Karlsruhe Institute of Technology

Where did my data go?
Evaluation of Distributed Ledger Technologies' Suitability for Personal Data Provenance

Bachelor's Thesis of
Aleksandar Bachvarov

Reviewer: *Prof. Dr. Hannes Hartenstein*
Second reviewer: *Prof. Dr. Ali Sunyaev*

Advisor: *Oliver Stengele M.Sc.*
Second advisor: *Jan Bartsch M.Sc.*

Contents

Contents

1.1 Fundamentals

- Data Provenance (DP)

> *"The provenance of a data item includes information about the processes and sources that lead to its creation and current representation"*

[p.3;1]

1.1 Fundamentals

- Data Provenance (DP)

> *"The provenance of a data item includes information about the processes and sources that lead to its creation and current representation"*

[p.3;1]

- Personal Data

> *“Any information relating to an identified or identifiable natural person ('data subject')”*

[p.1;2]

1.2 Fundamentals

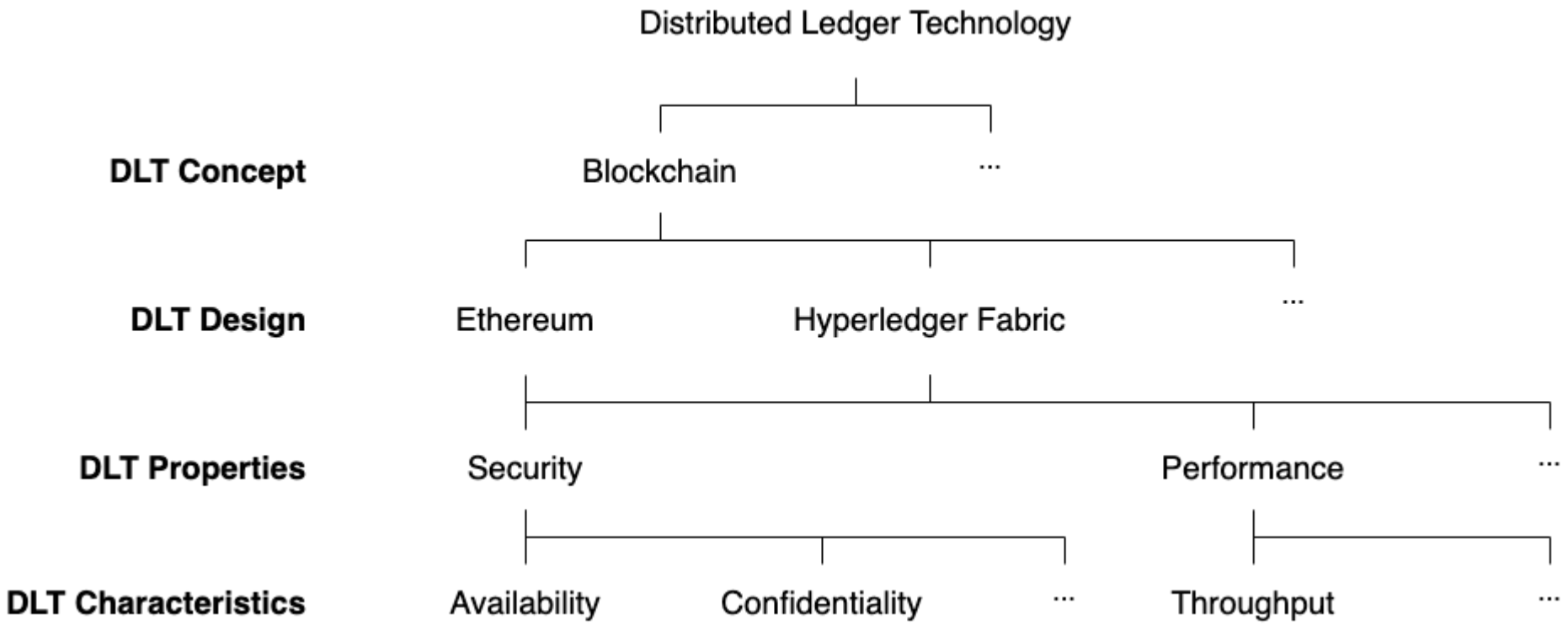
- Distributed Ledger Technology (DLT)

Consensus of replicated, shared, and synchronized digital data geographically spread across multiple sites, countries, or institutions. Unlike with a centralized database, there is no central administrator.

[3,4]

  - public
  - private
  - permissioned
  - permissionless

1.2 Fundamentals



based on fig.1 (p.42:4) from N. Kannengießer et al. [5]

1.2 Current State

- Healthcare industry - 30% of the world's data volume generated [6,7]
- Countries globally adopting FinTech [8]
- Big Data Technology in finance [9]
- Fines are growing (2021) - €1,345,200,000 [10]

1.3 Potential Solution

- DLT [11-14]

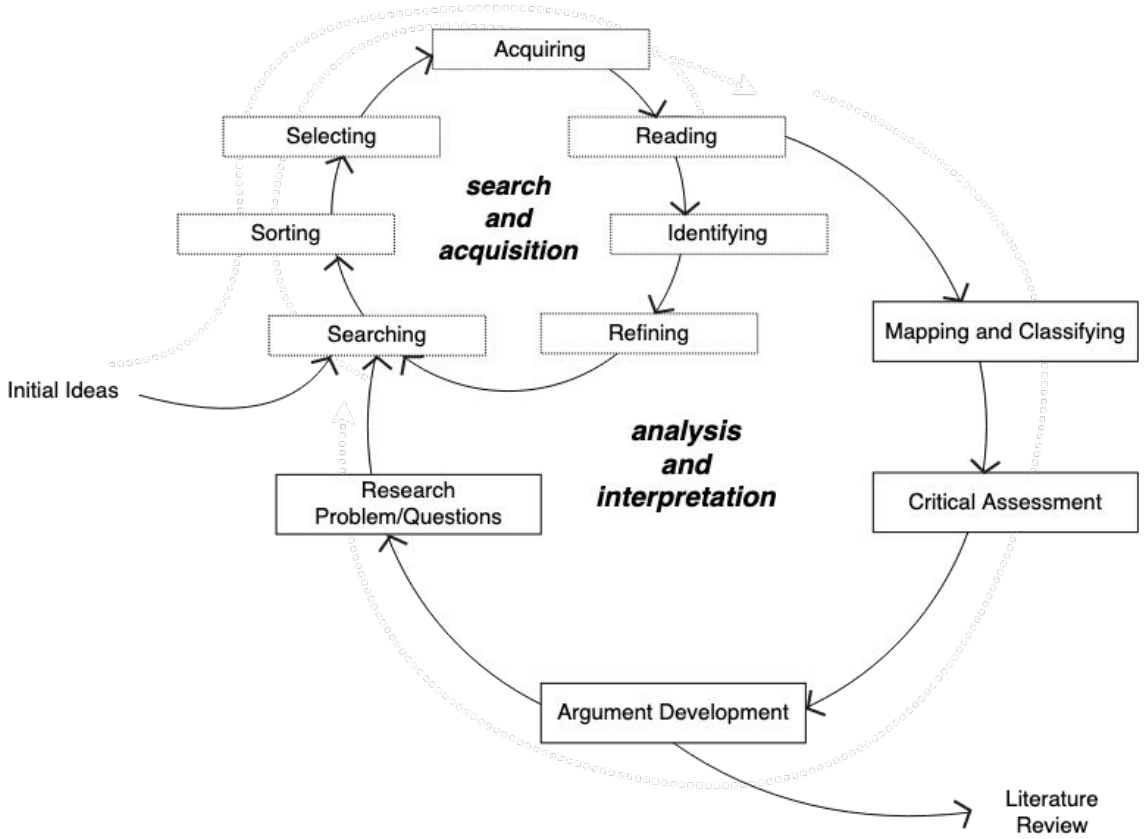
- Integrity
- Authenticity
- Transparency
- Accountability
- Trustworthiness

through

- Immutable record of transactions
- Lack of single authority
- Consensus mechanisms
- Smart contracts
- Tamper-proof storage of data

Contents

2.1 Framework and Objectives



based on fig.1 p.264 from S. K. Boell et. al. [15]

RQ: *What are the characteristics of DLTs that make them suitable for personal DP in healthcare and finance?*

Q1: What are the fundamental requirements for a personal DP approach?

Q2: What is the importance of individual DP requirements in terms of healthcare and finance?

