Blockchain Technology
for
Scientific Data?!

Peter Wittenburg
Max Planck Computing & Data Facility

wrote a paper together with Wolfgang Kuchinke,
Heinrich-Heine University Düsseldorf
Wittenburg and Kuchinke: Blockchain and Data (2017)

Speakers & Discussion

14.00 Peter Wittenburg
>   Intro to BCT in Science and link to RDA/FAIR

14.20 Edwin Morley-Fletcher, Lynkeus
>   Applying blockchain in Health sector
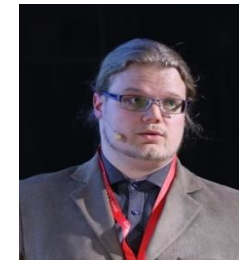>   (One of the co-chairs of new BCT RDA group)

14.40 Monique Morrow, Humanized Internet
>   Providing Self-Verification Capabilities via Blockchain

15.00 Visa Vallivaara, VTT
>   Applying blockchain to increase trust in AI

15.00 Discussion

15.30 End

Data in Science

- Data since about 2010 a top priority

  - Riding the Wave document kicked it off in EU

- Complexity and volumes create challenges for management, access and analytics

- Sharing & reuse lead to challenges for trust

  - EUDAT Sensitive Data Group: integration of sensitive and open data

- Quite some steps since then

  - Much push to Open Data, but there is
    protected data (sensitive, commercial interest, etc.)

  - 2008 DSA/WDS certification initiatives
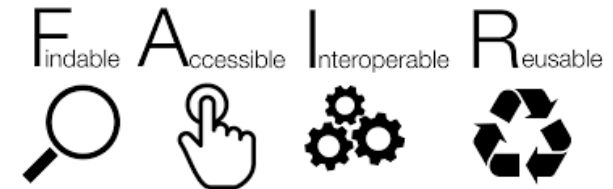   to increase trustworthiness of repositories
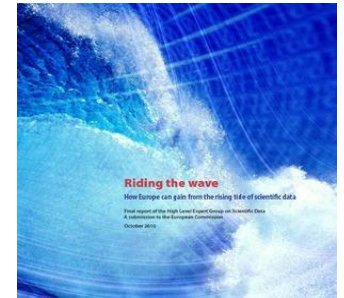
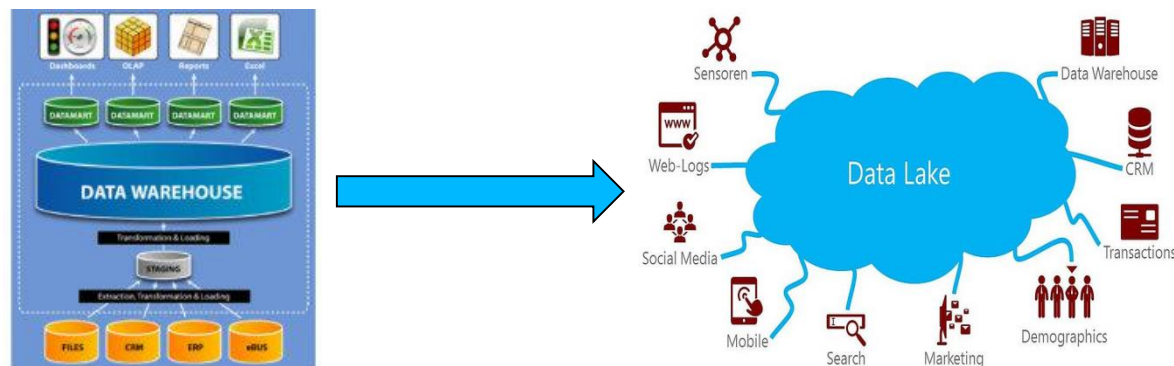# Data in Science

- some more steps since then

  - 2014 RDA DFT Core model
    making Digital Objects the core
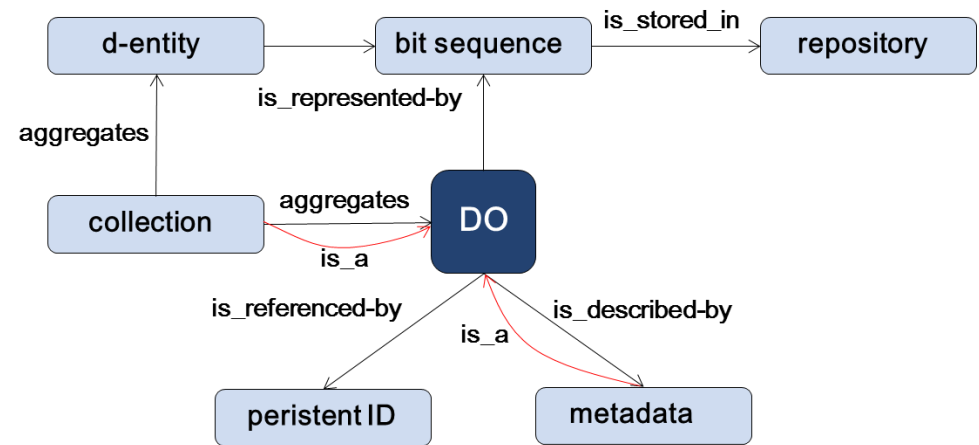
  - 2014/5 FAIR principles
    how to facilitate exchange and reuse

