

ChemChain @

open **chemical** identifier on the blockchain -
InChI
a decentralized and authoritative identifier

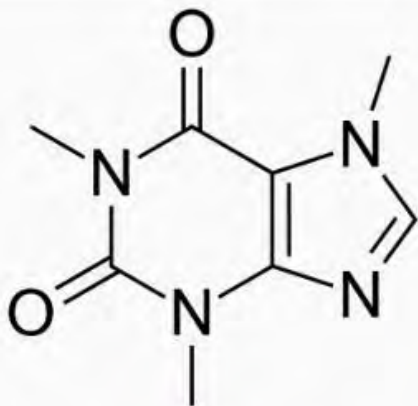
Executive summary:

















open algorithmic identifiers
+ blockchain

= open authoritative
identifiers

The present chemical identifier ecosystem:

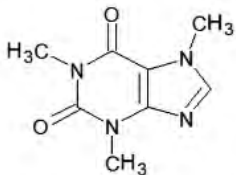
Caffeine



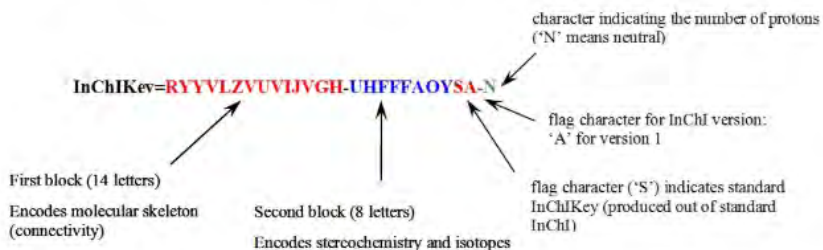
Identifiers		Chemical and physical data	
IUPAC name		Formula	$C_8H_{10}N_4O_2$
	1,3,7-Trimethylpurine-2,6-dione	Molar mass	194.19 g/mol $g \cdot mol^{-1}$
CAS Number	58-08-2  ✓	3D model (JSmol)	Interactive image 
PubChem CID	2519 	Density	1.23 g/cm ³
IUPHAR/BPS	407 	Melting point	235 to 238 °C (455 to 460 °F) (anhydrous) ^{[8][9]}
DrugBank	DB00201  ✓	SMILES	<code>CN1C=NC2=C1C(=O)N(C(=O)N2C)C</code> [hide]
ChemSpider	2424  ✓	InChI	InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3 [hide]
UNII	3G6A5W338E 	Key:	RYYVLZVUVIJVGH-UHFFFAOYSA-N ✓
KEGG	D00528  ✓		
ChEBI	CHEBI:27732  ✓		
ChEMBL	ChEMBL113  ✓		
PDB ligand	CFF (PDB  , RCSB PDB )		
CompTox Dashboard (EPA)	DTXSID0020232  		
ECHA InfoCard	100.000.329  		

Algorithmic identifiers: InChI

InChI string and key: example caffeine



InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H,1-3H,1-3H,1-3H (caffeine)



Considerations regarding **safety** and **authority** of the InChI system:

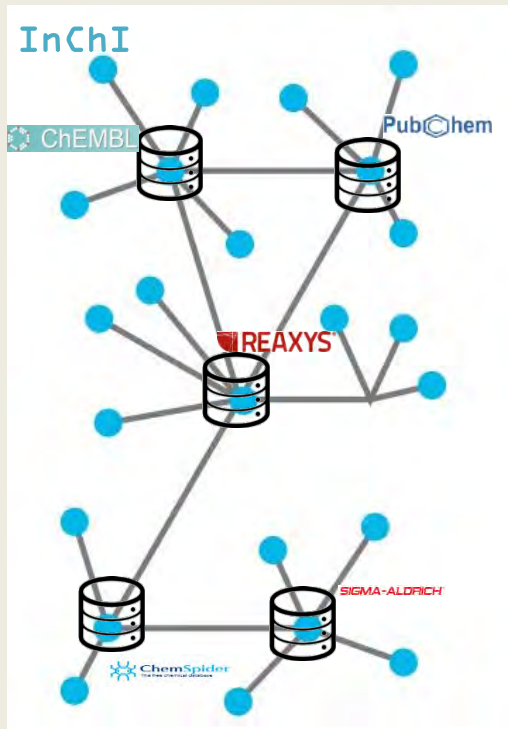
- InChI versions
- Hash collisions
- Trusting the InChI sender/provider is required.

InChI is an **open** textual identifier for chemical substances, designed to provide a standard way to encode molecular information and to facilitate the search for such information in **decentralized**

The InChI system: from a network perspective

Decentralized DBs using

InChI



Inconsistency of InChI and other systematic

identifiers in different data

Table 3 Consistency of MOLs and systematic identifiers (in % agreement) within databases

Database	MOL-InChI	MOL-SMILES	MOL-IUPAC
DrugBank	98.2	98.5	90.0
ChEBI	96.5	96.5	75.3
HMDB	89.3	37.2	55.7
PubChem	97.7	97.8	87.2
NPC	-	93.4	

Akhondi et al. *Journal of Cheminformatics* 2012, 4:35
<http://www.jcheminf.com/content/4/1/35>



RESEARCH ARTICLE

Open Access

Consistency of systematic chemical identifiers within and between small-molecule databases

Saber A Akhondi^{1*}, Jan A Kors¹ and Sorel Muresan²

Abstract

Background: Correctness of structures and associated metadata within public and commercial chemical databases greatly impacts drug discovery research activities such as quantitative structure–property relationships modelling and compound novelty checking. MOL files, SMILES notations, IUPAC names, and InChI strings are ubiquitous file formats and systematic identifiers for chemical structures. While interchangeable for many cheminformatics purposes there have been no studies on the inconsistency of these structure identifiers due to various approaches for data integration, including the use of different software and different rules for structure standardisation. We have investigated the consistency of systematic identifiers of small molecules within and between some of the commonly used chemical resources, with and without structure standardisation.

Results: The consistency between systematic chemical identifiers and their corresponding MOL representation varies greatly between data sources (37.2%–98.5%). We observed the lowest overall consistency for MOL-IUPAC names. Disregarding stereochemistry increases the consistency (84.8% to 99.9%). A wide variation in consistency also exists between MOL representations of compounds linked via cross-references (25.8% to 93.7%). Removing stereochemistry improved the consistency (47.6% to 95.6%).

Conclusions: We have shown that considerable inconsistency exists in structural representation and systematic chemical identifiers within and between databases. This can have a great influence especially when merging data and if systematic identifiers are used as a key index for structure integration or cross-querying several databases. Regenerating systematic identifiers starting from their MOL representation and applying well-defined and documented chemistry standardisation rules to all compounds prior to creating them can dramatically increase internal consistency.