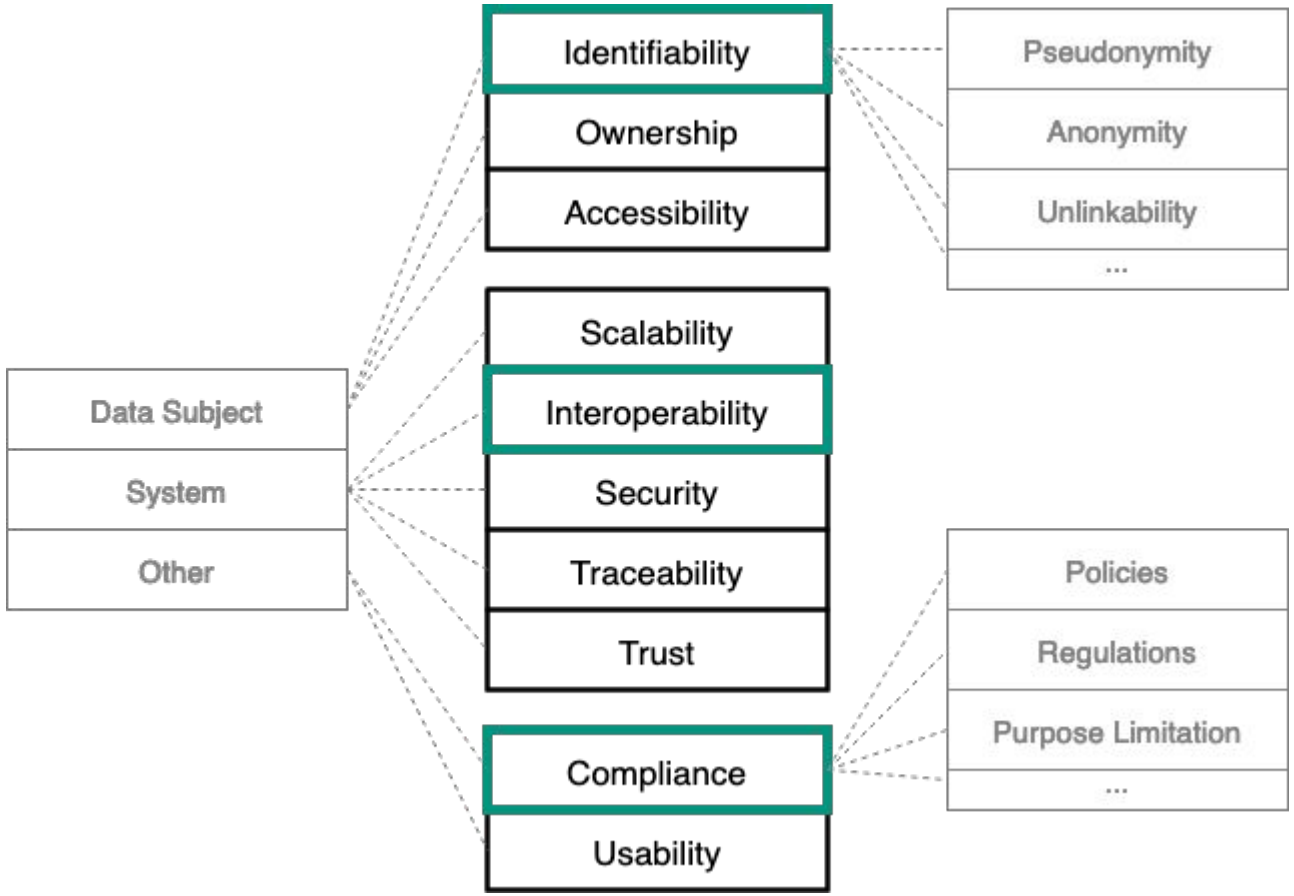
Q3: Which are the preferred DLT approaches for DP in healthcare and finance and why?

Q4: Which requirements are considered important by DLT-based DP approaches in healthcare and finance?

Contents

3.1 Requirements

Q1: What are the fundamental requirements for a personal DP approach?



3.2 Use Cases

Q2: What is the importance of individual DP requirements in terms of healthcare and finance? [16-22]

- Identification of physicians vs Anonymisation of patients
- Potentially better treatment and patient safety in emergency situations
- Can paradoxically harm patients' well-being in the long run

Identifiability

Interoperability

Compliance

- Identifiability challenges privacy but Anonymity fosters crime
- Faster transactions and lower fees
- Laws and regulations benefit the consumer by ensuring transparency, security and trust

3.3 Considered Approaches

Q3: Which are the preferred DLT *approaches* for DP in healthcare and finance and why?

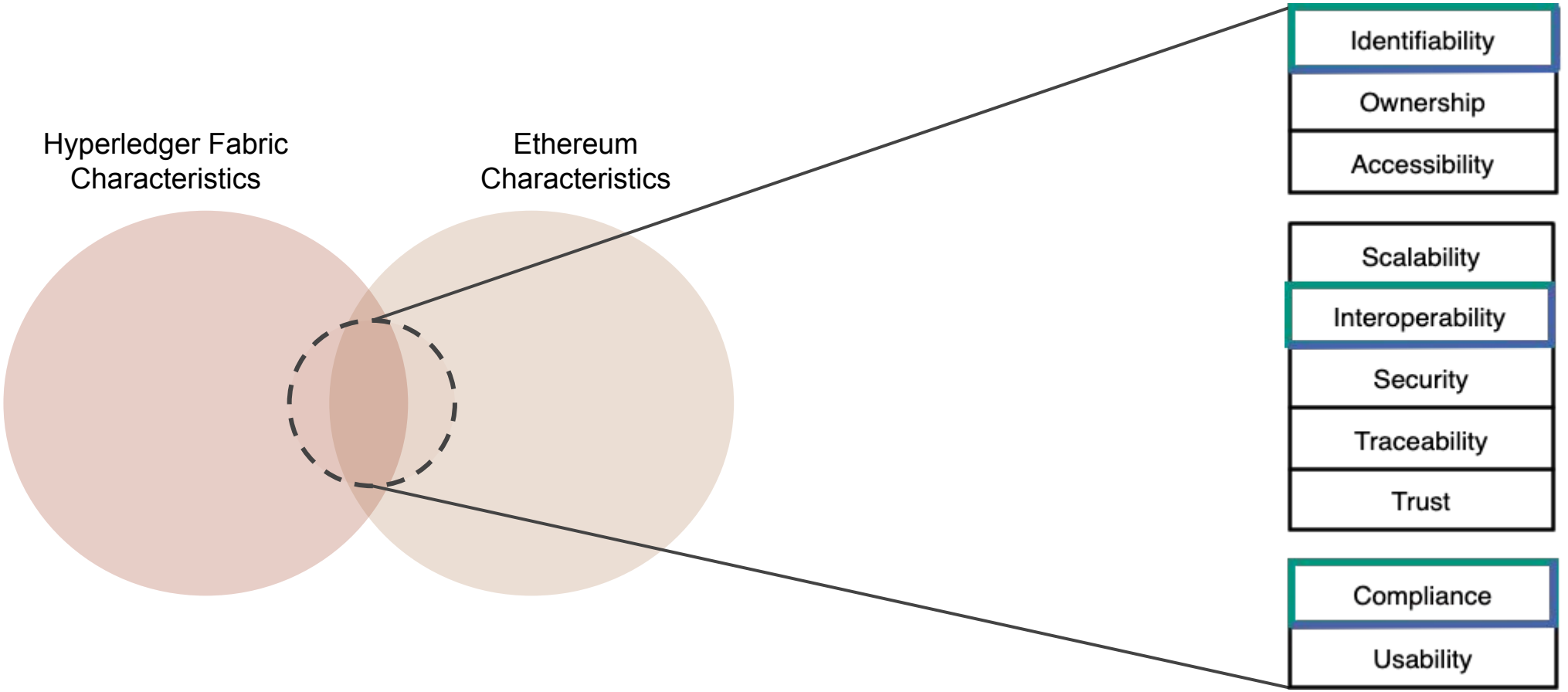
Hyperledger Fabric [23]

- Private
- Channelisation features
- Permissions and privacy
- Scalability
- etc.

Ethereum [24]

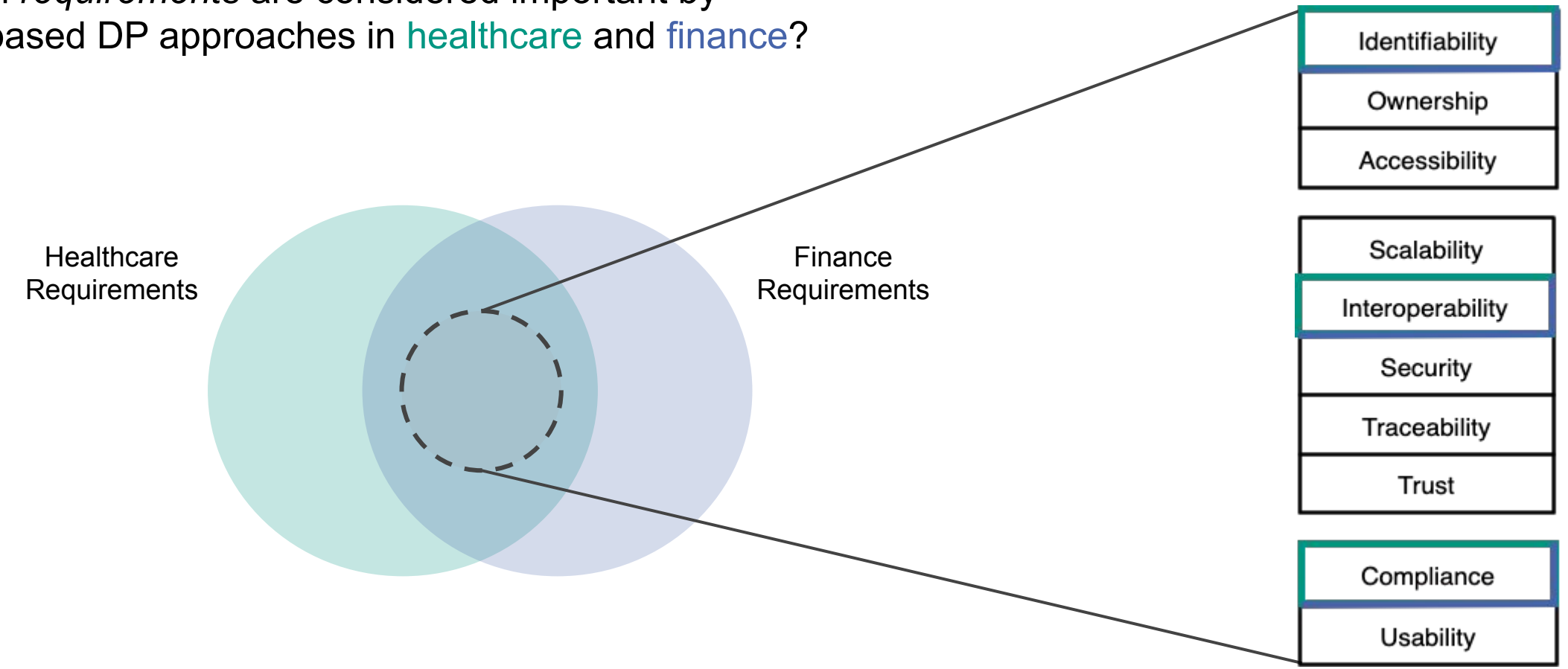
- Public
- High availability
- General purpose platform
- DAOs, DACs
- etc.

