THE JBBA

PEER REVIEWED RESEARCH

OPEN ACCESS
ISSN Online: 2516-3957
ISSN Print: 2516-3949
https://doi.org/10.31585/jbba-4-2-(3)2021

Blockchain-hosted Data Access Agreements for Remote Condition Monitoring in Rail

[1]Rahma A Alzahrani, [2]Simon J Herko, [3]John M Easton
[1,3]School of Engineering (EESE), University of Birmingham, UK
[2]TravelSpirit Foundation, UK

Correspondence: raa926@bham.ac.uk

Received: 23 February 2021 **Accepted:** 3 August 2021 **Published:** 12 August 2021

Abstract

Advances in sensor technologies, remote authentication, and high-bandwidth data networks mean that Remote Condition Monitoring (RCM) systems are now an essential "Internet of Things" (IoT) resource for the efficient operation of railway infrastructure. However, the full potential of the big data generated by these systems has yet to be realised. RCM data within the industry is typically collected and used in silos, with limited possibility of exploitation across system boundaries. In 2013, the Rail Safety and Standards Board (RSSB), on behalf of the GB rail industry, established a cross-industry research programme, T1010, which aimed to build stronger cooperation between stakeholders and to enable sharing of RCM data. Building on the outputs of T1010, this work explores the use of blockchains and smart contracts (SC) in the automation, in an auditable and tamper-proof way, of commercial agreements for RCM data transfers in rail. By removing the limitations of paper-based agreements, we aim to enable innovation in shared business processes and stimulate the market for RCM data in rail. Leveraging existing smart contract-based schemes for trading and sharing IoT data over blockchain networks, we identify suitable methods for the enforcement of agreements and ensure fair cost attribution between stakeholders, without a trusted third party. The outline of a blockchain-based RCM data audit framework is presented, appropriate data access agreements and accounting models are specified in detail, and three permissioned blockchain platforms (Hyperledger Fabric, Sawtooth, and Iroha) have been analysed for their suitability for implementation. Finally, the chapter outlines planned future work around validation of the tools based on two industrial use cases: monitoring systems for unattended overhead line equipment and axle bearings.

Keywords: *big data, blockchain, Remote Condition Monitoring, cost attribution, process automation*
JEL Classifications: L92, O31

1. Introduction

The pursuit of higher quality services in the railway sector is a continuous process, and the availability in recent years of affordable, reliable, digitally enabled additions to traditionally mechanical-based infrastructure systems has provided a fruitful avenue for advancement. Remote Condition Monitoring (RCM) systems are one example of a tool that has been widely deployed to improve the standards of maintenance, reliability, and safety across the rail network. The advanced warnings of incipient faults provided by RCM data enable preventative maintenance to be performed before service-impacting failures arise, leading to reduced costs of disruption and increased passenger satisfaction. The perceived benefits of RCM have led the industry to install sensors on an ever-higher proportion of its assets, with a corresponding increase in the volume of data generated. In general, and according to [1], railway RCM operations can be divided into four major divisions (quadrants), which are defined by the location of the monitoring sensors and the assets being monitored: train monitoring train, infrastructure monitoring infrastructure, train monitoring infrastructure, and infrastructure monitoring train. In countries such as the UK, where the vast majority of the mainline rail infrastructure is maintained by a single Infrastructure Manager (IM), sensors that are mounted on assets belonging to one stakeholder but are being used to monitor assets related to another will, by definition, fall into the train monitoring infrastructure or infrastructure monitoring train quadrants; an example of this would be sensors mounted on the tracks that are used to detect wheel flats on the rolling stock [2]. Although this type of cross-interface monitoring of assets may be the most technically practical solution to many industry-wide problems, commercially they can prove complex as the business paying to install, maintain, and operate the sensing device is not the party benefitting from the data collected. As a result, it can be hard to generate business cases for the purchase, installation, and operation of cross-interface monitoring systems that would have recognised industry-wide benefits.

In order to address this issue, it is widely recognised within GB rail that either closer collaborations must be established between stakeholders to enable more effective cross-interface business cases to be developed or there must be a trusted audit process that can enable costs of data collection to be fairly attributed based on business benefits accrued by individual stakeholders. To investigate these issues the Rail Safety and

Standards Board (RSSB) set up a Cross-Industry RCM (XIRCM) research programme, which in turn acted as sponsor to the T1010 research project [3] from 2013 onwards. The stated aim of T1010 was to overcome the barriers for rail companies to use RCM systems across company boundaries, with the first round of findings presented by RSSB and Network Rail at the IET RCM conference in 2014 [4].

A key component of business case generation for cross-interface RCM is the assignment of value to the data generated by one party but used by another. In order to address the cost issue, it was suggested in project T1010 that commercial agreements could be established between all the actors in a new condition monitoring workflow before installation of the system began [5]. However, there are issues with this approach; commercial agreements do not remove the need for a trusted third party (arbiter) to ensure compliance with the terms of the agreement, and they do not inherently include any ongoing audit mechanism that would act as evidence should issues arise. In combination, these two issues act as a barrier to the full exploitation of XIRCM data and cost sharing between stakeholders.

