# DORA

*e le nuove esigenze di attestazione crittografica*

## Sintesi esecutiva

Il Digital Operational Resilience Act, pienamente applicabile dal 17 gennaio 2025, produce una domanda strutturale di evidenza: gli enti finanziari devono conoscere, documentare, provare, ricostruire e notificare.

Questa richiesta incontra però una realtà bancaria molto diversa dal modello dell'ente unitario, fatta di aggregazioni stratificate, joint venture, catene ICT multilivello e concentrazioni di quarta parte. A questa complessità si aggiunge l'intelligenza artificiale agentica, che DORA non poteva incorporare nel testo del 2022 e che oggi richiede un livello di tracciabilità (modello, configurazione, dataset, strumenti, interventi umani, momento dell'esecuzione) per il quale i log applicativi non bastano.

Le attestazioni temporali tradizionali risolvono una parte del problema, ma possono introdurne un altro. Una Time Stamping Authority esterna non vede il contenuto, ma può osservare pattern operativi, volumi, cadenze e correlazioni temporali del soggetto attestante. Per una banca, la frequenza delle attestazioni legate a pagamenti, trading, incidenti, riconciliazioni o cicli AI può diventare di per sé informazione strategica. La risposta deve quindi separare verificabilità e osservabilità: produrre prove robuste senza generare nuove superfici di osservazione laterale sui flussi del soggetto regolato.

CertiSigma e Census, due servizi complementari erogati da Ten Sigma Sagl, costituiscono la risposta progettuale a questa tensione. Census opera localmente nel perimetro del cliente come strumento di inventario crittografico, manifestazione e versionamento degli artefatti, inclusi dataset, configurazioni di agenti AI, dossier documentali.

CertiSigma riceve esclusivamente digest crittografici e produce un'evidenza a tre livelli: firma CertiSigma T0, marca temporale qualificata eIDAS T1 su radice Merkle aggregata, ancoraggio T2 su blockchain Bitcoin tramite OpenTimestamps. Il contenuto originale non lascia mai il perimetro del cliente. La Time Stamping Authority esterna vede Ten Sigma Sagl come operatore della radice aggregata, non i singoli clienti finali né i loro flussi individuali. Restano naturalmente nel perimetro contrattuale e tecnico di CertiSigma i metadati

operativi necessari all'erogazione del servizio, da gestire secondo il modello documentato, le policy applicabili e le misure di sicurezza previste.

La tesi conclusiva si riassume in una formula: prova fuori, contenuto dentro, significato sotto controllo del cliente. La regolazione europea sta spingendo il mercato verso un nuovo tipo di infrastruttura, capace di produrre prova dei dati senza necessariamente vedere i dati. È in questo spazio che CertiSigma e Census si collocano, come strato di evidenza componibile con audit, conservazione a norma, firma qualificata e governance AI, senza pretendere di sostituirli.

## Nota di posizionamento

CertiSigma e Census non certificano la conformità a DORA, non determinano il perimetro soggettivo dell'ente finanziario, non sostituiscono il framework di ICT risk management, il Registro delle Informazioni, le attività di audit, le funzioni di controllo interno o la valutazione delle autorità competenti. Forniscono uno strato di evidenza crittografica content-blind su digest, manifest e radici aggregate, utilizzabile come supporto nei processi di governance, incident response, gestione dei fornitori ICT, audit trail e AI governance.

# 1. DORA e il cambio di paradigma per il settore bancario

## 1.1 Genesi e razionale del regolamento

Il Digital Operational Resilience Act, Regolamento (UE) 2022/2554, è stato negoziato dalle istituzioni europee tra il 2020 e il 2022, pubblicato nella Gazzetta Ufficiale dell'Unione europea il 27 dicembre 2022 e applicabile a partire dal 17 gennaio 2025. Il regolamento nasce da una constatazione che la crisi finanziaria del 2008 e la digitalizzazione dei servizi finanziari hanno reso evidente: banche, assicurazioni e intermediari finanziari sono diventati organizzazioni tecnologiche regolate. La continuità dei servizi finanziari dipende da infrastrutture ICT, cloud, software, reti, processori, piattaforme dati e fornitori specializzati.

Prima di DORA, la gestione del rischio tecnologico e cibernetico nel settore finanziario era disciplinata da norme nazionali frammentate, linee guida delle autorità europee di vigilanza e prassi divergenti tra Stati membri. DORA introduce un quadro armonizzato a livello europeo, direttamente applicabile, con un perimetro che include banche, assicurazioni, intermediari di mercato, gestori di fondi, fornitori di servizi di pagamento, emittenti di moneta elettronica, fornitori di servizi cripto e una parte rilevante della filiera ICT del settore finanziario.

La logica del regolamento non è soltanto impedire gli incidenti. DORA assume che gli incidenti accadranno e chiede agli enti finanziari di essere in grado di assorbirli, contenerli, ricostruirli, segnalarli e ripristinare i servizi entro soglie compatibili con la stabilità del sistema. È un passaggio culturale importante: la conformità tecnologica diventa resilienza operativa documentabile.

## 1.2 I cinque pilastri operativi

### 1.2.1 Governance del rischio ICT

Il primo pilastro riguarda l'organo di amministrazione dell’ente finanziario. La responsabilità della resilienza operativa digitale arriva al livello apicale. Il consiglio di amministrazione approva la strategia ICT, definisce la tolleranza al rischio, riceve reportistica strutturata e risponde dell'adeguatezza dell'impianto. La materia tecnologica smette di essere un tema confinato alla funzione IT e diventa parte della governance strategica.