3.3 Considered Approaches



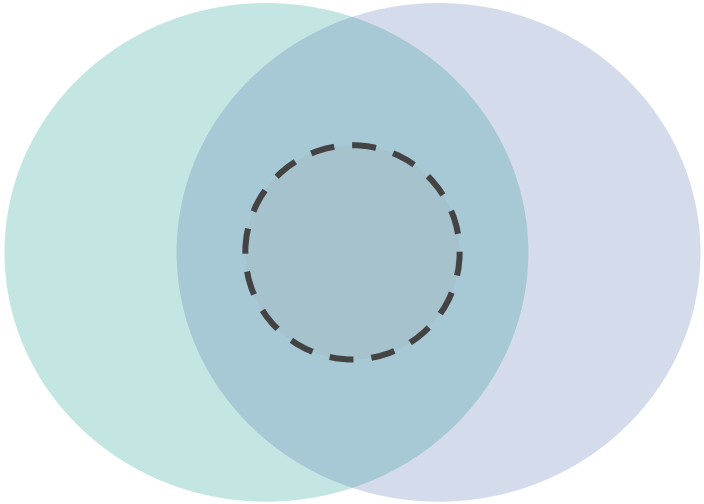
3.4 Considered Requirements

Q4: Which *requirements* are considered important by DLT-based DP approaches in healthcare and finance?



3.4 Considered Requirements

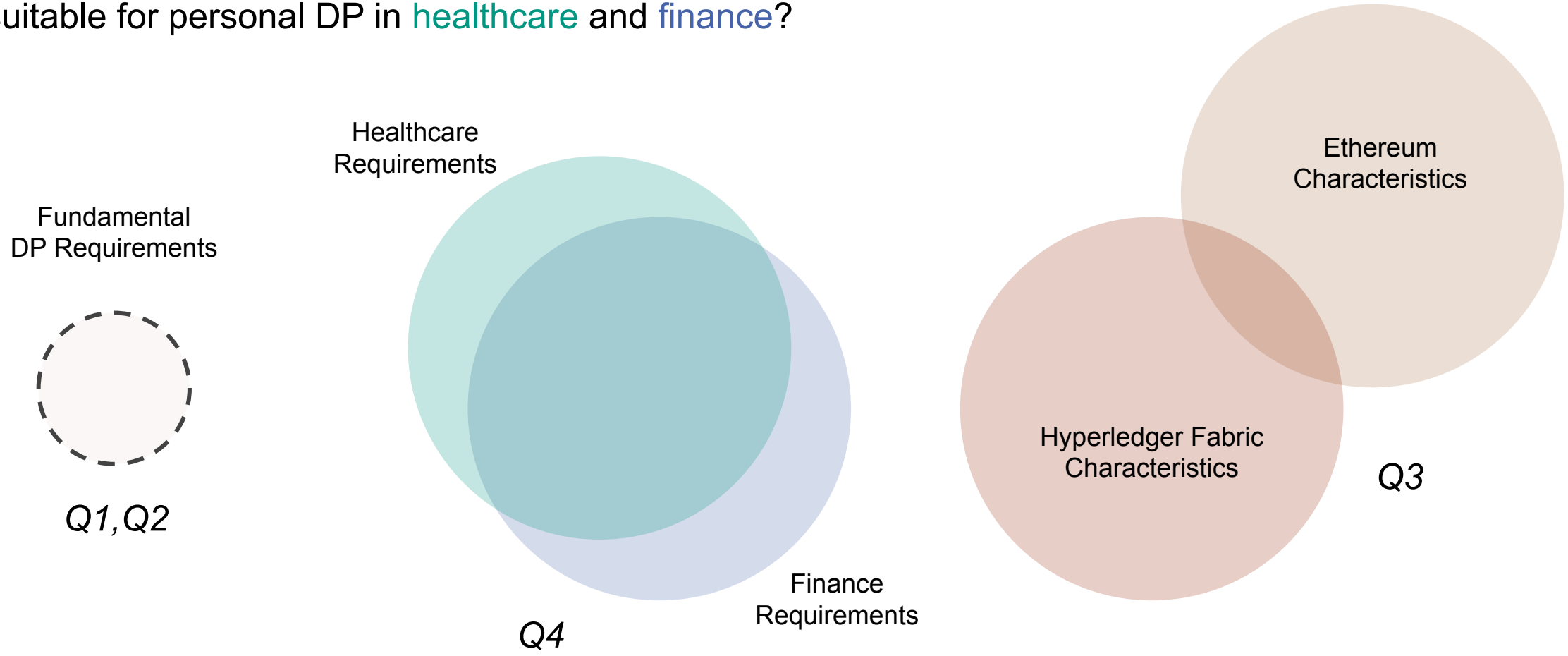
[25-30]



Healthcare (3)	Both (12+10)	Finance (5)
• Emergency Availability • Physician Responsibility • Patient Engagement	• Smart-contracts • Non-repudiation • Standardisation • etc.	• Cost, Corruption, Crime Reduction • Transaction Speed • Disintermediation

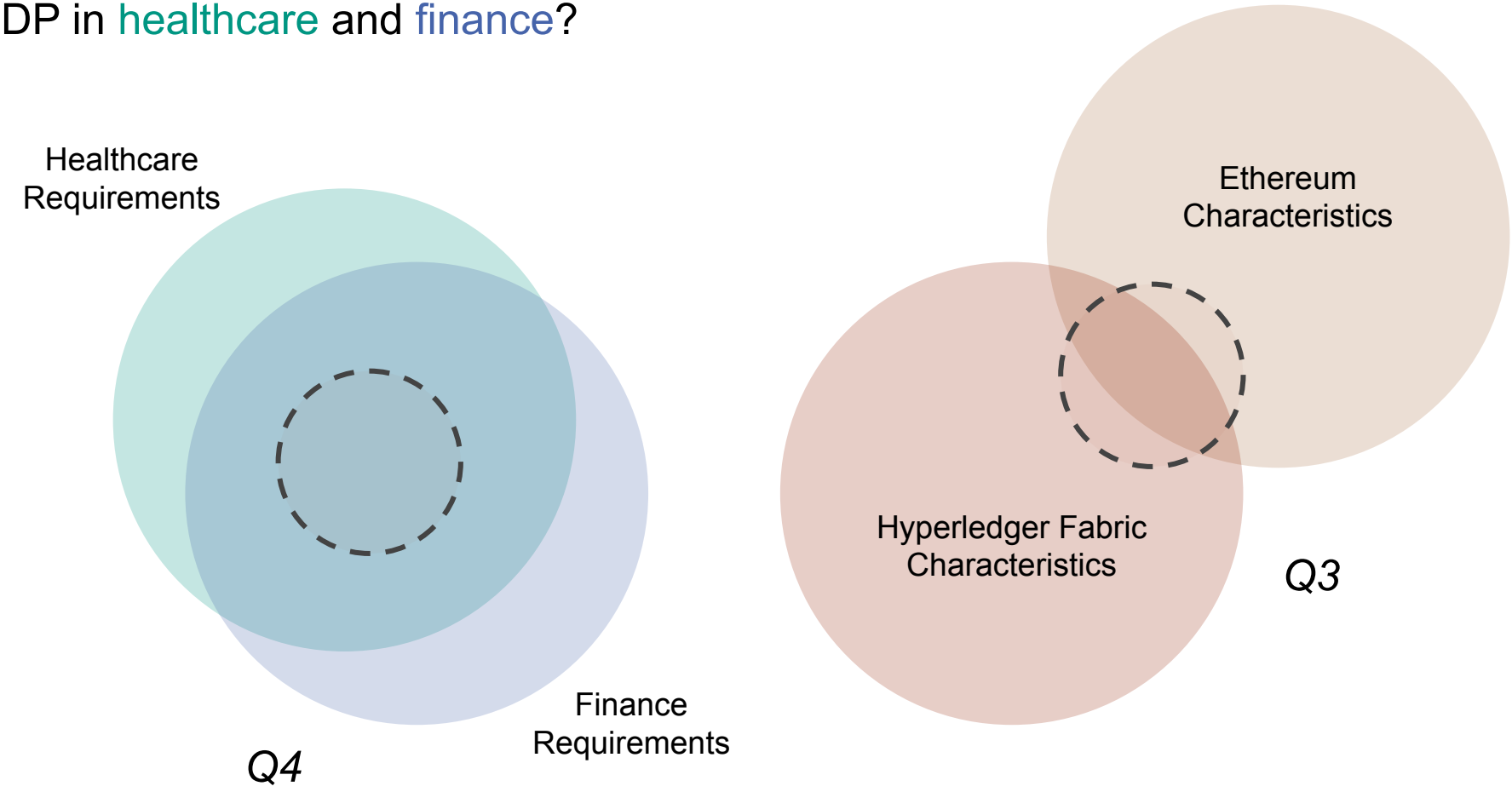
3.5 Mappings

RQ: What are the *characteristics* of DLTs that make them suitable for personal DP in healthcare and finance?