Distributed Ledger Technologies (DLTs) have several features which can be leveraged to address the issues outlined. The benefits offered to the industry through improved system-wide asset information and decision support are clear, but for those benefits to be realised in a privatised rail system where the separation of business functions is the main architectural driver, the commercial implications of the operation of cross-industry systems for each actor must be clear. Further to this, existing investments in specific RCM systems made by the industry are currently only in their mid-life stages, meaning a method to deliver a clear understanding of operational costs must be cognizant of, and compatible with, the methods of operation of these existing assets. DLTs are one possible solution to these issues, offering the potential for traceability of data flows between industry actors with a minimum restructuring of the current systems. By understanding the flows of data between actors, and the ultimate costs/benefits accrued by the installation and use of the system (for which mechanisms are already in place), it will be possible to accurately assign costs to the relevant parties, to cut down on the operational inefficiencies associated with manual attribution and trusted third parties, and to enable improved understanding of data provenance via the decentralised and immutable record in the ledger.

Blockchains are a specific type of DLT constructed from structured sequences of blocks connected via cryptographic hashes, providing a tamper-proof ledger that leads to a traceable and auditable log of all activities between stakeholders. In industrial environments, the implementation of this technology facilities greater integration of business processes and stakeholder data, with the blockchain delivering three major protocols: decentralisation, cryptography, and consensus [6]. Due to the censorship-resistant and tamper-proof digital networks of distributed trust created by this revolutionary technology, blockchain-driven technologies help to enhance transactions and make them more reliable and safer. Industrial deployments of the blockchain are still in the early stages of development, and further work is required to establish the full extent of the value the technology offers. However, substantial efforts have been made to investigate its applicability and future penetration in numerous industries, including the industrial sector, as the new technology continues to mature [7], [8]. The transformative potential of blockchain technology in industry settings has already been established in the literature [9], and in the rail industry specifically, blockchain-based applications for ticket sales, invoicing, and freight distribution, among others, have also been investigated [10].

In this chapter we present early findings from the European Union (EU)-funded B4CM project, a study commissioned to investigate the value that blockchain technology offers the rail industry as a ledger of RCM data transfers (section 2), along with a discussion of related work in the literature. The proposed blockchain framework will be presented in depth in section 3, with plans for future work and concluding comments detailed in sections 4 and 5, respectively.

2. Background and related work

Large volumes of data are generated daily by RCM systems installed on the GB rail network. While this data is already utilised to improve performance within the context for which the system was initially specified, in many cases, opportunities exist for the realisation of additional benefits by sharing this data between stakeholders and across system boundaries, enabling it to be used in problems that cross traditional industry interfaces (primarily the separation between the infrastructure and vehicles). The continuous improvement of system performance through RCM-informed operations and maintenance is a field of intensive research, and many projects focusing on this area have been initiated [11]. At present, the industry is still on an upward performance trend in this area, and localised sensor systems used in isolation are still providing operational benefits. However, moving forward, the industry is expecting these systems to coalesce into fewer, multiparty and sensor environments, essentially evolving the network's current RCM capability into an "Internet of Railway Things" (IoRT) [12] requiring new ways of managing, processing, and accounting for data. This amalgamation of the state-of-the-art IT, cloud computing, and big data, presented as an Internet of Things (IoT) paradigm, will ultimately lead to a viable "smart railway" fit for the next century [13].

Depending on the nature of the sensors deployed, the data produced by RCM systems takes many forms, including audio, video, pictorial, continuous analogue measurements, and digital signals. In order for the raw datastreams to have operational value, they must first be processed, cleaned, and aligned to the point where they can be reliably used as the basis for analytics. As shown in Figure 1[14], there are six recognised levels of data analysis in condition monitoring, ranging from raw data collection (at the lowest levels), through the generation of alarms in response to defined alert criteria, to a full diagnostic function that involves sending prognostic information to the operations and maintenance team to instruct them to repair a particular asset before it fails. The data used as the input to each level of the stack (or indeed the analytics process itself) may originate from multiple stakeholders, and as the level of data processing increases, the inherent value of data becomes higher as a result of the additional knowledge associated with it. According to [5], unless specific contractual provisions say otherwise, it is typical for the Intellectual Property Rights (IPR) to the data recorded by RCM systems to be held by the party that collected it, while the IPR for derived data (data the results from a processing chain and is considered "enhanced") belongs to the party who performed the processing.

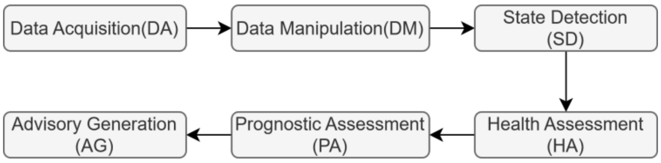


Figure 1: The six processing levels of ISO 13374. Source: [14].

As is the case in any trading environment, successful RCM deployments require that both the providers and the consumers of the data gathered comply with any contractual arrangements made around the system, and particularly when ensuring the quality and reliability of the data and advisory information produced. To this end, it is desirable for a traceable mechanism to exist within the system that monitors the provenance of the RCM data; this provenance information provides evidence that directly affects payment, compensation, or refund processing. In current RCM deployments, a Trustworthy Third Party (TTP) such as a bank, third escrow mediator, or conflict board may be a requirement to manage these needs.