  - Since 2015: In industry opening from Data Warehouse to Data Lake

Digital Objects are Core

DOs are the „atoms" of our digital data world, since it makes sense to associate relevant characteristics with them. It's conceptual and powerful.

FAIR & DFT

- associate metadata and PIDs with atomic entities and collections of such entities
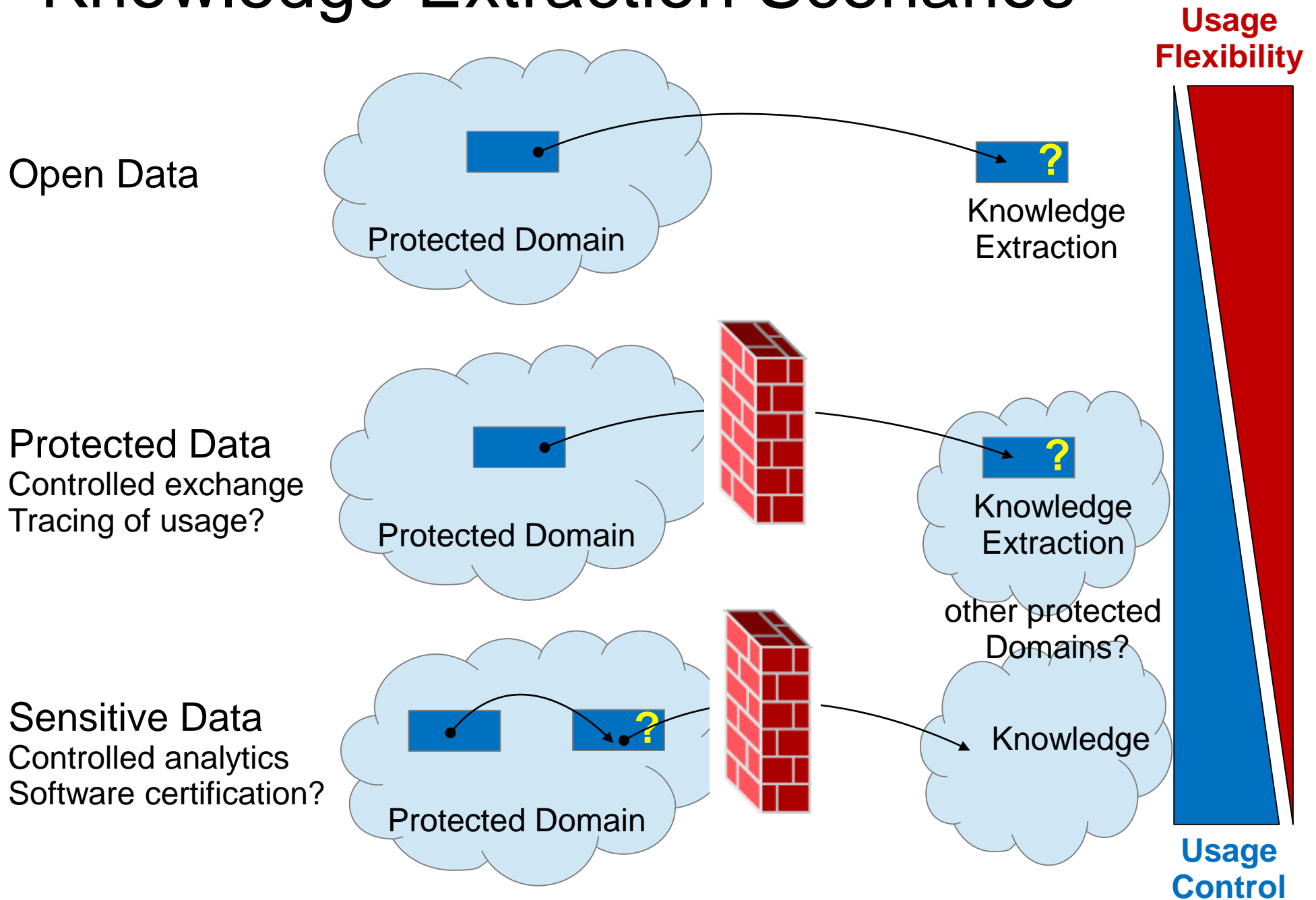- make properties of DOs explicit (from PIDs to semantics)

Data Lake

- maintain different data entities as part of the whole dynamic data set

Cloud Systems

- a local implementation of the DO concept

Knowledge Extraction Scenarios

Usage Flexibility

Open Data

Protected Domain

Knowledge Extraction

Protected Data
Controlled exchange
Tracing of usage?

