Data Governance and Retrieval Through Decentralized Systems

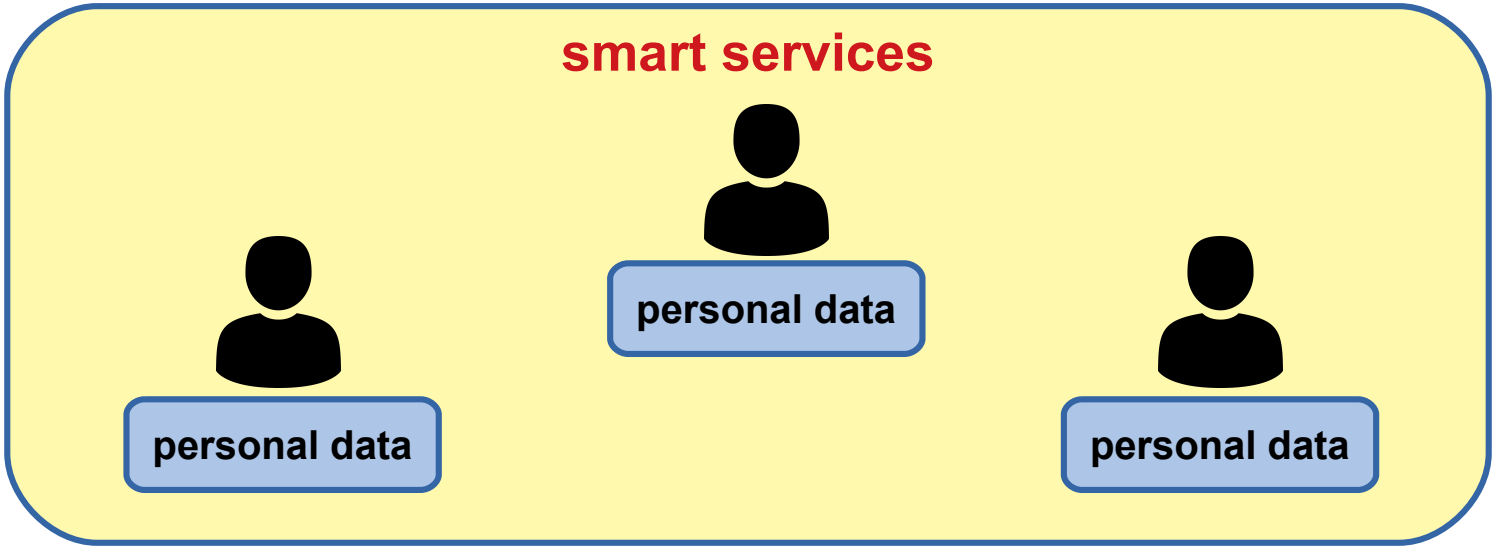
Prof. Stefano Ferretti
stefano.ferretti@uniurb.it

Department of Pure and Applied Sciences
University of Urbino Carlo Bo

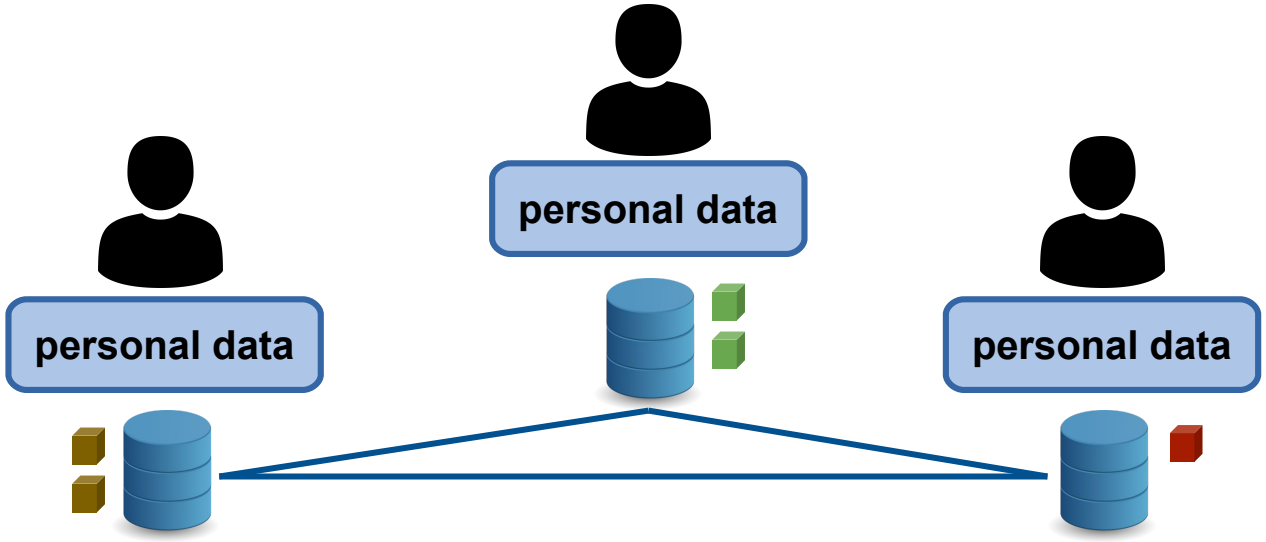
AnaNSi Research Lab @ (UniBo, UniUrb)

Smart Systems from Crowd-sourced/sensed Data

Smart Systems from Crowd-sourced/sensed Data



Why Centralized is "Bad"

