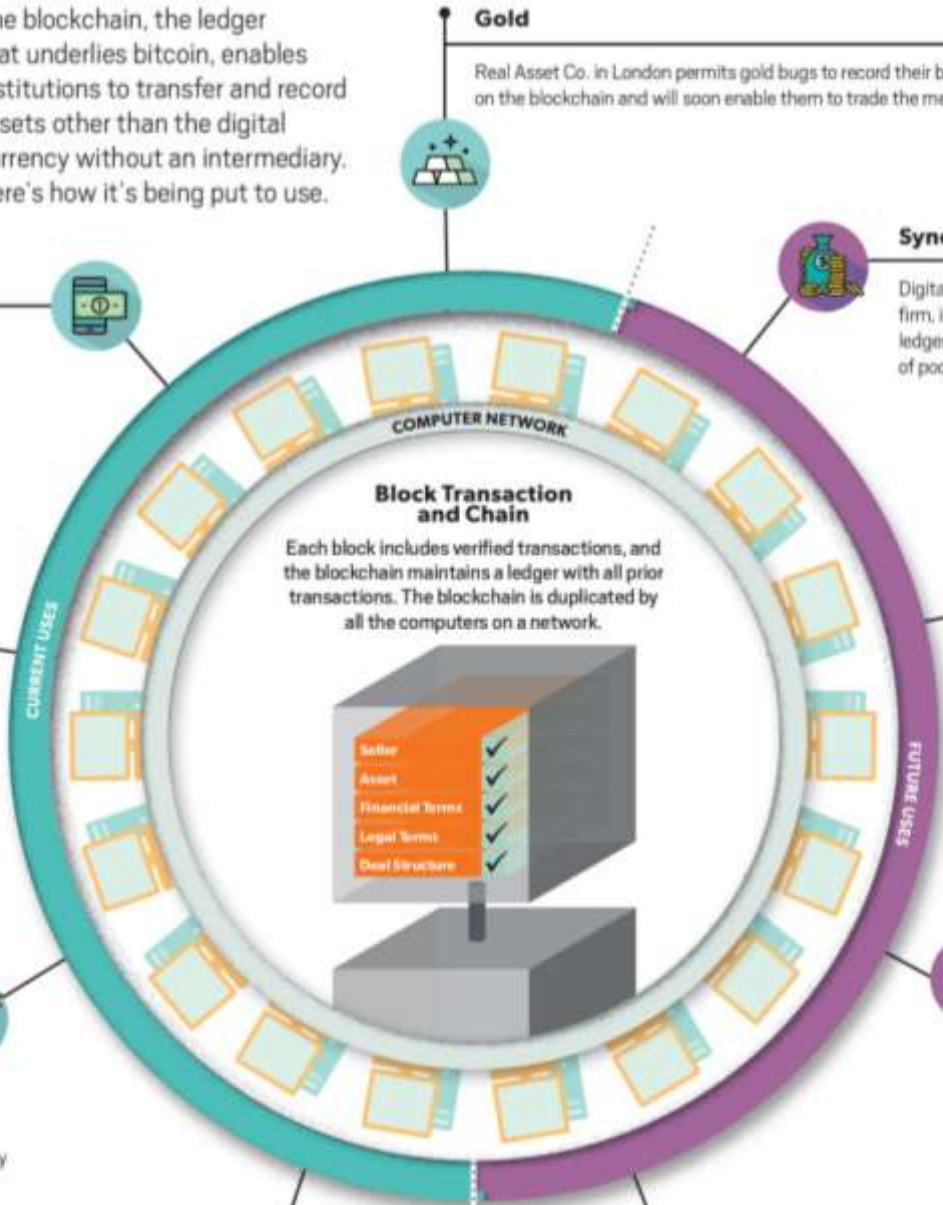
U.S. Treasury Repos

Digital Asset is developing a way to record and settle these short-term government bond trades on a distributed ledger.



Airline Miles

Chain, a San Francisco firm, offers software tools for developers to build apps to transmit anything of value on the ledger, including loyalty points.



Block Transaction and Chain

Each block includes verified transactions, and the blockchain maintains a ledger with all prior transactions. The blockchain is duplicated by all the computers on a network.



