Building a Trusted Environment

Event-based Attribute Assurance

White Paper

Written by Ben Helps, Factern Ltd

November 2019

FACTERN

OPEN IDENTITY EXCHANGE
White Paper

PROJECT PARTICIPANTS

The Open Identity Exchange (OIX) is a non-profit, technology agnostic, collaborative cross sector membership organisation with the purpose of accelerating the adoption of digital identity services based on open standards. OIX's broad membership and independent nature have seen it develop a significant body of digital identity research, and it is a significant influencer working towards the development of a digital identity market.

The following individuals and organisations participated in the OIX Discovery Project that formed the basis of this Whitepaper:

- DWP: James Randall, Callum Denyer
- Factern: Ben Helps
- GDS / Cabinet Office: Alastair Treharne, Livia Ralph
- HMRC: Nick Davies, Ross McDonald
- Idemia: David Rennie, Dawn Wray
- Inidsol: Romek Szczesniak, Eleanor McHugh
- Mydex CIC: David Alexander
- Santander: Mitesh Savjani, Antonio Gomez Sotelo
- tScheme: Richard Trevorah
- Zonafide: Paul Worrall, Salena Worrall

OIX and the authors would like to thank all the Discovery Project team members for their considerable contribution to the work.

EXECUTIVE SUMMARY

In this OIX Whitepaper, we argue that giving a Claimant the ability to provide Events as evidence to support their claim to hold certain Attributes has the potential to transform the economics of data assurance, combat fraud and improve access to services.

The basic concept of an Event is not a new one: it is in many ways analogous to a digital 'witness statement' or 'signed receipt'. For the purposes of this Whitepaper, we define an Event as a trustable record of universal content that can be linked to other Events and is made available as a Digital Resource for re-use by one Entity (the Event Provider) for re-use by either the Claimant or other authorised Entities (the Event Consumers).

However, Events have yet to be fully harnessed for the purpose of data assurance. Although many Entities record audit trails of activity for compliance purposes, very few – if any – set out to capture and manage such records as a valuable, shareable resource (i.e. as Events). Similarly, very few Entities use reasoning models that consume Events as part of the due diligence that delivers the data assurance outcomes they require.

In this paper, we seek to articulate what it would take for an Entity to set itself up as an Event Provider and/or Event Consumer, what the motivations for doing so might be and how it would benefit not only the Claimant but also the Ecosystem as a whole.

We highlight the functionality that would be required both to streamline the interactions and to establish an environment of trust between the Claimant, Event Provider and Event Consumer. We investigate a sub-set of potential options for the implementation of such exchange infrastructure, and point to a minimum set of standards that would enable interoperability between different implementation solutions. An initial set of technical artefacts were developed to test and visualise the use of such standards.

Finally, we set out a modest proposal for next steps, anchored on an OIX Alpha Project to develop a common set of requirements to generate, exchange and consume Events for the purposes of data assurance within an Ecosystem of otherwise independent actors, tested and refined through the development of a set of prototype capabilities.

CONTENTS

PAGE

1 PROJECT PARTICIPANTS

2 EXECUTIVE SUMMARY

3 CONTENTS

4 INTRODUCTION

Main body

8 CHAPTER 1: WHAT IS THE 'TRUSTED ENVIRONMENT'?

11 CHAPTER 2: WHAT DO WE MEAN BY AN 'EVENT'?

16 CHAPTER 3: HOW DO EVENTS BENEFIT DATA ASSURANCE?

20 CHAPTER 4: WHAT MOTIVATES THE EXCHANGE OF EVENTS?

25 CHAPTER 5: WHAT INFRASTRUCTURE IS REQUIRED?

30 CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

Appendices: Supporting material

35 APPENDIX A: GLOSSARY OF TERMS

38 APPENDIX B: WHAT DO WE MEAN BY 'DATA ASSURANCE'?

44 APPENDIX C: DISCOVERY PROJECT ARTEFACTS

INTRODUCTION

Almost every organisational Entity (such as a business, government department or academic institution) stores data that relates the customers, citizens, students, beneficiaries, suppliers, employees, etc. that it deals with. And many such Entities – particularly those that are regulated in some way – have to be confident that the data they hold is, in fact, true.

Entities need to be confident in the data they hold because it is fundental to their ability to exercise control over delivery of the product, service, benefit or entitlement (i.e. good) that they responsible for. Entities assure data quality so they can use the resulting data to determine eligibility (e.g. access to a state benefit), assess risk (e.g. when making a loan), comply with policy or legal requirements (e.g. seeking parental approval for child use of an app or web-site), etc.

A Verified Attribute is a piece of data associated with a given Subject that has been investigated via a due diligence process which itself is sufficiently robust for the data to be considered 'true' by the Entity that is consuming or using it. That investigation may have been undertaken either by the Entity itself or by a third party (sometimes referred to as the Attribute Provider) on which the Entity relies (i.e. acting as a Relying Party).

If an Entity consumes a Verified Attribute from a third party Attribute Provider, their trust in that piece of data is entirely a function of their trust in the Attribute Provider themselves. The Entity can be confident that the data is true precisely *because* it came from a "trusted" or "authoritative[1]" source. The Entity may have a commercial agreement in place that enables it to hold the Attribute Provider accountable for the quality of its due diligence, but it does not need to know the details of the investigations undertaken in each instance.

If an Entity that undertakes its own investigations were ever challenged to explain *why* it trusts the data that it holds, it might point to audit records which show that exactly what form of due diligence has been undertaken to assure the data. The audit records that

[1] Note that an "authoritative source" should not be confused with the concept of the Author of a piece of data. "Authoritative sources" are often commercial intermediaries that aggregate and process data.

capture who/when/how such an investigation took place are elemental units of the Entity's corporate memory and form the basis of the Entity's confidence in that data.

In contrast, the Subject of that data (i.e. the customer, citizen, student, beneficiary, supplier, employee, etc.) does is not generally able to access or exercise control over either the Verified Attributes or audit records that underpin confidence in the data associated with them. As a result, Subjects start afresh as Claimants every time they deal with a different Entity that is seeking to assure that data.

The granular foundations on which trust is based are either not explicitly captured or not made accessible for re-use. Even though Subjects/Claimants know that their data has been through the same basic investigatory process steps elsewhere (probably more than once), they have no corporate memory to draw on, and cannot use the due diligence efforts of one Entity as an input into the due diligence efforts of another.

But data assurance systems that focus solely on Verified Attributes (whether the due diligence was performed in-house or by a third party) are limited by the nuances of what might be understood by "verified": a singular definition simply cannot satisfy all potential Use Cases.

Even when it can (e.g. in reduced contexts, such as industry-based schemes), the resulting systems tend to place an enormous onus on the small number of Attribute Providers who act as Trust Anchors for the system as a whole. Such systems risk ossification, as bad actors move faster to undermine the (implicit) basis of trust than the system can reconstitute them. Furthermore, such systems are by definition siloes: they do not provide Subjects/Claimants with a mechanism to export trust outside of the specific context for which the system was designed.

In this OIX Whitepaper, we argue that the humble audit record has far greater value than it has been accorded to date: it not only has value for the Entity that records it, but also for the Subject to whom it relates and for other Entities who view the Subject as a Claimant and are looking for evidence that can be used to support their claim and assure their data. It is a granular unit of trust that can be made explicit and accessible, re-used in multiple different

contexts, and combined or aggregated to satisfy the different interpretations and requirements for a piece of data to be "trusted", "verified" or "assured".

Data assurance systems do not have to draw on Events to underpin every instance of trust – there are plenty of circumstances in which Verified Attributes might be both available to and sufficient for the Use Case in question. However, data assurance systems that have the *ability* to call on Events to supplement, explain or anchor the basis of trust will be better placed to adapt to changing circumstances and deliver collective benefits on a much wider scale.

In Chapter 1, we seek to establish a baseline for some of the terminology used throughout the Whitepaper, by describing the "trusted environment" of the title in terms of an Ecosystem of Participants who play the different roles required to provide, exchange and consume Events as a form of evidence to support data assurance.

In Chapter 2, we describe the concept of an Event in more detail, highlighting its key features. These include an Event's essential form as a witness statement ("At time T, X says that Y is true") which makes it inherently trust*able*; an Event's focus on granular activities and relationships that – when linked together – collectively provide a solid basis for data assurance; an Event's availability as a machine readable Digital Resource that does not depend on human interpretation; and the governance model associated with an Event that provides a basis for value exchange between Event Providers and Event Consumers.

In Chapter 3, we argue that Events can and should be made available to Subjects to be used as a form of evidence to support their claims (i.e. assure their data):

- Events can be used as inputs into models that use different forms of reasoning to validate and verify claims, providing a powerful and agile defence against fraud
- Relative to alternatives, Events are cheap to capture, store, provide, exchange and consume, and therefore directly address the costs of duplicated effort
- Events capture the individual activities which collectively form aggregated and competing quality standards, enabling overlaps to be identified and re-used
- An Event-based approach federates trust to any Entity with which a Subject may have interacted, thereby diversifying away from a reliance on a few large scale Trust Anchors

- Events offer a mechanism for 'thin file' Claimants to provide third party evidence to support their claims even in the absence of traditional forms of proof

In particular, Events which record the interaction of organisational Entities with their customers, citizens, suppliers, employees, beneficiaries, etc. in the course of *day-to-day* activity – i.e. Events not normally be associated with due diligence – have the potential to transform data assurance, minimising the need for dedicated investigative processes and strengthening risk-adjusted assurance outcomes.