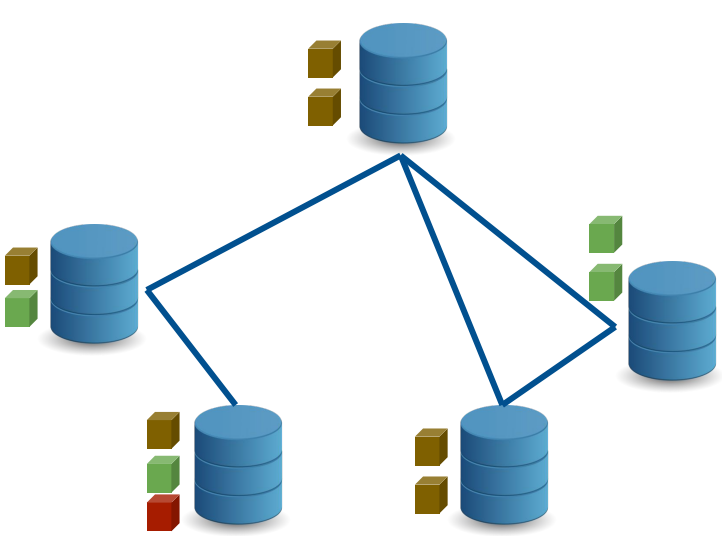


Users lose the sovereignty over their data
- rely on the data controller

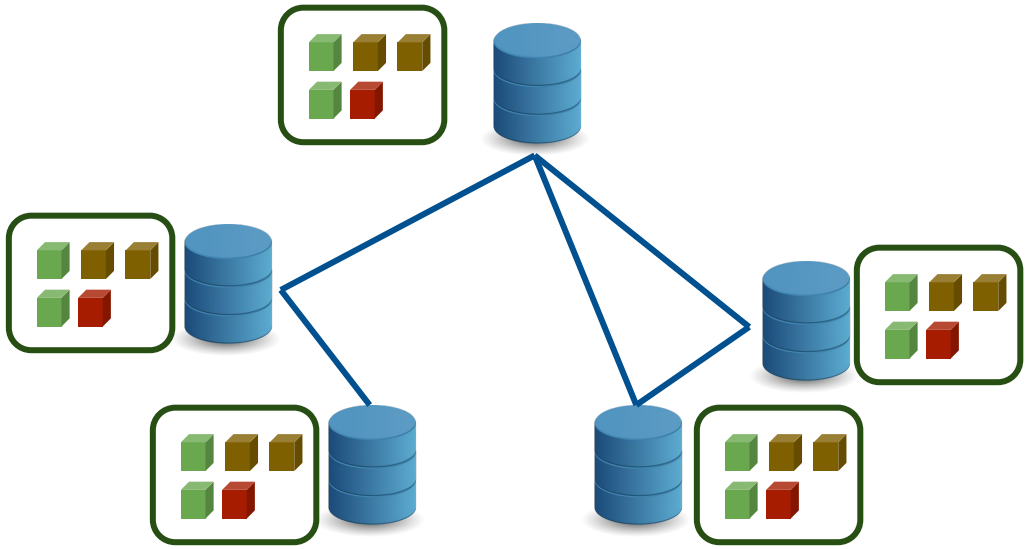
Limits of a C/S system
- single point of failure
- prone to censorship

Decentralized Storage

- High data availability
- Integrity
- Auditability
- Confidentiality
- Automate and enforce processes (through smart contracts)