### 1.2.2 Gestione e segnalazione degli incidenti

Il secondo pilastro impone agli enti finanziari di classificare gli incidenti ICT secondo criteri armonizzati e di notificare quelli qualificati come maggiori alle autorità competenti. Gli standard tecnici attualmente applicabili prevedono una initial notification il prima possibile e comunque entro quattro ore dalla classificazione dell'incidente come major ICT-related incident, nonché non oltre ventiquattro ore dal momento in cui l'ente finanziario ne è venuto a conoscenza. Seguono un intermediate report entro settantadue ore dalla initial notification e un final report non oltre un mese dall'intermediate report o, se del caso, dall'ultimo updated intermediate report. La logica è consentire alle autorità una lettura aggregata del rischio sistemico e una risposta più coordinata in caso di eventi rilevanti.

### 1.2.3 Testing della resilienza

Il terzo pilastro impone test periodici e proporzionati delle capacità di resilienza. Per gli enti più significativi è previsto il Threat-Led Penetration Testing, ispirato al framework TIBER-EU della Banca Centrale Europea. Il test deve simulare scenari realistici, coinvolgere funzioni critiche e, ove rilevante, includere le interfacce con fornitori ICT critici.

### 1.2.4 Gestione del rischio di terze parti ICT

Il quarto pilastro ha un impatto operativo particolarmente forte. Gli enti finanziari devono mappare i fornitori ICT, mantenere un Registro delle Informazioni in formato strutturato, condividere i dati richiesti con le autorità, inserire clausole contrattuali obbligatorie nei rapporti con i fornitori, presidiare subappalto, audit, exit strategy, localizzazione dei dati, livelli di servizio e continuità operativa. I fornitori ICT qualificati come critici rientrano inoltre in un regime europeo di oversight da parte delle autorità di vigilanza.

### 1.2.5 Condivisione delle informazioni sulle minacce

Il quinto pilastro incoraggia la condivisione volontaria di informazioni e intelligence sulle minacce cibernetiche. L'obiettivo è ridurre l'asimmetria tra attaccanti coordinati e difensori isolati, creando canali settoriali di collaborazione e apprendimento reciproco.

## 1.3 Implicazioni di fondo

Tre implicazioni sono rilevanti per questa relazione. La prima è il passaggio dalla prevenzione alla resilienza provabile: il sistema deve dimostrare come reagisce, con quali evidenze e con quale catena di responsabilità. La seconda è l'aumento

della responsabilità del management: il rischio ICT entra nel perimetro di governance e può generare conseguenze organizzative, reputazionali e sanzionatorie. La terza è la ridefinizione del rapporto con i fornitori tecnologici, inclusi gli hyperscaler, che devono accettare clausole, audit, livelli di trasparenza e controlli prima non presenti nei contratti standard.

Queste implicazioni convergono in un punto: DORA produce una fame di evidenza. Non basta avere procedure. Occorre poter provare che una procedura era attiva, che un fornitore era mappato, che un incidente è stato rilevato in un certo momento, che una configurazione era quella dichiarata, che un log o un report non è stato alterato. Il regolamento rende la prova tecnica parte della postura di resilienza.

## 2. La banca non è un monolite: la realtà delle aggregazioni stratificate

### 2.1 Il mito implicito del soggetto coerente

Quando il regolatore scrive che l'ente finanziario deve mappare i propri sistemi ICT, mantenere un registro unico dei fornitori e applicare politiche coerenti di gestione del rischio, presuppone un soggetto con perimetro chiaro, governance integrata e architettura tecnologica riconducibile a decisioni centrali. La realtà del sistema bancario europeo è spesso diversa. Una banca commerciale di medie dimensioni può essere il risultato di fusioni stratificate, migrazioni incomplete, contratti ereditati, sistemi legacy mantenuti in vita per ragioni operative e società prodotto con autonomia gestionale.

A questa stratificazione interna si aggiunge una rete di partecipazioni, joint venture, outsourcer, società consortili, processori specializzati e fornitori SaaS che a loro volta utilizzano cloud provider, subfornitori e componenti open source. Il perimetro effettivo della banca, dal punto di vista del rischio operativo digitale, eccede il perimetro giuridico della singola entità autorizzata.

### 2.2 Anatomia tecnologica tipica di una banca commerciale europea

Per fissare il problema, si può descrivere l'anatomia tipica di una banca commerciale europea nata da successive aggregazioni. Il core banking può essere distribuito su due o tre sistemi: un mainframe legacy ancora in produzione per una parte del portafoglio retail, un sistema più recente per la rete principale e una piattaforma ereditata da un'acquisizione non ancora migrata. I pagamenti istantanei transitano attraverso un consorzio interbancario. Le carte di pagamento sono processate da un operatore specializzato. L'home banking è sviluppato da un partner tecnologico. L'antiriciclaggio gira su una piattaforma SaaS. Il trading utilizza sistemi di mercato globali. Gli eventuali servizi cripto passano attraverso un custodian regolamentato, HSM, nodi blockchain, cloud e strumenti di monitoraggio esterni.

