Evolving Blockchain Applications: Semantic Models and Distributed Databases with an Accounting Application to Virtual Organizations

Daniel E. O'Leary

University of Southern California

Basic Questions …

- Why not use a database rather than a blockchain?
  - Is it possible?
  - What kind of databases would / could work?
  - Advantages? Disadvantages?
- How can I use blockchain or a “blockchain like approach” in the capture of actual accounting transactions?
  - Blockchain really does not work for accounting transactions – why have all transactions visible to others?
  - Is there an accounting setting where it would make sense?
  - Where does it make sense to allow others to see your transactions?

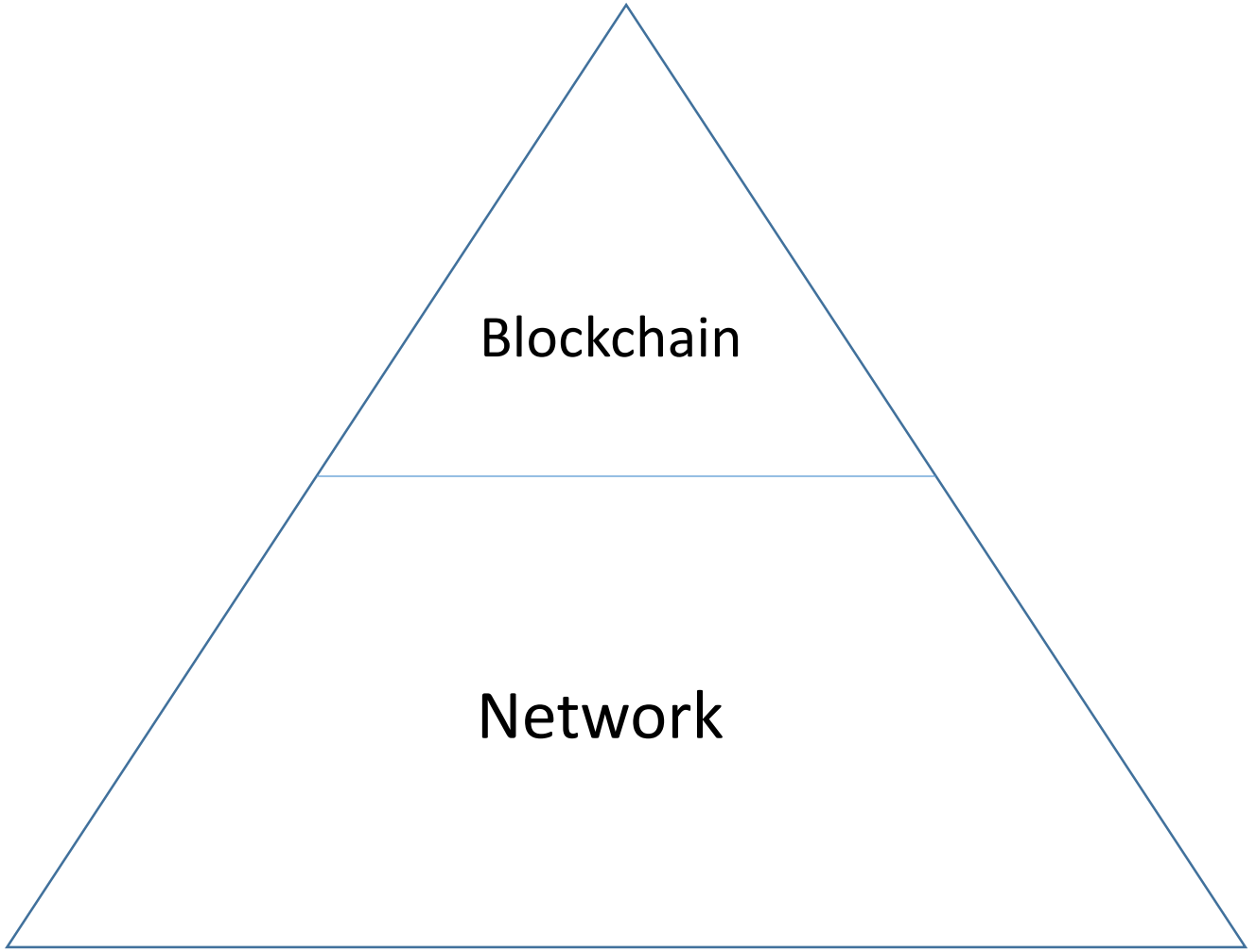
Outline

- Blockchains – background and discussion
  - Emerging Applications
- Emerging Issues
  - Background
  - Multiple Semantic Models
  - Blockchain Governance
  - Distributed DB / Multiple – Tiers and Data Independence
- Blockchain Application Matrix (Uses vs. Data Demands)
- Application to Virtual Organizations
  - Distributed Databases (e.g., Big Chain DB)
- Summary

Blockchains

- Have captured the “control” imaginations of potential users
  - Immutability preserves each of the transactions on it in a ledger
- Are particularly useful, when you want to preserve documents or information
  - Deeds, pollution information, anything where I want to keep the information in the specific context
- Are particularly useful when information visibility is an asset
  - When you want to share information with a group of affiliated users
  - Because they leverage network capabilities

Architecture



Blockchains …

- However, blockchains have been criticized for
  - Limitations with throughput, latency, scalability and power requirements
  - Providing data in a ledger format
    - Can’t query
    - Basically a flat file, so limited data complexity and modeling capabilities
  - Making information publicly available to others on the blockchain
    - Allows people to “see” transactions, number of transactions, etc.
    - There are a limited number of organizational settings where people want others to “see” transactions
  - Biased governance – in the case of some private blockchains
    - Biased semantic models (Arrow’s Impossibility Theorem)