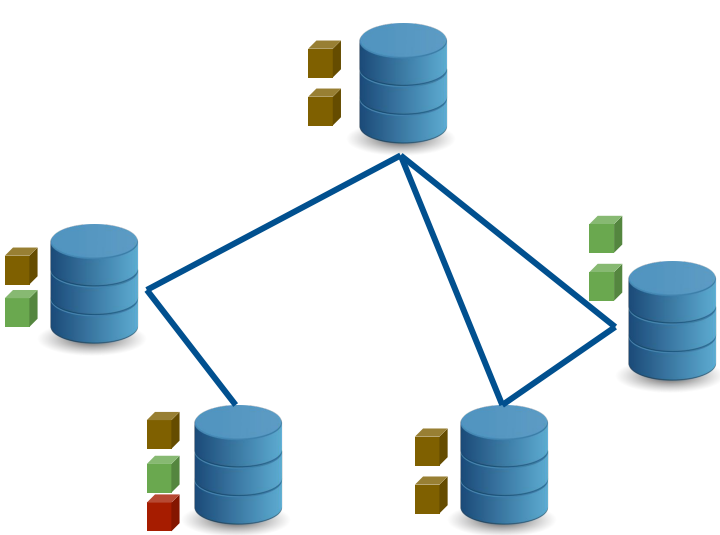
Decentralized File Storage



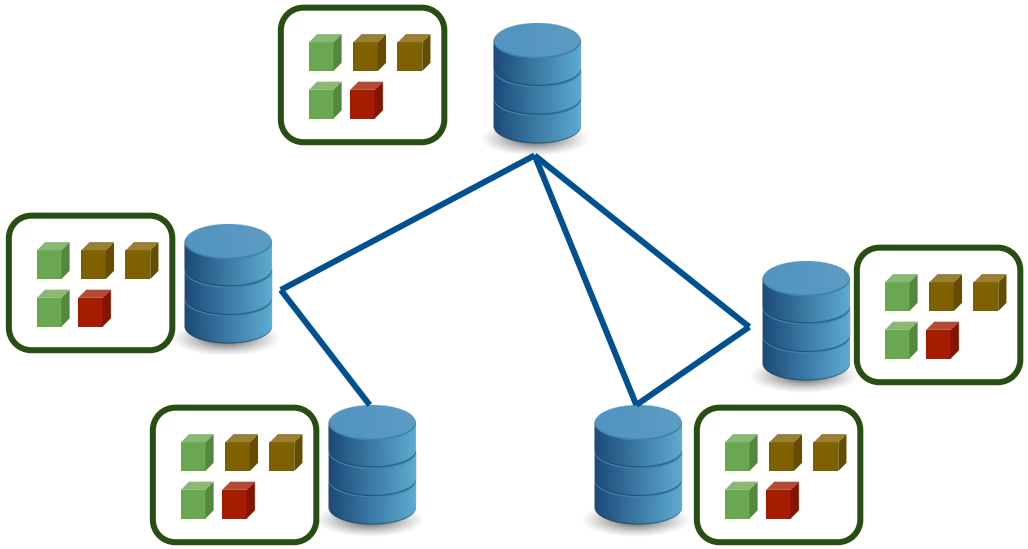
Decentralized Ledgers

High data availability

- Through replication

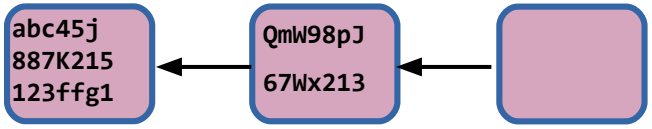


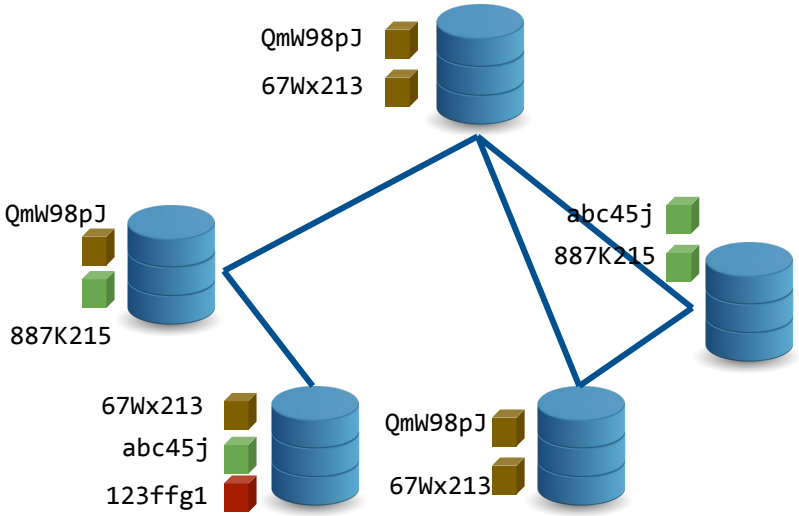
Decentralized File Storage



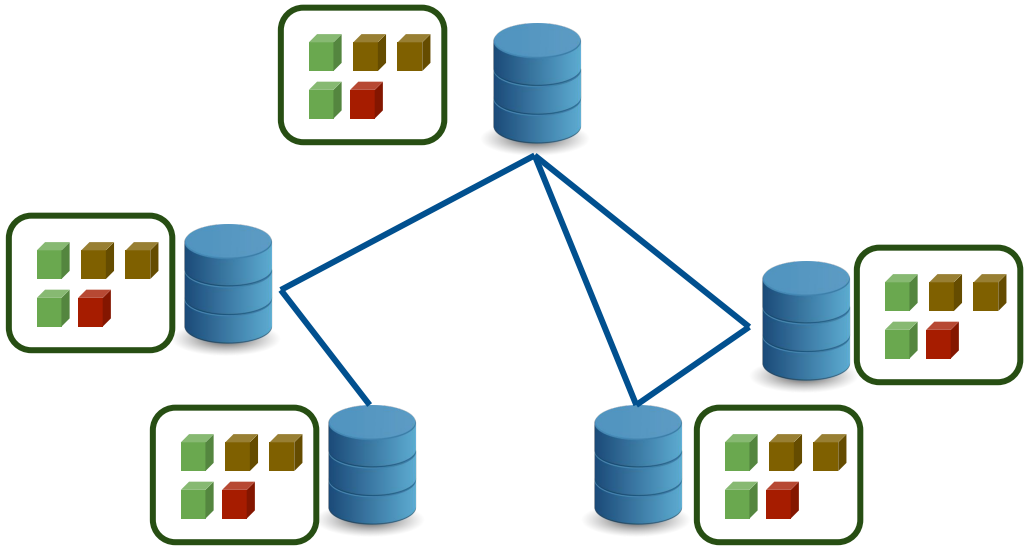
Decentralized Ledgers

Integrity, auditability (GDPR issues to be considered)