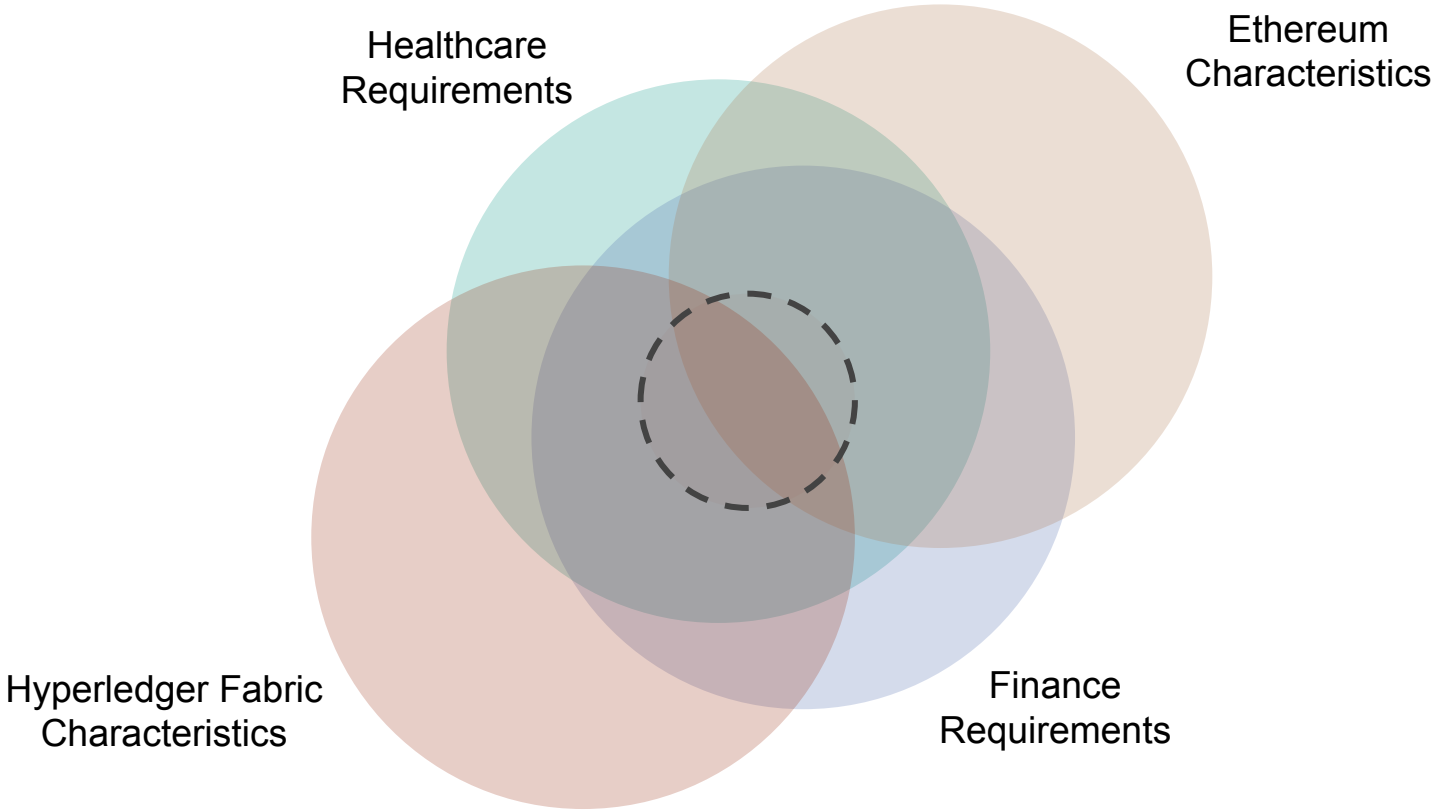
3.5 Mappings

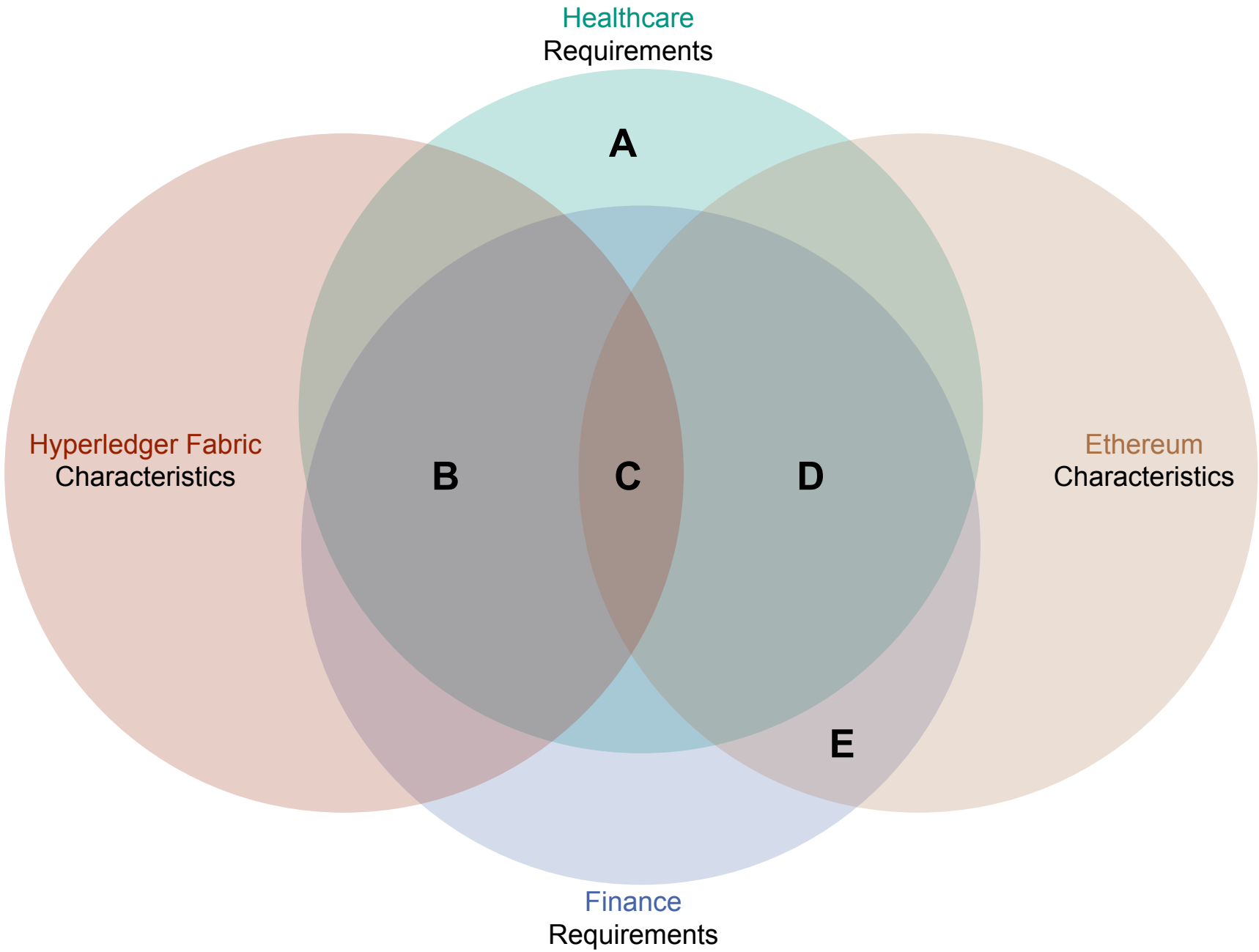
RQ: What are the *characteristics* of DLTs that make them suitable for personal DP in healthcare and finance?



3.5 Mappings

RQ: What are the *characteristics* of DLTs that make them suitable for personal DP in healthcare and finance?





A Responsibility

B Efficiency
Comp. Resources
Cost Reduction
Trans. Speed

Interoperability

Scalability

C Smart Contracts
Transparency
Standardisation
Immutability
Verifiability
Non-repudiation
Stability
Consistency
Authenticity

Identifiability

Ownership
Security
Traceability
Usability

D Decentralisation

Compliance

Trust

E Disintermediation

Contents

4.1 Implications for Practice and Research

- Implications for Practice
  - Viability of DLTs for personal DP
  - Aid selection of DLT designs
  - Benefits and drawbacks
  - Technology inline with international data regulations

4.1 Implications for Practice and Research

- Implications for Research
  - Underlining important aspects and relationships
  - An easier comparative analysis
  - Overcoming the explored drawbacks
  - Beyond Hyperledger Fabric and Ethereum
  - Beyond healthcare and finance

4.2 Limitations and Future Work

- Limitations
  - *Blockchain* and *personal* DP
  - Breadth-first search methodology
  - Requirements definitions
  - Predominant focus on potentially positive influences

4.2 Limitations and Future Work

- Future Work
  - Degree of influenced or satisfaction
  - Requirements' influence on each other
  - DLTs' further application in healthcare and finance
  - More prototypes, proof-of-concepts
  - *Interoperability, Scalability*

Conclusion

- Personal DP in Healthcare and Finance
- Private Permissioned vs Public Permissionless
- No one-size-fits-all solution
- Worth consideration

5. References

[1] Buneman P, Khanna S, Tan W-C. Data Provenance: Some Basic Issues. In: Kapoor S, Prasad S, editors. FST TCS 2000: Foundations of Software Technology and Theoretical Computer Science [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2000 [cited 2021 Jun 5]. p. 87–93. (Goos G, Hartmanis J, van Leeuwen J, editors. Lecture Notes in Computer Science; vol. 1974). Available from: http://link.springer.com/10.1007/3-540-44450-5_6

[2] TrueVault. What is personal data? - TrueVault [Internet]. 2018; Available from: https://www.truevault.com/learn/what-is-personal-data

[3] Maull R, Godsiff P, Mulligan C, Brown A, Kewell B. Distributed ledger technology: Applications and implications. Strategic Change. 2017;26(5):481–9.

[4] Distributed ledger technology: beyond block chain [Internet]. GOV.UK; Available from: https://www.gov.uk/government/news/...block-chain

5. References

[5] Niclas Kannengießer et al. “Trade-offs between distributed ledger technology characteristics”. In: ACM Computing Surveys (CSUR) 53.2 (2020), pp. 1–37.