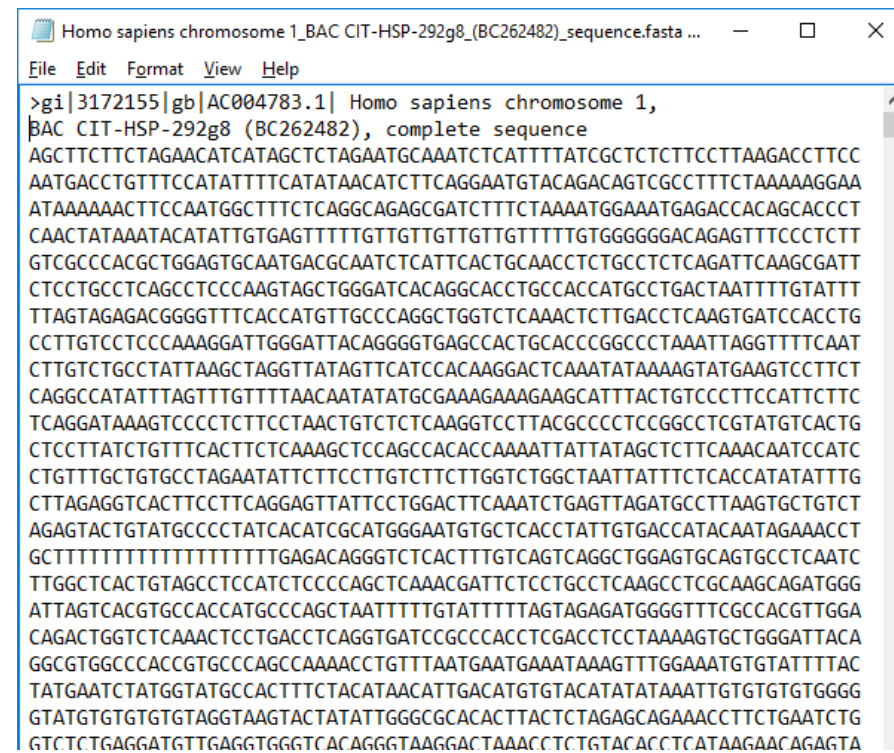
COMPUTER NETWORK

CURRENT USES

FUTURE USES

So how could a blockchain be used to secure and transact genomic data?

- Upfront – big current problem with the bitcoin blockchain!
 - Blocks are size constrained (1MB). Cannot (within the current bitcoin blockchain design) store large amounts of "raw" data in a block.
 - 1 full genome → 1 block is not possible ... yet
- For example: Complete sequence of Human Chromosome 1, BAC CIT-HSP-292g8 (BC262482) ...
 - 96,654 bp linear DNA
 - <https://www.ncbi.nlm.nih.gov/nuccore/3172155?report=fasta>
 - File size (FASTA): 96kB



```
Homo sapiens chromosome 1_BAC CIT-HSP-292g8_(BC262482)_sequence.fasta ...
File Edit Format View Help
>gi|3172155|gb|AC004783.1| Homo sapiens chromosome 1,
BAC CIT-HSP-292g8 (BC262482), complete sequence
AGCTTCTTCTAGAACATCATAGCTCTAGAATGCAAATCTCATTTTTATCGCTCTCTTCCCTAAGACCTTCC
AATGACCTGTTTCCATATTTTCATATAACATCTTCAGGAATGTACAGACAGTCGCCTTTCTAAAAAGGAA
ATAAAAACTTCCAATGGCTTTCAGGCAGAGCGATCTTCTAAAAATGGAAATGAGACCCACAGCACCTT
CAACTATAAATACATATTGTGAGTTTTTGTGTTGTTGTTGTTTTGTGGGGGGACAGAGTTTCCCTCTT
GTCGCCCCACGCTGGAGTGCAATGACGCAATCTCATTCACTGCAACCTCTGCCTCTCAGATTCAGCGGATT
CTCCTGCCTCAGCCTCCCAAGTAGCTGGGATCACAGGCACCTGCCACCATGCCTGACTAATTTTGTATTT
TTAGTAGAGACGGGGTTTACCATGTTGCCAGGCTGGTCTCAAACCTTTGACCTCAAGTGATCCACCTG
CCTGTCTCCCAAAGGATTGGGATTACAGGGGTGAGCCACTGCACCCGGCCCTAAATTAGGTTTTCAAT
CTTGCTGCCTATTAAGCTAGGTTATAGTTTATCCACAAAGACTCAAATATAAAGATGAAGTCCCTTCT
CAGGCCATATTTAGTTTTTTTTAACAATATATGCGAAAGAAAGAAGCATTTACTGTCCCTTCCATCTTC
TCAGGATAAAGTCCCTCTTCTAACTGTCTCTCAAGGCTTACGCCCTCCGGCCTCGTATGCTACTG
CTCCTTATCTGTTTCACTTCTCAAAGCTCCAGCCACACCAAATTTATAGTCTTCAAACAATCCATC
CTGTTTGTCTGTGCCTAGAATATTTCTTCTTGTCTTCTTGGTCTGGCTAATTTATTTCTCACCATATATTTG
CTTAGAGGTCACCTTCCAGGAGTTATCTTGGACTTCAAATCTGAGTTAGATGCCTTAAGTGTCTGTCT
AGAGTACTGTATGCCCCATCACATCGCATGGGAATGTGCTCACCTATTGTGACCATAACAATAGAAACCT
GCTTTTTTTTTTTTTTTTTTGTGAGACAGGGTCTCACTTTGTGACAGTCAAGGCTGGAAGTGCAGTGCCTCAATC
TTGGCTCACTGTAGCTCCATCTCCCCAGCTCAAACGATTCTCTGCTCAAGCTCGCAAGCAGATGGG
ATTAGTCAGTGCACCATGCCAGCTAATTTTTGTATTTTGTAGTAGAGATGGGGTTTCCGCCAGTTGGA
CAGACTGGTCTCAAACCTCTGACTCAGGTGATCCGCCACCTCGACCTCTCAAAGTGTCTGGGATTACA
GGCGTGGCCACCCTGCCCCAGCCAAAACCTGTTTAAATGAATGAAATAAAGTTTGGAAATGTGATTTTAC
TATGAATCTATGGTATGCCACTTCTACATAACATTGACATGTGACATATATAAATGTGTGTGGGG
GTATGTGTGTGTAGGTAAGTACTATATTGGGCGCACACTTACTCTAGAGCAGAAACCTTCTGAATCTG
GTCCTGAGGATGTTGAGGTGGGTACAGGGTAAGGACTAAACCTCTGTACACCTCATAAAGAACAGAGTA
```