Evolving Scope of Blockchain Applications

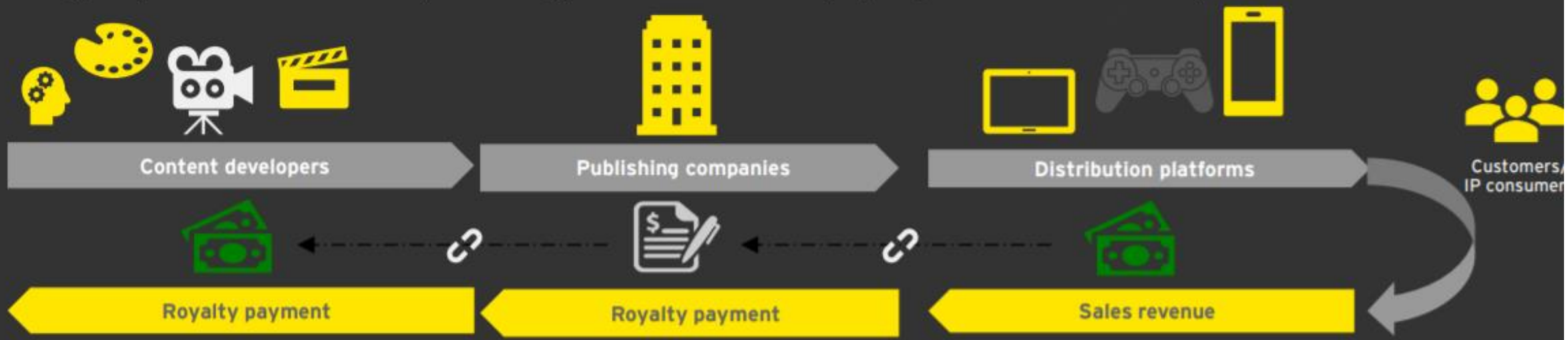
- Original scope of blockchain capabilities
  - Information put on the blockchain as part of a market, Bitcoin
  - Transaction information, basically open to everyone
  - Effectively, the original source document is put on the blockchain and made immutable
- Increasingly, blockchain is being used for other applications, such as information creation and distribution
  - Getting royalty information to participants in real time.
  - But, there are multiple similar models to do the same thing
  - There is a need to query the data in the ledgers
- This is leading to potential issues such as
  - Multiple and differing semantic models of the same process (e.g., timing)
  - The need for a distributed database rather than a ledger to capture the information
  - Blockchain governance
- The purpose of this paper is examine these issues and examine the impact of those issues.

In any case, there are a number of applications that reportedly use blockchain

- Capture pollution information in Los Angeles, Shanghai, etc. and let others know what that pollution information is.
- Food traceability
- Supply chain
  - IBM and Maersk's TradeLens
  - Former head of Deloitte's version
  - Others in play
- Distribution of royalty earnings information -> next

Gaming industry is a complex value chain

A long, complex value chain with many interacting parties can create multiple pain points for benefits and royalties distribution.



Developers

"I don't trust the calculations."

"Most of my **time is spent reconciling** the payment with various sources."

"Once the game is live, it's **hard to understand what's going on**. It's a long waiting game."

"**I want a self-service option**. I put the statement in a format in which it's easy for me to read."

Publishers

"Royalty calculation processes are a black box. There's no visibility into calculations and underlying data."

"There is **no mechanism to reconcile payments** or trace back to terms of the agreements."

"**Often, accrual amounts are restated** post-period close, requiring manual adjustments"

"Often, the **statements are received too close to the close period**, causing operational issues."

Distributors

"It's a nightmare to reconcile data among various sources."

"My current **processes are rigid** and often IT systems and processes dictate contracts terms."

"We have to **make accruals within a matter of hours** before close."

"At any point, we have five to 10 audits going on. It's a **huge burden on the resources.**"

"Manual processes **limit my ability to scale.**"

* Gaming industry survey by independent third party.

Supply chain and trade applications: TradeLens

- Best known is the IBM Maersk blockchain
- “Some of the largest supply chain companies in the world have begun the process of moving their work to a blockchain. A previously unnamed collaborative effort between the world’s largest shipping company, Maersk, and IBM, has now grown to 92 participants and been dubbed TradeLens.
- Far from an early prototype, the blockchain platform has been quietly orchestrating global trade with less reliance on middlemen for a year, resulting in 154 million shipping events in ports around the world, and is now growing at a rate of one million per day.
- The cooperative effort now also includes Germany-based Hamburg Sud, which Maersk bought last year for $4 billion, and Singapore-based Pacific International Lines, along with numerous customs authorities, cargo owners and freight forwarders.”
- Likely that this primarily captures Maersk and Maersk’s partners

Organizations are Hesitant to Join TradeLens

- So why are most carriers opting out of joining this revolutionary new system? The simple answer is competition. The underlying IP behind the blockchain joint venture, in fact, belongs to Maersk and IBM, which puts rival shipping carriers like CMA CGM, and Hapag-Lloyd in a difficult situation. Both carriers have publicly rejected the blockchain solution, with CMA CGM general manager Peter Wolf stating,
- *“Technically the solution (by Maersk and IBM) could be a good platform, but it will require governance to make it an industry platform and not just a platform for Maersk and IBM. And this is the weakness we’re currently seeing in many of these initiatives, as each individual project claims to offer an industry platform that they themselves control. This is self-contradictory”.*
  - This means that there can be multiple models
- The need for centralized parties to establish control over decentralized systems is a primary impediment to enterprise blockchain adoption.