Ogni anello di questa catena è stato spesso acquistato, negoziato e governato da una funzione diversa. Procurement, IT, conformità, risk, legal e linee di business possono avere viste parziali. Prima di DORA il rischio aggregato di queste architetture veniva spesso percepito per segmenti. DORA costringe a vederlo in modo unitario.

## 2.3 Il Registro delle Informazioni e ciò che ha rivelato

Il Registro delle Informazioni previsto da DORA impone di censire gli accordi contrattuali ICT con terzi a un livello di dettaglio significativo: identificativo del fornitore, funzione supportata, criticità operativa, localizzazione geografica di dati e trattamenti, subfornitori, entità del gruppo coinvolte, servizi erogati, termini contrattuali, exit strategy e altri parametri rilevanti. La prima compilazione rigorosa ha mostrato a molte organizzazioni quanto fosse incompleta la loro conoscenza operativa.

Il Registro delle Informazioni non è una semplice tabella contrattuale. Gli standard tecnici di implementazione europei ne definiscono template strutturati, da mantenere e aggiornare in relazione agli accordi contrattuali con fornitori terzi ICT, e ne fanno uno strumento centrale del framework di gestione del rischio ICT di terze parti e della supervisione da parte delle autorità.

Sono emersi contratti ICT non censiti, sistemi considerati interni ma ospitati da terzi, catene di subappalto più profonde del previsto, concentrazioni cloud invisibili, fornitori formalmente distinti ma dipendenti dalla stessa region di uno stesso hyperscaler. DORA ha funzionato come una radiografia dell'anatomia tecnologica reale delle banche.

## 2.4 Il problema della concentrazione nascosta

Il rischio di concentrazione di quarta parte è uno degli aspetti più delicati. Una banca può ritenersi diversificata perché utilizza cinque fornitori distinti per cinque funzioni critiche. Se quattro di quei fornitori girano nella stessa region cloud, una singola interruzione può colpire simultaneamente funzioni che sulla carta apparivano indipendenti. Lo stesso vale nei pagamenti, nella messaggistica finanziaria, nelle piattaforme antifrode, nei servizi di identità e nei sistemi di sicurezza gestiti.

La concentrazione nascosta è pericolosa perché appare solo quando i dati vengono collegati. Il singolo contratto non la mostra. Il singolo fornitore non la dichiara sempre in modo operativo. Il rischio compare nell'aggregazione. Per questo il Registro delle Informazioni non è solo un adempimento amministrativo: è un dispositivo di scoperta del rischio.

## 2.5 Joint venture, società prodotto e crypto-asset

La complessità aumenta con joint venture e società prodotto. Molte banche hanno entità separate per asset management, bancassurance, leasing, factoring, credito al consumo, pagamenti digitali e servizi cripto. DORA richiede di trattare queste entità secondo la loro funzione effettiva. Se sono enti finanziari soggetti al regolamento, hanno obblighi propri. Se erogano servizi ICT alla banca, devono essere trattate come terze parti ICT, anche quando appartengono allo stesso gruppo.

Il caso dei crypto-asset mostra bene il problema. Una banca che offre custodia o trading cripto può utilizzare un Crypto-Asset Service Provider regolamentato, un custodian tecnico, moduli hardware di sicurezza, nodi blockchain, cloud provider, strumenti di transaction monitoring e servizi di conformità. Ogni anello introduce rischio operativo e richiede evidenze. Il confine tra servizio finanziario, infrastruttura ICT e componente crittografica diventa meno netto.

## 2.6 Risposte organizzative emergenti

Le risposte organizzative che si stanno consolidando seguono tre direttrici. La prima è la creazione di funzioni centralizzate di Third Party Risk Management, capaci di unire approvvigionamento, IT, conformità, ufficio legale e rischio. La seconda è l'investimento in strumenti di intelligence sui fornitori e monitoraggio continuo, per seguire certificazioni, incidenti pubblici, assetti proprietari, localizzazione dei servizi e solidità finanziaria. La terza è la rinegoziazione dei contratti esistenti, perché molte clausole richieste da DORA non erano presenti nei contratti storici.

Nel breve periodo DORA aumenta la complessità operativa, perché obbliga le banche a formalizzare ciò che prima restava opaco. Nel medio periodo questa visibilità può aumentare la resilienza. Il passaggio intermedio richiede però strumenti capaci di produrre evidenze affidabili senza aggiungere ulteriore esposizione informativa.

## 2.7 Dalla mappatura alla prova

Il Registro delle Informazioni non esaurisce il fabbisogno probatorio. Sapere che un fornitore esiste e che un contratto è censito è solo il primo livello. Occorre poter provare che una configurazione era attiva, che una versione di un contratto era quella vigente, che un report di audit non è stato modificato, che una catena di subfornitura era stata dichiarata in un certo momento, che un incidente è

stato documentato con una sequenza temporale coerente. La resilienza operativa digitale si traduce quindi in una domanda di attestazioni ripetute, granulari, verificabili e minimizzate.

Questa domanda diventa ancora più forte quando sistemi AI e agenti software entrano nei processi bancari. La catena di fornitura ICT non è più composta solo da applicazioni statiche. Include modelli, prompt, policy, strumenti, dataset, embedding, orchestratori e agenti capaci di azione autonoma. La banca deve poter ricostruire una catena che cambia nel tempo.