DLTs, in the form of blockchains and smart contracts (SC), have the potential to offer great value to industry in this context enabling operators of RCM systems to dispense with the need for a TTP and inherently prevent the RCM data generated from being falsified, altered, or corrupted without the changes being evident. Further to this, in order to both quantitatively and qualitatively monitor and manage the flows of data between providers and consumers, SC may be deployed on the blockchain. Deployed SC are essentially distributed executable scripts running in the blockchain [15], and this combination of traceability (as provided by the chain itself) and transformation/transaction of data (as provided by the SC) provides an environment in which the whole value chain around items of data may be audited and understood. As pointed out by Christidis and Devetsikiotis [16], in a traditional relational database management system, an SC would essentially be used as a stored process, but by using an SC within the underlying execution framework offered by the blockchain, a wide range of applications can be created.

Within the literature, a range of examples of the use of blockchains in partial solutions to the problems seen in XIRCM may be found. Existing studies on the use of micropayments between stakeholders linked to IoT data exchange, for example, have suggested that SC-based frameworks would form an appropriate basis for that use case. In the Saranyu system [17], Nayak et al. created a cloud tenant and service management system using Quorum (a private blockchain network) as a platform but ultimately failed to capture appropriate information on charging tenants. A subscription-based model for trading data on cloud platforms was also introduced by Al-Zahrani [18]. In the proposed model, the ledger tracked all subscriptions and orders, and this included those on which the request has not been concluded and finalised, providing potentially useful information to forensic investigators should problems occur. A blockchain-based solution using Ethereum was launched in [19], which regulated both payments to and access by the owners of data-generating IoT devices. When subscribing to a particular IoT device and before accessing the data processed in the MQTT broker, which represented a single point of failure within the system, data owners paid a deposit in ether (the "currency" of the chain).

With the exception of [17], none of the work identified provided a mechanism for the suspension or revocation of malicious actors/account subscriptions, other than the removal of the associated data from the cloud platform used. Typically, the authors assumed that data providers acted honestly in all the systems surveyed, and did not address the issues raised by the presence of falsified or garbage data that may have been deliberately inserted into the platform to deceive customers. The payment companies BitPay [20], BitHalo [21], and DCSP [22] have considered the issue of dishonest actors, and all have previously proposed the use of double deposit escrow. In all three proposals, both the buyer and the supplier use SC to create an escrow for the deposited values, but the actual transfer of assets is made off-chain. Both parties must acknowledge the SC that the transaction is successfully made in order to unlock the escrow. Should confirmation not be given, both forfeit their deposits. A dual-deposit escrow mechanism identical to the previous three schemes was suggested by Asgaonkar and Krishnamachari [23] but offered a subsequent dispute resolution stage (potentially preventing deposit loss) and involving the main payment transaction. However, this system was only suitable for one-time usage scenarios, and the buyer was required to review every transaction and provide a reply to open the escrow and process the payment. The seller received no compensation if the customer did not respond (regardless of the presence or absence of malicious intent) and would forfeit their deposit and right to payment. A different data-sharing mechanism is proposed in [24], in which data hash values are encrypted with a symmetrical key and deposited in a secure location off-chain by the data provider before the transaction is actioned. In the cloud, all providers are able to promote their data services and public keys. To enable consumers to gain one-time access to the appropriate records, SC were generated on the fly and the activity was logged on the chain to be used in the resolution of any potential disputes.

In this chapter, the framework proposed will build on the escrow proposals discussed above but will additionally include litigation solutions that ensure escrow locking or payment/compensation loss do not take place.

There are several known limitations of blockchain technology; the blockchain trilemma [25], for example, states that the interrelated properties of scalability, decentralisation, and stability cannot be achieved simultaneously on the same chain, meaning that compromises must be made in terms of desired functionality based on the specific use case. Furthermore, all blockchain-based applications must make a trade-off between the size of any on-chain storage and operational performance; in practice this is manifested by a significant increase in processing time as the overall size of the ledger increases, a process that is naturally much more rapid if data being exchanged is recorded within a transaction alongside the record of the transaction itself. Scalability of storage has significant implications for the usage of blockchains in RCM contexts, and essentially enforces an architectural choice on the designer to use a hybrid approach that combines off-chain storage of data, with on-chain storage of provenance. Data integrity and immutability are ensured through the use of a checksum of the raw data, which when computed, stored, and verified within the blockchain record, can be used as evidence of data ownership, to automate integrity checks and to check latency claims.

3. Proposed framework

In this section, the authors present their proposed framework for the audit of RCM data in industrial systems. The framework replaces the TTP typically involved in these systems with a permissioned blockchain architecture, leaving data producers/data owners (providers), data users (consumers), and SC as the key actors in the system. Figure 2 illustrates this change; Figure 2 (a) shows a typical trust arrangement that would apply in a none DLT-based RCM network; in this case all parties must trust that the other producing/consuming parties will honour their obligations under the agreement defining the distribution of system costs; the TTP reviews local financial cost assessments provided by the other actors in order to confirm adherence to the applicable terms. This process will henceforth be referred to as “local cost monitoring.” As the local cost monitoring of both providers and consumers is dependent on the data they report, even with the TTP in place there is no guarantee of strict adherence to the terms of the contractual agreements between the parties.