How to resolve the block size problem for genomic data?

- One way would be to “hash” the complete genomic data file.
 - Store the genomic data file itself “elsewhere”.
- Store the file hash plus information about the file’s location in a blockchain transaction using an OP_RETURN script.
 - See: “Mastering Bitcoin”, by Andreas M. Antonopoulos, O’Reilly Media, 2015, ISBN: 978-1-449-37404-4, pp132-134

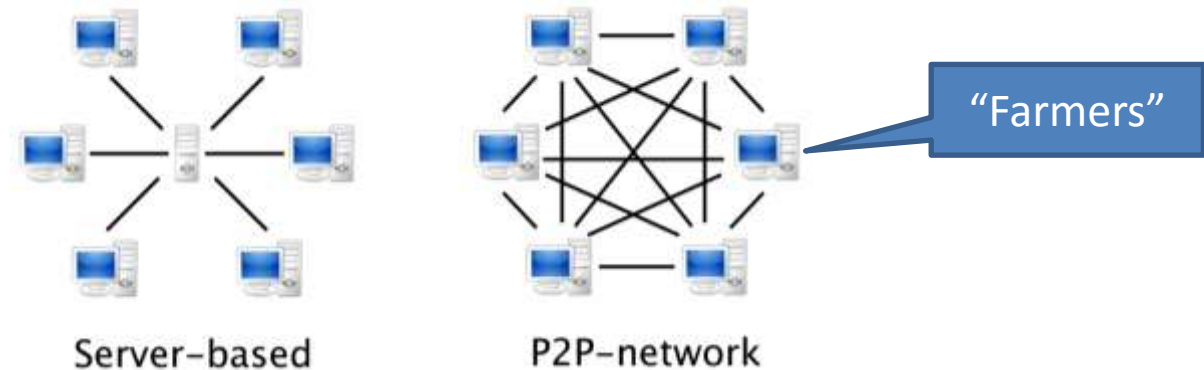
... BUT

- Still got the issue of storing the genomic data file securely, somewhere.
 - NB: This approach is used by Proofofexistence.com “to anonymously and securely store an online distributed proof of existence for any document”
 - Useful for inventions and patents.

Not resolved the “Jennifer Lawrence” issue

Another way to resolve the block size problem for genomic data?

- Divide the genomic data across multiple blocks.
- Approach pioneered by [Storj.io](https://storj.io) for “data”
 - Using a peer-to-peer storage network



- <http://cointelegraph.com/news/blockchain-decentralized-cloud-storage-storj-and-competitors>

How might it work?



- Subdivide/shred genomic datafiles into "shards"
- Encrypt/hash the shards uniquely
- Set up peer-to-peer storage of shards
- Store file metadata on a blockchain
 - File hash; network locations of the copies of the shards; Merkle roots
 - File itself is not stored remotely
- Reconstitute file by retrieving file metadata from the "farmers" (peers); decrypt to get back to the shard locations; retrieve shards and rebuild file.