Keywords: Molecular structure, Chemical databases, Systematic chemical identifiers, Quality control, InChI, SMILES, IUPAC

Conclusion: algorithmic

identifiers like InChI,

SMILES need to be

consistently applied among

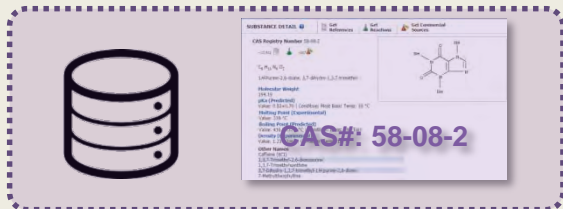
the different chemistry

sources. However,

inconsistencies exist among

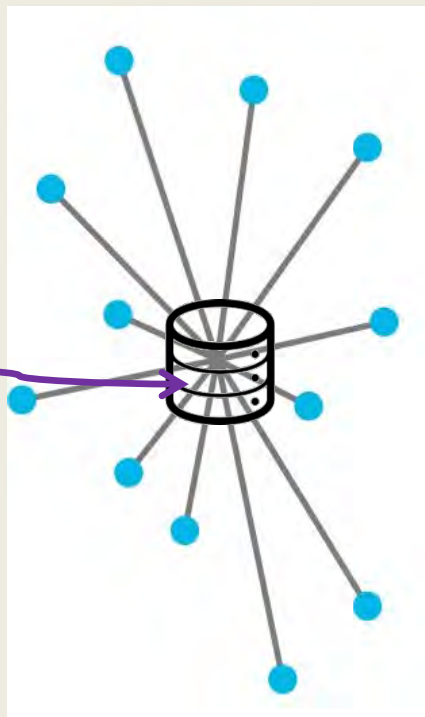
Centralized reference identifier system: network perspective

centralized



The image shows a screenshot of a chemical database interface. On the left is a database icon. The main window displays 'SUBSTANCE DETAILS' for 'CAS Property Number 58-08-2'. It includes a chemical structure diagram, molecular weight (216.16), and other properties. A purple arrow points from the database icon to the search result.

CAS#: 58-08-2

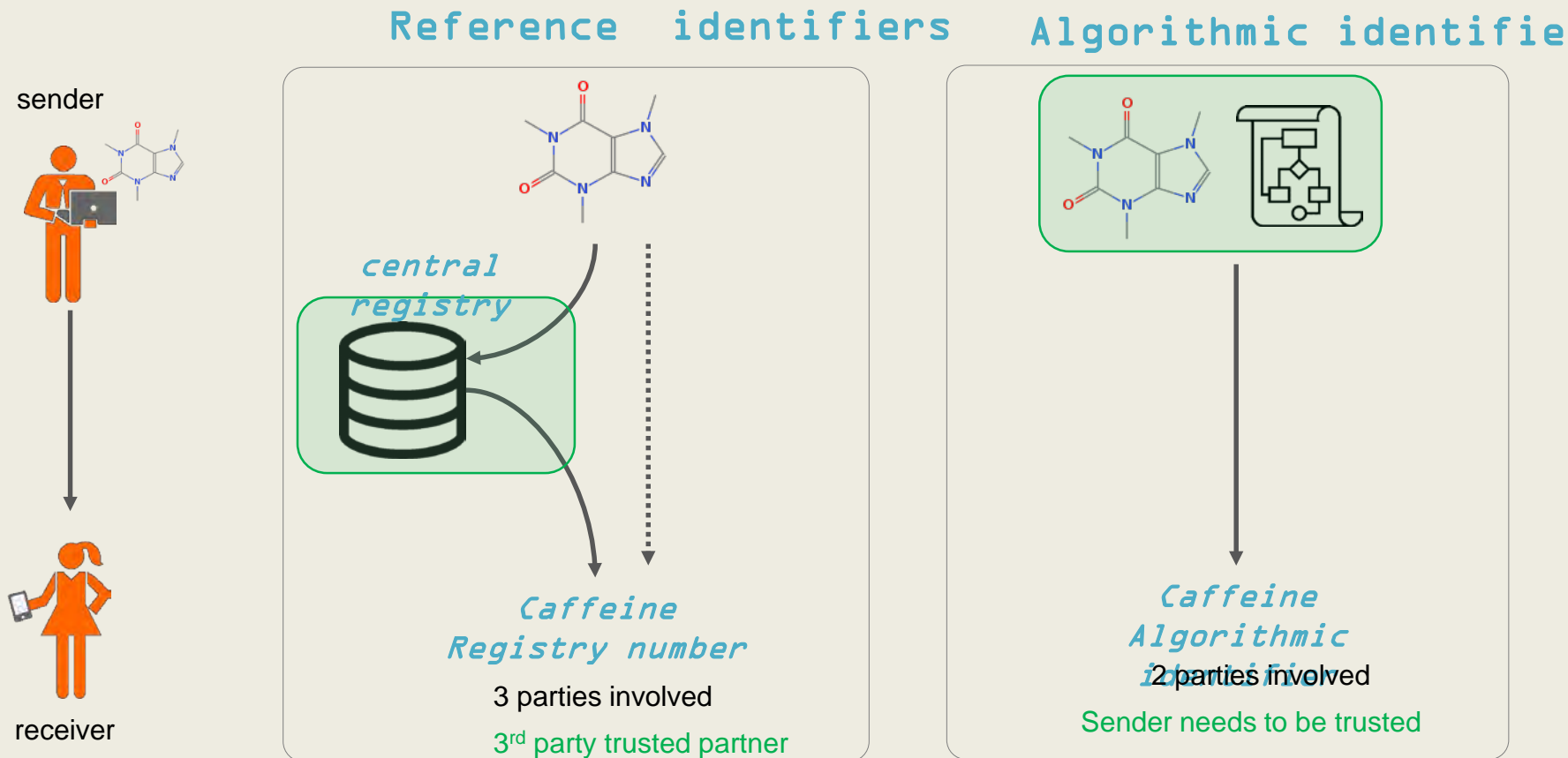


Note:

Reference identifiers are hosted on centralized systems.

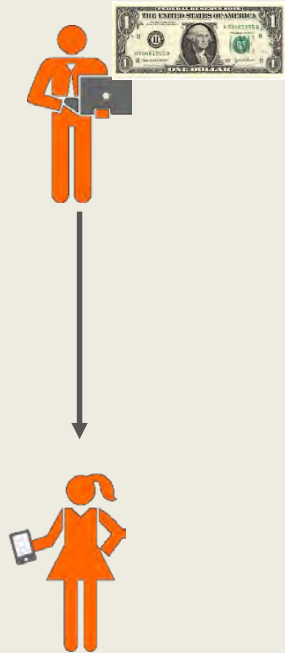
The owner of such important DB has enormous power over the users because of the dependency.