[6] Coughlin et al International Medicine Journal article "Looking to tomorrow's healthcare today: a participatory health perspective". IDC White Paper, Doc# US44413318, November 2018: The digitalisation of the World - From Edge to Core

[7] Healthcare data volume globally 2020 forecast. en. url: https://www.statista.com/statistics/1037970/global-healthcare-data-volume/

[8] Tech Nation 2020; EY 2019. Consumer fintech adoption rates by country en. url: https://technation.io/global-consumer-fintech-adoption-rates-tech-nation-report-2020/

5. References

[9] Frederic Boissay et al. “Big techs in finance: on the new nexus between data privacy and competition”. In: The Palgrave Handbook of Technological Finance. Springer, 2021, pp. 855–875.

[10] Niall McCarthy. The Biggest GDPR Fines of 2021.Updated: 03.03.2022 https://www.eqs.com/compliance-blog/biggest-gdpr-fines-2021/

[11] Scardovi C. Restructuring and Innovation in Banking. 1st ed. 2016. Cham: Springer International Publishing : Imprint: Springer; 2016.

[12] Ray S. The Difference Between Blockchains & Distributed Ledger Technology [Internet]. Medium. 2021 [cited 2021 Jun 5]. Available from: https://towardsdatascience.com/the-difference...92

5. References

[13] Blockchain Basics: Introduction to Distributed Ledgers [Internet]. IBM Developer. [cited 2021 Jun 5]. Available from: https://developer.ibm.com/technologies/blockchain/...x-trs/

[14] Blockchains & Distributed Ledger Technologies [Internet]. BlockchainHub; Available from: https://blockchainhub.net/...in-general/

[15] Sebastian K Boell and Dubravka Cecez-Kecmanovic. “A hermeneutic approach for conducting literature reviews and literature searches”. In: Communications of the Association for information Systems 34.1 (2014), p. 12.

[16] Sue M Evans et al. “Anonymity or transparency in reporting of medical error: a community-based survey in South Australia”. In: Medical Journal of Australia 180.11 (2004), pp. 577–580.

5. References

[17] Use Case private data use - XG Provenance Wiki. url: https://www.w3.org/2005/Incubator/prov/wiki/Use_Case_private_data_use (visited on 2021- 11-14)

[18] Andrea Margheri et al. “Decentralised provenance for healthcare data”. en. In: International Journal of Medical Informatics 141 (Sept. 2020), p. 104197. issn: 1386-5056. doi: 10.1016/j.ijmedinf.2020.104197. url: https://www.sciencedirect.com/science/article/pii/S1386505619312031 (visited on 2021-06-05)

[19] Claes-Goran Westrin and Tore Nilstun. “The ethics of data utilisation: a compar- ison between epidemiology and journalism”. In: BMJ 308.6927 (1994), pp. 522– 523.

[20] Dennis D Steinauer, Shukri A Wakid, and Stanley Rasberry. “Trust and trace-ability in electronic commerce”. In: StandardView 5.3 (1997), pp. 118–124.

5. References

[21] Steffen Hoernig and Marc Bourreau. Interoperability of mobile money International experience and recommendations for Mozambique. 2017.

[22] Muzammil Hussain et al. “Security and Privacy in FinTech: A Policy Enforcement Framework”. In: Research Anthology on Concepts, Applications, and Chal- lenges of FinTech. IGI Global, 2021, pp. 372–384.

[23] Dannen, Chris. Introducing Ethereum and solidity. Vol. 318. Berkeley: Apress, 2017.

[24] Dhillon V., Metcalf D., Hooper M. (2017) The Hyperledger Project. In: Blockchain Enabled Applications. Apress, Berkeley, CA. Available from: https://doi.org/10.1007/978-1-4842-3081-7_10

5. References

[25] Anton Hasselgren et al. “Blockchain in healthcare and health sciences—A scoping review”. In: International Journal of Medical Informatics 134 (2020), p. 104040.

[26] Maged N Kamel Boulos, James T Wilson, and Kevin A Clauson. Geospatial blockchain: promises, challenges, and scenarios in health and healthcare. 2018.

[27] Igor Radanović and Robert Likić. “Opportunities for use of blockchain technology in medicine”. In: Applied health economics and health policy 16.5 (2018), pp. 583–590.

[28] Soonduck Yoo. “Blockchain based financial case analysis and its implications”. In: Asia Pacific Journal of Innovation and Entrepreneurship (2017).

5. References

[29] Martina Bettio et al. “Hyperledger fabric as a blockchain framework in the financial industry”. In: The Impact of Digital Transformation and FinTech on the Finance Professional. Springer, 2019, pp. 29–44.

[30] Samuel Fosso Wamba et al. “Bitcoin, blockchain and fintech: a systematic review and case studies in the supply chain”. In: Production Planning & Control 31.2-3 (2020), pp. 115–142.

- Emergency Availability
- Physician Responsibility
- Patient Engagement

- Smart-contracts
- Verifiability
- Standardisation
- Transparency
- Decentralisation
- Immutability
- Efficiency
- Authenticity
- Comp. Resources
- Non-repudiation
- Consistency
- Stability

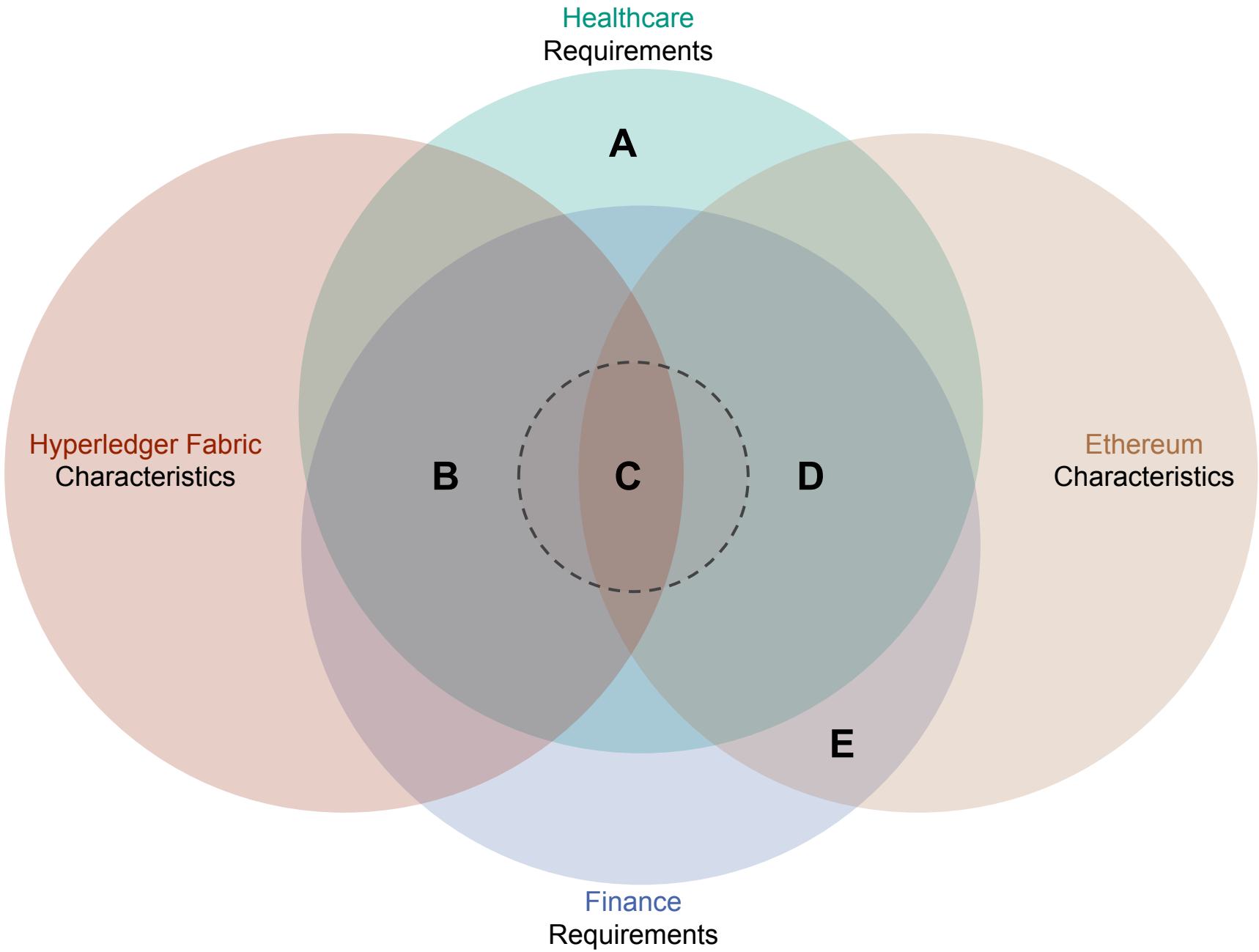
- Identifiability
- Ownership
- Accessibility
- Scalability
- Interoperability
- Traceability
- Security
- Trust
- Security
- Trust

- Cost Reduction
- Corruption Reduction
- Crime Reduction
- Transaction Speed
- Disintermediation

- Node Controller Verification
- Resource Consumption
- Scalability
- Throughput
- Interoperability
- Maintainability
- Confidentiality

- User Unidentifiability
- Traceability
- Smart Contracts
- Non-Repudiation
- Durability
- Integrity
- Consistency
- Str. of Cryptography
- Auditability
- Ease of Use
- Transaction Fee

- Transaction Content Visibility
- Token Support
- Availability
- Censorship Resistance
- Compliance
- Degree of Centralisation
- Incentive Mechanism
- Ease of Node Setup