As an example of the requirement for trust, consider the Quality of Service (QoS) criteria placed on a data provider. Honest providers could choose to comply with the terms of the signed agreement and offer the requested level of service that they initially advertised; this would result in an estimated cost calculation for the data as delivered and an associated attribution of the cost to the consumer. The consumer, on the other hand, will have their own interpretation of the quality of the service they have received; this may tally with that of the provider, or may be impacted by external factors such as network latency resulting in a different view of the fair attribution of the cost from the consumer’s side. To reinforce their point of view, both parties will provide evidence, but as there is no confidence between them, there will be no trust in the correctness of their evidence. The presence of the TTP goes some way to mediating these issues but still requires that the evidence as presented by the provider and consumer is fundamentally accurate, or that the TTP can identify when that evidence is incorrect and (ideally) who is in error. By comparison, the relationships and trust between actors required in the proposed framework are shown in Figure 2 (b). A trust relationship between the provider and the customer is no longer necessary, although both sides do need to trust the DLT and the SCs that implement the accounting logic, data access/delivery agreements, and cost allocations. Subsections 3.1 and 3.2 will explain these procedures in detail.

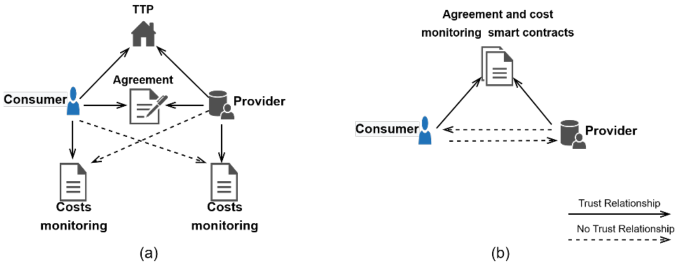
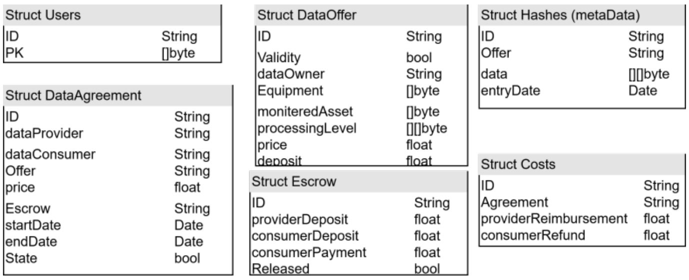


Figure 2: Trust relationship between actors.

3.1. Access agreement model

The commercial agreements originally outlined in project T1010 [5] have driven the definition of the components used in the SC for the access agreement and cost estimation process between provider and consumer as shown in Figure 3. Two new records, “DataAgreement” and “Escrow,” will be automatically generated by SC and be appended to the ledger each time a new data access request is made by a consumer to a producer. The DataAgreement will hold information on the new agreement between the data consumer and data provider, including the data offered by the provider, the unit price, and the period of validity. The Escrow record will form the basis for enforcement of access to the data and exchange of payment on release.



Struct Users	
ID	String
PK	[]byte

Struct DataAgreement	
ID	String
dataProvider	String
dataConsumer	String
Offer	String
price	float
Escrow	String
startDate	Date
endDate	Date
State	bool

Struct DataOffer	
ID	String
Validity	bool
dataOwner	String
Equipment	[]byte
moniteredAsset	[]byte
processingLevel	[][]byte
price	float
deposit	float

Struct Escrow	
ID	String
providerDeposit	float
consumerDeposit	float
consumerPayment	float
Released	bool

Struct Hashes (metaData)	
ID	String
Offer	String
data	[][]byte
entryDate	Date

Struct Costs	
ID	String
Agreement	String
providerReimbursement	float
consumerRefund	float

Figure 3: Data structure.

Recall that the IPR for the RCM data belongs to the provider, thus, no other party in the system will be able to advertise an offer for exactly the same data (although they may be able to advertise derivative forms) and this mechanism is protected by hash values. Both data providers and data consumers must be registered with the trustworthy authority (in this case the permissioned blockchain) in the set-up process of the system, and must have their IDs and public/private key pairs before participating.

The overall flow of the access agreement process is as follows:

1- The customer will submit a request to the SC in which they will specify the offer they are interested in, along with the subscription duration and all payments.

2- The authenticity of the submitted request will be tested by the SC. If it is not legitimate, so the request will be denied. A payment mechanism is triggered if the offer is still available; this process is addressed in depth in section 3.4.

3- After completing the payment process, the SC will automatically create a new agreement between the provider and consumer in addition to building an escrow to hold the payment. Both provider and consumer will be informed of the establishment of the agreement.

4- Prior to uploading the original data onto the external storage, the provider's private key and the consumer's public key will be used to sign and encrypt data respectively as follows: consumerPublicKey (providerPrivateKey (D)).

5- The consumer will decrypt the data they gain access to on the off-chain storage and compare its hash with the hash value provided in the on-chain record to validate its integrity.

In this proposed model, two types of malicious behaviour on the part of the data provider can be proven by the consumer:

a. Sending falsified or incomplete data;

b. Undue delay in uploading evidential hash values to the on-chain record.

If the QoS by either party is found to violate the terms of the agreement, both provider and consumer can revoke the agreement before the stated expiry date. This action is permanent, i.e., the agreement cannot be revived once revoked; instead, a new agreement must be entered into from the beginning. Figure 4 shows the sequence of creating the data access agreement.