Trust models of chemical identifiers



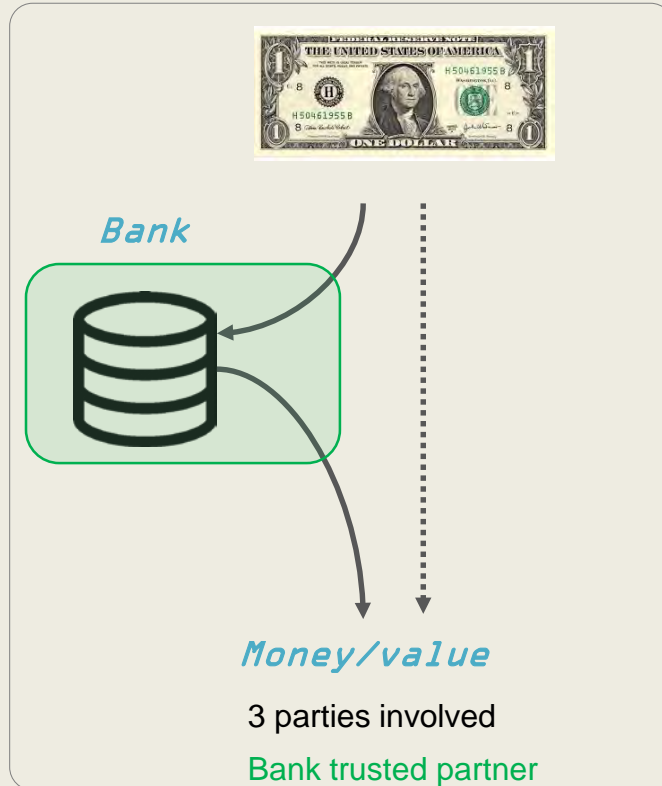
Transfer of value

sender

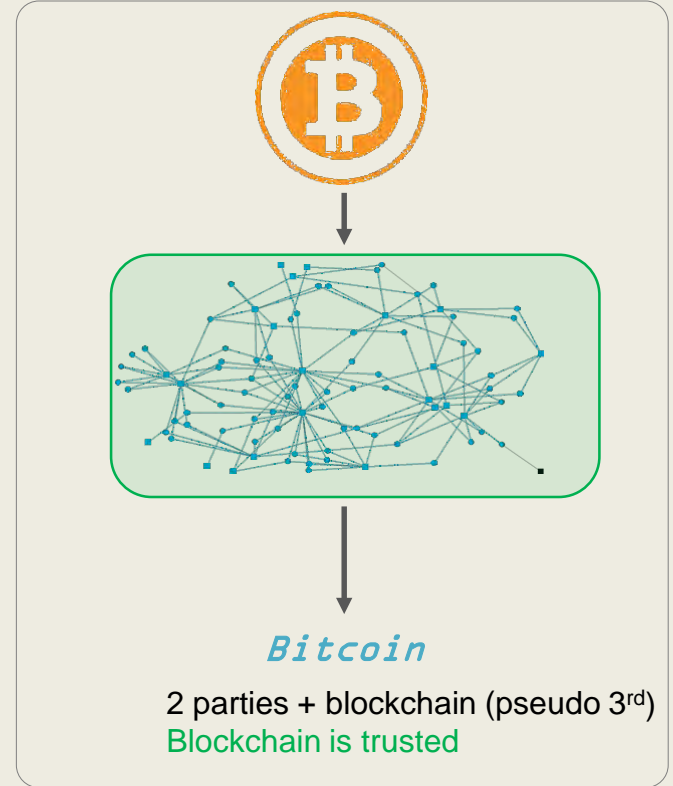


receiver

3rd party Bank



Blockchain based



Inspiration for ChemChain - Bitcoin explorer

Search for things like address, transaction, block

All Blockchains Search

Latest blocks [View more blocks](#)

Height	Hash	Mined	Miner	Size
591448	000000000000000001add5284fc5c22c73411...	15:50 PM	Unknown	1,325,534 bytes
591447	00000000000000000079f4f2f32bb1bc275277e...	15:20 PM	F2Pool	1,244,431 bytes
591446	00000000000000000019b992abdb0d5fb863a29...	15:17 PM	F2Pool	1,319,782 bytes
591445	00000000000000000008f17c81008fc3b8a772af...	14:54 PM	BTC.com	584,874 bytes
591444	00000000000000000003b4aa388985daacaa34...	14:49 PM	BTC.TOP	458,524 bytes
591443	000000000000000000033c0ecbb201398b69c2...	14:46 PM	BTC.com	1,009,367 bytes
591442	000000000000000000025edb7abd13863da67b...	14:40 PM	Unknown	1,211,424 bytes
591441	0000000000000000000fb8fa7cc77536c8f2f2f4b...	14:30 PM	Unknown	1,318,843 bytes
591440	00000000000000000002d817fbdda2667d62830...	14:17 PM	BTC.TOP	1,188,895 bytes
591439	0000000000000000000936c725985391d7477a...	14:12 PM	Unknown	1,385,586 bytes

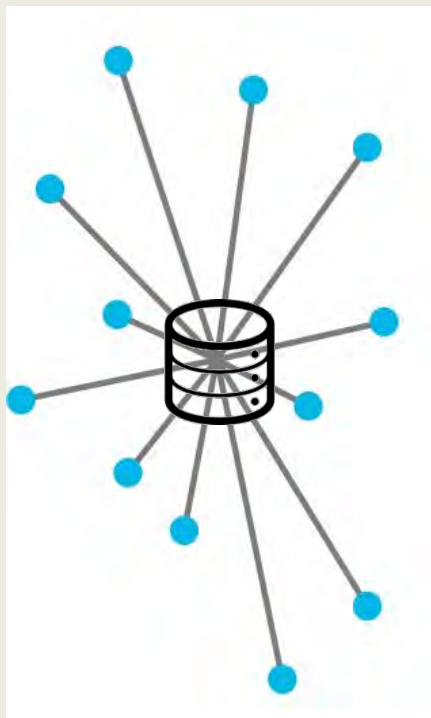
Hypophysis:

A future-proof chemical identifiers system needs to be:
authoritative, open and scalable

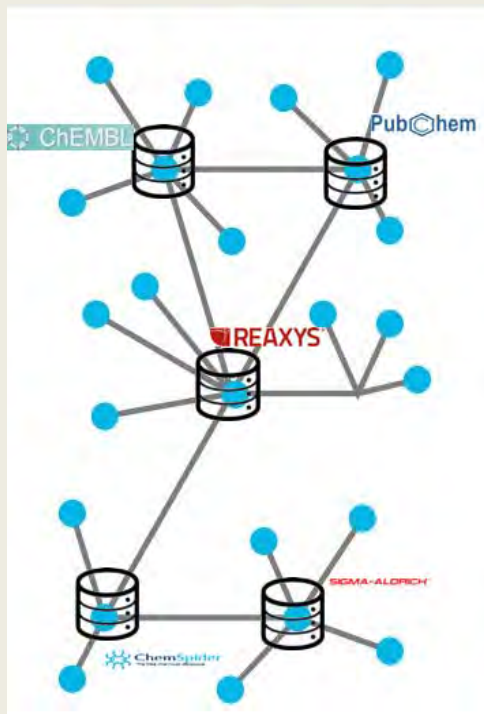
	Reference identifiers	Algorithmic identifiers	ChemChain	
- Open	NO	YES	YES	} <i>open code</i>
- Scalable	NO	YES	YES	
- Safe	YES	NO	YES	} <i>cryptography</i>
- Authoritative	YES	NO	YES	
- Decentralized	NO	YES	YES	} <i>distributed ledger P2P</i>
- Cost effective	NO	YES	YES	