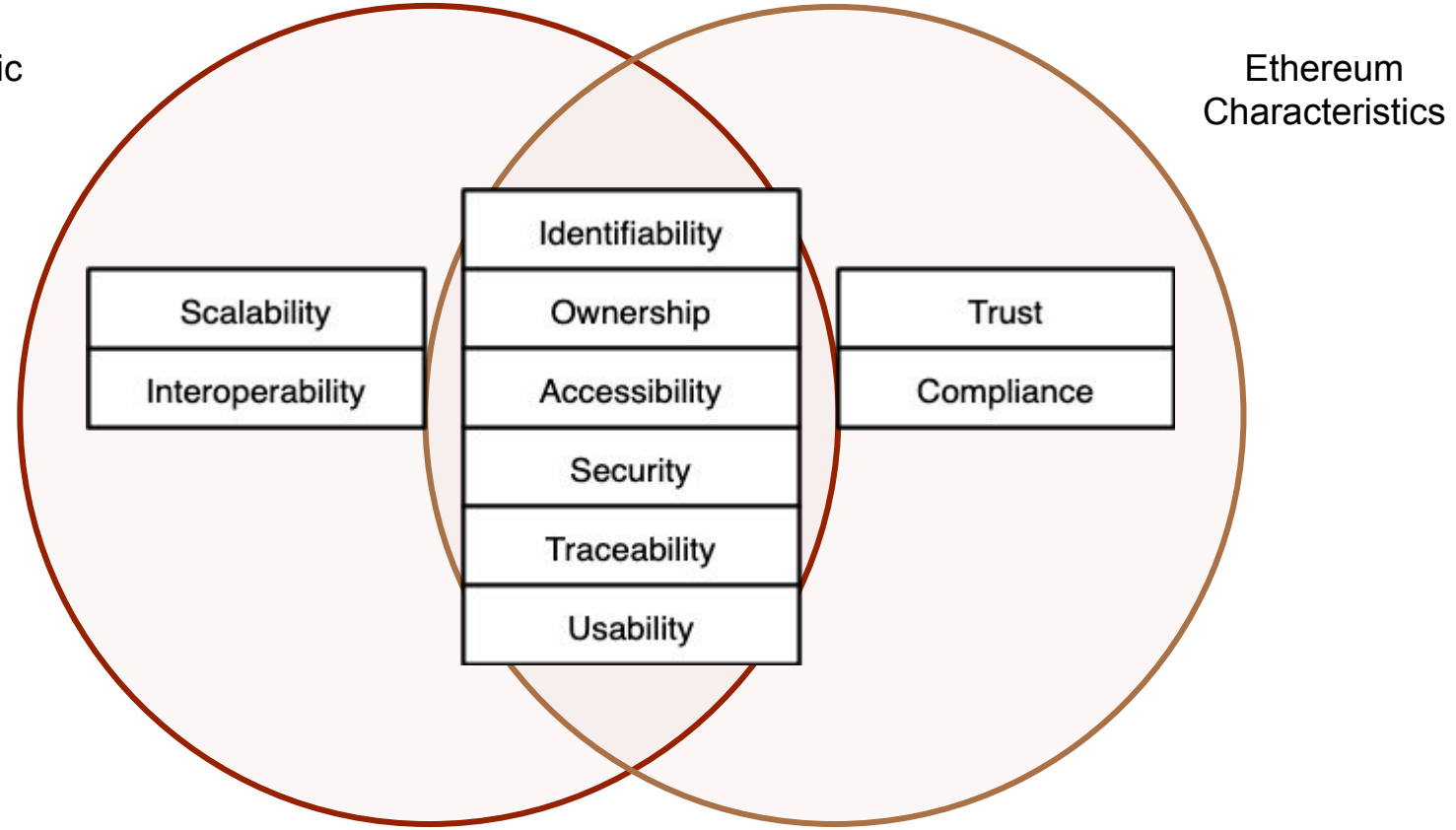
A Responsibility

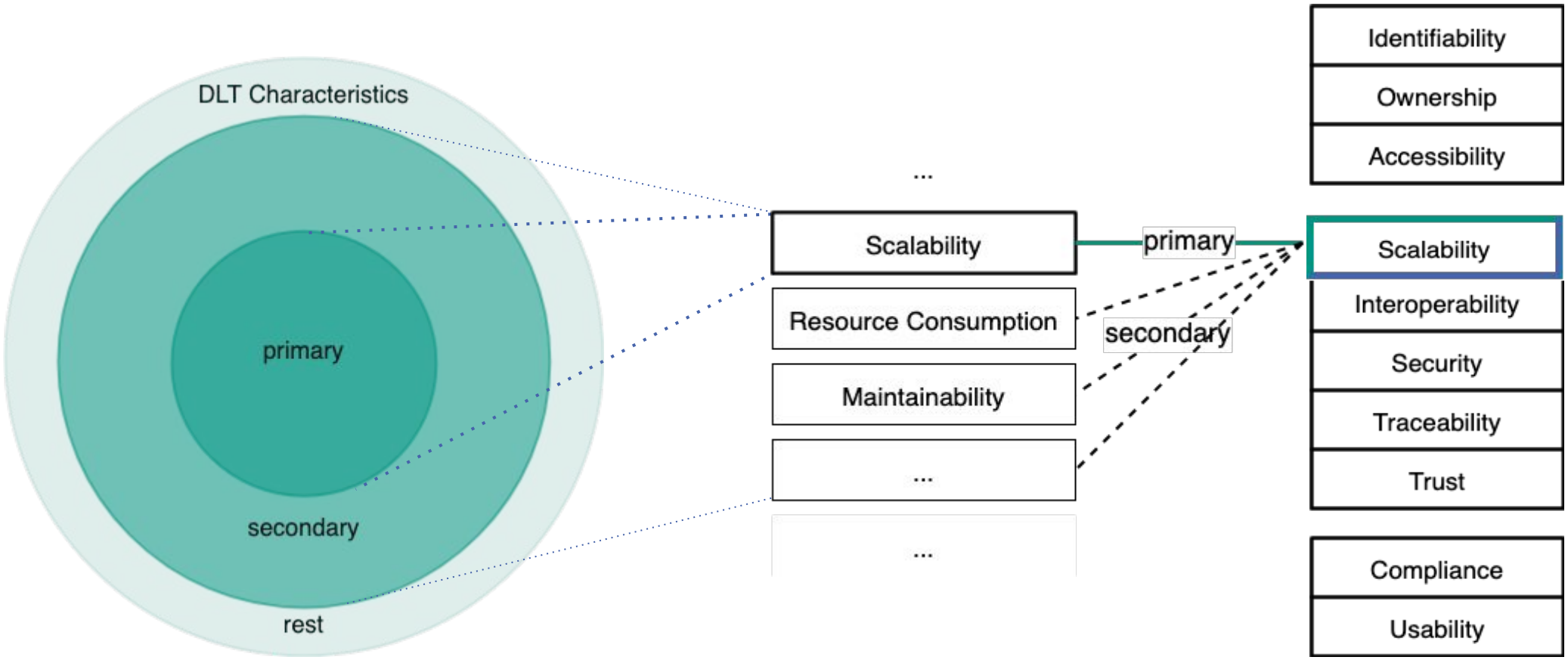
B
- Efficiency
- Comp. Resources
- Cost Reduction
- Trans. Speed
- Scalability
- Interoperability

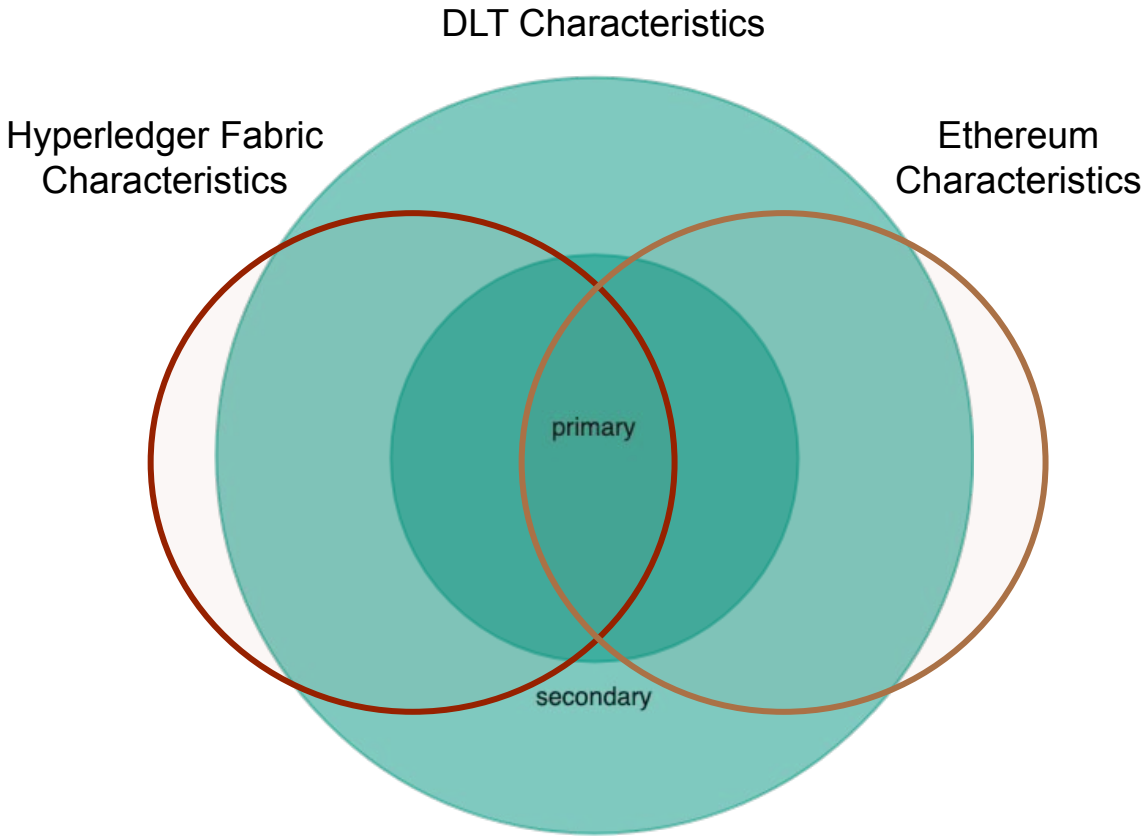
C
- Smart Contracts
- Transparency
- Standardisation
- Immutability
- Verifiability
- Non-repudiation
- Stability
- Consistency
- Authenticity
- Identifiability
- Ownership
- Security
- Traceability
- Usability