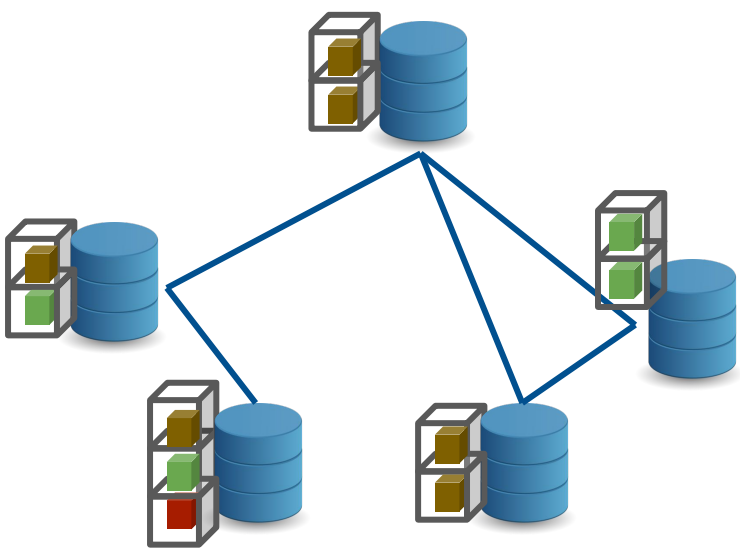


Decentralized File Storage

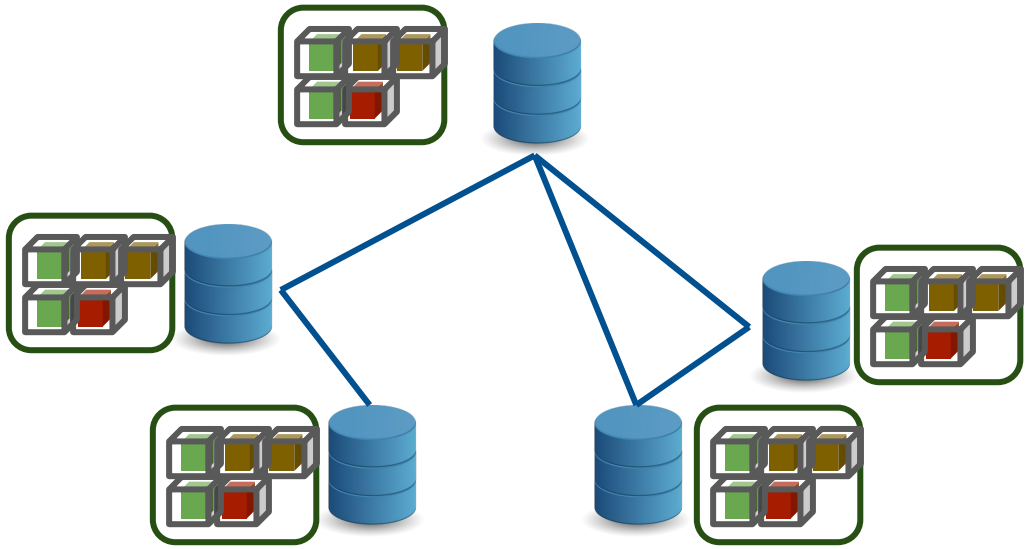


Decentralized Ledgers

Confidentiality

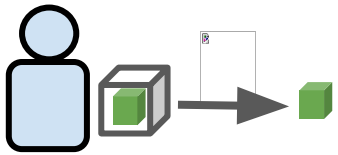


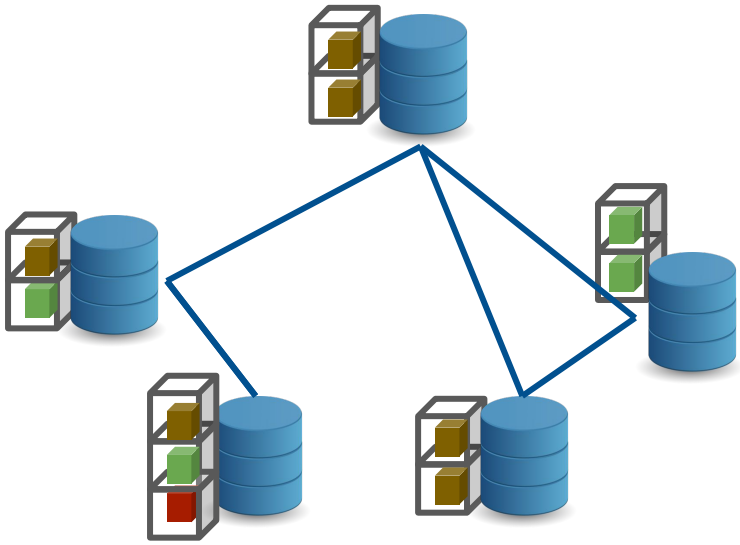
Decentralized File Storage

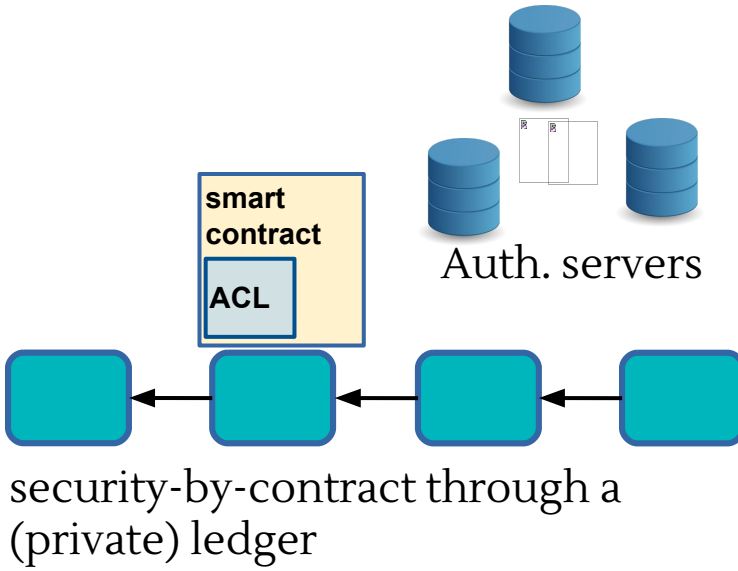


Decentralized Ledgers

Confidentiality



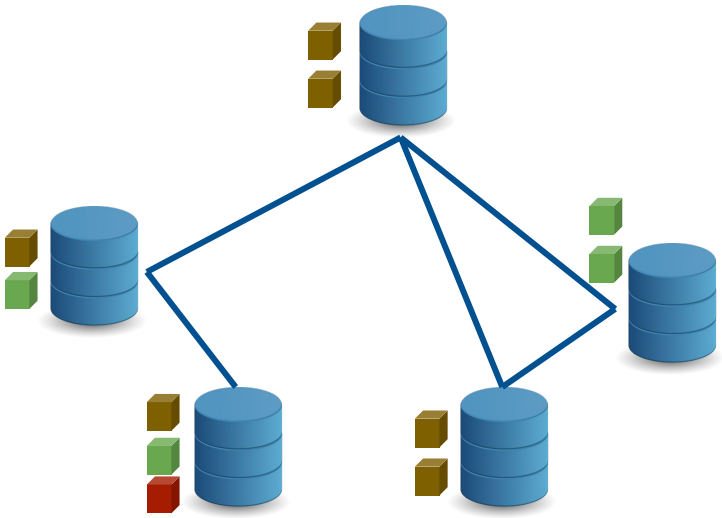




security-by-contract through a (private) ledger

How to lookup for data?

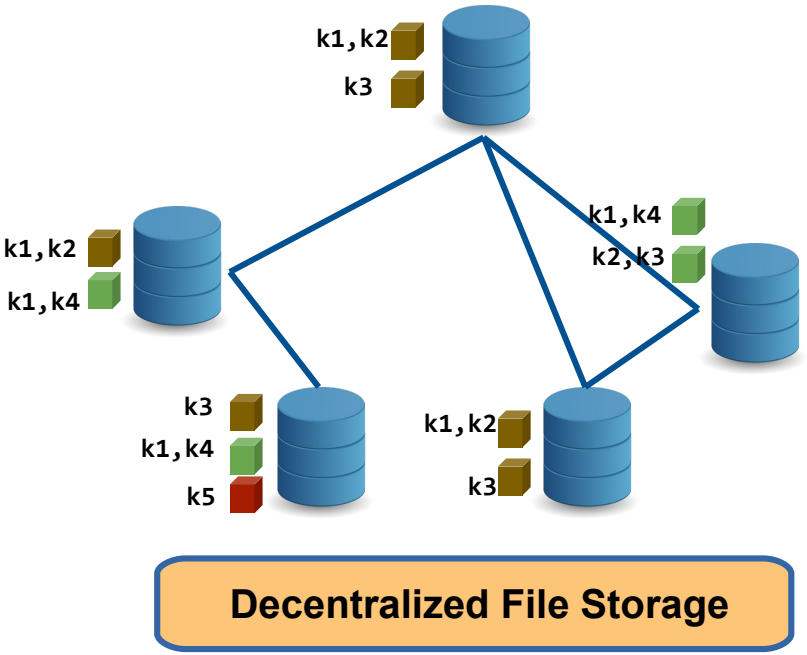
- Data stored in DLTs and DFS are usually unstructured
- Need to be filtered and indexed before any **complex query**