Emerging Issues in Emerging Blockchain Applications

- Semantic Models
- Distributed Databases
- Distributed Systems Architecture
- Governance

Background - Semantic Model

- Semantic models of data sources represent the implicit meaning of the data by specifying the concepts and the relationships within the data.
- Semantic data model is a high-level semantics-based database description and structuring formalism for databases.
  - This database model is designed to capture more of the “meaning” of an application environment than is possible with contemporary database models.
- Different models provide different ways, different timing, etc. of doing things
- Companies may need to use multiple models from different sources

Problem of Multiple Semantic Models

- Organizations could have multiple occurrences of roughly the same semantic models
  - For example, the IBM / Maersk model and the “Deloitte” model, or the EY model, or from other sources, e.g., Singapore
  - Data elements are not likely to be “exactly” the same, so these differences would need to be accounted for in processing data from each system.
- One approach is to use Open Standards when available (TradeLens)
  - “We believe this innovative approach based on open standards and open governance can benefit the entire industry while ultimately benefitting our customers who rely on the world’s shipping industry to transport global container volume … across international borders each year.”
  - But open standards are not always optimal, are likely to be dated, not complete for special cases or used, etc.

Background - Distributed Databases

- A distributed database is a database that consists of two or more files located in different sites either on the same network or on entirely different networks. Portions of the database are stored in multiple physical locations and processing is distributed among multiple database nodes.
- A distributed database is a database in which not all storage devices are attached to a common processor. It may be stored in multiple computers, located in the same physical location; or may be dispersed over a network of interconnected computers.
- Some distributed databases have begun to include properties typically attached to blockchain, e.g., BigChainDB

Blockchain and Distributed Database Characteristics (McConaghy et al. 2016)

Characteristic	Typical Blockchain	Typical Decentralized DB	BigchainDB
Decentralization	X	X	X
Immutability	X		X
Owner-Controlled Assets	X		X
High Throughput		X	X
Low Latency		X	X
Indexing and Querying of Structured Data		X	X
Rich Permissioning		X	X

It is becoming clear that some blockchain applications MUST be using database capabilities

- In the case of IBM / Maersk's TradeLens … reportedly users can make queries:
- "TradeLens gives users access to their own blockchain node similar to those on the bitcoin blockchain that lets users send money without the need of banks. But in the case of TradeLens, a shipper can cut out as many as five middlemen, even for simple queries such as identifying the location of a shipping container."
- As a result, either they are using a distributed database to capture the data or they are spinning the data out to a database after capturing it on the ledger (which would be an inefficient approach) – on or off chain data base

More from TradeLens

- “Track shipments in real time with greater precision via 120+ dynamic events published direct from the source - without the back and forth.”
  - “Dynamic Events” are database events
- “Improve data sharing and collaboration with shipping partners through blockchain encryption and powerful permission-based sharing.”
  - Could be attained using a distributed DB
- So there is a question if the implementation is truly blockchain or say, just blockchain encryption integrated with a database.
  - There is a strong need for databases

On-Chain vs. Off-Chain Data Storage

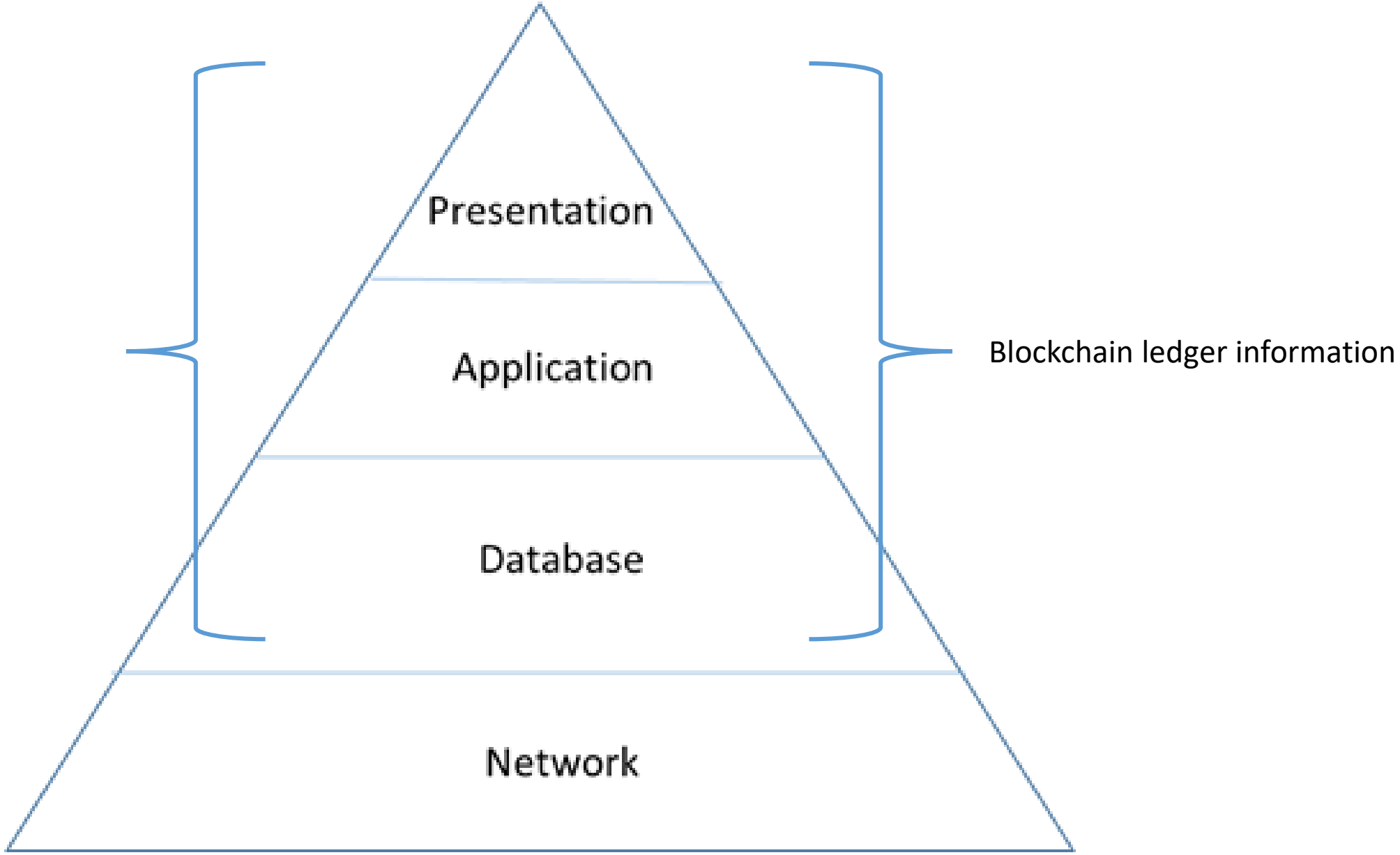
- The biggest problem of storing data on a public blockchain is the amount of data you can store. This is either because the amount is limited by the protocol or because of the huge transaction fees you would have to pay or because of the need to replicate data across multiple machines.
- Data can be stored on-chain or off-chain.
  - In a public chain, we typically must limit information communicated
  - Off-chain data storage reduces the storage requirements that each node is expected to store on-chain. ...
- A peril of off-chain data storage is data availability is no longer guaranteed since the data is not part of the blockchain but only a copy of the data is stored on-chain
  - In addition, data stored off line is not subject to the same level of encryption or control