In Chapter 4, we explore the motivation to participate in an Ecosystem by playing the role of an Event Provider or Event Consumer, and how those motivations differ from the more commonly recognised roles of Identity Provider, Attribute Provider and Relying Party. We also take the perspective of the Subject, demonstrating how a focus on Events (as opposed to the data those Events describe) increases their motivation to participate as well.

However, the production, exchange and consumption of Events requires a set of capabilities that already exist, but – generally speaking – have not been used for this purpose. Although many Entities record a subset of their activities to support operational efficiency, audit or compliance, few have implemented the infrastructure required to share such information with either the Subject involved or other authorised third parties. Chapter 5 therefore sets out what would be required to implement an Event-based data assurance ecosystem.

Finally, in Chapter 6, we summarise the key conclusions of the OIX Discovery Project that formed the basis of this Whitepaper, and make our recommendations for next steps.

Appendix A contains a glossary of the key terms that are used in this Whitepaper, and Appendix B sets out in more detail what is meant by "data assurance", to ensure common understanding. Appendix C contains a description of – and links to – the artefacts that were developed during the course of the Discovery Project to help test and visualise the concepts now summarised in this Whitepaper[2].

[2] The project team would like to extend their thanks to Paul Worrall of Zonafide for investing his time and skills into the development of these artefacts on their behalf.

CHAPTER 1: WHAT IS THE 'TRUSTED ENVIRONMENT'?

In a digital environment, what is the basis on which different Entities establish the trust they need to interact with one another? The overall objective of this OIX Whitepaper is to contribute to the wide ranging public debate on this topic, with a focus on data assurance.

1.1 DATA ASSURANCE

The rich tapestry of products, services, benefits, opportunities, etc. that are increasingly available through digital channels are only very rarely accessible to everyone. The Entities that control these goods (referred to here as Controlling Entities) usually restrict access to them, using some form of qualification criteria to screen potential recipients of the goods.

Where the goods are of higher import, value or risk, the Controlling Entities seek to enforce the qualifying criteria by investigating whether the entities that claim to meet them are telling the truth or not. This involves both a presentation of evidence to support the claim and an investigation of the evidence to ensure it is valid, genuine and relevant.

For the purposes of this paper, "data assurance" refers to both the due diligence process of investigating a claim and the assignment of a status to that claim, regardless of whether that status is implicitly or explicitly assigned[3]. For example, a successful claim (now held as pieces of data by on the systems of the controlling entity) may simply been assumed to be "true".

1.2 ROLES

From the perspective of the Controlling Entity, the Claimant is the entity that wants to (continue to) access the goods in question, claims to meet the qualifying criteria and presents evidence sourced from another Entity to support that claim. However, the same Entity that plays the role of the Claimant is also (claiming to be) the Subject of the evidence that they use to support their claim.

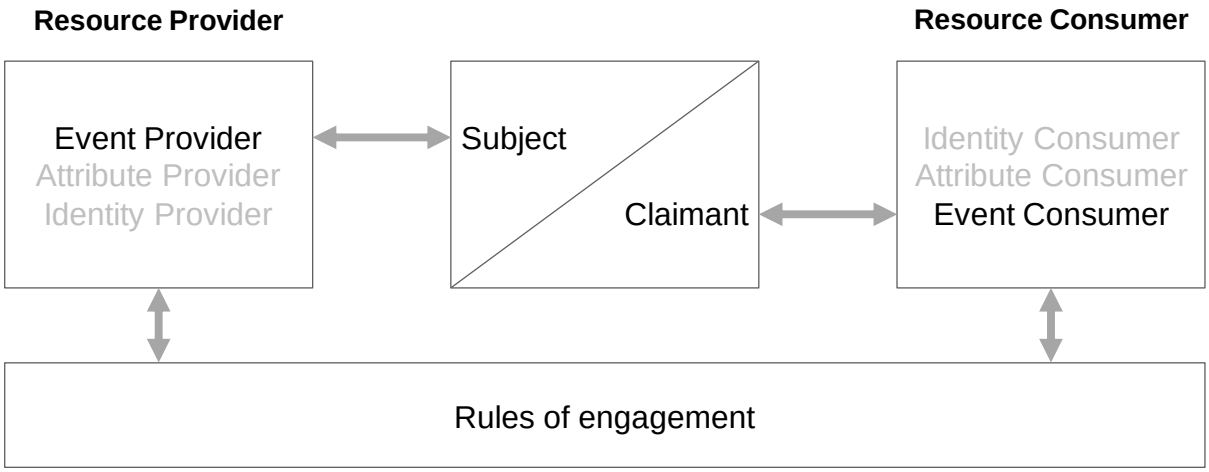
[3] For more detail, see Appendix B: What do we mean by 'data assurance'?

Controlling Entities "consume" the resources that they need to assure the Claimant's claim to the standard that they require. These resources include the claim itself, the evidence presented by the Claimant to support that claim, the operational activity to investigate it and the reasoning (e.g. a set of policy rules) used to assign a status or reach an outcome. In the context of data assurance, therefore, Controlling Entities act as Resource Consumers.

The entity from which the evidence is sourced is acting as a Resource Provider[4] to the Resource Consumer. The Resource Provider may or may not have a relationship with the Subject (e.g. as a customer, citizen, employee, etc.). It is possible for the same Entity to play the role of both Resource Provider and Resource Consumer, even in the same instance of data assurance. Equally, when the roles of Resource Provider and Resource Consumer are played by separate entities, those entities do not need to have a formal relationship with each other. When there is a formal relationship, "rules of engagement" are required (e.g. in the form of a contractual framework).

The evidence provided can take many different forms, but this Whitepaper focuses on the presentation of Events as a potentially new and powerful form of evidence. The paper therefore refers to the role of Event Provider and Event Consumer, to differentiate the capabilities to provide/consume Events as opposed to other forms of evidence.

Figure 1: Relative roles involved in data assurance

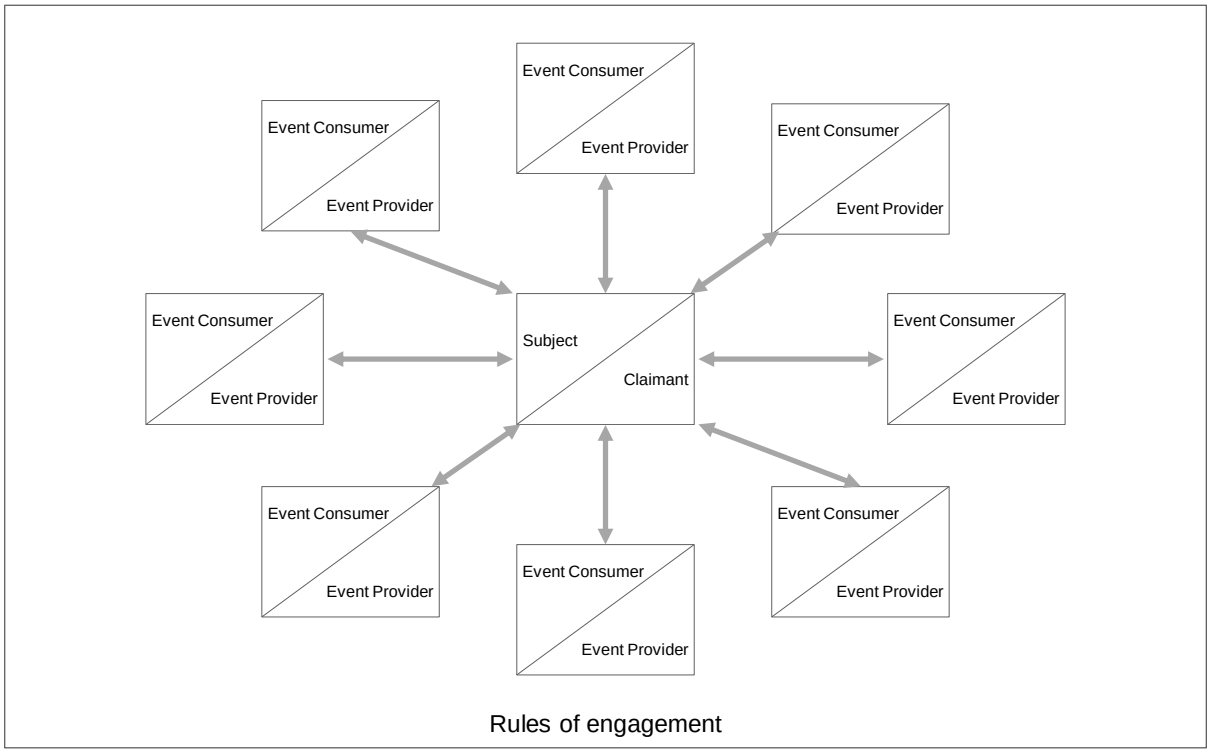


[4] For simplicity, this Whitepaper assumes that the Entity providing the resource is also its Issuer or Author.

1.3 ECOSYSTEM

The “trusted environment” from the title of this Whitepaper refers to an ecosystem in which Participants collaborate to provide, exchange and consume events for the purposes of data assurance, thereby establishing a basis for trust within a digital environment.

Figure 2: Collaborating to provide, exchange and consume Events



The central hypothesis of this Whitepaper is that the fundamental nature of an Event – as set out in Chapter 2 – means that it is easy for many different Entities to become Event Providers to a given Subject, particularly when they have some form of relationship with them. Equally, there are several good motivations for Entitis to act as an Event Provider – see Chapter 4 – making it worthwhile for them to do so, even as a by-product of their primary business activity.