## 3. AI agentica, dataset e nuovo debito evidenziale

### 3.1 La cronologia di un disallineamento

DORA è stato chiuso nel 2022. Il rilascio pubblico di ChatGPT è avvenuto a fine novembre 2022, quando il testo del regolamento era ormai definito. L'intelligenza artificiale agentica, intesa come sistema capace di pianificare, decidere, invocare strumenti e agire attraverso API o ambienti esterni, non era ancora un fenomeno operativo diffuso nel settore finanziario. Di conseguenza, DORA tratta l'AI dentro la categoria generale del rischio ICT: asset, servizio, fornitore, componente software, processo esternalizzato.

Questa impostazione resta utile, ma parziale. Un agente AI non si comporta come un applicativo tradizionale. Può produrre output differenti a parità di input, comporre azioni non anticipate dai progettisti, cambiare comportamento a seguito di aggiornamenti del modello, usare strumenti esterni, accedere a basi dati e generare artefatti. Il rischio non è soltanto tecnologico. È procedurale, probatorio, contrattuale, organizzativo e, in alcuni casi, regolatorio.

### 3.2 Le assunzioni implicite di DORA che l'AI agentica mette sotto pressione

L'impianto di DORA presuppone, in molti passaggi, che il software sia relativamente stabile, testabile e riconducibile a un perimetro contrattuale chiaro: un sistema, un fornitore, un proprietario interno, una funzione supportata. Gli agenti AI mettono sotto pressione questa semplificazione. Un agente può appoggiarsi a un modello base di un provider, girare su cloud di un altro provider, usare un framework open source, interrogare un vector database, invocare API di terzi, produrre documenti, scrivere ticket, aggiornare record e avviare flussi di lavoro.

La catena di responsabilità non sparisce, ma diventa più difficile da ricostruire. Chi ha prodotto l'evento? Il modello? L'orchestratore? Il tool invocato? Il dato recuperato dal sistema interno? L'operatore umano che ha approvato? Il fornitore SaaS che ha introdotto una funzione agentica in un aggiornamento? Per rispondere servono evidenze tecniche granulari e temporalmente ordinate.

## 3.3 Il problema della decisione delegata

DORA presuppone che le decisioni rilevanti siano prese da persone, eventualmente supportate da sistemi automatici. Gli agenti AI modificano questa assunzione. Nei processi di onboarding, classificazione AML, supporto clienti, gestione documentale, riconciliazione, generazione di report, flussi di lavoro di credito o incident response, un agente può suggerire, compilare, classificare, inoltrare, decidere entro soglie predefinite o attivare passaggi successivi.

La banca resta responsabile verso clienti, autorità e mercato. Il punto critico è dimostrare che l'agente ha operato dentro una configurazione autorizzata. Questo significa provare versione del modello, prompt di sistema, regole di policy, tool abilitati, base dati consultabile, soglie operative, eventuale approvazione umana e artefatto finale. La spiegabilità del ragionamento può restare limitata. La ricostruibilità forense, invece, deve essere progettata.

## 3.4 AI Act e sovrapposizione regolatoria

L'Unione europea ha introdotto il Regolamento (UE) 2024/1689, noto come AI Act, entrato in vigore il 1° agosto 2024. Il calendario applicativo è progressivo: alcune disposizioni sono già applicabili, le regole per i modelli di general-purpose AI hanno una propria traiettoria, e le regole per i sistemi ad alto rischio seguono scadenze differenziate. Poiché il calendario applicativo dell'AI Act è progressivo ed è oggetto di sviluppi interpretativi e di semplificazione, ogni qualificazione operativa deve essere verificata sul testo vigente, sugli atti di esecuzione e sulle linee guida applicabili al momento dell'uso.

Per il settore finanziario, l’AI Act è rilevante soprattutto quando il sistema AI rientra negli usi ad alto rischio previsti dall'Allegato III, per esempio nella valutazione del merito creditizio o della solvibilità di persone fisiche, fatte salve le eccezioni previste. Altri casi bancari, come AML, trading, customer service o gestione della liquidità, possono essere critici sotto DORA, MiFID, CRR, normativa antiriciclaggio, governance interna o responsabilità contrattuale, ma non vanno descritti automaticamente come ad alto rischio ai sensi dell’AI Act senza una qualificazione puntuale.

La sovrapposizione resta comunque concreta. Una banca che usa AI in processi regolati deve coordinare DORA, AI Act, GDPR, discipline settoriali, regole di outsourcing e obblighi di audit. Questi quadri normativi non sono stati scritti come un unico sistema. La sintesi operativa resta in capo all’ente vigilato.

## 3.5 Tre frontiere critiche per il settore bancario

La prima frontiera riguarda gli agenti customer-facing capaci di influire su operazioni o decisioni verso il cliente. Anche quando l'agente non ha piena autonomia dispositiva, può orientare scelte, raccogliere informazioni, precompilare istruzioni o innescare processi successivi. In caso di contestazione, diventa necessario ricostruire la sequenza dell'interazione.

La seconda frontiera riguarda gli agenti di back office. Qui il rischio principale è l'automation bias: operatori umani travolti dal volume finiscono per approvare raccomandazioni senza reale controllo. La supervisione umana richiesta sulla carta deve essere dimostrabile in concreto. Occorre provare chi ha visto cosa, quando, con quale livello di informazione e con quale esito.

La terza frontiera riguarda l'introduzione silenziosa di funzioni AI nei prodotti SaaS già utilizzati dalla banca. Un fornitore può integrare classificatori, copiloti, agenti di processo, funzioni generative o sistemi di scoring in una piattaforma esistente. Il rischio AI entra dalla porta dell'aggiornamento software. Se la banca non ha una capacità di inventario e attestazione delle configurazioni, la governance arriva tardi.