Distributed Systems Architecture

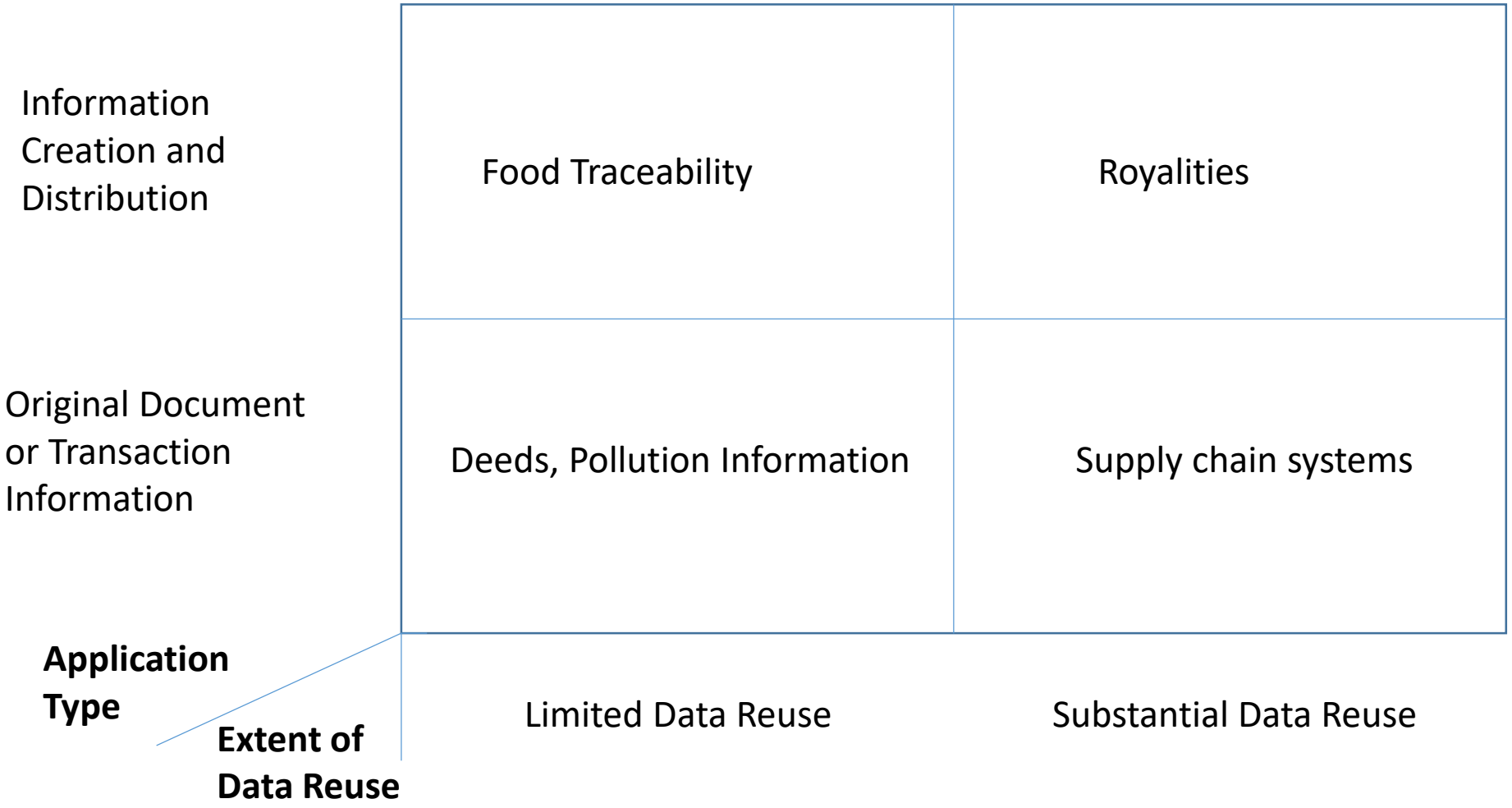
- Generally, distributed systems design typically employs a four tier architecture consisting of the *network, a database, an application and a presentation* tier (e.g., Tiwana 2013).
- System and database design promulgate that a database is independent of the applications that use it.
- As noted by Wikipedia, “Data independence … refers to the immunity of user applications to changes made in the definition and organization of data. Application programs should not, ideally, be exposed to details of data representation and storage.”
- Although the notion of data independence was originally generated and used at enterprise level, it likely is even more important when there are multiple enterprises involved because of the potential for multiple definitions of the same variable in different settings.