The potential range and diversity of Event Providers is therefore much greater than that of Identity Providers and Attribute Providers whose business model is predicated on the provision of Verified Attributes (or Identities). This gives the Subject the ability to present Events sourced from many different Event Providers as evidence, making it easier for them to satisfy the Event Consumer’s need to assure their claim to the standard it requires.

CHAPTER 2: WHAT DO WE MEAN BY AN 'EVENT'?

An Event is a trustable record of universal content that can be linked to other Events (an effect that we call chaining) and is made available as a Digital Resource for re-use by authorised third parties. Events can be used as evidence to help assure that a particular piece of data associated with a given Subject (a relationship referred to as an Attribute) is in fact true at a given point in time.

For example, an individual (the Subject) might claim to live at a particular residential address. A delivery company might generate a digital record (i.e. Event) of the fact that they made a successful delivery to the individual at that address. This Event is useful to the delivery company (e.g. as an audit trail of its own business activity) but – potentially – also to other Entities that are looking to draw on independent evidence that to support the individual's claim to live at that address.

To ensure a common understanding of what we mean by an Event, the following sections break down our definition of an Event into its constituent parts.

2.1 A TRUSTABLE RECORD

To be a *trustable* record, every Event must be expressed as – in effect – a signed and dated witness statement. An Event is therefore an assertion whose essential form is "At time T, X says that Y is true", where X is a uniquely identifiable Entity (the witness) that is demonstrably the speaker of the assertion. The speaker, timing and content of the assertion are integral parts of an Event – i.e. an assertion without a speaker or a date is not an Event.

An Event is a *trustable* record because the Event Consumer knows just enough about the provenance of the assertion to provide a basis on which to make an assessment of the risk of the assertion being wrong. The fact that an assertion has been made is therefore itself a significant "event" (in the common understanding of the term).

An Event cannot be *fully* trusted as it doesn't guarantee that "Y is true". An Event only guarantees that "At time T, X *says* that Y is true", as X's signature makes it an undeniable fact that X made the assertion when they did. It may turn out that X is in fact wrong or lying.

The Entity consuming the Event determines how much to trust the assertion, based on how much they trust X (e.g. how often does X make assertions that turn out to be wrong?), when the assertion was made (e.g. could things have changed since then?) and on how much other information they have (e.g. have *other* Entities also asserted that "Y is true"?).

For example, our initial Event might record that "X says that the claimant's legal name is Bob Smith". This Event alone might be sufficient for an Event Consumer to believe that the content of the assertion is true if X (the Event Provider) is the Author or official register of that content (e.g. "The Registry Office says that the claimant's legal name is Bob Smith") or a Trust Anchor in the eyes of the Event Consumer (e.g. "The Scheme Identity Provider says that the claimant's legal name is Bob Smith").

2.2 UNIVERSAL CONTENT

In circumstances where the Event Provider is neither the Author of the Attribute nor a Trust Anchor to the Event Consumer, the Event Consumer may not take X's assertion alone as a sufficient basis on which to place trust. The Event Consumer might therefore solicit further information from X that help to answer the question "How does X know that Y is true?"

The Event Provider X might respond with information that is also constituted as an Event:

- "X says that Z checked the claimant's passport" (the assertion of an activity)
- "Z says that the passport contains the name Bob Smith" (the assertion of an Attribute)
- "Z says that the claimant is the subject of the passport" (the assertion of a relationship)

A second distinguishing feature of an Event, therefore, is that it is a universal concept: its content may refer to any Attribute, a relationship or an activity, and such Attributes, relationships and activities can be about anything.

As a result, Events can be used not only to describe the aggregate outcomes of large scale processes, but also to describe the individual elements of which those processes comprise. Events can reference any point in a hierarchy or communicate information from anywhere within an information graph.

2.3 SUPPORTS CHAINING

A third distinguishing feature of Events is that they can be linked to other Events, to create what we call *chaining effects*. Most systems of data assurance already depend on chaining effects. However, they are very often based only on major links in the chain, and neglect – or fail to make explicit – the minor links and sub-chains that are either assumed implicitly or "taken on trust". This heuristic places a significant burden of trust on the Trust Anchor (hence its name), and consequently only a few institutions are equipped to play that role.

Events can be used to make detail chaining effects explicit and transparent to the Event Consumer in a machine readable format. These chains of activity provide Event Consumers with a basis on which to place their trust. From our previous example, an Event Consumer might want to know how Z knows that "the claimant is the subject of the passport":

- "Z says that the passport contains a photo of its subject"
- "Z says that the claimant was present in person"
- "Z says that the claimant's face matched the photo in the passport"

The Event Consumer can use a Reasoning Engine to assure the status of the original claim ("the claimant's legal name is Bob Smith"), depending on the detailed chain of activities[5] that have been undertaken, taking into account the Event Consumer's confidence in Z's ability to undertake those activities and X's to stand witness to activities having occurred.

There are many circumstances in which the Event Provider is neither the Author of the Attribute nor a Trust Anchor to the Event Consumer, and where chaining becomes a valuable tool with which to establish trust. Entities can use Events to create a granular 'corporate memory' which remembers *why* they trust the data that they hold.

[5] At their purest, Events can be seen as stateless descriptions of the functions that transform data from one state to another state.

2.4 AVAILABLE AS A DIGITAL RESOURCE

A third feature of an Event is that it is made available as a Digital Resource that is machine readable and does not depend on human interpretation.

Many Entities issue evidence, but they often do so in non-digital form: for example, an official document (such as a passport), a certificate conferring status (such as a degree certificate) or a receipt of transaction (such as a purchase receipt). All of these contain information of the form "X says that Y is true", but none of them are directly consumable as a Digital Resource.

Similarly, Entities that rely on such evidence may record the chaining effects that collectively form their due diligence investigation, but often do so in ways which are not directly machine-readable, keeping – for example – paper-based records or unstructured digital notes of what has happened.

An Event must be constituted as a Digital Resource that can be consumed directly – and at scale – by automated Reasoning Engines. The richer the set of Events that are made available, the better an Event Consumer can assess the risk inherent in the content of any individual Event being false. Event Consumers can then reason over a set of Events to derive more (or less) confidence in the information that they contain.

2.5 RE-USE BY AUTHORISED THIRD PARTIES

The final defining feature of an Event is that its structure (as a form of witness statement) potentially confers a different set of "rights" than might normally be associated with the underlying data. In other words, Events should be treated as primary data objects that are subject to separate governance, while still adhering to laws that protect personal data.

Consider the Events:

- "The claimant says that the claimant's name is Bob Smith"
- "The witness says that the claimant's name is Bob Smith"

Both Events contain the original Attribute (and are clearly there contain personal data), and the claimant has some rights over who should access them.

But nonetheless, the two Events themselves are obviously distinct. Although the claimant has rights over both statements, the witness clearly also has rights over the second Event: after all, this Event is fully comprised of their own assertion, and they have explicitly gone to the trouble and cost of recording, signing, and making that assertion available for re-use.

Consider also the following Events:

- Event #1: "The claimant says that the claimant's name is Bob Smith"
- Event #2: "The witness says that Event #1 is a true statement"

It is clear that – in isolation – Event #1 and #2 have very distinct content. Event #2 does not in and of itself contain any personal data, and is only understandable if the recipient of Event #2 (i.e. Event Consumer) also has access to Event #1.

This Whitepaper does seek to determine whether or not Event #2 constitutes personal data under the EU's GDRP or the laws of other jurisdictions. However, we postulate that there are grounds for investigating whether Event #1 and Event #2 should receive differential treatment under data privacy law.

If so, a focus on Events rather than the data that Events assure could also bring a different perspective to the issues of privacy and liability that have historically constrained data exchange. For example, it may be easier to share Events which *refer* to instances of personal data (but do not contain that data) without compromising the personal privacy rights of the Subject, offering a alternative heuristic for the value exchange between Event Provider and Event Consumer.

2.6 EXPRESSING EVENTS IN PRACTICE

Many readers may already have noted that the Semantic Web provides standards to deliver most of these key features. The Discovery Project used an open-source tool to develop a test ontology of Events (expressed in RDF) which could be converted and exchanged as a Verifiable Credential. Details of the key artefacts can be found in Appendix C.

CHAPTER 3: HOW DO EVENTS BENEFIT DATA ASSURANCE?

Events constitute a valuable new kind of Digital Resource that has yet to be fully harnessed[6] for the purposes of data assurance. In this chapter, we set out the arguments for making Events available to their Subjects, so that Events can be offered as evidence to support claims being made by the Subject/Claimant and used as key input to data assurance.

3.1 DEFENCE AGAINST FRAUD

Today, most systems of data assurance conflate both the process that undertakes the investigation of a claim and the reasoning model that is used to assign an assurance status. Data is considered to be "true" because it has successfully passed through the prescribed process, which itself has been designed to test whether the data satisfies a predetermined set of policy conditions.

However, this approach creates an enormous dependency on what are often industrial-scale processes that are themselves difficult and slow to evolve. As a result, bad actors are able to exploit either weaknesses in the process prescribed by each Entity or – more often – gaps between the separate processes of different Entities.

The use of Events very explicitly *de-couples* the process that undertakes an activity from the Reasoning Engine that assesses the risk of a piece of data being wrong.

This de-coupling allows Event Consumers to take a much wider set of activity into account when assuring data. It does not matter if the activity is spread across multiple *different processes*, or if those processes are operated by *different Entities* or have been executed at *different points in time*. All activity – if captured as Events – can be taken into account.