3.2. Accounting model

Payments on any trading site may be realised using post-paid or pre-paid models. The post-paid model requires the provider to place trust in the consumer (buyer) that the payment will be made as agreed after the data is obtained correctly. The pre-paid model requires that the consumer places trust in the provider that the data will be delivered once the payment has been made as agreed. Neither model guarantees both consumer and provider satisfaction, and both bear some risk if the other party breaches the terms of the agreement. There is also a requirement for a TTP to provide both the provider and the consumer with an escrow service.

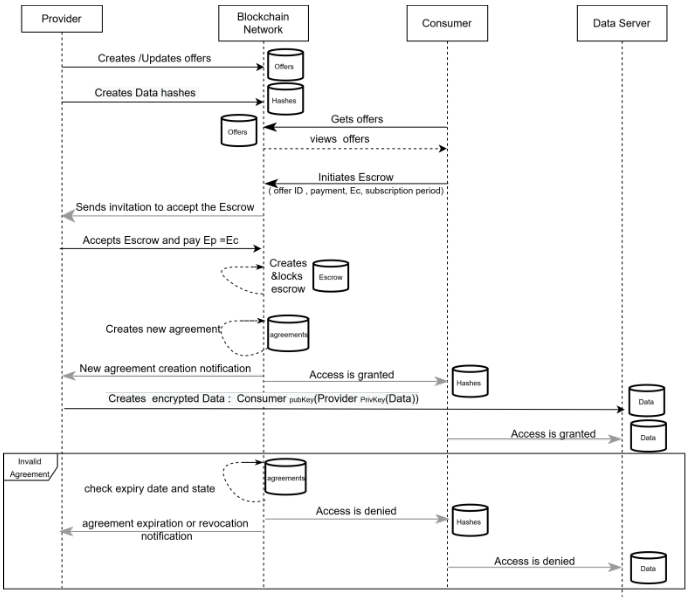


Figure 4: Data access agreement sequence model.

In the proposed framework, SC will be used to provide escrow, removing the need for a TTP and ensuring the payment is released to the provider after the data is delivered and the consumer agrees that it meets the stipulations of the agreement, assuming a revocation request is not made. The escrow SC is also responsible for managing any penalty payments required by the agreement, and these would be charged in advance of any data exchanging process by both provider and client.

The provider is expected to deploy the following attributes and values with the offer they are advertising as shown in Figure 3:

D_{price}: Denotes the data price of a certain offer in a specified period.

E: Denotes the deposit both consumer (E_{cns}) and provider (E_{prd}) should pay to build an escrow. The deposit will act as the penalty in case of any breach of the terms, and therefore must be set at a level that acts as a deterrent for both parties.

h(D): Denotes the hash value of the shared data.

The flow of the payment process is as follows:

1- An escrow SC will be initiated once the consumer responds to a published offer. The escrow details the offer being responded to and triggers payment of the corresponding charge and deposit by the consumer. On receipt, the SC will then direct the request to the provider.

2- On receiving the request, the provider will check if the payment and deposit detailed in the escrow are matched with their offer. Then, in order to lock up the escrow, the provider must pay their deposit, which may not be less than the deposit of the consumer. If the provider determines that the size of the payment or the deposit does not match with the terms of their offer, the provider can reject the request and the consumer will get back their payment.

3- The process of locking the escrow will trigger an SC to initiate an agreement, in which the period over which the consumer has access to the provider's data is specified.

4- The cost of data consumption will be monitored via the SC when the escrow is released. The escrow will be released automatically if either of the two states below is realised:

a. The agreement's expiry date is reached, or

b. The agreement is revoked.

In both cases, if there is a claim of inappropriate activity from either side, it should be evaluated before calculating the final cost attribution. The deposits that have been charged would then be used in settling any penalties due if maleficence has been proven on either side. Figure 5 summarises all the possible outcomes of an investigation into QoS breaches between a provider and a consumer. Costs are calculated based on each scenario, which are outlined in equations 1–4. The terminology below is used in the equations:

$\text{Cns}_{\text{Payment}}$: Denotes the payment that the consumer should pay when initiating the offer request. It represents the total of D_{price} and E_{cns}.

$\text{Act}_{\text{Payment}}$: Denotes the actual payment of the consumed data based on the period of use; this value should be less than or equal to $\text{Cns}_{\text{Payment}}$.

$\text{Prd}_{\text{Reimbursement}}$: Denotes the final cost that will be transferred to the provider based on the status of the agreement and the raised claims.

$\text{Cns}_{\text{Refund}}$: Denotes the refunds that will be transferred to the consumer based on the status of the agreement and the raised claims.

To calculate the $\text{Act}_{\text{Payment}}$ three different dates will be considered:

Rvc_{Date}: Denotes the revocation date.

$\text{Start}_{\text{Date}}$: Denotes the beginning of the agreement, as declared in the agreement.

Exp_{Date}: Denotes the end date of the agreement, as declared in the agreement.

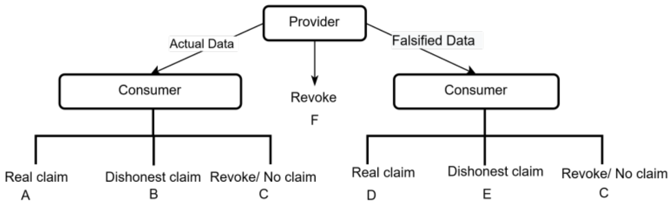


Figure 5: All possible scenarios in trading data.