Mixing Tiers in Blockchain Systems

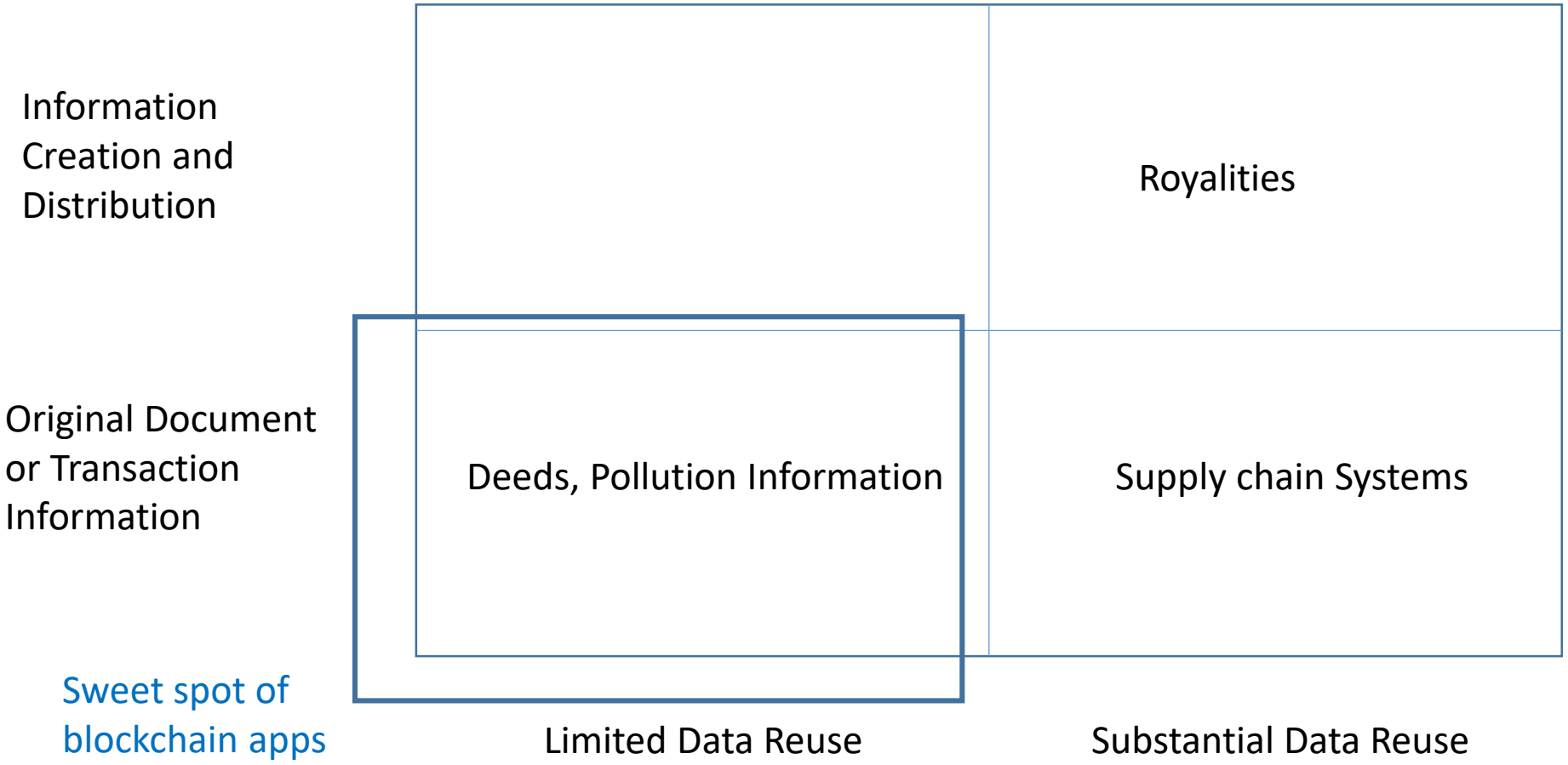
- Further, blockchain models typically integrate the presentation tier with both the data and application tiers.
- With the emerging blockchain applications, there appears to be a mixing of architectural components in the ledgers.
  - The blockchain may capture data but not in an application neutral database.
  - Unfortunately, that can limit the ability to use the data in other applications and it can limit the ability to structure and query the data for other decision making uses.