## 3.6 Il dataset come oggetto regolatorio e probatorio

Nei sistemi AI il modello finale è solo una parte della storia. La base informativa da cui il modello apprende o da cui recupera contenuti diventa un oggetto probatorio centrale. Training set, validation set, test set, corpus di fine-tuning, dati sintetici, dataset annotati, embedding store e knowledge base per retrieval-augmented generation sono componenti diverse, con rischi diversi e con responsabilità diverse.

Una banca o un fornitore AI deve poter dimostrare non solo quale modello ha usato, ma anche quale corpus ha alimentato una certa fase del ciclo di vita. Serve sapere quali file sono stati inclusi, quali esclusi, quali duplicati rimossi, quali trasformazioni applicate, quali versioni congelate, quali controlli eseguiti, quali report di valutazione prodotti. Senza una genealogia tecnica del dataset, la governance AI resta dichiarativa.

Questo punto collega DORA e AI Act in modo concreto. DORA chiede resilienza, gestione del rischio ICT, controllo dei fornitori e ricostruibilità degli incidenti. L'AI Act chiede documentazione, tracciabilità, qualità dei dati, robustezza e

monitoraggio per i sistemi rientranti nel suo perimetro. Entrambi convergono su un'esigenza operativa: poter collegare un artefatto AI a una storia verificabile.

## 3.7 Dall'AI agentica all'evidenza per progettazione

L'intelligenza artificiale agentica non genera solo un problema di supervisione. Genera un problema di evidenza. Un agente che pianifica, invoca strumenti, accede a basi dati, produce output e attiva flussi di lavoro lascia dietro di sé una sequenza di eventi tecnici che, in caso di audit, incidente, contestazione o verifica regolatoria, deve poter essere ricostruita. Il punto non è promettere che l'agente diventi pienamente interpretabile in senso cognitivo. Il punto è rendere verificabile la catena degli eventi.

L'evidenza non può essere aggiunta a posteriori come semplice esportazione di log. Deve essere progettata nel ciclo operativo dell'agente. Ogni evento rilevante dovrebbe poter generare una prova crittografica proporzionata e minimizzata: versione del modello, configurazione, prompt template, tool disponibili, policy attiva, input trattato, output prodotto, chiamata esterna, approvazione umana, artefatto finale. La ricostruibilità forense deve diventare una proprietà architetturale, non una promessa documentale.

Questo è il passaggio dalla registrazione dei log all'evidenza per progettazione. Il log resta utile, ma è modificabile, frammentato e spesso custodito nello stesso ambiente operativo che potrebbe essere compromesso. L'evidenza crittografica separa la prova dell'esistenza e dell'integrità dell'evento dal contenuto integrale dell'evento. In questo modo un sistema AI può avere memoria probatoria senza imporre una memoria informativa centralizzata presso un terzo.

## 3.8 Implicazioni per attestazioni, audit e incident response

Le implicazioni pratiche sono rilevanti. In caso di incidente AI, non basta conservare una conversazione o un log applicativo. Occorre dimostrare quale agente era in esecuzione, quale versione del modello era attiva, quale knowledge base era accessibile, quali strumenti erano abilitati, quale policy limitava l'azione, quali dati sono stati trattati, quale output è stato generato e quale essere umano, se previsto, ha validato il passaggio.

Lo stesso vale per audit e contestazioni. Un cliente può contestare una decisione assistita da AI. Un'autorità può chiedere di ricostruire un flusso AML. Un auditor può verificare il dataset usato per una classificazione. Un fornitore

può sostenere che una configurazione era diversa. In tutti questi casi la risposta migliore non è un deposito massivo di contenuti presso terzi. È una catena di impronte, manifest, prove di inclusione, timestamp e ancoraggi che consenta di ricostruire ciò che serve, quando serve, con il contenuto ancora sotto controllo del soggetto responsabile.

# 4. Il problema delle attestazioni temporali e dei metadati identificativi

## 4.1 Cosa attestano le marche temporali e perché sono necessarie

Le attestazioni temporali servono a provare che un determinato dato, documento, log o artefatto esisteva in una certa forma in un certo momento. Nel modello tradizionale, una Time Stamping Authority riceve l'impronta crittografica del dato, produce un token conforme allo standard applicabile e firma l'associazione tra impronta e tempo. Nel contesto europeo, il Regolamento (UE) 910/2014, eIDAS, attribuisce alle marche temporali elettroniche qualificate una presunzione di accuratezza della data e dell'ora indicate e di integrità dei dati ai quali data e ora sono associate.

Nel contesto DORA, NIS2, eIDAS 2.0, AI Act e discipline settoriali, l'uso di attestazioni temporali diventa sempre più importante per log di sistema, evidenze di incidenti, registri di audit, backup, configurazioni, contratti, report di sicurezza, dataset e artefatti AI. L'obiettivo è garantire integrità, anteriorità e ricostruibilità. In caso di incidente o contestazione, il soggetto regolato deve poter dimostrare che l'evidenza non è stata prodotta o modificata ex post.