Protected Domain

Knowledge Extraction

other protected Domains?

Sensitive Data
Controlled analytics
Software certification?

Protected Domain

Knowledge

Usage Control

Trust Federations

Trust Federations are collaboration based on formal agreements forming a closed domain – activities linked to clear IDs.
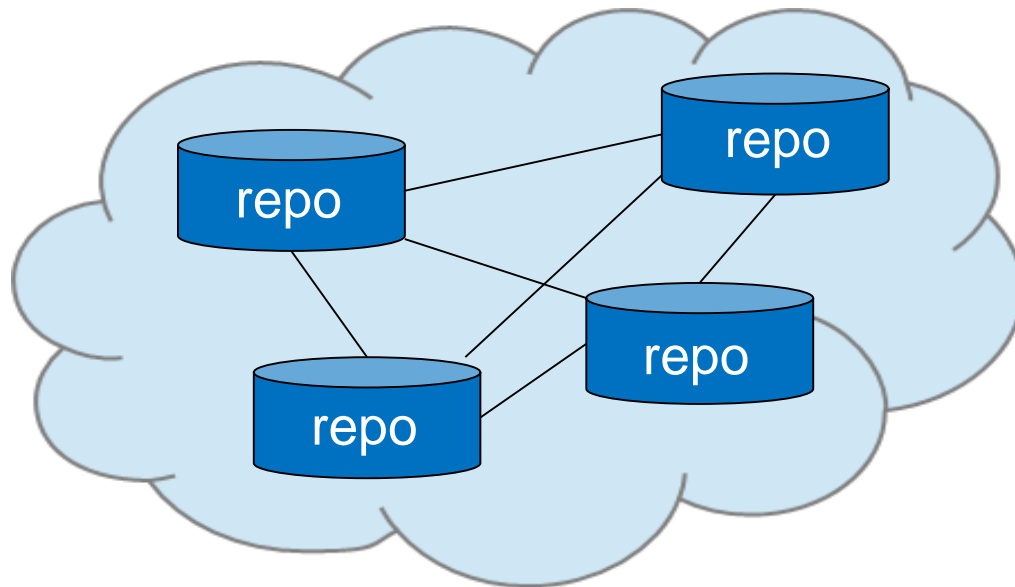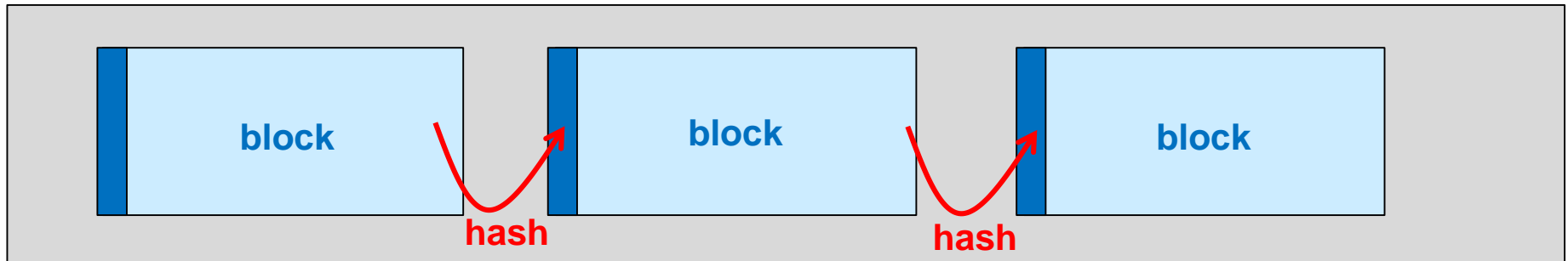

most probable protected & encrypted
- exchange of credentials
- exchange of information
- exchange of data, code
- etc.

- Example EUDAT
  - shared use of storage capacity
  - transparent replication of data (KB up to TB)
  - certification between nodes

- Other examples: PRACE, GRID, Cloud Systems, etc.