Presentation
Application
Database
Network
Blockchain ledger information

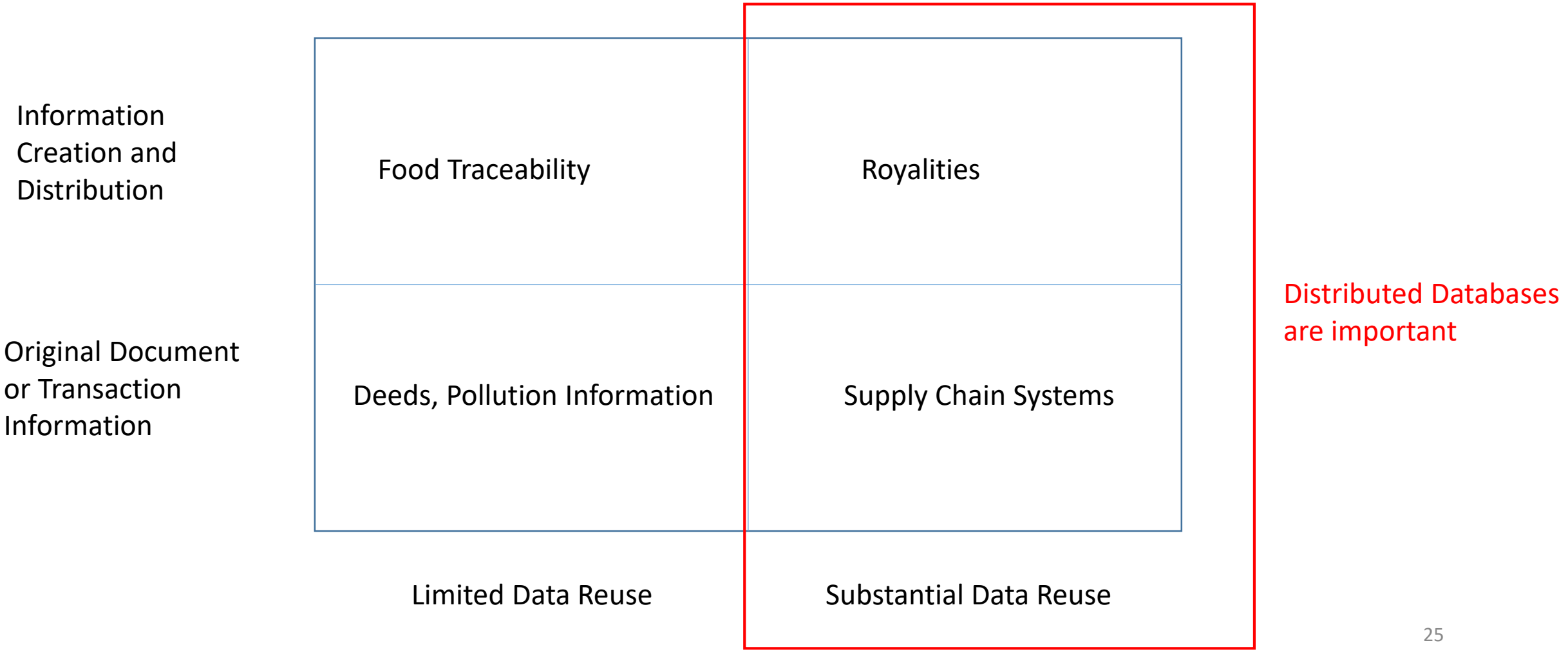
Blockchain Application Matrix: Use vs. Data



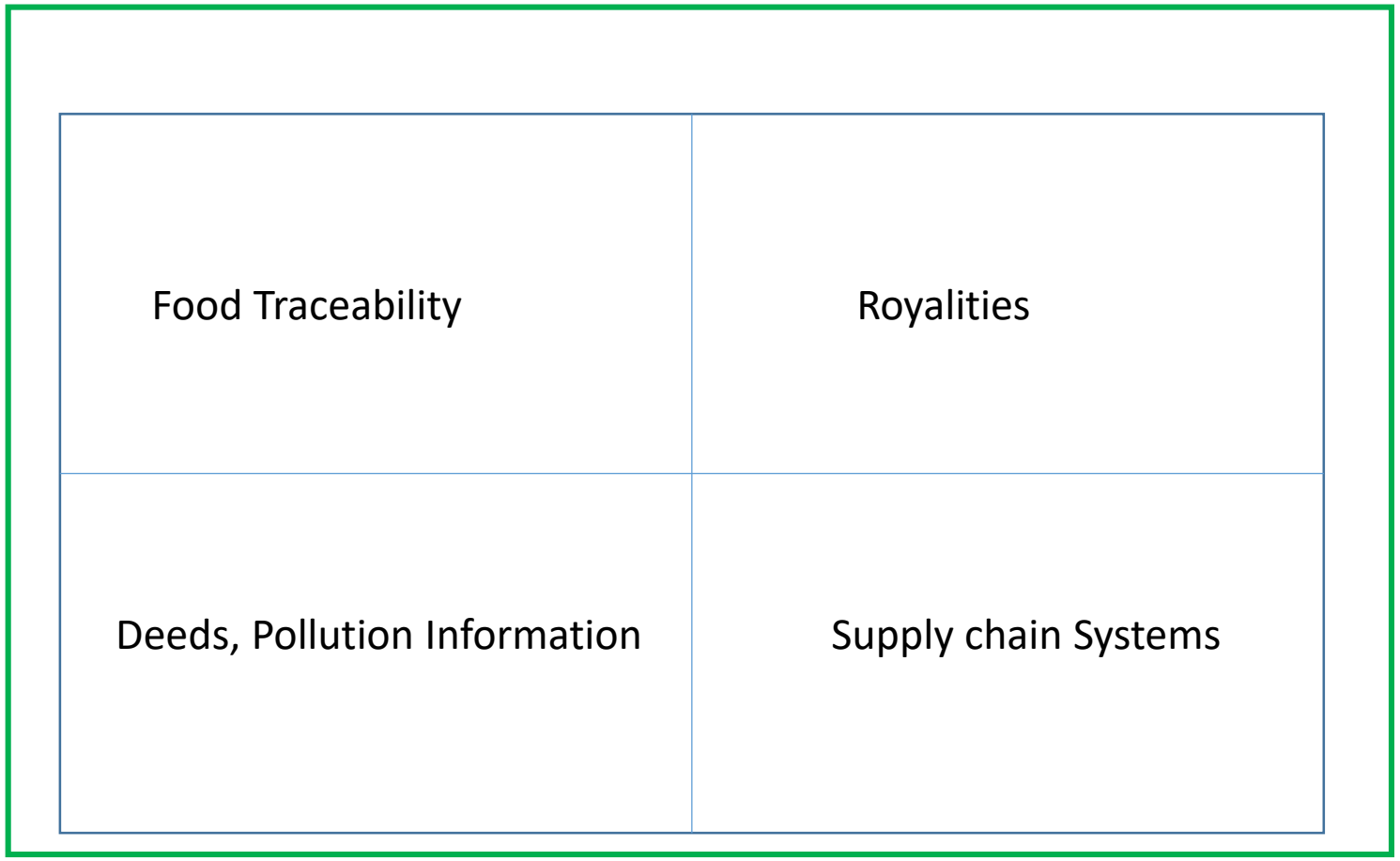
Blockchain Applications – Original Sweet Spot