Scenario A: The consumer receives the requested data as agreed but raise a genuine complaint about the latency in providing the hashes to the network. The cost SC will evaluate this claim by checking the dates of appended hash values on the chain, using the block's timestamp. As the consumer's claim is genuine, the agreement will then be revoked, triggering the calculation of costs as follows:

$$\text{Act}_{\text{Payment}} = D_p \times (\text{Rvc}_{\text{Date}} - \text{Start}_{\text{Date}})$$

$$\text{Prd}_{\text{Reimbursement}} = \text{Act}_{\text{Payment}} \qquad (1)$$

$$\text{Cns}_{\text{Refund}} = (\text{Cns}_{\text{Payment}} - \text{Act}_{\text{Payment}}) + E_{prd} + E_{cns}$$

Scenario B: The consumer falsely claims the data is corrupted or incomplete, or that the hash values are not appended to the chain in a timely fashion. In this case, the cost SC will evaluate both cases to validate the claim. The former is evaluated by requesting the received data which is signed using the provider's private key that verifies the data source, and then the SC will perform a hashing process to the data, enabling it to be compared with the hashed value that is stored on-chain. The latency in appending hash values will be validated as mentioned before in scenario A. In this scenario, the consumer's claim will be found to be false by the SC, and as a result the agreement will be revoked and the cost will be calculated as follows:

$$\text{Act}_{\text{Payment}} = D_p \times (\text{Rvc}_{\text{Date}} - \text{Start}_{\text{Date}})$$

$$\text{Prd}_{\text{Reimbursement}} = \text{Act}_{\text{Payment}} + E_{prd} + E_{cns} \qquad (2)$$

$$\text{Cns}_{\text{Refund}} = \text{Cns}_{\text{Payment}} - \text{Act}_{\text{Payment}}$$

Scenario C: The consumer revokes the agreement without raising any claim. In this case the agreement will be revoked and the cost will be calculated as follows:

$$\text{Act}_{\text{Payment}} = D_p \times (\text{Rvc}_{\text{Date}} - \text{Start}_{\text{Date}})$$

$$\text{Prd}_{\text{Reimbursement}} = \text{Act}_{\text{Payment}} + E_{prd} \qquad (3)$$

$$\text{Cns}_{\text{Refund}} = \text{Cns}_{\text{Payment}} - \text{Act}_{\text{Payment}} + E_{cns}$$

A similar process will be triggered when the agreement reaches the expiry date without any revocation or complaints from the consumer's side:

$$Act_{Payment} = D_p \times (Exp_{Date} - Start_{Date})$$

$$Prd_{Reimbursement} = Act_{Payment} + E_{prd} \qquad (4)$$

$$Cns_{Refund} = Cns_{Payment} - Act_{Payment} + E_{cns}$$

Scenario D: The provider sends falsified data to the consumer. In this case, the consumer raises a claim providing the received data to the SC, which compares it to the hash value stored on the chain. As a result of the provider's actions, the agreement will be revoked, triggering the calculation of costs according to equation (1).

Scenario E: The consumer raises a genuine claim against the provider, but attaches the wrong evidence leading the SC to evaluate the claim as false. Such a situation may occur if, for example, the provider uploaded the right hash values to the network at the right time, but sent the wrong data to the consumer on the external storage. When the consumer identifies the mismatch between the hash values, there is a risk of raising a latency claim rather than a claim resulting from the mismatched hash. Were the consumer to raise a latency claim in this situation then the SC would prove the claim false and process the cost according to equation 2. In this scenario, resolution and reimbursement of the consumer would be possible if the consumer provided the signed original data to a dispute board. The provider won't be able to show the hash value that matches with the provided signed data that has been uploaded to the network on the same date. This would of course require such a board to be in place and may reduce the overall financial benefit of the blockchain implementation.

Scenario F: The provider chooses to revoke the request as they can no longer provide the data as advertised or are unwilling to provide the data for another reason. In this case, costs will be calculated according to equation 1. Such a scenario could arise if the consumer was suspected of data reselling, which is against the terms of the agreement with a provider. Proof of data reselling would be achieved by comparing hash values uploaded to the chain as part of a data offer. Such a case would require the intervention of the dispute board and may lead to legal action.

4. Future work

In this chapter, a proposed architecture for the delivery of a data audit chain for RCM in GB rail and other industrial contexts has been presented. The next step is for the proposed architecture to be implemented and trialled with real-world data. As there are no one-size-fits-all platforms for blockchain projects, identifying the most suitable deployment platform is critical to the success of this work. A trade-off study was carried out that compared four of the most commonly adopted blockchain platforms: Ethereum [26], [27], Fabric [28], Sawtooth [29], and Iroha [30], based on the parameters set out in Table 1.

Table 1: Trade-off analysis between Ethereum, Fabric, Sawtooth, and Iroha.

Criteria	Ethereum	Fabric	Sawtooth	Iroha
Supports SC	✓	✓	✓	✓
Consensus algorithm modularity	✗	✓	✓	✗
Built-in components for managing identities	✗	✓	✗	✓
Supports payment in fiat currency	✗	✓	✓	✓
Proficient in maintaining different privacy levels between users	✗	✓	✓	✓

Of particular interest was the fact that for any SC execution, the Ethereum chain incurred costs (gas) in its native payment currency (Ether), while the Fabric, Sawtooth, and Iroha Hyperledger systems are cryptocurrency-independent, and payment was possible in fiat currencies. Further to this, because of the voting-based consensus algorithms adopted in Hyperledger platforms, there is no requirement for time- and power-consuming consensus algorithms. The associated performance characteristic ensures quick access to provider information, which would be a key criterion for most RCM use cases.