A Subject with the ability to disclose Events from different Event Providers can offer a combination of evidence to support their claim that is very hard for a bad actor to replicate or fake. This in turn allows Event Consumers to deploy different kinds of reasoning to

[6] There are a few examples – such as in Healthcare – where Events are already being captured, stored and accessed to create value for different Participants in the Ecosystem

investigate those claims. Rather than a static rules-based logic, Event Consumers can build reasoning models (based on correlation or consensus, for example) that not only exploit the power of network effects but also improve over time as they learn from subsequent Events, providing a powerful and agile defence against ever more sophisticated frauds.

3.2 DE-DUPLICATION OF EFFORT

How much effort should Entities put into their data assurance?

This decision is usually made based on a trade off between the cost of undertaking the investigative process, the risk of making a mistake when assigning assurance status (i.e. believing something to be true that is in fact false), and the consequences of making such a mistake.

Events represent a very cheap and easy way to re-produce, exchange and consume the value of processes that have already been executed, and which can be re-used or re-purposed as evidence of activity that can be used to assure data quality.

As Event Consumers, Entities are able to make a different set of trade offs: for example, they may choose to spend $10 undertaking an investigative process step themselves (e.g. an in person passport check) or $3 to access three instances of the same Event (i.e. an in person passport check) undertaken by three different Event Providers.

For the Event Consumer, the cumulative effect of the three externally sourced Events may produce the same level of assurance as a single internally sourced Event, but for less than a third of the cost. At the same time, the Event Providers have each recovered 10% of the cost of undertaking their instance of the investigation.

3.3 INTEROPERABILITY OF STANDARDS

In some industries, a regulator may articulate minimum *standards* for the data assurance required for a particular Use Case. To meet these standards, Entities might adopt a set of *guidelines* agreed on by the industry (even if they are free to choose how they *implement* their own data assurance in line those guidelines). A *scheme* may provide the “rules of

engagement" that enable Entities to rely on data assurance undertaken by another Entity, on the basis that it conforms to the standards, having been implemented to the guidelines.

But in this paradigm, Entities operating in different industries and dealing with different Use Cases cannot rely on each other: the standards and schemes are not themselves interoperable because they focus on data assurance *outcomes*.

In contrast, Events foster a common understanding of the individual assertions and activities of which the guidelines and implementations comprise. Standards or schemes can therefore be described as the set of Events that collectively need to be fulfilled to deliver a given data assurance outcome. This approach allows overlaps in the Event sets required by competing standards to be easily identified, increasing interoperability between them.

3.4 FEDERATION OF TRUST

The cryptographic signature that Entity X associates with an assertion "At time T, X says that Y is true" is the basis of the trust framework that underpins Event-based data assurance.

This means that any Entity can become a valid Event Provider. Furthermore, each Entity will be the *Author* (and therefore authoritative source) of certain Events that are unique to the interaction between the Entity and a given Subject.

For example, when the delivery company makes a successful delivery to a Subject who is identified by a given name at a given address, it is – uniquely – the Author of that activity. Similarly, an individual that uses a social networking site to assert their friendship with the same Subject is the Author of that relationship. Both of these can be packaged as Events, with the delivery company and friend each acting as Event Providers.

Neither the delivery company nor the friend would be considered as a Trust Anchor in the traditional sense. However, an Event Consumer that reasons over their Events in the context of a set of Events from other Event Providers (e.g. other delivery companies, other friends), may discover patterns and network effects that provide compelling evidence.

An Event-based approach to data assurance therefore federates trust to any Entity which acts as an Event Provider, thereby diversifying away from a reliance on a small number of Trust Anchors that have historically dominated the market for data assurance.

3.5 IMPROVED ACCESS TO GOODS

At the same time, Events offer a mechanism for 'thin file' Subjects to provide third party evidence to support their claims in the absence of traditional forms of evidence.

A Subject who can call on a wide range of Events which attest to their own interactions with different Entities over a period of time is clearly equipped with evidence that uniquely identifies them as an individual and can be used to assure some of their primary Attributes (such as their name, residential address, age, occupation, etc.) to a high level of assurance.

The 'formal' documentation (like a passport) which is relied upon to assure data quality may not be available to all Subjects, and this reduces the ability of the latter access to goods. Events provide an alternative source of evidence that all Subjects should be able to access, as Events can be drawn from Event Providers of all kinds.

CHAPTER 4: WHAT MOTIVATES THE EXCHANGE OF EVENTS?

In this chapter, we explore the motivation to participate in an Ecosystem of Entities that are prepared to collaborate with each other, whether playing the role of an Subject, Event Provider or Event Consumer.

We start by arguing that Events which record interactions that occur in the course of day-to-day activity – i.e. Events not normally be associated with due diligence – have perhaps the biggest potential to transform the way that Entities approach data assurance, illustrating a major part of the rationale for the exchange of Events.

4.1 VALUE OF DAY-TO-DAY 'EVENTS'

'Events' (in the common understanding of the word) occur in their trillions every day. As we go about our day-to-day lives – as citizens, consumers, employees, office holders, etc. – we interact with each other both in person and online, undertaking innumerably different types of action, each of which could be represented as an Event.

For the most part, day-to-day activity goes unrecorded. Only a tiny fraction of it is actually 'written down' as having taken place, and only a subset of those records is made available to the Subject involved in the interaction, as a form of 'receipt'[7]. The Entities that do share a 'receipt' of activity with the Subject are motivated to do so for good reason: for example, access to a shared record is valuable in the event of a dispute between the Participants.

And that record can also add value in a completely different context: confirmation of having attended an appointment could be used to help prove that you are a real person (useful in a digital context); a track record of having recently taken deliveries at a given residential address is a useful way of supporting a claim to live there (particularly if there were

[7] A courier might record the fact that you took a delivery; a shop might record the fact that you made a purchase; a registrar might record the fact that you got married. And many couriers send out confirmation of shipping status as a service enhancement, most (but not all) shops issue receipts for purchases to facilitate the customer's ability to return an item and all Registrars of course issue a marriage certificate to the newlyweds.

different deliveries from different companies); and a shop receipt proves that you – or perhaps your credit card – was at a particular location at a particular point in time (a fact that is valuable for combating fraud).

A more widely recognised example might be the utility bill: produced by an Entity (the utility provider), it records activity involving a Subject (the customer) to the benefit of the relationship between them. And – out of context – utility bills are regularly presented by the Subject as evidence to support their claim to hold a given name and address as Attributes.

4.2 VALUE TO EVENT PROVIDER

For the Event Provider (i.e. Entities that make their activity records accessible to the Subject and authorised third parties), the cost of the activity is a sunk cost – it has already taken place as part of its day-to-day activity – and the only incremental cost that they bear is for capturing and managing access to the digital Event record.

Event Providers may be motivated to do this for the simple reason that it reduces their own operating costs and improves their service offer. Giving Subjects (i.e. their customers, citizens, employees, etc.) access to a digital record of activity is a way of avoiding more expensive alternatives, such as paper-based records. Equally, keeping Subjects informed can be a highly valuable addition to the service offer overall[8].

4.3 VALUE TO EVENT CONSUMER

For the Event Consumer (i.e. third party Entities that are authorised to access Event records), the value of using an Event record a different context may be higher still, if – for example – access to that record allows them to avoid a more expensive alternative investigative process, to manage their risks better, to streamline the experience they offer or to improve their own service levels.

[8] The UK Passport Office offers a good example of this, providing ongoing updates of progress through the passport replacement process via text and email to the Subject (i.e. the passport holder).

As we have already seen in Chapter 3, reasoning models that leverage Events can deliver enormous benefits in the context of data assurance, enabling a more agile defence against fraud, de-duplicating effort, improving interoperability, federating trust and reducing barriers to access to goods for those that cannot provide traditional forms of evidence to support their claims and assure their data.

4.4 VALUE EXCHANGE BETWEEN EVENT PROVIDER AND CONSUMER

As a result, there is an obvious opportunity for a value exchange between Event Providers and Event Consumers, allowing the former to recover some of the cost of their day-to-day activity and the latter to benefit from access to a valuable Digital Resource. Since the structure of an Event is in the form of a witness statement (see Section 2.5), both the Event Provider and Subject have rights over who accesses an Event.

The Event Provider is therefore in a position to determine how Event Consumers access an Event and may choose to make access – for either Subject[9] or third party Event Consumers – contingent on a set of commercial terms. Since the Event Provider is not obliged either to record Events or make them accessible to authorised third parties, there is a justifiable economic rationale for to charge a fee for access, should they choose to do so.

It is important to note that charging for access is distinct from charging for accountability. An Event Consumer should be able to access an Event without holding the Event Provider accountable for whether the *content* of the Event is true or not[10]. When something like a utility bill is presented as evidence of someone's name and address, the Entity receiving it uses it as part of its risk assessment and data assurance, but does not seek to hold the utility

[9] Shops do not make the cost of providing a receipt explicit to its Subjects (i.e. customers), as the provision of a receipt is part of the shops overall value proposition. Utility providers are increasingly charging Subjects (i.e. customers) that opt for a physical bill. The Passport Office and Registry Office explicitly charge Subjects (i.e. citizens) for the provision of a new passport or marriage certificate.

[10] Recall that an Event only guarantees that "At time T, X *says* that Y is true"

provider to account for the accuracy of its content. An Event Provider may choose to make accountability for the accuracy of Event content contingent on commercial terms, but this arrangement would be separate to a charge for access[11].

4.5 VALUE TO SUBJECT / CLAIMANT

What is the motivation for Subjects (that is, you and me) to engage in a process that facilitates the exchange of Events between Event Provider and Event Consumer?