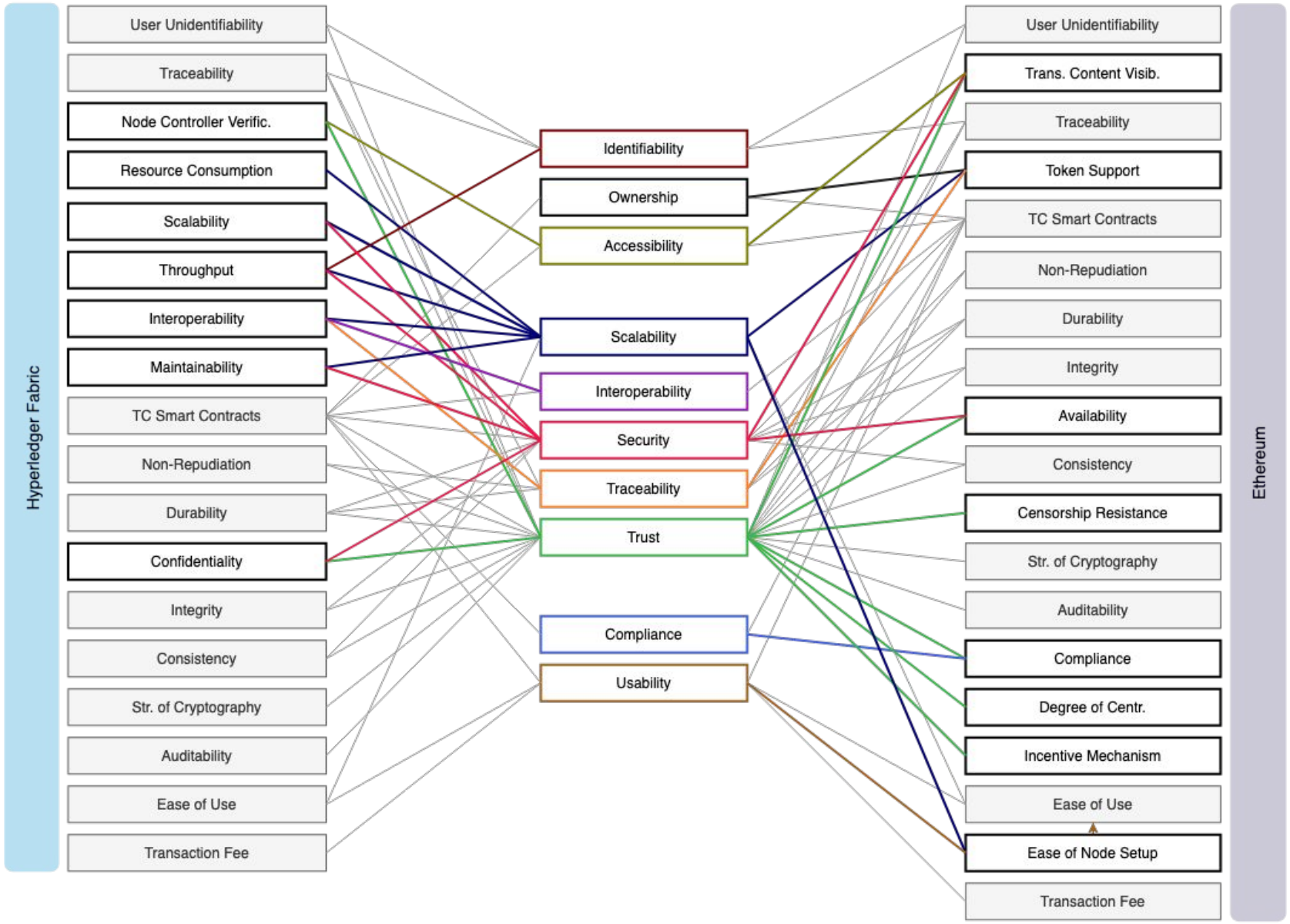
D Decentralisation
- Compliance
- Trust

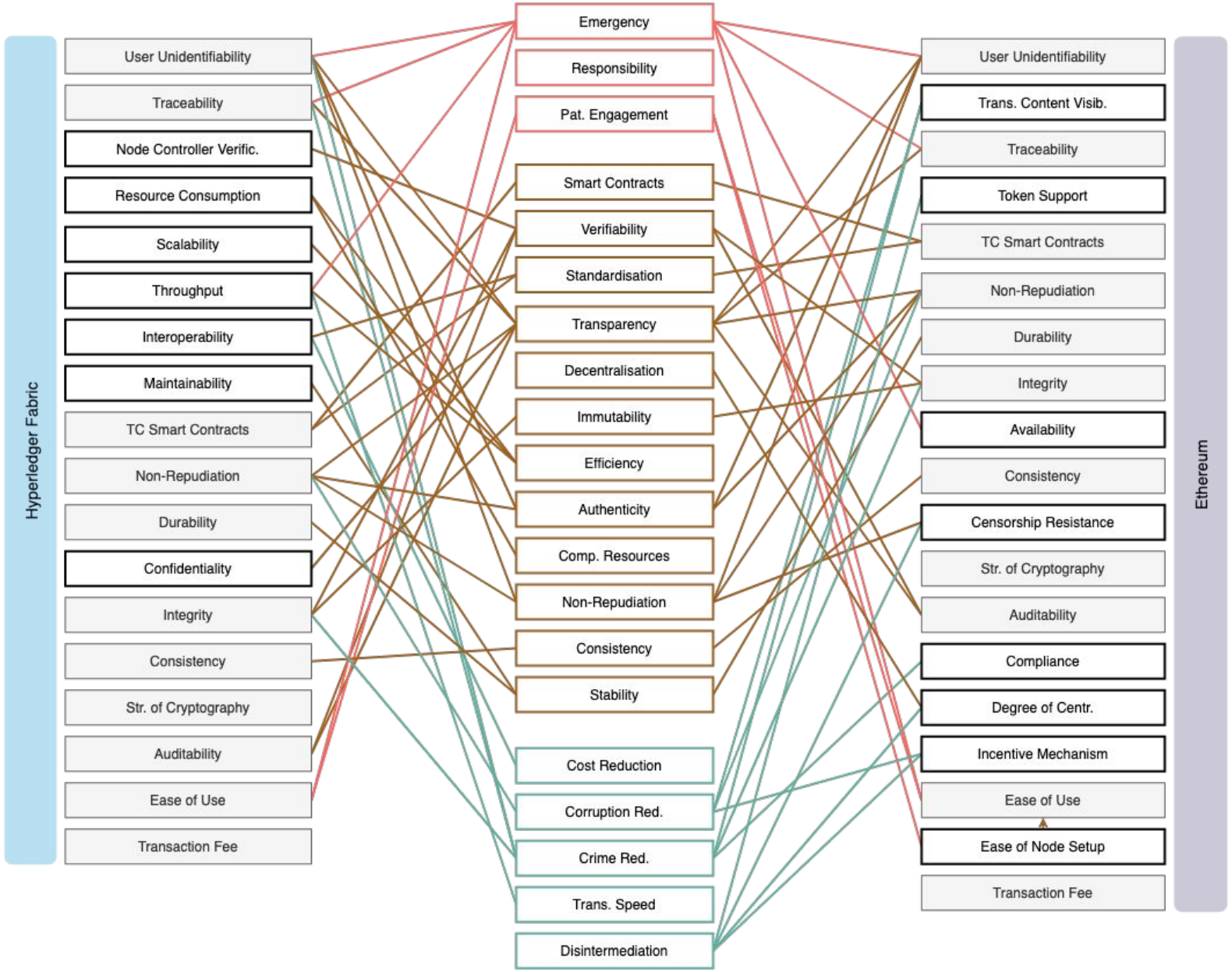
E Disintermediation

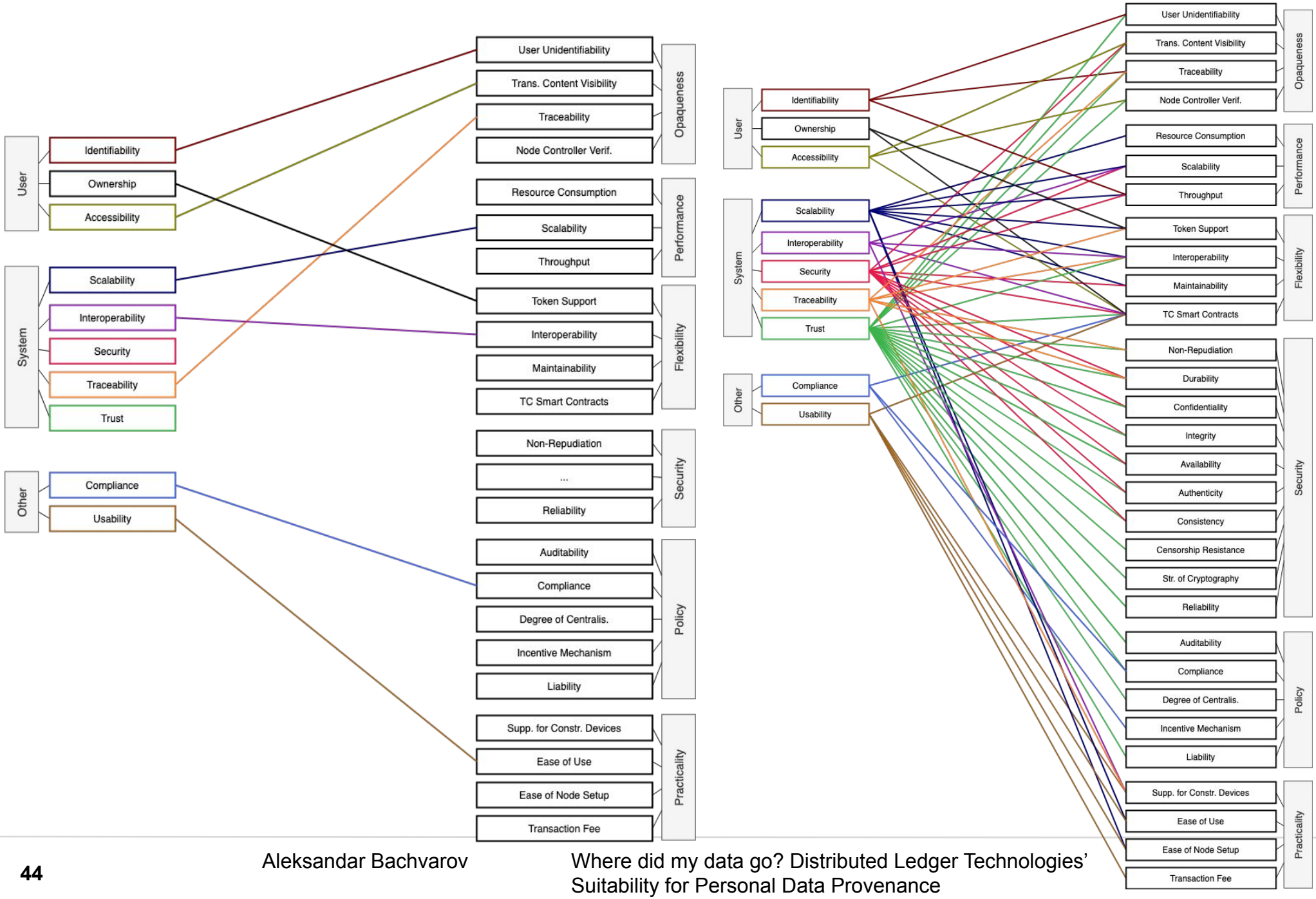

Hyperledger Fabric Characteristics
Ethereum Characteristics
Identifiability
Scalability
Ownership
Trust
Interoperability
Accessibility
Compliance
Security
Traceability
Usability


DLT Characteristics
primary
secondary
rest
...
Scalability
Resource Consumption
Maintainability
...
...
primary
secondary
Identifiability
Ownership
Accessibility
Scalability
Interoperability
Security
Traceability
Trust
Compliance
Usability


DLT Characteristics
Hyperledger Fabric Characteristics
Ethereum Characteristics
primary
secondary


Hyperledger Fabric
User Unidentifiability
Traceability
Node Controller Verific.
Resource Consumption
Scalability
Throughput
Interoperability
Maintainability
TC Smart Contracts
Non-Repudiation
Durability
Confidentiality
Integrity
Consistency
Str. of Cryptography
Auditability
Ease of Use
Transaction Fee
Identifiability
Ownership
Accessibility
Scalability
Interoperability
Security
Traceability
Trust
Compliance
Usability
User Unidentifiability
Trans. Content Visib.
Traceability
Token Support
TC Smart Contracts
Non-Repudiation
Durability
Integrity
Availability
Consistency
Censorship Resistance
Str. of Cryptography
Auditability
Compliance
Degree of Centr.
Incentive Mechanism
Ease of Use
Ease of Node Setup
Transaction Fee
Ethereum


Hyperledger Fabric
User Unidentifiability
Traceability
Node Controller Verific.
Resource Consumption
Scalability
Throughput
Interoperability
Maintainability
TC Smart Contracts
Non-Repudiation
Durability
Confidentiality
Integrity
Consistency
Str. of Cryptography
Auditability
Ease of Use
Transaction Fee
Emergency
Responsibility
Pat. Engagement
Smart Contracts
Verifiability
Standardisation
Transparency
Decentralisation
Immutability
Efficiency
Authenticity
Comp. Resources
Non-Repudiation
Consistency
Stability
Cost Reduction
Corruption Red.
Crime Red.
Trans. Speed
Disintermediation
Ethereum
User Unidentifiability
Trans. Content Visib.
Traceability
Token Support
TC Smart Contracts
Non-Repudiation
Durability
Integrity
Availability
Consistency
Censorship Resistance
Str. of Cryptography
Auditability
Compliance
Degree of Centr.
Incentive Mechanism
Ease of Use
Ease of Node Setup
Transaction Fee


User
Identifiability
Ownership
Accessibility
System
Scalability
Interoperability
Security
Traceability
Trust
Other
Compliance
Usability
User Unidentifiability
Trans. Content Visibility
Traceability
Node Controller Verif.
Opaqueness
Resource Consumption
Scalability
Throughput
Performance
Token Support
Interoperability
Maintainability
TC Smart Contracts
Flexibility
Non-Repudiation
...
Reliability
Security
Auditability
Compliance
Degree of Centralis.
Incentive Mechanism
Liability
Policy
Supp. for Constr. Devices
Ease of Use
Ease of Node Setup
Transaction Fee
Practicality

Group	Requirement	Description
Data Subject	**Identifiability**	An unique identifier allows identification and lays the ground for accountability [Lee+13]. However anonymity, pseudonymity and unlinkability are as important. [HPH11; Sen].
	Ownership	Allows Data Subjects to get an overview, request or perform changes and deletion of the data that they own. [ZN+15]
	Accessibility	Allows Data Subjects with access to view, store, retrieve, move or manipulate data, based on their access rights [ZN+15; BKB16].
System	**Scalability**	With the increase of the data volume, number of operations or participants, it should be possible to store and process provenance information efficiently and without risk of information loss [TBA16; Fre+08, p. 16].
	Interoperability	The ability of different information systems, devices or applications to connect, in a coordinated manner, within and across organisational boundaries to access, exchange and cooperatively use data amongst Data Subjects [SecPROV, IntOp].
	Security	Ensures non-disclosure of data traveling over the network to unauthorised Data Subjects (confidentiality) [Asg+12]. Ensures that the Data Receiver may detect unauthorised changes made to the data (integrity) [Tsa+07]. Ensuring that data and is provenance is available to Data Subjects,when and where they need it (availability) [Lia+17].
	Traceability	In this work we consider traceability and transparency synonymous to each other. It means providing information on what transmitting principle was used, what type of data, for what purpose and to whom the information was sent. How data is collected; how, when, where it is stored [Fre+08; ZN+15, p. 13].