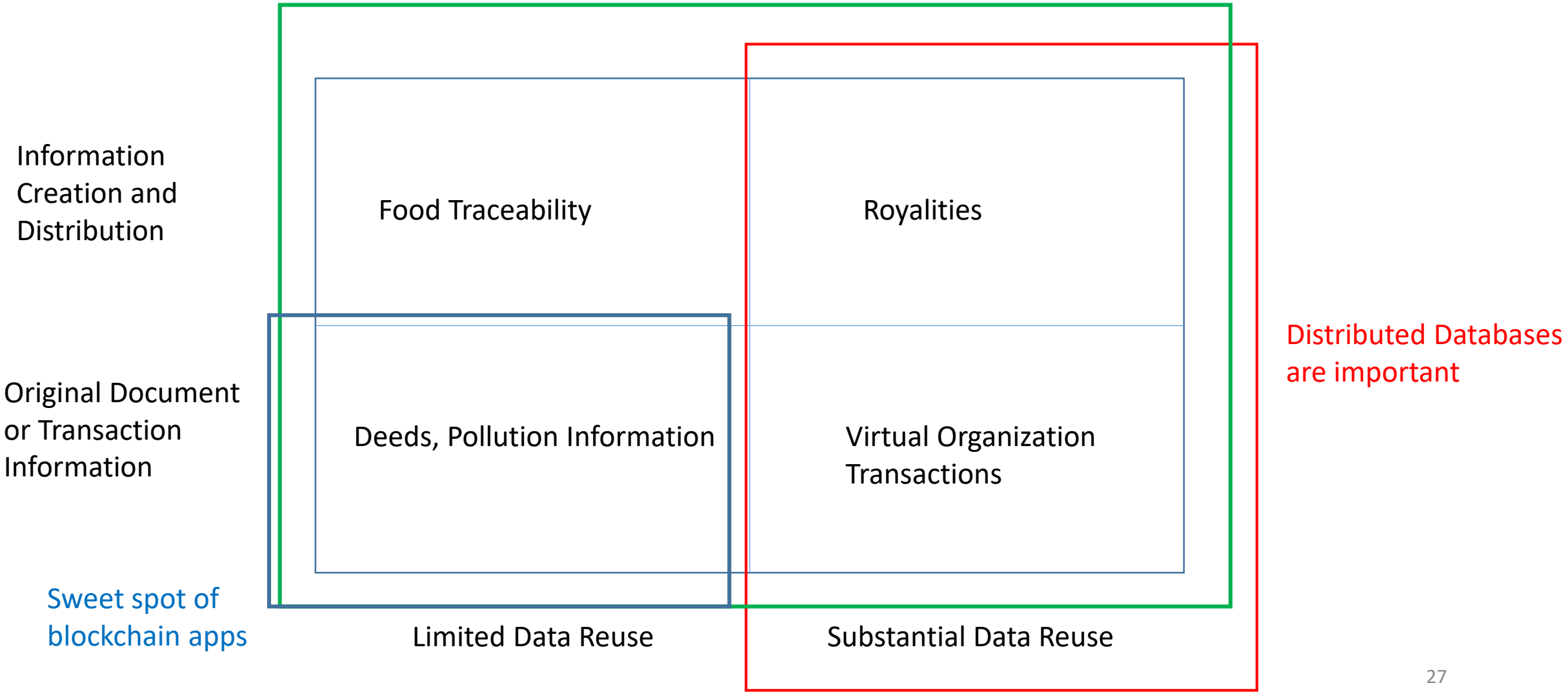
Blockchain Applications – Could benefit from distributed database



Blockchain Applications – Potentially Multiple Semantic Models



Blockchain Applications



And I still want a good application of “blockchain” (or “blockchain-like”) in accounting

- Generally, firms do not want accounting information to be public
  - But blockchains make much information public, if only the number of transactions
- In addition, accounting systems need to be able to query the accounting information.
  - But blockchain is a ledger – basically a flat file
  - Blockchain does not provide a query capability
- As a result, it is not clear what role straight accounting transactions can play in a blockchain environment.
- This paper presents one such application, where a “blockchain like” approach is used for virtual organizations.

Paper discusses building a distributed database

- Model uses BigChainDB
- Uses Jason – an Xml replacement
- Generates digital representation of assets
- …

Sale of an Asset

```
function createAssetsale() {
  // Construct a transaction payload
  const txCreateAssetsale = BigchainDB.Transaction.makeCreateTransaction(
    // Asset field
    {
      warehouse1,
    },
    // Metadata field, contains information about the transaction itself
    {
      datetime: new Date().toString(),
      location: 'Las Vegas',
      value: {
        value_dol: '$2000000',
      }
    },
    // Output. Create Ed25519 condition
    [BigchainDB.Transaction.makeOutput(
      BigchainDB.Transaction.makeEd25519Condition(star.publicKey))],
    // Issuers
    star.publicKey
  )
  // The owner of the building signs the transaction
  const txSigned = BigchainDB.Transaction.signTransaction(txCreateAssetsale,
    star.privateKey)
  // Send the transaction off to BigchainDB
  conn.postTransactionCommit(txSigned)
    .then(res => {          document.body.innerHTML += '<h3>Transaction created</h3>';
      document.body.innerHTML += txSigned.id
      // txSigned.id corresponds to the asset id of the warehouse1
    })
}
```

Both examples are based on https://www.bigchaindb.com/developers/guide/tutorial-piece-of-art/ . Ed25519 is a public key signature system. See Section 9 for discussion.