## 4.2 Token RFC 3161 e metadati operativi

Il punto delicato riguarda i metadati. In un token RFC 3161 sono presenti elementi tecnici come message imprint, policy, serial number, generation time, identificazione della TSA ed eventuali campi opzionali. Il token, in sé, non deve essere descritto come se contenesse necessariamente l'identità del richiedente finale. Tuttavia, l'interazione operativa con una TSA esterna può generare metadati osservabili al di fuori del token: account cliente, credenziali API, indirizzi di rete, log di accesso, volumi, frequenza delle richieste, serialità dei token ottenuti, orari di picco, pause e correlazioni temporali.

Questa distinzione è essenziale. Il contenuto può restare protetto perché viene trasmesso solo un hash. Il pattern operativo, però, può restare leggibile. Se una banca attesta direttamente ogni log, ogni batch di pagamento, ogni evento di trading o ogni artefatto AI presso una TSA esterna, il fornitore può osservare la cadenza delle richieste. Anche senza vedere i contenuti, può inferire periodi di attività intensa, momenti di crisi, cicli batch, finestre operative, incidenti o comportamenti anomali.

## 4.3 Incompatibilità con l'opacità operativa bancaria

Le banche proteggono dati, clienti e operazioni. Proteggono anche la forma temporale dei propri flussi. Sapere quando una banca esegue netting interbancario, processa grandi pagamenti corporate, aggiorna modelli di rischio, effettua stress test, riconcilia posizioni, muove liquidità infragiornaliera o risponde a un incidente può avere valore informativo per competitor, attaccanti e attori statuali ostili.

La granularità temporale dei flussi finanziari è informazione sensibile. Può alimentare inferenze su esposizioni, posizioni di liquidità, relazioni di credito, stato operativo e vulnerabilità. La protezione dell'opacità temporale non è un eccesso di riservatezza. È parte della stabilità operativa e della sicurezza del sistema.

## 4.4 Una contraddizione regolatoria latente

Il quadro europeo produce una tensione tra esigenze legittime. DORA chiede tracciabilità, integrità, evidenza e reporting. Il GDPR chiede minimizzazione e protezione dei dati personali. Il segreto bancario nazionale protegge clientela e operazioni. Le normative antiriciclaggio richiedono conservazione e ricostruibilità. La sicurezza nazionale richiede protezione dei flussi finanziari critici. Ogni vincolo è razionale. La difficoltà nasce quando le prove richieste da un vincolo diventano metadati sensibili per un altro vincolo.

Una TSA esterna basata in un paese terzo introduce esposizioni giurisdizionali e possibili ordini di divulgazione. Una TSA esterna europea riduce il rischio geopolitico, ma può concentrare in un unico soggetto metadati strategici su molti operatori. Una TSA interna tutela maggiormente il controllo informativo, ma riduce l'indipendenza probatoria. Il problema richiede quindi un'architettura che separi il più possibile verificabilità e osservabilità.

## 4.5 Il nodo della giurisdizione tecnica

La giurisdizione formale del prestatore non esaurisce il problema. Una TSA qualificata eIDAS opera sotto il regime europeo dei servizi fiduciari, ma lo stack tecnico sottostante può includere componenti cloud, moduli hardware, storage, servizi gestiti o sistemi di monitoraggio soggetti a giurisdizioni diverse. Il punto non è sostenere che ogni componente extraeuropea sia inaccettabile. Il punto è riconoscere che la sovranità formale del servizio e la sovranità operativa dello stack possono divergere.

Per questo la coerenza giurisdizionale è un fattore progettuale. In contesti bancari, sanitari, pubblici o di sicurezza, il cliente non valuta solo la firma o il certificato. Valuta dove transitano le richieste, chi gestisce la terminazione TLS, chi vede i log, chi ospita l'applicazione, chi opera l'HSM, chi controlla il proxy, quale legge può imporre accesso a metadati e quali elementi restano verificabili anche se un singolo fornitore diventa indisponibile.

## 4.6 Tecniche di mitigazione e limiti correnti

Esistono diverse tecniche di mitigazione. Il batching crittografico aggrega molti hash in strutture Merkle e sottopone alla TSA solo la radice. Questo riduce la frequenza di esposizione verso l'esterno e attenua il canale laterale verso la TSA, ma richiede una gestione corretta delle prove di inclusione. Il timestamping decentralizzato su blockchain pubbliche, per esempio tramite OpenTimestamps su Bitcoin, riduce la dipendenza da una singola autorità, ma il suo valore probatorio va presentato come componente tecnica complementare, non come sostituto automatico delle marche qualificate. Il timestamping interno mantiene controllo informativo, ma ha minore indipendenza. Schemi avanzati zero-knowledge sono promettenti, ma spesso non ancora standardizzati per l'uso operativo di massa in settori regolati.

La direzione più robusta è un'architettura multilivello. Ogni livello ha una funzione diversa: immediatezza operativa, opponibilità qualificata, persistenza distribuita, riduzione dei metadati, prova di inclusione, verificabilità indipendente. L'obiettivo non è trovare un singolo punto di fiducia perfetto. È distribuire il trust e minimizzare ciò che ogni soggetto esterno può osservare.

## 4.7 Requisiti di una risposta adeguata

Una risposta adeguata al problema deve soddisfare cinque requisiti. Primo: lavorare su impronte e manifest, evitando l'esportazione del contenuto quando non necessaria. Secondo: ridurre la visibilità del cliente finale verso i prestatori esterni attraverso aggregazione e separazione dei ruoli. Terzo: mantenere prove verificabili in modo indipendente, anche a distanza di tempo. Quarto: dichiarare con precisione ciò che il servizio attesta e ciò che resta fuori dal suo perimetro. Quinto: integrarsi nei flussi di lavoro reali, inclusi agenti AI, dataset, incident response, audit trail e gestione dei fornitori.