The proposed framework requires differentiation between users to ensure the privacy of their transactions, i.e., not all agreements and payment processes are open to all network users. Any consumer may opt to have a private contract with a provider, and to keep the costs of sharing the data secret from those not participating in that agreement. The Ethereum chain treats all users identically, and all transactions are open and available to all network participants. Hyperledger networks by comparison are able to fulfil this criterion by one of several mechanisms; Fabric, for example, establishes a different channel to isolate parties requiring private agreements and cost allocations; changing the identity namespace in the transaction family on the Sawtooth chain would limit access to specific identities; and specifying guidelines for access management in Iroha would retain easy role-based access at different stages.

In the future, we seek to trial our proposal against representative use cases from GB rail and to evaluate its performance in terms of promoting trust, simplifying cost attribution, delivering a workable payment mechanism for RCM data, and implementing ad-hoc data access agreements between parties. To this end, two representative case studies will be developed, one around the Unattended Overhead Line Equipment Monitoring System (UOMS) and a second around RailBAM, an acoustic axle bearing monitoring system. Both case studies involve systems that require collaboration across the rail sector between multiple stakeholders, and the data generated is of interest to multiple actors, perfectly illustrating the cross-interface scenario that is the target of the system.

5. Conclusion

RCM is a critical technology in the evolution of the smart railway, enabling improved reliability at a reduced cost. As sensors attached to fixed and mobile assets are increasingly used to inform the operational decision making of the industry, it is becoming critical that the business processes that distribute the costs and benefits of such systems across stakeholders within the industry are aligned in a way that is fair to all parties. The ability to trade in RCM data offers a net market advantage to the industry, as this enables easy access to data by any party that believes they have a use case, while also ensuring that data providers are adequately reimbursed.

Traditional approaches to the management of costs associated with cross-stakeholder RCM deployments in rail have relied on specific business-to-business commercial agreements and predefined costs. These lack the flexibility required to fully exploit the data generated in the "big data" age, where automated model development often requires access to a wide range of data resources from across an industry. Furthermore, the specific use cases being investigated are unlikely to have been foreseen at the time the RCM systems were procured, meaning the initial agreements would need to be modified to support new usage scenarios, an expensive and time-consuming process. Some legacy collaboration arrangements are not wholly defined or explicit and are thus open to misinterpretation or may not be enforceable.

The B4CM project aims to provide the rail industry with an alternative to the traditional model for the attribution of RCM costs. This chapter has introduced a new architecture based on blockchain technology which ensures the rights to data are allocated to the data provider as long as they supply the blockchain network with evidential hash values. The architecture simplifies the mechanism for coordination between a data provider and data users, while also allowing automation of the underlying business agreements and cost distribution. A service quality agreement between provider and consumer is established enabling both actors to prove some violating behaviours; for example, a consumer may claim low service quality, prove their claim, and be paid for; otherwise, for making dishonest claims, the consumer would be fined. Fundamentally, the proposed system allows all stakeholders to contribute, and realise revenue from, their data while enabling cross-industry use cases that are currently not easily realised.

The next stage of the work is to validate the framework by trialling it with real-world industry use cases, and the results of these will be reported to the community in the near future.

Competing interests
None declared.

Ethical approval
Not applicable.

Author's contribution
RA was the lead author of the paper, including delivery of all aspects of the work presented and preparation of the initial manuscript. SH contributed to the initial design of the framework and to the editing of the manuscript. JE is the lead academic on the project and contributed to the design and delivery of the work presented alongside the structuring, editing, and proofreading of the draft manuscript.

Funding
This project has received funding from the Shift2Rail Joint Undertaking (JU) under grant agreement No 826156. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Shift2Rail JU members other than the Union. While this paper is an output of work funded by the Shift2Rail Joint Undertaking (JU), the content is solely reflective of the author's views. The Shift2Rail JU is not responsible for the findings presented within this paper, or for any use that may be made of its contents.

Acknowledgements
The authors would further like to acknowledge Imam Abdulrahman Bin Faisal University and the Saudi Government for funding RA, the first author.

References

[1] C. P. Ward et al, "Condition monitoring opportunities using vehicle-based sensors," Proceedings of the Institution of Mechanical Engineers. Part F, Journal of Rail and Rapid Transit, vol. 225, (2), pp. 202–218, 2011.

[2] A. Alemi, F. Corman, and G. Lodewijks, "Condition monitoring approaches for the detection of railway wheel defects," Proceedings of the Institution of Mechanical Engineers. Part F, Journal of Rail and Rapid Transit, vol. 231, (8), pp. 961–981, 2017.

[3] Sparkrail, Cross-industry remote condition monitoring (T1010). [Online]. Available: http://www.sparkrail.org/ Lists/Records/DispForm.aspx? ID=8096.

[4] G. J. Tucker and A. Hall, "Breaking down the barriers to more cross-industry Remote Condition Monitoring (RCM)," in 6th IET Conference on Railway Condition Monitoring (RCM 2014), Birmingham, UK, September 2014, pp. 1–6.