Decentralized File Storage

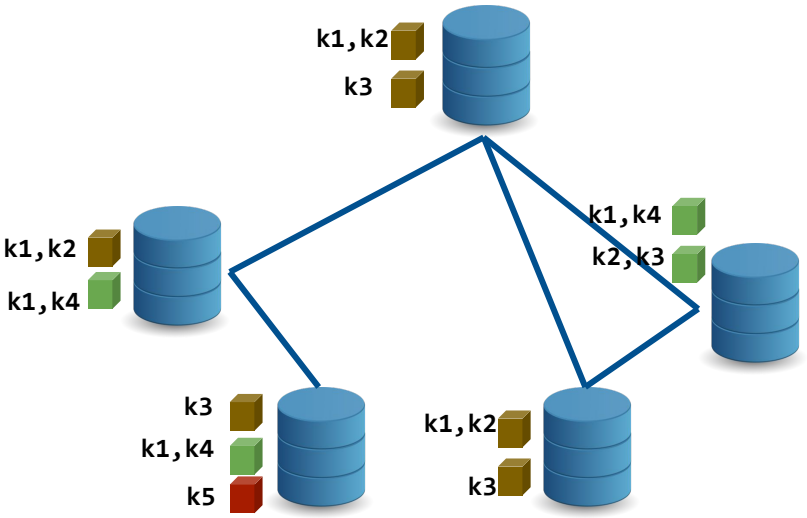
Keyword based lookup

- Keywords associated to data

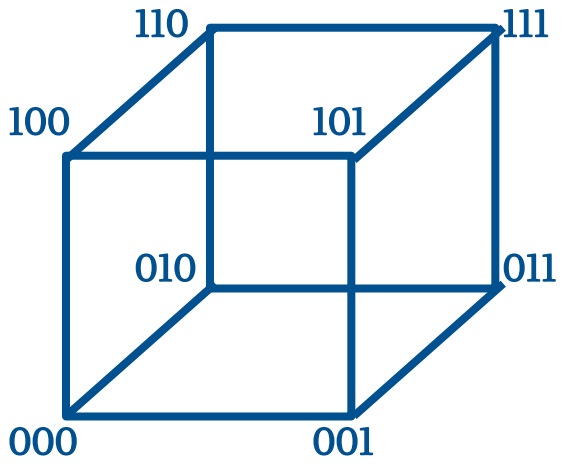


Hypercube based DHT

- We use these r-bit strings to identify logical nodes in a **DHT** network
- Net topology: $H_r(V, E)$ r-dimensional hypercube
  - **V**: set of vertices that represent logical nodes
  - **E:** set of edges formed when two vertices differ of only one bit (they are also network neighbors), e.g. 1011 and 1010



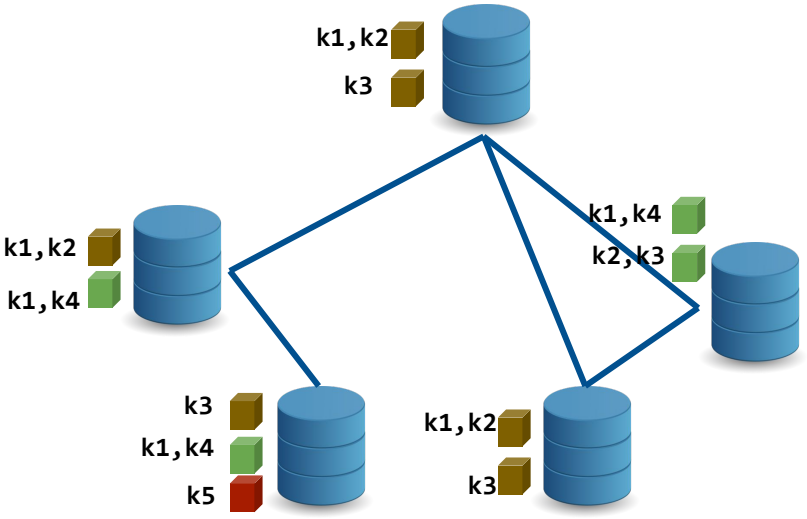
Decentralized File Storage



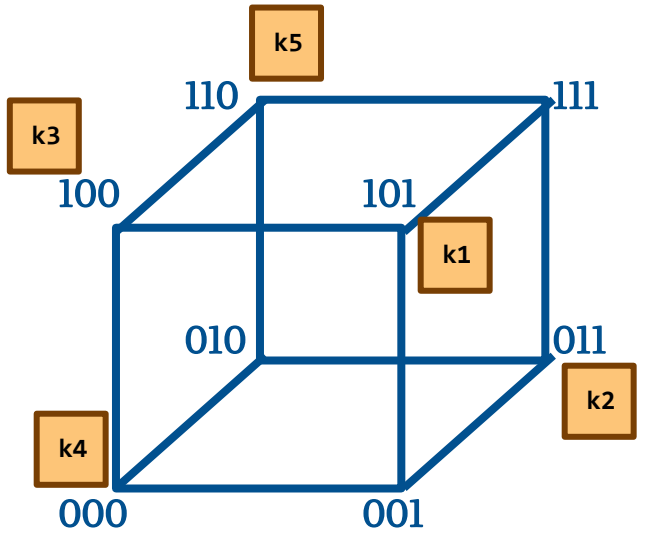
Hypercube DHT

Hypercube based DHT

- Keywords are associated to corresponding nodes of the DHT
  - e.g.: h(k3) = 100
    - node 100 is in charge maintaining info about data with with keyword k3