We see three strong, valid reasons: convenience, confidence and continuity. However, we also acknowledge the obvious challenge of overcoming inertia and mistrust.

Because we want to access goods controlled by different organisational Entities, we regularly play the role of a Claimant: i.e. we seek to qualify ourselves as being eligible for the good in question and submit ourselves to the due diligence processes that the Controlling Entity imposes on us, which involves producing evidence to support our claims.

This is clearly an administrative burden for us: it involves the fulfilment of essentially the same investigative process steps time and again, for different Controlling Entities. It would be much more convenient for us if we had the tools to satisfy these requirements by providing a combination of Verified Attributes and – when required – Events as further evidence, at the touch of a button.

We would also be more confident of being able to meet the data assurance standards of the Entity we are dealing with, as we would be able to draw on a much wider range of 'witnesses' able to support our claims than we might have been able to historically. In the role of Subject (i.e. as consumers, citizens, employees, etc.), we already have relationships with dozens of other Entities that can act as Event Providers, and whose Events we can

[11] In traditional terminology, the utility provider is acting as an Attribute Provider. If they were to take accountability for they would be acting as an Identity Provider, and the Entity receiving the utility bill would be acting as a Relying Party (i.e. able to rely on the accuracy of the name and address contained in the utility bill).

produce as evidence – either instead of or in addition to evidence from traditional Trust Anchors.

Finally, our ability to access Events across a period of time means that we can present a longitudinal view that can be used to support data assurance. This continuity enables an accumulation of evidence over time, allowing ‘thin file’ customers in particular to present a track record of their interactions.

CHAPTER 5: WHAT INFRASTRUCTURE IS REQUIRED?

The production, exchange and consumption of Events requires a set of capabilities that already exist, but – generally speaking – have not been used for this purpose. Although many Entities record a subset of their activities to support operational efficiency, audit or compliance, few have implemented the infrastructure required to share such information with either the Subject involved or other authorised third parties.

For an Ecosystem to emerge, Participants in that Ecosystem will need to develop the capabilities to *provide* and *consume* the Events that are specific to their own circumstances. The infrastructure to orchestrate an *exchange* of Event information may take different forms and may be provided by a number of different Solution Providers.

This Whitepaper advocates that a standards-based approach is developed that is agnostic both to the implementation choices of Event Providers and Consumers, and to the various choices for exchange infrastructure that may be used.

5.1 STANDARDS-BASED APPROACH

5.1.1 Common implementation standards

The use or development of common and open standards will both lower barriers to entry for participation and enable interoperability across the different technical implementations that Event Providers and Consumers might choose to adopt.

Our hypothesis is that the W3C Semantic Web standards (including RDF, OWL, SPARQL and Verifiable Credentials) provide a solid basis on which to build a trusted environment for Event-based data assurance. The existence of open-source tools that have already implemented these standards mean that Participants in the Ecosystem can quickly adopt these standards, while the established W3C governance helps to ensure their longevity.

5.2.1 Metadata model

Ecosystem Participants must be able to express Events in a way that can be understood by other Ecosystem Participants. This suggests the need for a shared ontological framework

whose fundamental elements are common to all Ecosystem Participants, and which can evolve over time to include new Event types as required.

Different Ecosystem Participants may undertake a different set of activities to be recorded as Events. Each Participant will therefore need to describe the Events that are specific to them and publish their definitions to other Participants according to this shared framework.

5.3.1 Rules of engagement

A basic set of "rules of engagement" may be needed to govern behaviour within the Ecosystem. This Whitepaper advocates that such rules are kept as light touch as possible, to minimise barriers to entry and ensure that the Ecosystem remains as open to new Participants as possible.

Such rules might include a commitment to adopt the common standards as a default requirement of participation; contribute to the shared ontologies; involve the Subject in authorisation protocols (e.g. consent) where the Subject has rights over the Events; use contractual frameworks to manage the commercial relationship between Event Providers and Event Consumers; etc.

Depending on these rules, it may be necessary to establish a vehicle with which to evolve and govern the Ecosystem over time (e.g. an open alliance of Ecosystem Participants), or to migrate governance of the Ecosystem to an established standards body.

5.2 "IN-HOUSE" CAPABILITIES

Each Ecosystem Participant will need to develop their own in-house capabilities to enable them to act as an Event Provider and Consumer. These capabilities include:

- Event store: the ability to capture, store and manage access to Event metadata, and the ability to transform Event metadata to conform to the requirements specified by the shared metadata model
- Reasoning engine: the ability to use Event metadata (which may have been sourced from other Ecosystem Participants) to assign an assurance status to suit their own specific Use Case requirements

– Pub-Sub: the ability to publish (notifications of) new Events to authorised subscribers and to become an authorised subscriber to Events published by other Ecosystem Participants, leveraging the agreed standards, metadata model and rules of engagement to interact with different potential exchange infrastructures

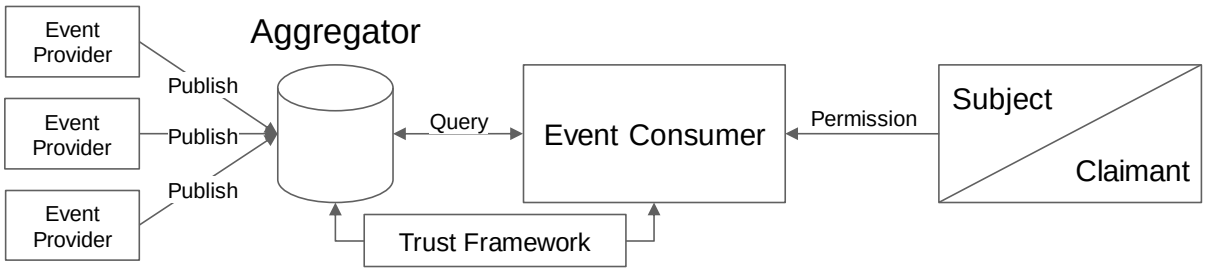
5.3 EXCHANGE INFRASTRUCTURE

Several different mechanisms exist through which Event information might be exchanged between Event Providers and Event Consumers, and there are already several solutions operating in the market today.

The Discovery Project which formed the basis of this OIX Whitepaper focused on three main models, which were labelled Aggregator, Point-to-Point and Self-Sovereign. However, the Discover Project did not seek to recommend the adoption of any one of these models over any of the others. As outlined above, this Whitepaper also advocates a standards-based approach that is agnostic to the choice of exchange infrastructure.

5.3.1 Aggregator

Figure 3: Aggregator



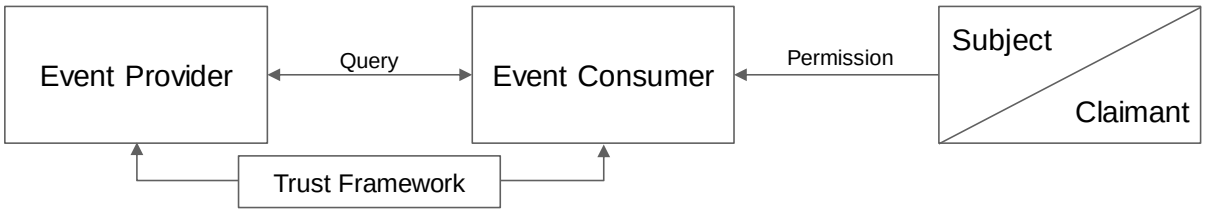
In the Aggregator model, one or more solution providers aggregate Event metadata from all the Event Providers participating in the Ecosystem. This gives Event Consumers a single point of access to Event metadata, subject to authorisation by the relevant Subject.

Because they trust the Aggregator to fulfil its role, Event Consumers can be confident of the provenance of the Events themselves and in the consistent application of the rules of engagement that govern behaviour in the Ecosystem – for example, the Aggregator would establish a single contractual framework with which to engage with Event Provider.

The role played by the Aggregator is similar to that played by credit rating agencies.

5.3.2 Point to Point

Figure 4: Point to Point



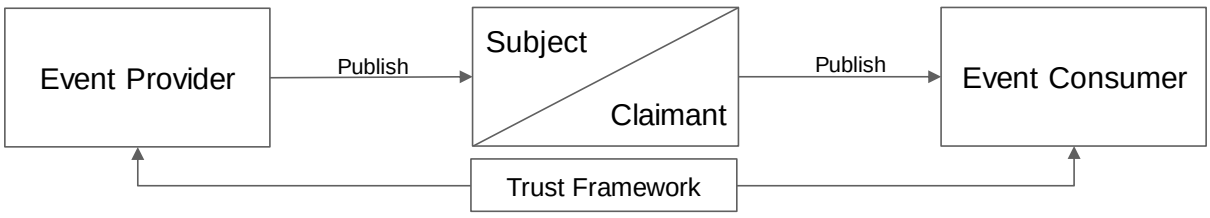
In the Point-to-Point model, Events are shared directly between the Event Provider and Event Consumer, subject to authorisation by the relevant Subject. Because of this, Event Consumers can be confident of the provenance of the Events themselves.

The rules of engagement are implemented bilaterally between each Event Producer and Event Consumer, meaning that – for example – every Ecosystem Participant will either negotiate a contractual framework with every other Ecosystem Participant, or will collectively operate under a standardised contractual framework (i.e. as a scheme).

There have been several examples of bilateral arrangements in the market (e.g. between accounting software packages and banks) as well as examples of schemes (e.g. Open Banking), leveraging standard authorisation protocols such as OAuth 2.0.