Questi cinque requisiti delineano due funzioni distinte ma complementari. Una funzione locale, nel perimetro del cliente, capace di inventariare artefatti,

calcolare impronte, costruire manifest strutturati e gestire versioni: è il ruolo di Census, presentato nei capitoli successivi. Una funzione esterna, indipendente e verificabile, capace di attestare digest e radici aggregate senza vedere i contenuti né osservare direttamente i flussi del cliente finale: è il ruolo di CertiSigma. L'efficacia operativa dipende dalla combinazione delle due.

# 5. CertiSigma e Census come risposta progettuale

## 5.1 Due servizi complementari

La risposta progettuale che si descrive in questo capitolo è composta da due servizi complementari, erogati entrambi da Ten Sigma Sagl, società a garanzia limitata di diritto svizzero con sede nel Cantone Ticino, e da intendersi come componenti di un'unica architettura di evidenza. Census opera nel perimetro locale del cliente come strumento di inventario crittografico, hashing, manifestazione e versionamento di artefatti. CertiSigma opera come servizio esterno di attestazione crittografica dell'esistenza temporale e dell'integrità di valori di hash. Il primo prepara l'evidenza nel contesto del cliente; il secondo la ancora a una catena verificabile indipendentemente. Il contenuto originale non lascia mai il perimetro del cliente, perché tra Census e CertiSigma circolano esclusivamente impronte crittografiche.

## 5.2 CertiSigma: la natura del servizio

CertiSigma riceve digest crittografici, tipicamente SHA-256 di 32 byte, calcolati dal cliente nel proprio ambiente a partire da file, documenti, log, manifest, dataset, configurazioni o artefatti tecnici. Il servizio non riceve il contenuto originale per attestare l'impronta. La prova viene separata dal contenuto. Il marchio CertiSigma è registrato presso l'Istituto Federale della Proprietà Intellettuale e protetto fino al 2035.

Questa separazione è la base del modello content-blind. Il servizio esterno non interpreta il documento, non classifica il dato, non conserva il corpus e non conosce il significato informativo dell'artefatto. Riceve un valore crittografico e produce una catena di evidenza associata a quel valore.

## 5.3 Content-blindness, limiti e buone pratiche

La content-blindness va descritta con precisione. Per file, documenti e artefatti con entropia sufficiente, la preimage resistance di SHA-256 rende computazionalmente impraticabile la ricostruzione del contenuto a partire dal digest. Questo è un limite tecnico forte, non una semplice promessa contrattuale.

Esistono però casi da gestire con attenzione. Se l'artefatto ha bassa entropia o appartiene a un insieme facilmente enumerabile, un soggetto che conosce le

possibili alternative può calcolare hash candidati e confrontarli. Esempi semplici sono valori binari, codici brevi, template molto standardizzati, etichette di stato o piccoli record prevedibili. In questi casi il cliente dovrebbe usare strategie adeguate: incapsulamento in un manifest più ricco, salting controllato, HMAC con chiave custodita localmente, aggregazione Merkle o altre tecniche coerenti con il livello di verificabilità richiesto. Census fornisce nativamente queste opzioni nella costruzione dei manifest.

Va distinta anche la content-blindness dalla metadata-blindness. CertiSigma può evitare di vedere il contenuto, ma l'operatività del servizio può comunque generare metadati: account, API key, indirizzo di rete, timestamp di ricezione, frequenza delle richieste, namespace o pattern di utilizzo. La tesi forte non è che nessun metadato esista. La tesi corretta è che l'architettura riduce l'esposizione del contenuto e impedisce alla TSA esterna di osservare direttamente i flussi individuali del cliente finale, spostando la relazione sensibile dentro un perimetro controllabile e contrattualizzabile con il servizio svizzero.

## 5.4 L'architettura a tre livelli

### 5.4.1 Livello T0: firma del servizio

Il primo livello consiste in una firma ECDSA su curva P-256 apposta da CertiSigma su un payload canonico che include il digest del cliente, un identificativo di versione del formato e un timestamp di registrazione in UTC. La firma T0 attesta che il servizio ha osservato quel digest e ha prodotto la firma sul payload dichiarato entro l'istante indicato. La verifica indipendente richiede il digest, il payload canonico, la chiave pubblica del servizio e la firma.

T0 ha valore operativo immediato. È il livello utile per attestazioni frequenti, integrazioni API, agenti AI, pipeline dati, log e manifest. Non pretende di essere una marca temporale qualificata eIDAS in proprio. Fornisce una prima prova crittografica firmata dal servizio, utile per costruire la catena successiva.

### 5.4.2 Livello T1: timestamp qualificato eIDAS su Merkle root

Il secondo livello aggrega digest individuali in una struttura Merkle e sottopone la radice a una Time Stamping Authority esterna qualificata ai sensi di eIDAS, conforme allo standard RFC 3161, iscritta nelle Trusted Lists europee e vigilata dall'autorità competente nazionale.