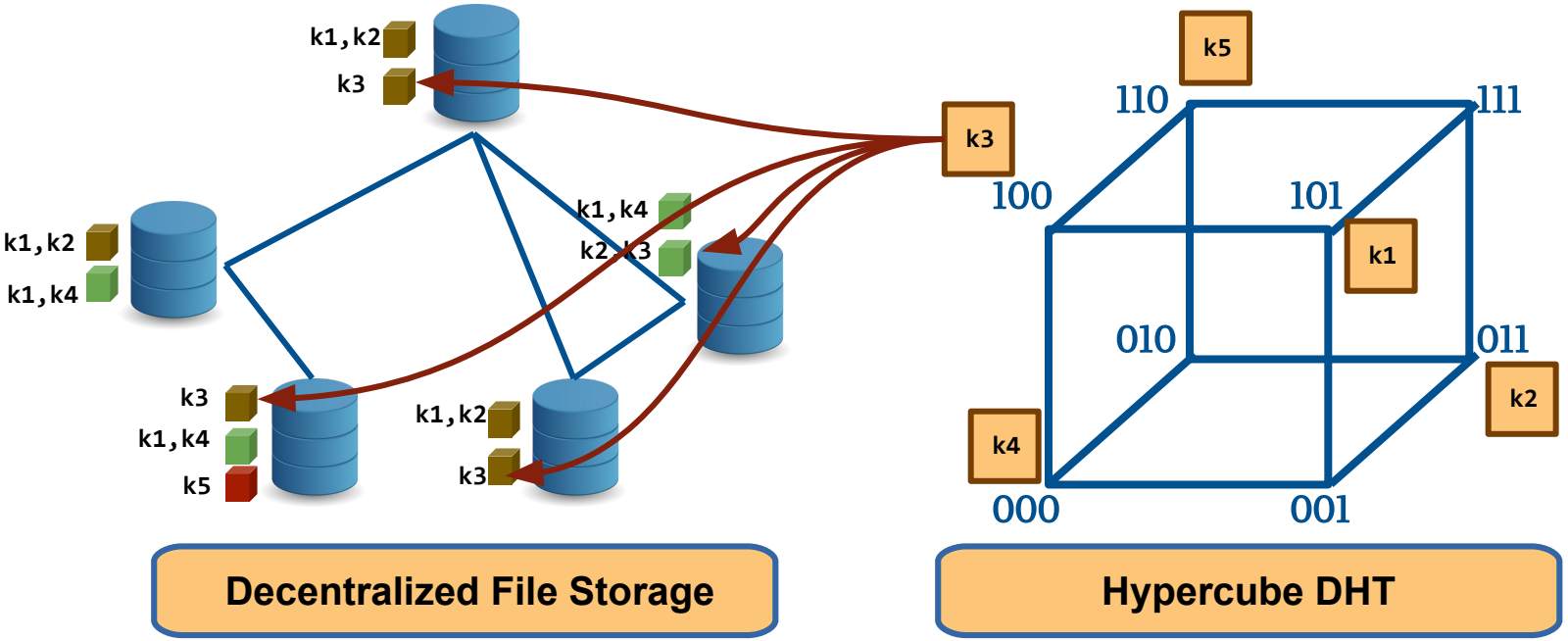
Decentralized File Storage



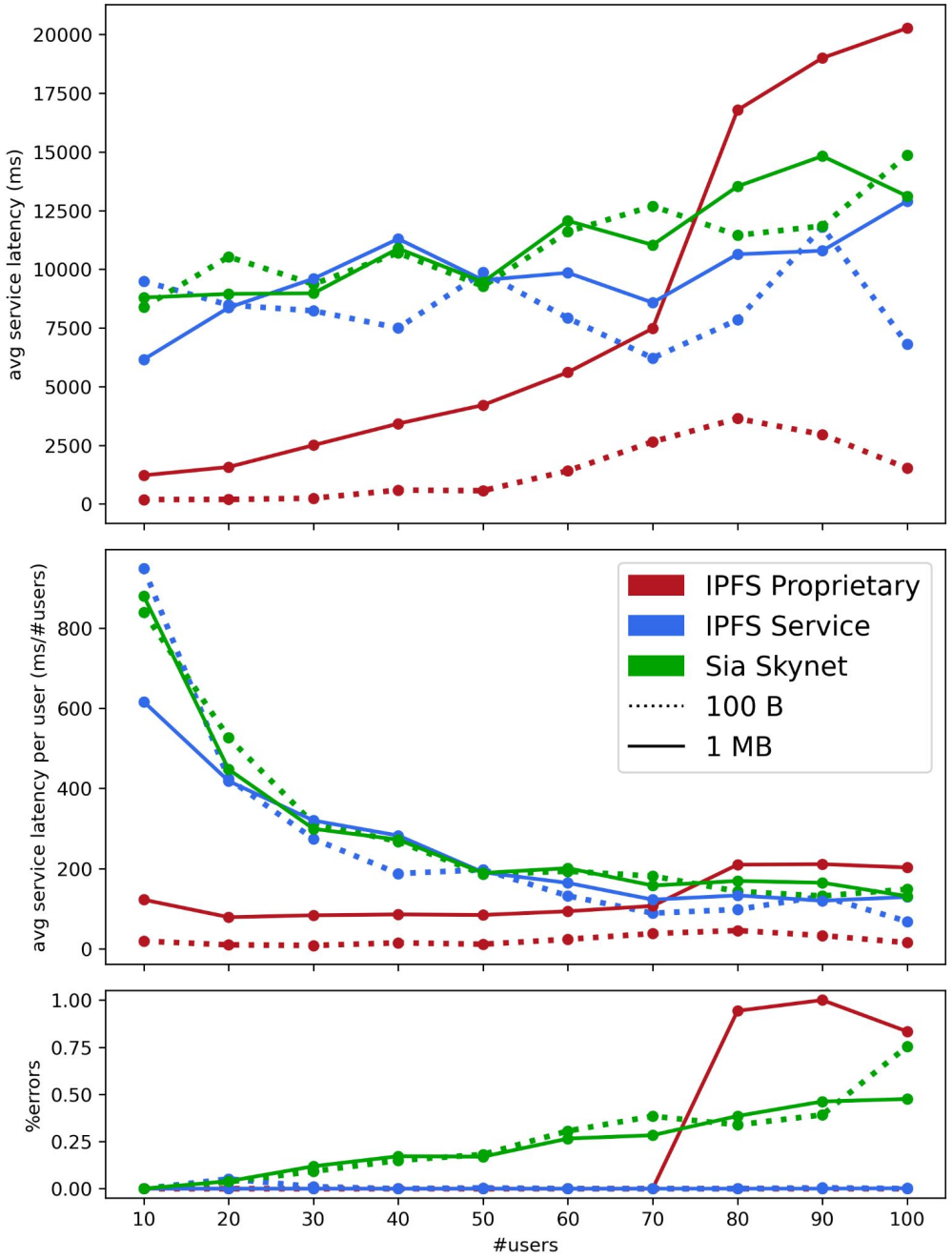
Hypercube DHT

Hypercube based DHT

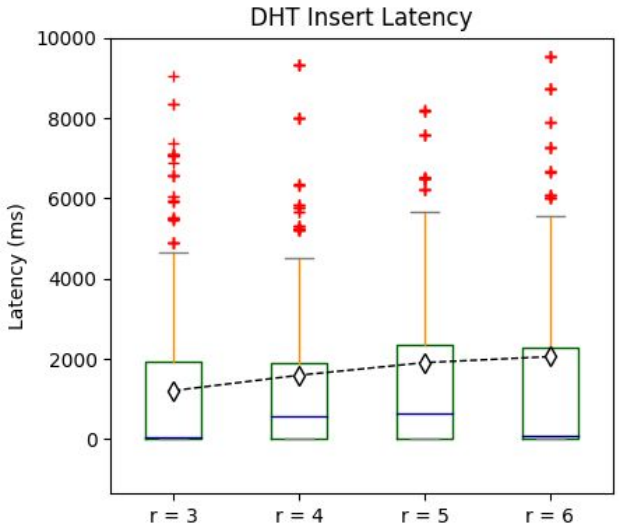
- Keywords are associated to corresponding nodes of the DHT
  - e.g.: h(k3) = 100
    - node 100 is in charge maintaining info about data with with keyword k3