5.3.3 Self-sovereign

Figure 5: Self-sovereign



In the Self-sovereign model, Events are passed from the Event Producer to the Subject, who is then responsible for storing them and – when requested – disclosing them to Event Consumers. The Subject is therefore the "bearer" of their own digital credentials, much as they are today the "bearer" of their documentary credentials (e.g. passport).

Because the Subject intermediates the exchange of Event metadata, the Event Consumer must rely on a trust framework to be confident of both the provenance of the Events and the integrity of the Events (i.e. non-repudiation).

Having passed the Event to the Subject, the Event Provider is not responsible for managing the access of the Event Consumer and has no way of making that access contingent on commercial terms. If the Event Consumer wants to rely on the Event Provider (i.e. hold them to account), then a bilateral commercial relationship needs to be established.

5.3.4 Key requirements of exchange infrastructure

Regardless of the infrastructure model that is used to exchange Events, that infrastructure will need to play an important orchestration role. For example:

- How does an Event Consumer discover the availability of Events associated with a Subject?
- How does the Subject present the availability of Events associated with them?
- How does the Event Consumer secure the Subject's consent (assuming consent is required to gain authorised access to the relevant Events)?
- How is the Event information transported between the System where it is produced, the System where it is stored and the System(s) that consume it?
- How does the Event Consumer contract, account for and settle with the Event Provider (assuming payment is required to gain authorised access to the relevant Events)?
- What is the mechanism for arbitration and redress in the event of dispute?

Importantly, all of the main infrastructure models that solution providers might offer as a way of implementing the exchange of Events should conform to the common standards and metadata model that form the basis of the Ecosystem, so that they can interoperate with the "in-house" capabilities developed by each Ecosystem Participant.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

This Whitepaper is based on an OIX Discovery Project that ran over the course of the summer 2019. The Discovery Project sought to explore an initial hypothesis that Entities can provide other Entities with a valuable service by recording the Events that they undertake in the course of their day-to-day business, and subsequently managing access to those events such that they can be used to assure the quality of the data of individual subjects.

There is clearly a huge amount of work that has taken place – and continues to take place – in areas such as data mobility, digital identity and standards development that is relevant to this hypothesis. The project team therefore undertook a landscape review both to test a 'straw man' version of an Event-based data assurance framework and to identify open (meta)data models, standards and protocols that might form the basis of such a framework.

A series of workshops were used to define the different roles that would be needed within an Ecosystem that collaborated to produce, exchange and consume Events; to agree the basic functional requirements of such an Ecosystem; and to discuss the motivations for Entities to play the different roles; and to articulate potentially viable economic models.

An additional workshop was used to discuss different Use Cases from private, public and third sectors, in an effort to identify a Use Case to be taken forward to a subsequent Alpha Project. Working on behalf of the project team, Paul Worrall from Zonafide developed a visualisation of how an Event-based data assurance framework might work, creating a number of artefacts that can be used to illustrate one particular implementation approach[12].

6.1 KEY CONCLUSIONS

The Discovery Project's starting point was the assertion that the current market for data assurance services is struggling, and that:

[12] See Appendix C

- Fraudsters and other bad actors are able to exploit gaps within and between the different solutions and schemes that have been developed to assure data and – as a result – idEntities in different contexts
- Over-reliance on any particular implementation of shared infrastructure can result in uncompetitive market dynamics and unnecessary costs
- Similarly, a dependency on a small number of Trust Anchors makes it difficult for some Subjects (e.g. 'thin file' customers) to access those goods which require a relatively high level of data assurance as part of their qualification criteria

The Discovery Project therefore concluded that there is likely to be enormous value in defining, recording and reasoning over Events as a new type of Digital Resource that can be used to assure data quality for eligibility, entitlement and identity management purposes:

- Events capture the value of processes and – perhaps more usefully – the elemental activities and assertions that collectively comprise processes. Events therefore allow the value of those activities and assertions to be shared, re-used and re-purposed across an Ecosystem of participating Entities
- Events are distinct from the underlying data that they describe, and should be governed accordingly to reflect the rights of the Event Provider. This important feature has the potential to provide a basis for value exchange between Event Providers and Event Consumers, without compromising the personal privacy rights of the Subject
- Events are highly amenable to the techniques of modern data science, and can be used to fuel powerful automated reasoning models. Adoption of such techniques will not only improve data assurance outcomes but also reduce the costs of doing so

Events occur in their trillions on a daily basis as different Entities go about their business. The Discovery Project concluded that Events therefore represent a hitherto largely unharnessed resource which can be used to complement existing data assurance solutions:

- The introduction of Event-based data assurance does not require Entities to change the way in which they operate – Events are an incremental resource that supplement and extend existing approaches to data assurance

- Because Events are – by definition – universal in nature and the "rules of engagement" that supports the exchange of Events are thin, Events can be integrated into existing initiatives and schemes (such as Verify, eIDAS, etc.)
- The elemental nature of Events can be used to improve interoperability between existing initiatives and schemes, reducing the room for mistakes and malfeasance
- The implementation of Event-based data assurance can draw on standards (e.g. RDF, OWL, SPARQL, Verifiable Credentials) and solutions (e.g. policy automation engines) that already exist in – or are emerging into – the market

However, the Discovery Project also recognised the introduction of Event-based data assurance would need to overcome a number of challenges. Because the concept of an Event is relatively obscure, business and technical leaders may need time and support to get to grips with the opportunities it represents:

- Very few organisational Entities have systems in place to record Events, meaning that most Entities will need to invest in some upfront development if they are to establish themselves as Event Providers. Such capabilities have not got a lot of attention or enthusiasm from internal sponsors, because the development costs can be shouldered by one part of the organisation while the benefits (i.e. the reduction in friction, effort, risk and cost and the improvement in outcomes) are often more widely distributed.
- Similarly, not every organisational Entity has implemented Reasoning Engines capable of automating data assurance decisions, and very few – if any – consume Events as an input. Their use may require a fundamental shift away from traditional operating models that rely on human-readable policy documents to express "reasoning" which is in turn implemented by long-established processes. Nonetheless, it is precisely the cost and rigidity of such operating models that present a business case for change.

It is also clear that few organisational Entities have established a track record of sharing data with other Entities. Indeed, there is a genuine resistance to being an Attribute or Event Provider, for a number of reasons (such as fear of loss of control of a customer or a sense of breaching information governance rules). The exchange of Event metadata will therefore require a shift in behaviour that is likely to be predicated on a better understanding of both the benefits and the costs of doing so.

Finally, the issue of data assurance is one of collective action. Schemes that serve specific sectors – and even national schemes – specify high standards of assurance that serve a limited audience for only a sub-set of the Use Cases that are relevant to Subjects/Claimants.

The issue can only be solved by an ecosystem that has the lowest possible barriers to entry, but which builds the foundations of trust between different participants incrementally and iteratively, with elemental units of trust being made available for re-use in different contexts and constantly re-tested by Reasoning Engines that adopt different methodologies.

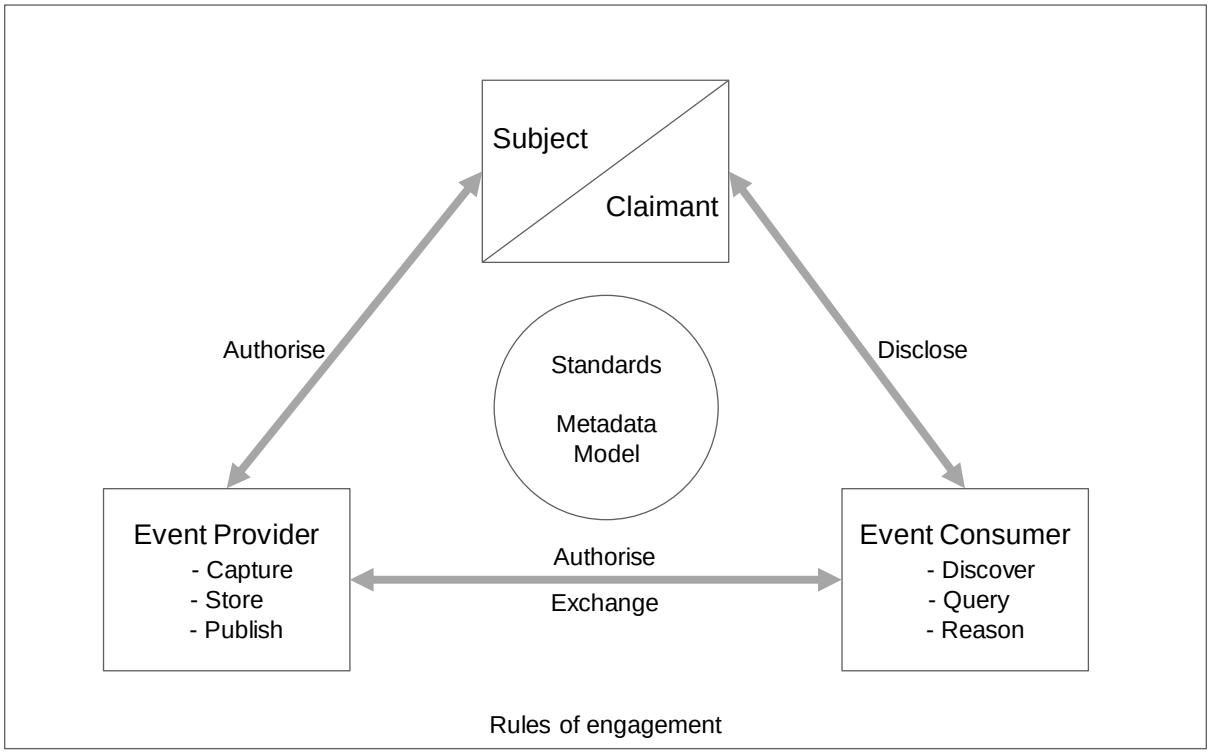
The Discovery Project therefore concluded that the form of "hypo-infrastructure" (i.e. an infrastructure layer that sits *underneath* existing capabilities) required to support the delivery of Event-based data assurance should leverage default standards and "rules of engagement" that involve the lightest touch that captures aggregate benefit for the ecosystem and demonstrates specific motivation for Entities to participate (particularly as Event Providers).