See <http://storj.io/storj.pdf> for more detail on how this is achieved

Sounds complicated!



- Yes, but it can be simply programmed.
 - Blockchain technology could have prevented Jennifer Lawrence's photos from being hacked
 - NB: Does not mitigate against phishing



- P2P file storage services are growing.
 - Other blockchain supported file storage services
 - Tierion.com; Sia.tech; Neuroware.io (Everstore.io)



- There are now blockchain services being offered for Pharma, e.g. Drug Development Supply Chain

- Blockrx.com



So what could happen in the future with genomic data stored on the blockchain ...?

What could blockchain technology enable for YOUR genomic data?

Dedicated blockchain(s) to store YOUR healthcare data and records securely.
e.g. Genecoin, blockchain storage service specifically for genomic data.



- Smart contracts could be set up to allow individuals to control what happens to THEIR genomic data, who can access it and work with it.
- Individuals could potentially be "paid" for allowing others to access/use THEIR data.

- With YOUR data stored decentralised over many peers and replicated in many blocks, you have built-in redundancy.
 - If one "peer" goes down you do not lose YOUR data.

Provided you do not lose your private key(s)!

What's the benefit? What's the risk?



Some of the Benefits

- Of taking back control
 - Choose who can have access to and use YOUR data
 - Choose what a 3rd party can do with YOUR genome
 - Choose what price to access YOUR data
- Of using the blockchain
 - "Unhackable" security
 - All transactions irrefutably tracked, managed, controlled and time-stamped

Some of the Risks

- Just how secure and "unhackable" is the blockchain?
- Will blockchain technology scale if everyone chooses to use it?
- If you lose your private key, you won't get your file(s) back.
 - Or if you reveal your private key(s) in a phishing attack.

Still early days for healthcare information on blockchain ... so watch this space!

What have we covered?

- Three famous women – connected
- The growth of YOUR online data and content
- The need for controlled, secure access to YOUR data and content
- A potential solution: blockchain technology
- What “the blockchain” is and what it is not
- How blockchain technology could be used to store, secure and transact genomic data
- Why securing YOUR genomic data using blockchain technology could be a good idea



Blockchain Technology could
securely help



take back control of



Thank you!

Any Questions?



www.curlewresearch.com

Enabling Life Science Success

Selected Additional Reading

- “Blockchain : Blueprint for a New Economy” by Melanie Swan, O’Reilly Media, 2015, ISBN: 978-1-4919-2049-7
- <http://blockstrap.com/en/a-complete-beginners-guide-to-blockchain-technology/>
- <http://radar.oreilly.com/2015/01/understanding-the-blockchain.html>
- <http://ledracapital.com/blog/2013/12/28/the-ledra-bitcoin-series>
- <http://www2.deloitte.com/uk/en/pages/innovation/articles/blockchain.html>
- <http://bitsonblocks.net/2015/09/09/a-gentle-introduction-to-blockchain-technology/>

The Genesis Block: Block #0

Height	0 (Main chain)
Hash	000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f
Previous Block	00
Next Blocks	00000000839a8e6886ab5951d76f411475428afc90947ee320161bbf18eb6048
Time	2009-01-03 18:15:05
Difficulty	1
Bits	486604799
Number Of Transactions	1
Output Total	50 BTC
Estimated Transaction Volume	0 BTC
Size	0.285 KB
Version	1
Merkle Root	4a5e1e4baab89f3a32518a88c31bc87f618f76673e2cc77ab2127b7afdeda33b
Nonce	2083236893
Block Reward	50 BTC
Transaction Fees	0 BTC

A Sample Block: #403000

Height	403000 (Main chain)
Hash	000000000000000002b3b820d8d9af0a81a22fce882f5e055448f18dcc4975fd
Previous Block	0000000000000000026f687f27adf4f09eb6dd89833f2de6438ccf904a9df43b
Next Blocks	00000000000000000005ba8ee291ee724d1fe9065924bd981eb0d5ea31a9367e
Time	2016-03-17 00:20:01
Difficulty	158,427,203,767.39
Bits	403108008
Number Of Transactions	1925
Output Total	67,929.77960121 BTC
Estimated Transaction Volume	2,766.07072324 BTC
Size	998.192 KB
Version	4
Merkle Root	eefb62e10871c58d5f840a780cbef8ecd5cb4f06eea347d2588c5ec674469d2e
Nonce	317549265
Block Reward	25 BTC
Transaction Fees	0.396025 BTC

<https://blockchain.info/block-height/403000>