Chemical identifier evolution: network point of view

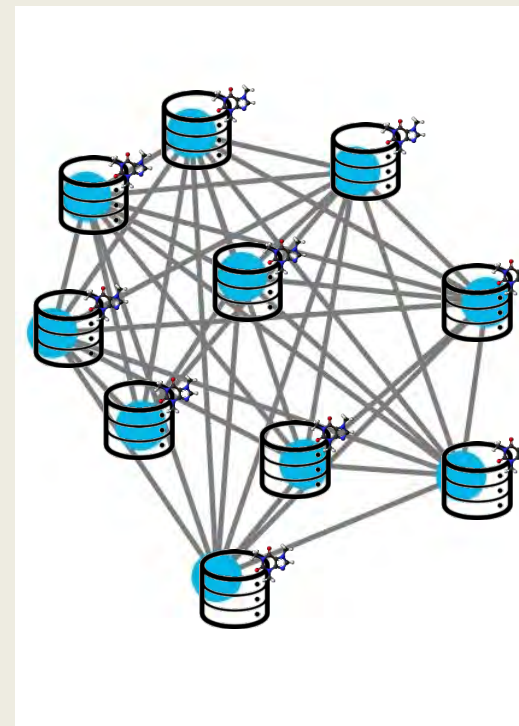
centralized



decentralized



distributed

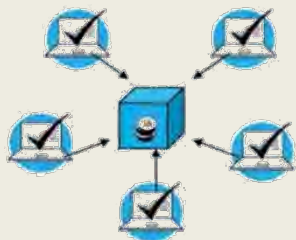


Three major blockchain components



ChemChain

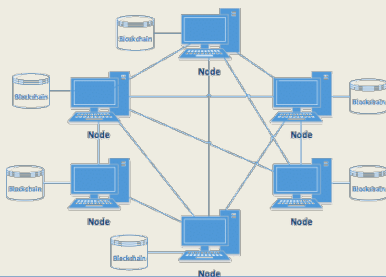
Consensus Mechanism (PoW, PoS)



Consensus mechanism prevents the double-spending problem. CM keeps the network clean and honest.

Consensus mechanism controls the uniqueness and relevance of structures in ChemChain.

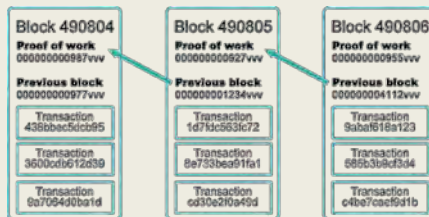
Peer-to-peer network



Distributed ledger technology enables every node to have the exact same copy of transactions.

Distributed ledger technology every chemist (organization) to keep a copy of structures.

Crypto/hashed blocks

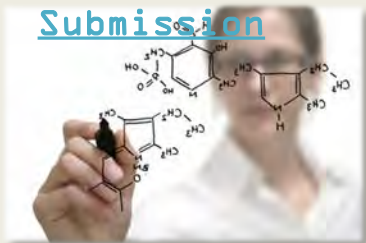


Hashed blocks keep all containing transactions immutable.

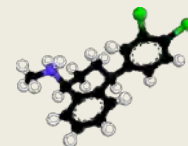
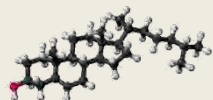
Hashed blocks make it impossible to change the registered structures.

ChemChain - chemical ID Registration process

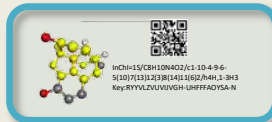
Submission



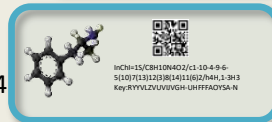
ChemChain ID-generation



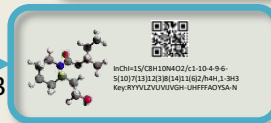
bl. ...



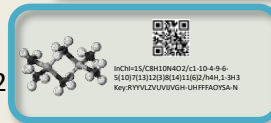
bl.4



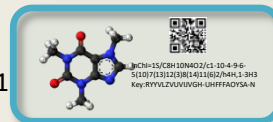
bl.3



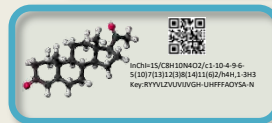
bl.2



bl.1



bl.0

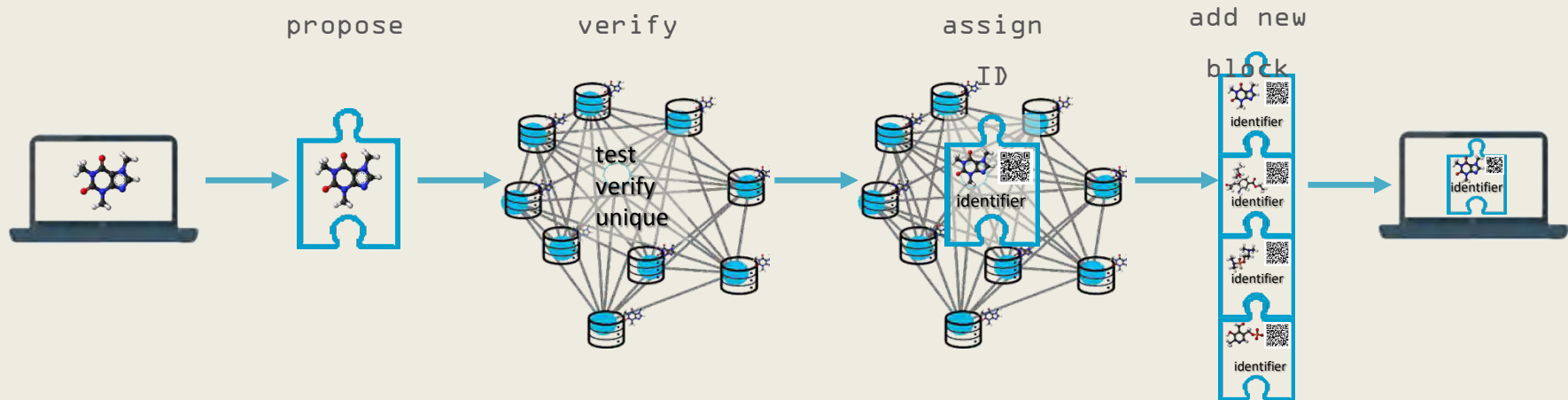
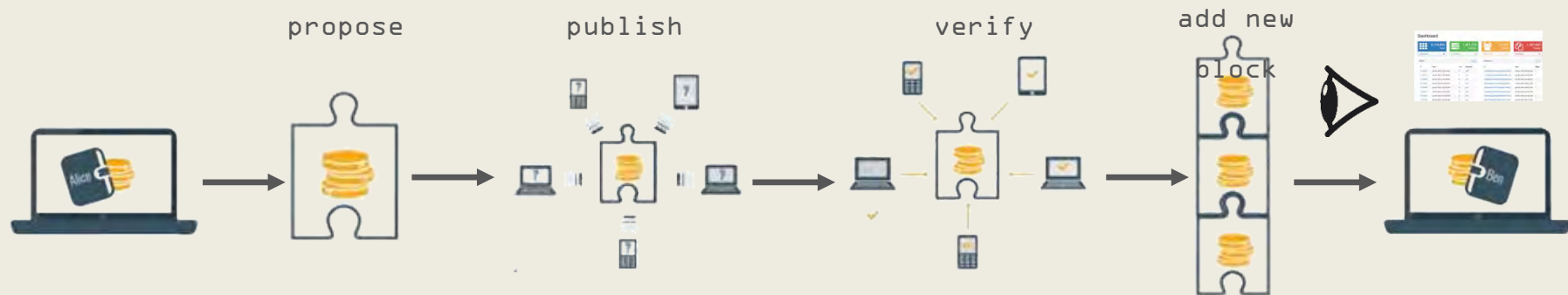