6.2 RECOMMENDATIONS

This Whitepaper therefore recommends the establishment of an OIX Alpha Project whose objective is to agree a set of common requirements to support the provision, exchange and consume Events, to be adopted as part of an open Ecosystem that shares the value of processes, activities and assertions to support data assurance. These requirements include default implementation standards, a metadata model and 'rules of engagement' between participants in the Ecosystem. The Alpha Project should – ideally – test and refine those requirements by building a prototype set of the capabilities required to deliver them.

Implementation of the "hypo-infrastructure" required could take different forms. Although Section 5.3 includes a highlevel description of three architypes of exchange infrastructure, it does not pre-suppose or advocate any particular solution.

Figure 6: Depiction of prototype capabilities to be developed by Alpha Project



In fact, it should be possible for different solutions to co-exist, and the Alpha Project should include the use or development of open standards to enable interoperability across technical implementation choices. This Whitepaper therefore recommends that the Alpha Project will be split into two phases, with the following deliverables:

– Phase 1: agree a default set of common standards, an initial metadata model and base-level 'rules of engagement' as a basis for interoperability
– Phase 2: develop a prototype of the set of capabilities required to generate, exchange and consume Events. These include: published ontologies, Event stores, Reasoning Engines, pub-sub capabilities and implementations of prototype exchange infrastructure that represent up to three of the main options

Phase 1 should be led by Entities that are willing to participate in the role of Event Provider and Event Consumer, with the common requirements that are defined tested through the development of prototype capabilities by Event Providers, Event Consumers and Solution Providers that provide exchange infrastructure solutions.

APPENDIX A: GLOSSARY OF TERMS

- *Actor:* an Entity that actively participates in the Ecosystem by playing a defined role. May also be referred to as an (Ecosystem) Participant.
- *Attribute:* information that is associated with a Subject (i.e. an Entity, usually identified by a unique identifier). For the purposes of this document, Attribute is used interchangeably with the term Data.
- *Attribute Provider:* a Resource Provider that makes attributes (i.e. data) associated with a Subject accessible to authorised Resource Consumers
- *Author:* the Author of a Digital Resource associated an Entity that is uniquely the right of the Author to author. For example: my bank is (uniquely) the Author of my bank account; the DVLA is the Author of my right to drive in the UK; HM Passport Office is the Author of my passport. For the purposes of this document, Issuer is used interchangeably with the term Author.
- *Claim:* a set of Data (or Attributes) with which an Entity claims to be a) true and b) associated with themselves (i.e. as the Subject).
- *Claimant:* a role an Entity can perform by claiming a set of Digital Resources are a) true and b) associated with themselves, as part of an interaction to complete a given Use Case. The use of the term Claimant usually (but not always) indicates the lack of an existing relationship between the Claimant and the Event Consumer – for example, applying to open a bank account or take out a loan; claiming eligibility for a Government benefit; onboarding as a new supplier; etc..
- *Controlling Entity:* an Entity that controls access to goods, and may need to assure the data received from Claimants that claim to qualify for access to those goods
- *Data:* information that is associated with an Entity (via an identifier). For the purposes of this document, Data is used interchangeably with the term Attribute.
- *Digital Resource:* an umbrella term to describe information (Data, Attributes, Identities, Events, etc.) that one Entity makes accessible to another Entity on a digital basis.
- *Due Diligence:* a set of Events used to determine whether the Digital Resources that comprise a Claim are a) accurate and up-to-date at the time of undertaking and b) associated with the Subject.

- *Ecosystem:* an umbrella term used to describe a collection of independent actors that interact to achieve a given outcome
- *Entity:* a real life 'thing' (e.g. person, organisation, System, device) that can be digitally represented with a unique identifier and may play a role within an Ecosystem
- *Event:* a trustable record of universal content that supports chaining and is made available as a Digital Resource for re-use by authorised third parties. See Chapter 2 for a fuller explanation.
- *Event Consumer:* a role an Entity (usually a System acting on behalf of an organisation) can play by requesting access to Events that, once access is granted, are used to execute Use Case (e.g. through the deployment of a Reasoning Engine)
- *Event Provider:* a Resource Provider that makes Events accessible to authorised Event Consumers as part of an interaction to achieve a given outcome. "Event Provider" is used to denote the provision of Events specifically, in contrast to the provision of other resources, as denoted by commonly used terms such as Attribute Provider and Identity Provider
- *Identity Provider:* a Resource Provider that confirms the physical or legal identity of an Entity that is identified by a unique identifier. An Identity Provider acts as Trust Anchor for the Relying Parties that can be hold it accountable for the service it provides
- *Issuer:* the Issuer of a Digital Resource associated an Entity that is uniquely the right of the Issuer to issue. For example: my bank is (uniquely) the Issuer of my bank account; the DVLA is the Issuer of my right to drive in the UK; HM Passport Office is the Issuer of my passport. For the purposes of this document, Issuer is used interchangeably with the term Author.
- *Participant:* an Entity that actively participates in the Ecosystem by playing a defined role. May also be referred to as an Actor.
- *Personal Data Store:* a System that stores and protects access to Digital Resources and acts directly of behalf of the Subject. A Personal Data Store can therefore be a special kind of Event Provider, in that the Subject's ability to authorise access to Events is not intermediated by a third party organisation acting as a custodian of those Events, and therefore dependent on the functionality that the third party may (or may not) provide.

- *Reasoning Engine:* the automated execution of a predetermined instruction set to combine Digital Resources into an outcome, usually to support the execution of policy
- *Resource Consumer:* an Entity that is authorised to access the digital resources associated with a Subject that are being made available
- *Resource Provider:* an Entity that makes digital resources associated with a Subject available to authorised Resource Consumers
- *Relying Party:* a Resource Consumer that can hold the Resource Provider accountable for the service that it provides (for example: the accuracy of the information exchanged)
- *Roles:* an umbrella term used to characterise how Ecosystem Participants interact to fulfil a given Use Case or achieve a specific outcome. It is important to note that a single Participant can fulfil multiple roles within the context of a single interaction. Similarly, a single Participant can fulfil different roles in different interactions
- *Subject:* the Entity about which Digital Resources – including Events – are associated. The use of the term Subject usually (but not always) indicates a existing relationship between the Subject and the Event Provider – for example, as a customer, citizen, beneficiary, employee, supplier, etc.
- *System (or System instance):* a single implementation of a computer programme used to execute a set of functions. A System is a uniquely identifiable Entity which participates in the Ecosystem, acting on behalf of the Entity which owns or controls it.
- *Trust Anchor:* an Entity that is trusted other Entities to have assured data quality on their behalf. Trust Anchors are not the Author or Issuer of the resource in question – they have undertaken an investigative process as due diligence
- *Use Case:* an umbrella term used to describe a generally recognisable set of outcomes (e.g. Customer Due Diligence, Customer Onboarding, Credit Risk Underwriting)
- *Verified Attribute:* a piece of information whose association with a given Subject has been investigated to a sufficient standard that the association between the information and the Subject can be considered as 'true' by the Resource Consumer

APPENDIX B: WHAT DO WE MEAN BY 'DATA ASSURANCE'?

In this OIX Whitepaper, we have argued that Events constitute a valuable Digital Resource that have yet to be fully harnessed for the purposes of data assurance.

In this Appendix, we set out to establish a common understanding of what is meant by "data assurance".

B.1 ELIGIBILITY AND ENTITLEMENT

We regularly seek to gain – or maintain – access to the products, services, benefits and opportunities that are part of the rich tapestry of our day-to-day lives: a car to drive, a bank account, a phone contract, a medical treatment, a new job, a disability benefit, etc.

These "goods" are often integral to our wellbeing, but we do not always have unrestricted access to all of them. Anyone with the right money can buy a bag of crisps, but only people who are both over 17 years old and hold a provisional licence can drive a car.

The Entities controlling the goods that we want access to are responsible for making sure that we are eligible for or entitled to those goods. The Controlling Entities qualify our application for access against the criteria have been set and, in the case of an ongoing service, may periodically check that we continue to qualify for it, in case our circumstances have changed relative to those criteria.

B.2 INVESTIGATING THE STATUS OF A CLAIM

Our ability to access the goods that we want is therefore based on our claim to meet the qualifying criteria, and in some circumstances no more than a self-assertion of the claim itself is sufficient: for example, access to age-restricted websites and apps is still largely predicated on a self-asserted claim to be old enough[13].

[13] Regardless of whether or not this is a good thing, it is nonetheless the case

However, in circumstances where the desired good is of higher import, value or risk, the Controlling Entities seek to enforce their qualifying criteria by investigating whether our claim to meet them is true or not. As outlined below, this involves asking us to present some form of evidence to support our claim and performing due diligence on the evidence.