[5] Sparkrail, Cross-industry remote condition monitoring, Commercial, Final report Appendix E Standard Form (Template) (T1010 Report Appendix).

[6] M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman, "Blockchain technology: Beyond bitcoin," Applied Innovation, vol. 2, pp. 6–10, 2016.

[7] M. Friedlmaier, A. Tumasjan, and I. M. Welpe, "Disrupting industries with blockchain: The industry, venture capital funding, and regional distribution of blockchain ventures," in Proceedings of the 51st Hawaii International Conference on System Sciences, 2016.

[8] M. Risius and K. Spohrer, "A blockchain research framework," Business & Information Systems Engineering, vol. 59, (6), pp. 385–409, 2017.

[9] B. Biswas and R. Gupta, "Analysis of barriers to implement blockchain in industry and service sectors," Computers & Industrial Engineering, vol. 136, pp. 225–241, 2019.

[10] P. Mcmahon, T. Zhang, and R. Dwight, "Requirements for big data adoption for railway asset management," IEEE Access, vol. 8, pp. 15543–15564, 2020.

[11] D. Galar, D. Seneviratne, and U. Kumar, "Big data in railway O&M: A dependability approach," in S. Kohli, A. V. Senthil Kumar, J. M. Easton, and C. Roberts (Eds.), Innovative applications of big data in the railway industry. IGI Global, Hershey, PA, pp. 1–26, 2017.

[12] J. M. Easton. "Blockchains: A distributed data ledger for the rail industry," in S. Kohli, A. V. Senthil Kumar, J. M. Easton, and C. Roberts (Eds.), Innovative applications of big data in the railway industry. IGI Global, Hershey, PA, pp. 27–39, 2017.

[13] Q. Y. Li et al, "Chapter 14 – Smart railway based on the Internet of Things," in H.-H. Hsu, C.-Y. Chang, and C.-H. Hsu (Eds.), Big data analytics for sensor-network collected intelligence. Elsevier Inc., pp. 280–297, 2017.

[14] ISO13374-2: "Condition monitoring and diagnostics of machines – Data processing, communication and presentation – Part 2: Data processing," 2007.

[15] M. Alharby, A. Aldweesh, and A. V. Moorsel, "Blockchain-based smart contracts: A systematic mapping study of academic research," in International Conference on Cloud Computing, Big Data and Blockchain (ICCBB), Fuzhou, China, 2018, pp. 1–6.

[16] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the Internet of Things," IEEE Access, vol. 4, pp. 2292–2303, 2016.

[17] S. Nayak, N. C. Narendra, A. Shukla, and J. Kempf, "Saranyu: Using smart contracts and blockchain for cloud tenant management," in 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), San Francisco, CA, 2018, pp. 857–861.

[18] F. A. Al-Zahrani, "Subscription-based data-sharing model using blockchain and data as a service," IEEE Access, vol. 8, pp. 115966–115981, 2020.

[19] A. Suliman, Z. Husain, M. Abououf, M. Alblooshi, and K. Salah, "Monetization of IoT data using smart contracts," IET Networks, vol. 8, pp. 32–37, January 2019.

[20] Bit-Bay, Double deposit escrow. [Online]. Available: https://bitbay.market/double-deposit-escrow

[21] D. Zimbeck, Two party double deposit trustless escrow in cryptographic networks and bitcoin. [Online], 2014. Available: https://bithalo.org/whitepaper_twosided.pdf

[22] G. Bigi, A. Bracciali, G. Meacci, and E. Tuosto, "Validation of decentralised smart contracts through game theory and formal methods," in C. Bodei, G. Ferrari, and C. Priami (Eds.), Programming languages with applications to biology and security. Springer, pp. 142–161, 2015.

[23] A. Asgaonkar and B. Krishnamachari, "Solving the buyer and seller's dilemma: A dual-deposit escrow smart contract for provably cheat-proof delivery and payment for a digital good without a trusted mediator," in 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Seoul, Korea (South), 2019, pp. 262–267.

[24] H. Desai, K. Liu, M. Kantarcioglu, and L. Kagal, "Adjudicating violations in data sharing agreements using smart contracts," in 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifax, NS, Canada, 2018, pp. 1553–1560.

[25] T. Ometoruwa, Solving the blockchain trilemma: Decentralization, security & scalability. [Online]. Available: https://www.coinbureau.com/analysis/solving-blockchain-trilemma

[26] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," Technical report. [Online]. Available: https://gavwood.com/paper.pdf

[27] V. Buterin, Ethereum white-paper. [Online]. Available: https://ethereum.org/en/whitepaper/

[28] E. Androulaki et al, "Hyperledger fabric: A distributed operating system for permissioned blockchains," in EuroSys'18: Proceedings of the Thirteenth EuroSys Conference, April 2018, pp. 1–15.

[29] K. Olson et al, Sawtooth: An introduction – White paper. [Online]. Available: https://www. hyperledger.org/wp-content/uploads/2018/01/Hyperledger_Sawtooth_WhitePaper.pdf

[30] Hyperledger Iroha Community, Iroha handbook: Installation, getting started, API, guides, and troubleshooting. [Online]. Available: https:// iroha.readthedocs.io/_/downloads/en/1.1.3/pdf/