Structure-ID

Lookup

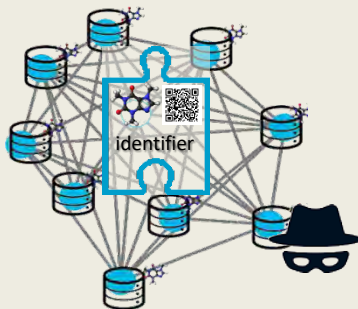


Blockchain principles of ChemChain

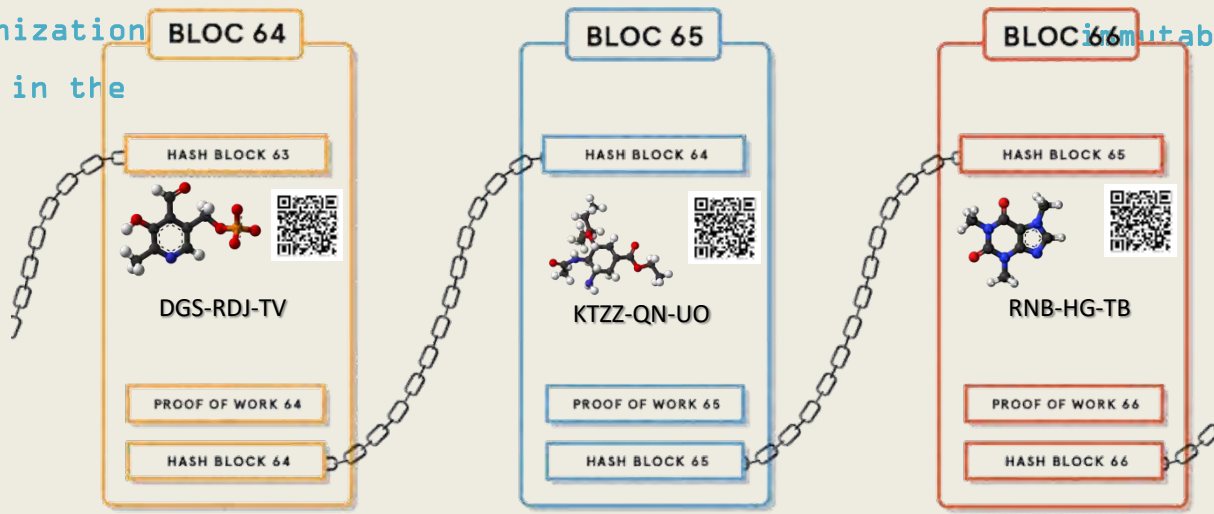


ChemChain is protected by cryptography

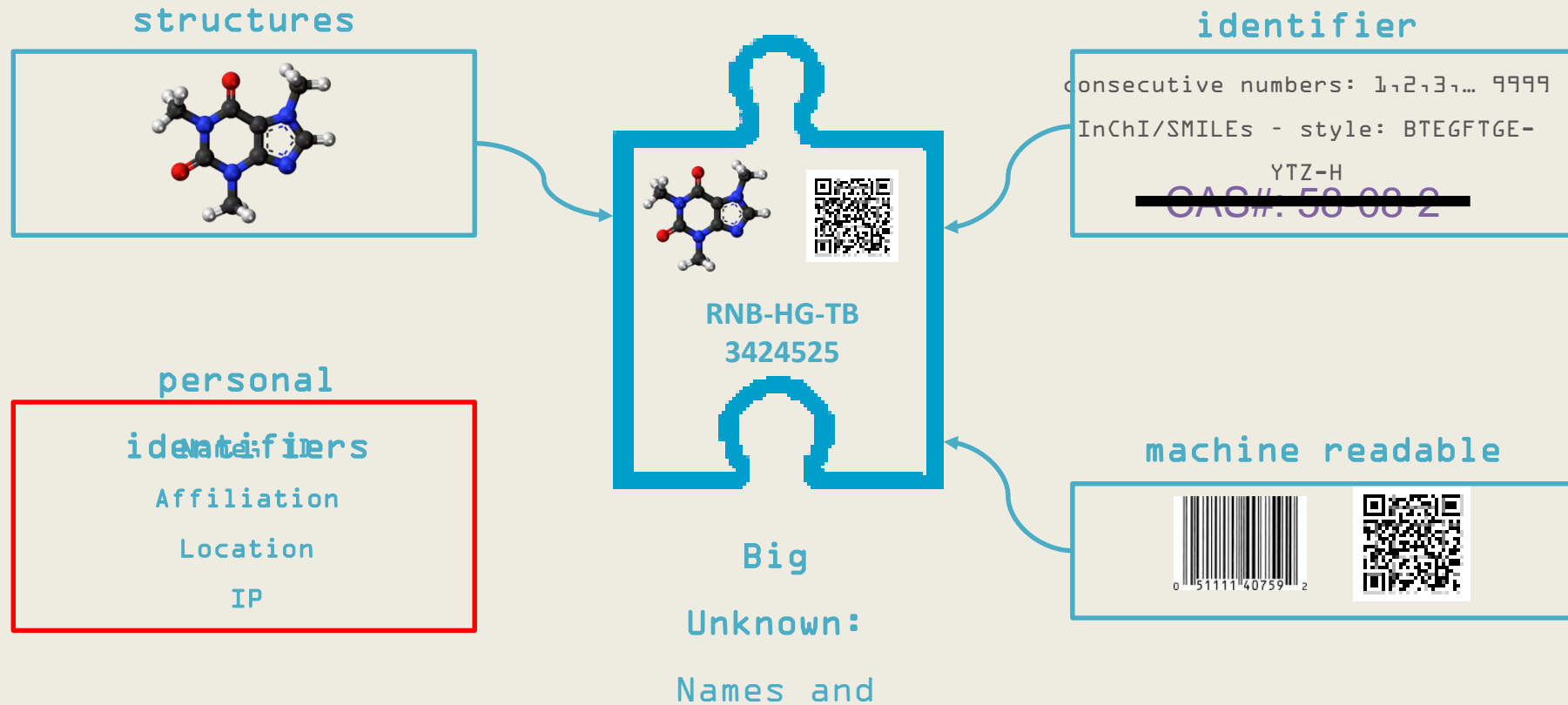
In a blockchain network all nodes have same copy of the ledger/DB. Cryptography and Peer-to-peer protocols assure the consistency and synchronization of the data in the ledger/DB.