The actual status that the Controlling Entity assigns to our claim to qualify for the good in question may never be *explicitly* expressed. Once we have successfully gone through the due diligence process it may simply be assumed that our claim (now held as pieces of data on the Controlling Entity systems) is "true".

In this document, we refer to "data assurance" as being both the due diligence process that leads up to the assignment of a status and – whether implicit or explicit – the assignment of that status to a claim held as data on the systems of the Entity controlling access to goods.

B.3 BASIC DYNAMICS OF DATA ASSURANCE

Usually, the Issuer (or Author) of the evidence to support our claim is a third party Entity that acts as a witness to the veracity of our claim[14]. The Controlling Entity is of course free to specify what sort of evidence they accept, the nature of the due diligence process and the rules or reasoning that determine whether they believe the evidence presented is sufficient to support our claim to the standards of data assurance required.

The key dimensions of a due diligence investigation are[15]:

- The Issuer of the evidence: Controlling Entities may place a higher or lower level of trust on the third party issuing the evidence

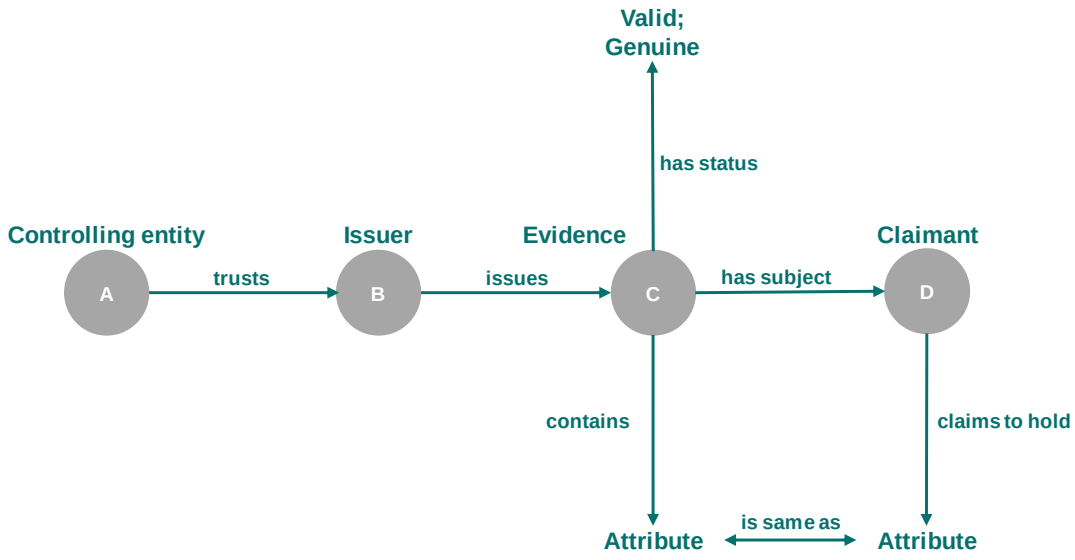
[14] It is of course possible for the evidence to be produced and consumed bilaterally, involving only the claimant (i.e. ourselves) and the Controlling Entity. For example, a retailer selling alcohol may agree that we are over 18 by seeing us in person; an animal welfare officer may agree that we offer a suitable home for a pet by coming to visit it.

[15] Both the form and holder of the evidence are relevant practical considerations for the investigations that are not considered here: for example, Controlling Entities may only be willing or able to deal with a given form of evidence (e.g. oral attestation, physical document, digital data object), and the Controlling Entity's access to the evidence might be via the claimant themselves, via someone holding the evidence on their behalf (e.g. a parent for a child) or directly from the issuer of the evidence

- Whether the evidence is genuine: Controlling Entities may seek to determine whether the evidence is counterfeit or has been tampered with
- Whether the evidence is valid: Controlling Entities may want confirmation that the evidence has not expired or been revoked
- Whether the evidence relates to the claimant: Controlling Entities may confirm that we are the subject of the evidence and that it was not issued in relation to someone else
- Whether the evidence supports the claim: Controlling Entities will compare the information contained in the evidence with the content of our claim

This can be expressed as a graph that captures the key Attributes and relationships:

Figure 1: Evidence to support a claim



The due diligence process comprises of the actions that are undertaken to support the assertion of each of these relationships. For example, the Controlling Entity may assert that "the evidence has the claimant as its subject" *because* the claimant looks like the person in the photo on the evidence.

The data assurance status that the Controlling Entity assigns to our claim – if it is ever made explicit – is based on the aggregation of these asserted relationships, which themselves may never be made explicit. Data assurance is more commonly thought of in procedural terms:

the due diligence was completed successfully, we were given access to the good, and therefore our claim must be true.

Clearly, however, it is not always possible to assert the relationships which underpin data assurance with the same level of confidence: a Controlling Entity might have less confidence in an assertion made by a new, untrained employee based on a cursory comparison of face and photo than on an assertion made by an experience, certified colleague using the latest biometric technology.

B.4 USE CASES, STANDARDS, GUIDELINES, IMPLEMENTATIONS AND SCHEMES

Even after a thorough due diligence process, it may not be possible for the Controlling Entity to establish that our claim is absolutely "true" – there may always be a risk that the claim is false at the point of due diligence, and with the risk increasing over time. For example, I may indeed live today at the address I claim to live at, but may have moved by next week.

How much effort should Controlling Entities put into their data assurance? The Entity controlling access to the goods in question makes this decision based on a trade off between the cost of the due diligence, the risk of making a mistake when assigning a status to a claim, and the consequences of such a mistake: i.e. the decision is always Use Case specific, and the assertion of a data assurance status is always a risk-adjusted exercise.

In some industries, a regulator may articulate minimum *standards* for data assurance where the industry Participants share a key Use Case. To meet these standards, the Controlling Entity might adopt a set of *guidelines* agreed on by the industry, but they are often free to choose how they *implement* their own data assurance in line those guidelines.

In this paradigm, in order for one Controlling Entity to rely on the data assurance of another Entity, they must share exactly the same Use Case (i.e. have exactly the same criteria to qualify access, either in part or aggregate) and the former must be confident that the data assurance of the latter meets their own standards.

Schemes enable this kind of reliance, usually by certifying that the Entities providing the data assurance have adhered a stated due diligence *implementation* and by ensuring that those Entities are accountable for the data assurance they produce. The Controlling Entities

that make use of third party data assurance as part of a scheme (often referred to as Relying Parties) can therefore not only rely on it but also hold the provider accountable if any of the relationships which underpin the data assurance turn out to be "false".

B.5 CHAINING EFFECTS, AUTHORSHIP AND TRUST ANCHORS

It should also clear that data assurance is a function of chaining: the issuers of the evidence that we use to support our claims have themselves – at some stage in the past – undertaken their own due diligence to qualify us and assign their own data assurance status.

Controlling Entities often ask claimants to present a passport as evidence in part[16] because the information it contains is widely trusted. This is because the Passport Office only issues passports having enforced its own qualification criteria by investigating the evidence that it needs: namely, an original birth certificate, a recent photo and a witness statement[17] confirming the applicant's (claimant's) identity.

The passport document itself asserts that its subject holds several Attributes, including name, date and place of birth and nationality. Of these, the Passport Office is only Author of our nationality (i.e. authorised to establish or revoke it), in the same way that a bank is the Author of our bank account. Although the Passport Office is clearly not the Author of our name, date and place of birth, it does act as a Trust Anchor for them (i.e. Controlling Entities place a high level of confidence on the Passport Office's assertion that these Attributes are "true") due to the chaining effect placed on its own due diligence, as outlined above.

Given the risk-adjusted nature of data assurance, it is important to distinguish between these different concepts. For example: the DVLA is the Issuer of UK Drivers Licences, which are regularly used as evidence to support claims. The DVLA is the Author of our right to

[16] A passport has several other advantages for the Controlling Entity: a large proportion of potential claimants have one; they are easy to access, as they are held by the claimant themselves; they are easy to check as being valid and genuine; and they contain mechanisms that link the passport to its subject (e.g. photo, biometrics, associated Attributes)

[17] Even a witness statement involves chaining at some level: an honest witness is relying on the evidence of shared experience to assert the applicant's identity; a dishonest witness – if caught out – may be required to *produce* evidence of the shared experience that supports their assertion

drive, and a Trust Anchor for our name and date of birth (because their due diligence relies on a UK Passport – or equivalent documentation – as evidence of identity). But the DVLA is neither Author of or a Trust Anchor for our address, which is also contained on UK Drivers Licences, and a Controlling Entity should place less confidence in this assertion, treating it as a corroborating evidence only, not least because it may well be out of date.

APPENDIX C: DISCOVERY PROJECT ARTEFACTS

Working on behalf of the project team, Paul Worrall from Zonafide created a number of artefacts to illustrate the way in which W3C Semantic Web standards and tools could be used to support the implementation of an Event-based data assurance framework. Paul also developed a visualisation of how Events might be exchanged under the Self-sovereign implementation model, using a Personal Data Store implemented as a SOLID Pod.

C.1 USE OF W3C SEMANTIC WEB TOOLS

Introductory video:

https://youtu.be/pgLFLrx_av8

Example ontology:

https://ontologies.interition.info/webprotege/

Model queries and tests:

https://github.com/pjworrall/trustedenvironment

Model query endpoint:

https://fuseki.interition.info/

C.2 EXCHANGE OF EVENTS (SELF-SOVEREIGN MODEL)

Introductory video:

https://youtu.be/ar-yof2Mkss

Visualisation:

https://github.com/pjworrall/oixonsolid

Solid server used with the Visualisation App:

https://solid.interition.info:8443/