BCT: Blockchain Technology (brief)

- **Core of BCT** is a special database structure:
    - hash value summarising previous block is part of next block and thus is included in hash value of next block – changing of a block would require to change all subsequent blocks
- **Other core elements**:
    - blockchain information is exchanged between all nodes in BC network (shared ledger) by secure protocols
    - complex decision algorithm between nodes to come to agreements – therefore no need for a central instance
    - „smart" actionable contracts to define all kinds of actions in BC networks

- BCTs are tightly controlled trust federations and thus closed domains
- BCT is turnkey technology for suitable applications with immutable data

Blockchain Technology (brief)

- Price can be high:
    - severe costs for complex decision algorithm in terms of CPU power – leading to a scaling problem
    - useful only for small data (such as transactions, consent statements, etc.)
    - science requires FAIRness and flexibility
    - research with sensitive data requires trust
    - scientific data to be re-used outside of closed trust federation

- But data has different components
    - scientific data content   can be large, to be FAIR, flexibility
    - metadata   to be open, to be FAIR
    - transactions   administrative info, different!

- BCT memorises safely transactions of data into other (trust) domains
- BCT useful for admin data where necessary (sensitive data, industry)
- BCT not so useful for scientific data and metadata
- BCT comes in different flavours (public, private, Hyperledger, Ethereum, etc.)
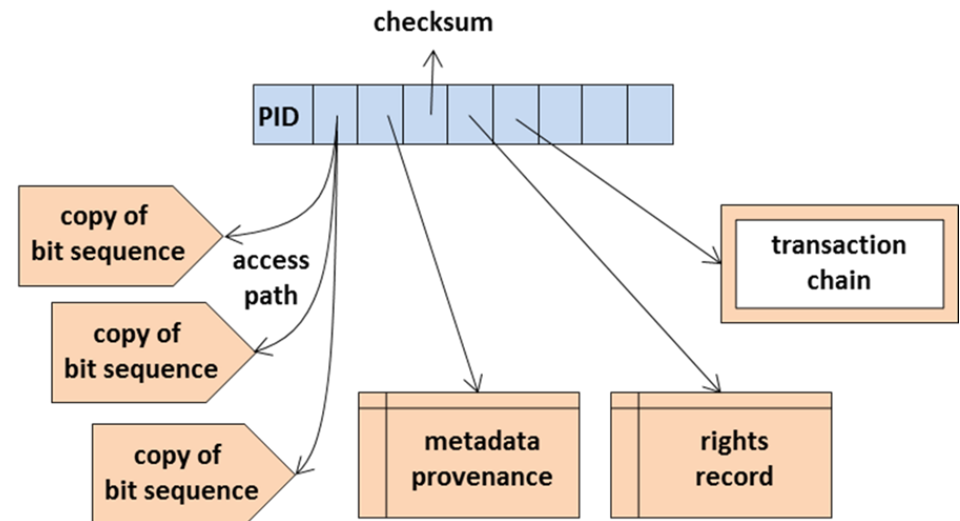
BCT and Digital Objects

- Should blocks with transactions be Digital Objects?

  - should they have a PID?

  - should they have metadata?
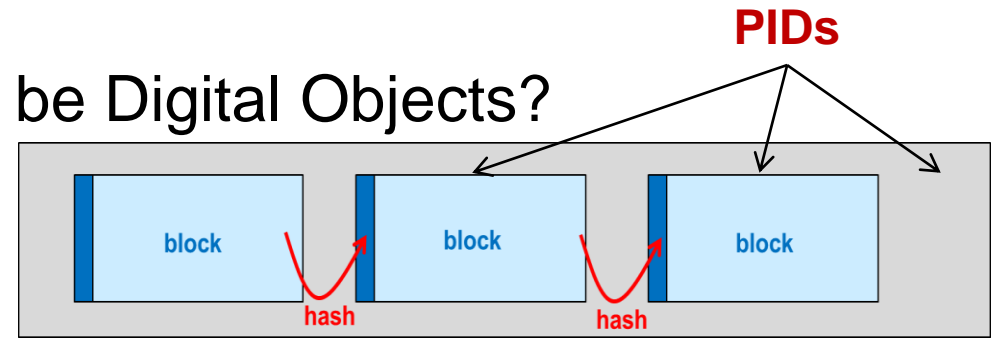
- How to bind transactions?

  - a scientific data entity has a PID

  - a PID record can be used to
    bind various information

Questions

- For what is BCT useful and for what not in data management & access (beyond all hype)?

- How can we use BCT in science without loosing flexibility where this is required?

- How can we integrate BCT functionality into our solutions helping to solve open issues?

- How to achieve full traceability when data leaves trusted space?

- What are the costs instead of using traditional technology?

Speakers & Discussion

14.00   Peter Wittenburg
>   Intro to BCT in Science and link to RDA/FAIR

14.20   Edwin Morley-Fletcher, Lynkeus
>   Applying blockchain in Health sector
>   (One of the co-chairs of new BCT RDA group)

14.40   Monique Morrow, Humanized Internet
>   Providing Self-Verification Capabilities via Blockchain

15.00   Visa Vallivaara, VTT
>   Applying blockchain to increase trust in AI

15.00   Discussion

15.30   End