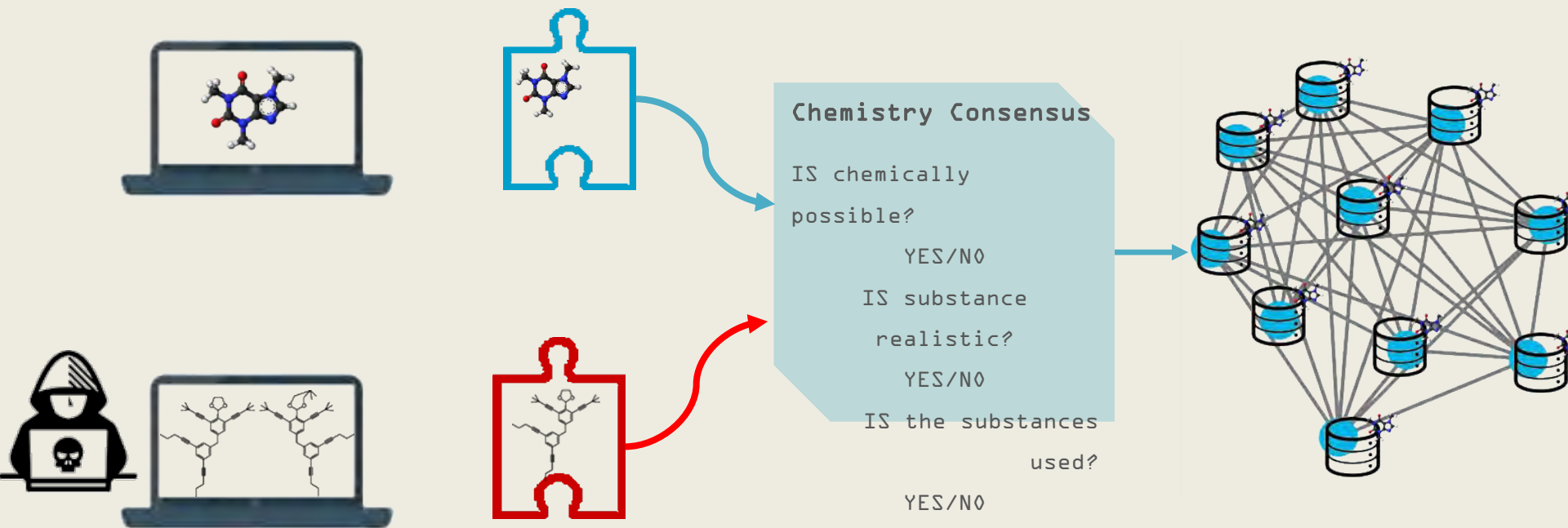
Corrupting data is practically impossible due to the economic burden of the cryptographic hash functions. Data stored in blockchains is immutable.



What information should be stored in ChemChain block?



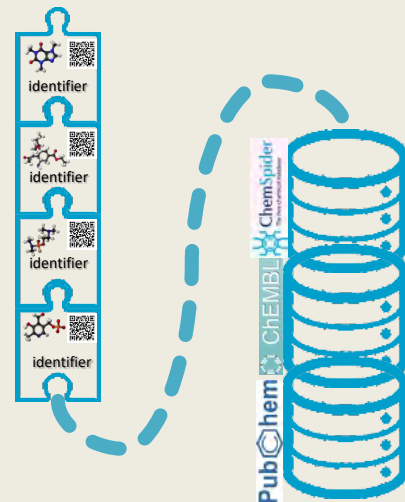
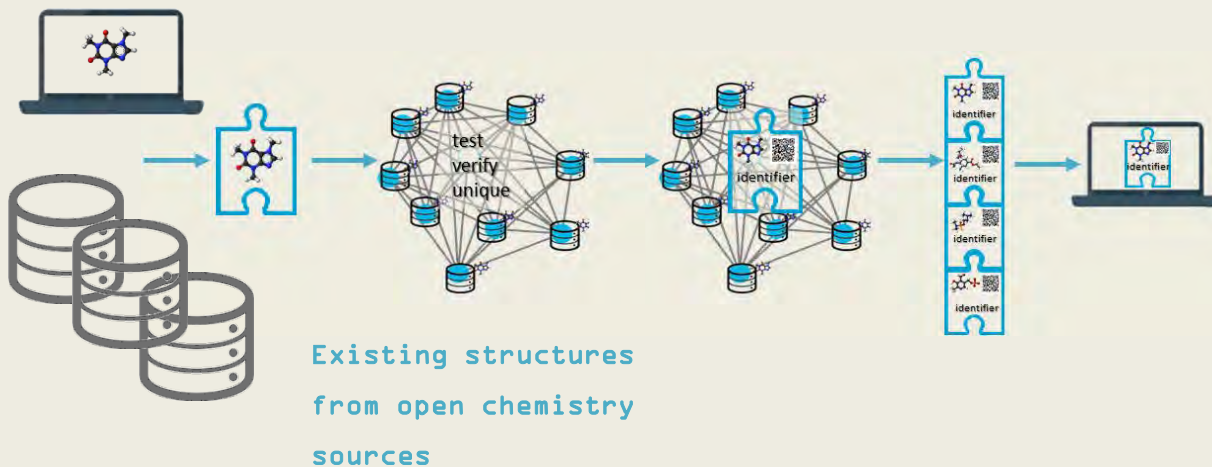
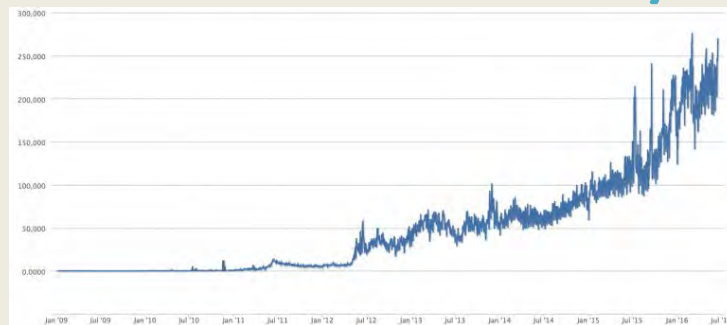
A chemistry-consensus mechanism controls ChemChain