	Trust	If the Data Subject trusts the system, they seem to be willing to share personal information [BHS02]. The willingness to share data can also increase if the Data Subject finds the advantages of engaging in such a transaction more valuable than the loss of privacy [BGS05; AG05].
Other	**Compliance**	Enforcing laws [GDPR], policies and regulations such as purpose limitation [FHS17], data minimisation [ASS17], etc.
	Usability	Provides clear interfaces and structures that display provenance information in an understandable way (usage of icons, graphs, etc.). Managing security (and privacy) is not the primary task of the user [Fre+08].

Primary Characteristic	Description
User Unidentifiability	The difficulty of mapping senders and recipients in transactions to identities
Transaction Content Visibility	The ability to view the content of a transaction in a DLT design
Traceability	The extent to which transaction payloads (e.g., assets) can be traced chronologically in a DLT design
Scalability	The capability of a distributed ledger to efficiently handle decreasing or increasing amounts of required resources
Interoperability	The ability to interact between distributed ledgers and with other external data services
Compliance	The alignment of a distributed ledger and its operation with policy requirements (e.g., regulations or industry standards)
Ease of Use	The simplicity of accessing and working with a distributed ledger

[16]

Secondary Characteristic	Description
Node Controller Verification	The extent to which the identity of validating node controllers is verified prior to joining a distributed ledger
Resource Consumption	The computational efforts required to operate a distributed ledger (e.g., for transaction validation, block creation, or storing the distributed ledger)
Throughput	The maximum number of transactions that can be appended to a distributed ledger in a given time interval
Maintainability	The degree of effectiveness and efficiency with which a distributed ledger can be kept operational
Token Support	The possible uses of tokens within a distributed ledger (e.g., security token, stable coin, or utility token)
Turing-complete Smart Contracts	The support of Turing-complete smart contracts within a DLT design
Confidentiality	The degree to which unauthorised access to data is prevented
Integrity	The degree to which transactions in the distributed ledger are protected against unauthorised (or unintended) modification or deletion
Availability	The probability that a distributed ledger is operating correctly at any point in time
Non-Repudiation	The difficulty of denying participation in transactions
Durability	The property that data committed to the distributed ledger will not be lost
Authenticity	The degree to which the correctness of data that is stored on a distributed ledger can be verified
Consistency	The absence of contradictions across the states of the ledger maintained by all nodes participating in the distributed ledger
Censorship Resistance	The probability that a transaction in a distributed ledger will be intentionally aborted by a third party or processed with malicious modifications
Reliability	The ability of a system or component to perform its required functions under stated conditions for a specified time
Strength of Cryptography	The difficulty of breaking the cryptographic algorithms used in the DLT design
Degree of Decentralisation	The number of independent validating node controllers reduced by the number of controllers that control more than average validating nodes divided by the total number of node controllers in the network.
Incentive Mechanism	A structure in place to motivate node behavior that ensures viable long-term operation of a distributed ledger (e.g., by contributing computational resources)
Liability	The existence of a natural or legal person that can be subjected to litigation with respect to the distributed ledger
Auditability	The degree to which an independent third party (e.g., state institution, certification authority) can assess the functionality of a distributed ledger
Support for Constrained Devices	The extent to which devices with limited computing capacities (e.g., sensor beacons) can participate in a distributed ledger
Transaction Fee	The price transaction initiators can or must pay for the processing of transactions
Ease of Node Setup	The ease of configuring and adding a new (or previously crashed) node to the distributed ledger

[16]