Virtual Organization

- Goldman et al. (1995) define a virtual organization as occurring when ”… complementary resources existing in a number of cooperating companies are left in place, but are integrated to support a particular product effort for as long as it is viable to do so. . . . Resources are selectively allocated to the virtual company if they are underutilized or if they can be profitably utilized there more than in the ‘home’ company.”
- Virtual organizations are designed to facilitate multiple types of capabilities, including:
  - Create or assemble productive resources quickly,
  - Create or assemble productive resources frequently and concurrently,
  - Create or assemble a broad range of productive resources (such as research, manufacturing, and design)
  - Provide resources to others in the form of partial or complete products.

Virtual Organizations

- Virtual organizations have been applied in a number of situations, ranging from manufacturing and supply chain, to white collar functions, such as accounting. For a detailed example, see Gunasekaran and Ngai (2004) who examined the use of virtual organizations in the supply chain.
- Virtual organizations attempt to employ slack resources, such as machines, buildings and people, in order to create additional value. They assemble resources in conjunction with other enterprises in loosely configured confederations of firms.
- As a result, many of the actions of interest in a virtual organization relate to organizational use of assets, not just the production of goods.
  - For example, in virtual organizations, firms' contributions can come from the amount of time that a facility, machine or personnel are involved in the confederation's project.
  - Further, it can be important for participating firms to inform others in the confederation of the planned and actual use of assets to mitigate asymmetries of information.

Virtual Organizations

- Generally, those confederations have limited centralized control and hierarchy, and are heavily decentralized.
- Thus, to ensure coordination in virtual organizations there needs to be a communication of plan and actual production information.
- As a result, any system designed for virtual organizations likely is necessarily a distributed system designed to ensure that members get timely information from other participants.

Virtual Organizations are facilitated by technology

- Typically, the coordination between the confederation enterprises is done using a range of technologies.
- Kurumluoglu et al. (2005) define a virtual organizations "… as a set of co-operating (legally) independent organizations, which to outside world provide a set of services as if they were one organization, supported by a computer network."
- However, virtual organizations do not just use networks.
- For example, a number of artificial intelligence applications have been built to help coordinate projects in virtual organizations and many intelligent agent applications have been built around the notion of virtual organizations (e.g., Cordchado et al. 2013).

Trust in a Virtual Organization

- In a world where self-interest can be assumed of most agents, trust is particularly important in virtual organizations because there is limited centralized and formal hierarchy and control. As noted by Handy (1995) in a discussion of virtual organizations, the notion of trust in virtual organizations raises some concerns:
  - “It is unwise to trust people whom you do not know well, whom you have not observed in action over time, and who are not committed to the same goals.”
  - “At its simplest, the managerial dilemma comes down to the question, ‘How do you manage people whom you do not see?’”

Trust in a Virtual Organization

- One approach to attaining trust in virtual organizations is to ensure transparency of asset use, by generating timely information flows between members of the virtual organization.
  - Public information flows
- Another approach to generating trust is to ensure that members provide the resources promised and that tasks are done at the times they were promised.
  - In this last setting, trust is attained because the members know what other members are doing and that they are doing their job for the project.
- A third approach to building trust is to provide each participating organization all of the accounting information relating to the virtual organization, eliminating the asymmetries of information of a centralized accounting system.
  - Thus, technologies designed to support virtual organizations must provide that transparency and capture information about member asset use, with information provided by a distributed, real-time accounting or supply chain system.

Virtual Organizations

- Because enterprises in virtual organizations employ their own resources or make expenditures on behalf of the virtual organization, there is a need to capture and forecast information about that resource use.
- Unfortunately, it can be quite challenging to measure, evaluate and track the work done by the different component organizations within the virtual infrastructure.
- Further, organizations with such confederations are likely to have their own accounting systems and those systems are not likely to be compatible with each other. In addition, component organizations need to coordinate efforts in order to create the anticipated value, and mitigate potential coordination problems.
- Finally, this all needs to be done in real time so that the component organizations can meet the various project needs.

Accounting Systems for Virtual Organizations

- Thus, virtual organizations require an accounting system that will integrate across each of the organizations in a timely manner.
- Unfortunately, as noted by Lee (2014) and others, such systems are a major challenge to virtual organizations.
- As a result, a peer-to-peer system based on a distributed database could be a critical contribution to virtual organizations.
- As discussed below, one approach is to have a design for the virtual organization at the overall public level

Rosetta Net as a Model

- Rosetta Net has long had the model of “public” and “private” transactions
- The Rosetta Net model would capture all of the “public” transactions and then individual firms transfer the transactions to their own private firms.
- The approach could use a blockchain among VO members but could also use a distributed database approach, as with BigChainDB.

Summary

- Blockchain models appear to use multiple semantic models.
  - In the case of private blockchains those models likely favor the developer. But the key is to know that there can be multiple models.
- Blockchain seems to ignore the approach of separating data from the models and the presentation
  - Data is basically captured as a flat file, potentially limiting complexity
- Distributed Databases, such as BigChainDB offer another alternative to blockchain with some distinct advantages.
- Use blockchain/BigchainDB for accounting for Virtual Organizations.
  - Transactions public to the entire virtual organization – facilitating trust

Questions?