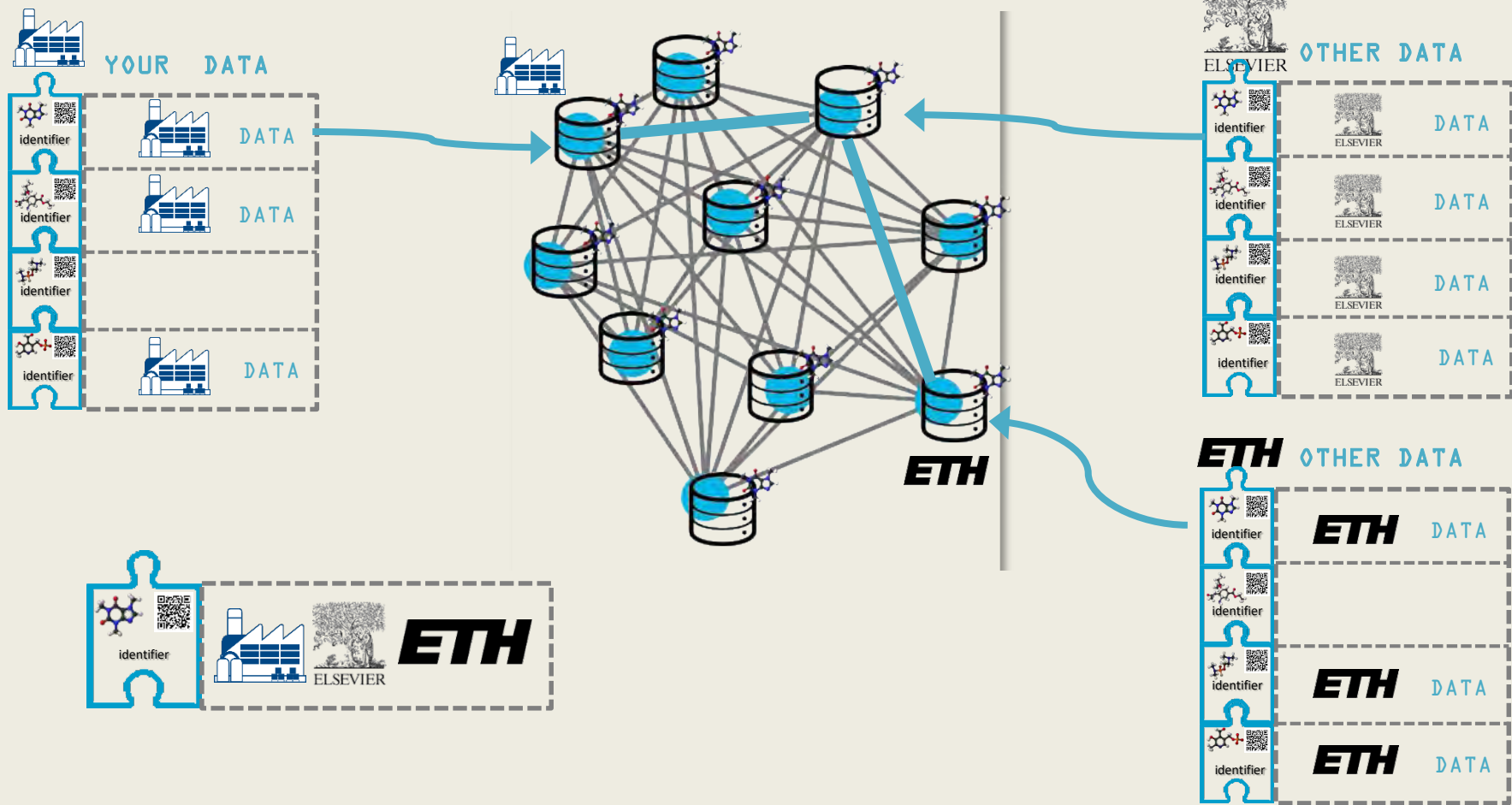
ChemChain could add existing structures quickly



Bitcoin transactions/day

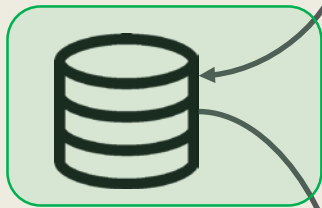
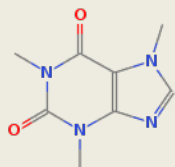


ChemChain semantically connects chemistry



ChemChain acts as a trusted authority

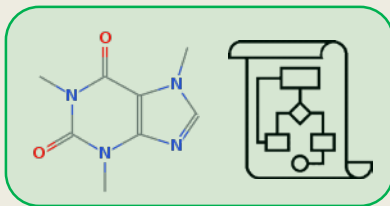
Reference identifier Algorithmic identifier Blockchain identifier



*Caffeine
Registry number*

3 parties involved

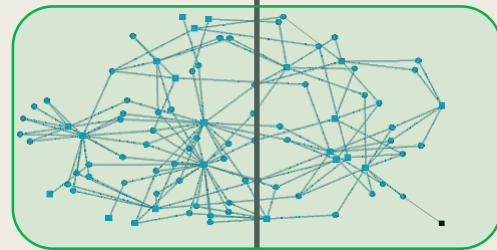
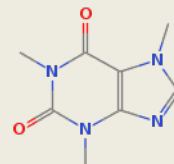
3rd party trusted partner



*Caffeine
algorithmic
identifier*

2 parties involved

Sender needs to be trusted



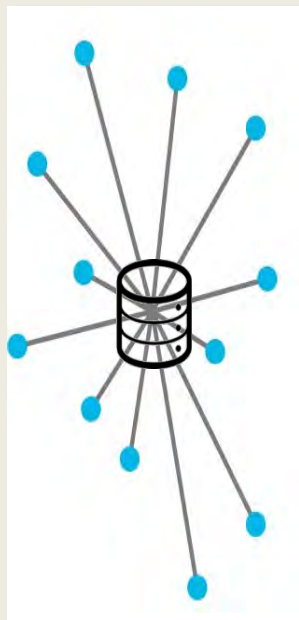
*Caffeine
blockchain
identifier*

2 parties and one BC network

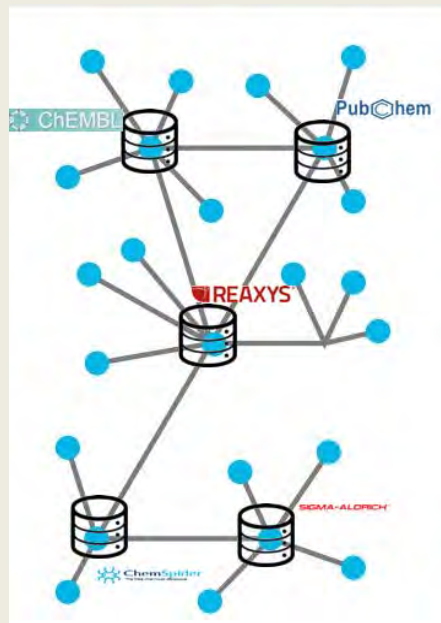
BC network is trusted

ChemChain is the better system, not the better identifier

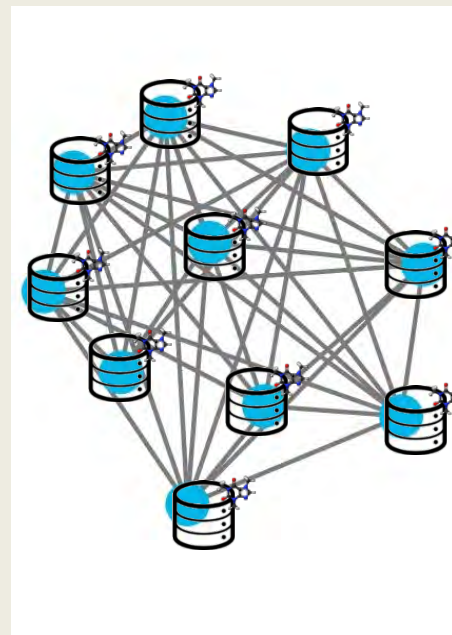
centralized



decentralized



distributed



Examples for chemical/pharma blockchain projects

Project Production Login




The world's first commodity-backed blockchain option

- Based on real production
- Complete transparency at every stage
- Unique, stable and innovative blockchain product
- Possibility to become a commodity token holder
- 100% profit for 18 months
- Over \$1 million in R&D