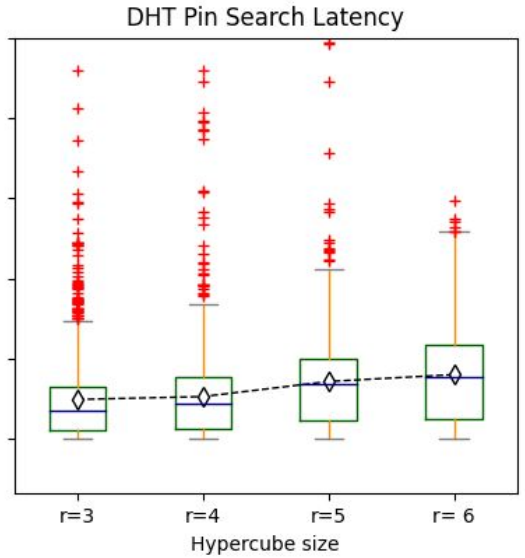


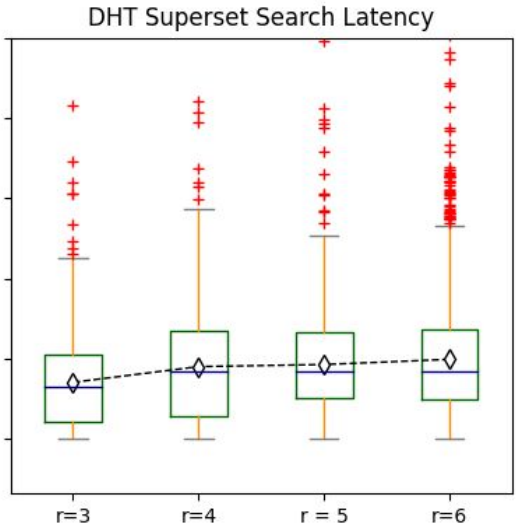
DFS Latencies



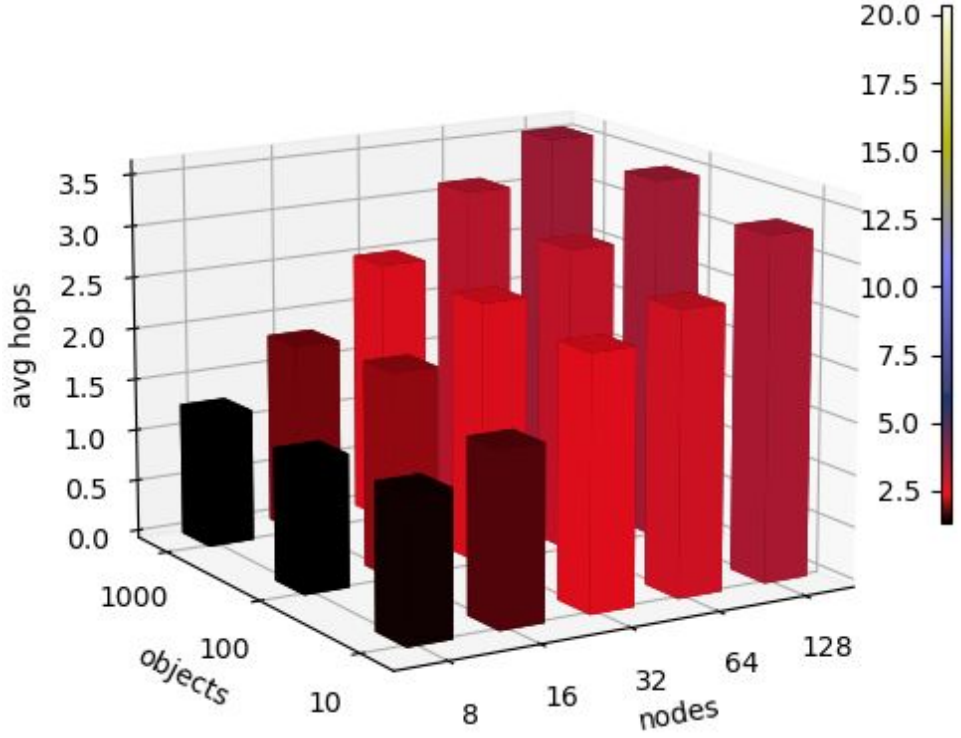
Hypercube Latencies



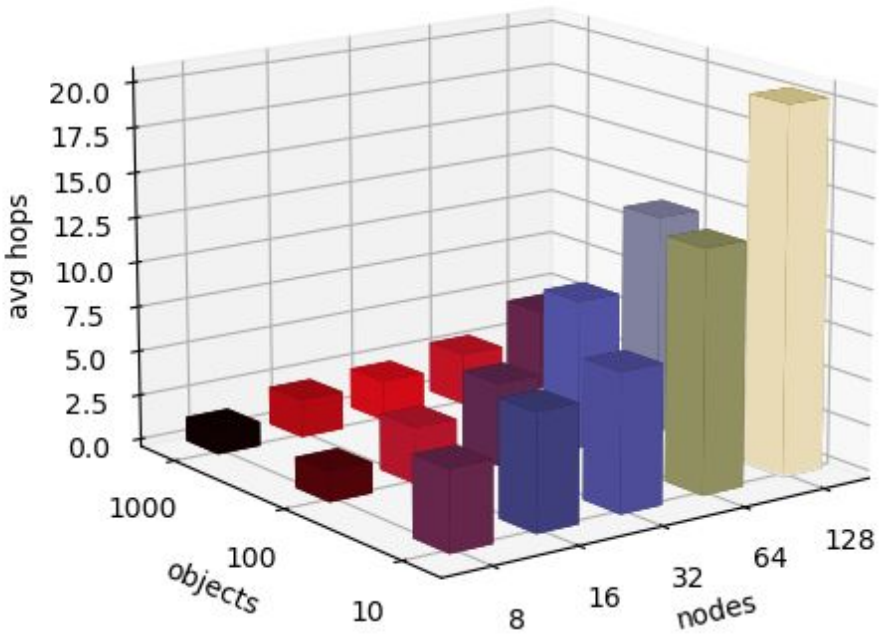




Hypercube Hops

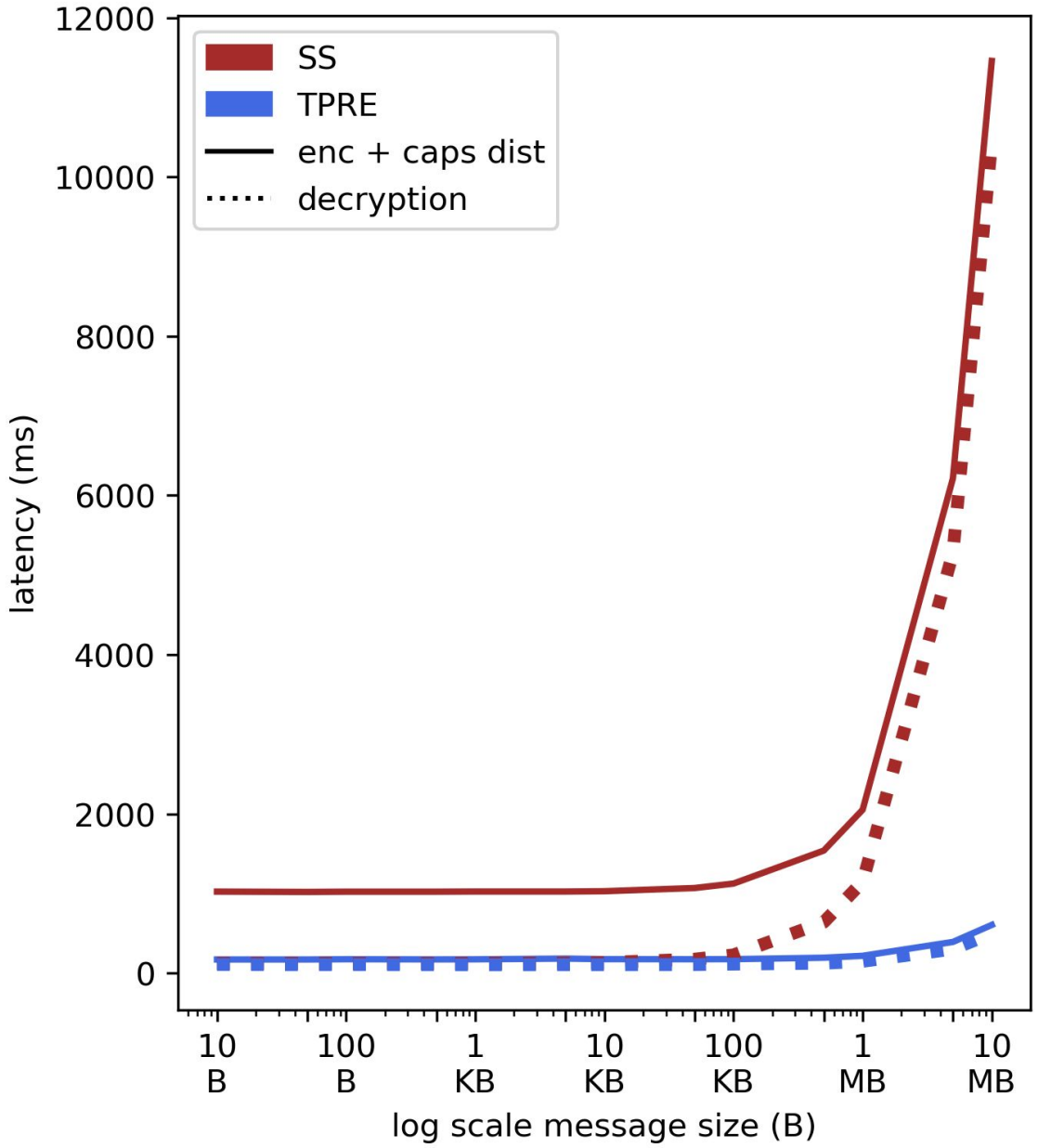


Pin search



Superset search

Encryption/ Decryption



Conclusion

- Use of a multi-layered architecture for (personal) data management of personal information based on the use of decentralized technologies
  - Distributed ledgers
  - Decentralized File Systems
  - Distributed authorization systems
  - Decentralized lookup mechanisms
- Implementation of the system: source code https://github.com/AnaNSi-research/
- Performance evaluation confirms viability of decentralized systems for data management and retrieval

References

- M. Zichichi, S. Ferretti, G. D'Angelo, V. Rodriguez-Doncel, Data Governance Through a Multi-DLT Architecture in View of the GDPR, in ***Cluster Computing***, Springer, August 2022, DOI:10.1007/s10586-022-03691-3
- M. Zichichi, L. Serena, S. Ferretti, G. D'Angelo, Complex queries over decentralised systems for geodata retrieval, ***IET Networks*** (2022)
- M. Zichichi, S. Ferretti, G. D'angelo, A framework based on distributed ledger technologies for data management and services in intelligent transportation systems, ***IEEE Access***, 8 (2020): 100384-100402

Stefano Ferretti

AnaNSi Research Lab @ (UniBo, UniUrb)

stefano.ferretti@uniurb.it

Department of Pure and Applied Sciences
University of Urbino "Carlo Bo"
P.zza della Repubblica 13
61029, Urbino
Italy