Blockchain can help transform supply chain networks in the chemicals and petroleum industry

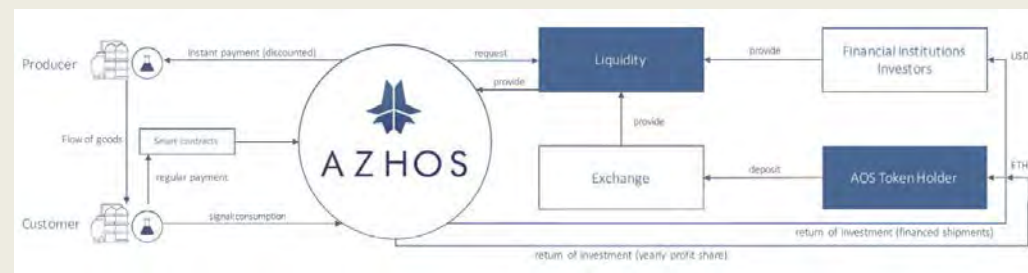
Shared-ledger technology drives substantial efficiency, fraud prevention and cost take out



Introducing Molecule

An open source ecosystem to accelerate innovation in drug development.

The Molecule Project is an open platform being developed by a Swiss non-profit foundation, led by an international team comprised of scientists, economists, software engineers and industry experts. The platform enables the economic incentives needed for distributed research and development via a unique open science and blockchain technology to create a fair and accountable market-based platform for discovery and funding of pharmaceutical research projects. The ecosystem aims to bring all stakeholders - academic pharmaceutical organizations, consumers, biotech investors, insurers, regulators and lawyers - to collaborate in an open, incentive-based market.

BASF and arc-net collaborate to use blockchain technology for livestock sustainability

FLORHAM PARK, NEW JERSEY, and BELFAST, NORTHERN IRELAND, April 17, 2018 – **BASF Corporation** and **arc-net**, are collaborating to use blockchain technology to capture and analyze sustainability parameters in livestock production along the value chain. arc-net is a technology company utilizing blockchain technology to provide transparency in the agri-food industry. As part of this collaboration, BASF will use its tailored calculation tool, **AgBalance™ Livestock**, to provide science-based information on environmental impacts along the animal production value chain.



Tony Graetzer
Walton & Pharmsa S&P
100 Park Ave, Florham NJ

With global pressure to reduce emissions from the agricultural and livestock sectors, consumers and regulators worldwide are becoming more focused on setting defined sustainability criteria. BASF and arc-net are collaborating to support the animal production value chain meet these demands. BASF's calculation tool AgBalance

Walmart considers blockchain technology for tracing chemicals

Potential to create 'a new era of transparency'

21 June 2018 | Confidentiality & right-to-know, Data, United States, Voluntary action

US retail giant Walmart is assessing whether the digital technology 'blockchain' can be used to trace chemicals across some of its products and packaging.

Blockchain is a digital record keeping system that enables the creation and maintenance of a growing number of records, allowing fast tracking of information. It was originally created to manage transactions through the crypto-currency Bitcoin, but has since shown potential for sharing and retrieving many other forms of data.




FarmaTrust

BLOCKCHAIN AND AI SOLUTIONS FOR THE PHARMACEUTICAL AND HEALTHCARE SECTOR

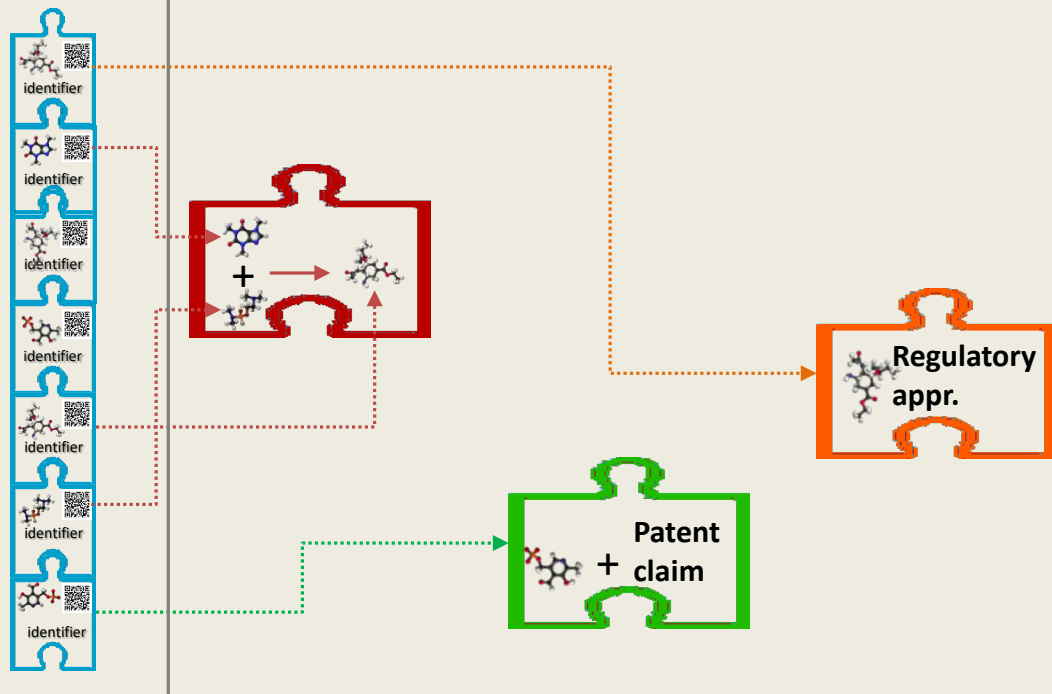
- PHARMACEUTICAL TRACKING & DATA SERVICES
- CGT & PERSONALISED MEDICINE SERVICES
- CLINICAL TRIALS SERVICES
- MEDICAL DEVICES SERVICES

ChemChain as reference layer for other service layers

Reference Layer

Service Layer

ChemChain ReactionChain PatentChain RegulatoryChain



ChemChain contact:

Dr. Markus Bussen

markus.bussen@protonmail.ch

m.bussen@elsevier.com