Il punto probatorio va espresso con precisione: la marca qualificata T1 si applica direttamente alla Merkle root, non a ogni contenuto originario del cliente. La

verifica completa richiede il file o artefatto originale, il digest, la prova di inclusione Merkle, la root marcata e il token della TSA. Una volta ricostruita correttamente questa catena, il cliente può collegare il proprio artefatto alla radice marcata temporalmente.

È importante chiarire perché questa catena indiretta può fornire, ai fini operativi, una continuità crittografica forte tra l'artefatto del cliente e la marca temporale qualificata apposta sulla Merkle root. La marca qualificata T1 si applica direttamente alla root; la Merkle proof collega il digest individuale a quella root. La valutazione del valore probatorio dell'intera catena, inclusa la prova di inclusione, resta rimessa al giudice, all'autorità competente o all'auditor secondo il diritto applicabile e le circostanze del caso.

In termini tecnici, alterare l'artefatto dopo l'attestazione senza modificare digest e proof richiederebbe di violare le proprietà della funzione hash o della struttura Merkle. In termini giuridici, tuttavia, l'effetto qualificato eIDAS resta riferito alla root timestampata, mentre l'estensione dell'evidenza al digest individuale deriva dalla verifica crittografica della proof.

La TSA esterna, per costruzione, vede Ten Sigma Sagl come soggetto che sottopone la root aggregata. Non vede i singoli clienti finali, i loro documenti, i loro dataset, i loro log o la frequenza delle loro attestazioni individuali. Questo riduce in modo sostanziale il canale laterale informativo verso la TSA.

### 5.4.3 Livello T2: ancoraggio su blockchain Bitcoin

Il terzo livello ancora la radice giornaliera del batch alla blockchain pubblica di Bitcoin tramite OpenTimestamps. Il file di prova può essere verificato contro la blockchain pubblica senza dipendere dal servizio. Questo livello non deve essere presentato come equivalente a una marca temporale qualificata eIDAS, né come servizio di registro elettronico qualificato ai sensi di eIDAS come modificato. Ha una funzione diversa: aggiunge un riferimento temporale distribuito, resistente alla dipendenza da un singolo prestatore fiduciario o da una singola giurisdizione operativa.

## 5.5 Il valore strategico dell'architettura multilivello

L'architettura T0, T1 e T2 non è una duplicazione ornamentale. È una distribuzione del rischio di trust. T0 offre attestazione immediata e integrazione operativa. T1 offre una validazione temporale qualificata sulla radice Merkle dentro il regime eIDAS. T2 offre un ancoraggio distribuito e verificabile nel tempo. Se un livello diventa indisponibile, contestato, deprecato o soggetto a

coercizione, gli altri livelli restano elementi di supporto alla ricostruzione probatoria.

Questa struttura è particolarmente adatta a un contesto regolatorio in movimento. DORA, AI Act, eIDAS 2.0, NIS2, GDPR e discipline settoriali continueranno a stratificarsi. Un servizio di attestazione progettato per essere componibile, verificabile e giuridicamente prudente è più robusto di un servizio che rivendica una singola qualificazione come soluzione universale.

## 5.6 Coerenza giurisdizionale dello stack

Un elemento specifico di CertiSigma è la cura della coerenza giurisdizionale dello stack. La società è svizzera. L'infrastruttura di hosting è presso Infomaniak Network SA, operatore svizzero con sede a Ginevra e data center ubicati in Svizzera. Allo stato dichiarato dell'architettura e per l'infrastruttura sotto il controllo di Ten Sigma Sagl, la progettazione evita l'interposizione ordinaria di CDN, reverse proxy o servizi di mitigazione gestiti da fornitori sotto giurisdizioni terze quando ciò introdurrebbe un nuovo osservatore dei flussi.

Il livello T1 utilizza una TSA europea qualificata. Il livello T2 esce dal perimetro CH/UE solo nella misura in cui si ancora a Bitcoin, infrastruttura pubblica distribuita e non riconducibile a un singolo operatore. La scelta non elimina ogni rischio, ma rende esplicita la funzione di ciascun livello e la sua giurisdizione operativa.

## 5.7 Disambiguazione giuridica: ciò che il servizio non è

La robustezza del servizio dipende anche da ciò che non rivendica. CertiSigma non opera come notaio o pubblico ufficiale rogante. Non produce atto pubblico né scrittura privata autenticata. Non opera come prestatore di servizi fiduciari qualificati eIDAS per la firma elettronica qualificata o il sigillo elettronico qualificato sul contenuto del cliente. Non opera come Time Stamping Authority qualificata in proprio: la qualifica eIDAS del livello T1 deriva dalla TSA esterna utilizzata. Non opera come servizio di registro elettronico qualificato, Certificate Authority, intermediario finanziario, operatore di comunicazione elettronica o servizio di conservazione qualificata dei documenti del cliente.

Questa disambiguazione non indebolisce la proposta. La rafforza. Evita overclaim, chiarisce il perimetro di responsabilità e consente di integrare CertiSigma con firme elettroniche, conservazione documentale, audit,

notarizzazione, sistemi GRC, SIEM, DMS, strumenti AI governance e flussi di lavoro contrattuali.

## 5.8 Posizionamento complementare nei flussi di lavoro regolati

CertiSigma si posiziona come componente di evidenza temporale e integrità dentro flussi di lavoro più ampi. Un documento firmato elettronicamente può avere il proprio digest attestato. Un atto notarile può integrare un'attestazione indipendente sul proprio hash. Una comunicazione PEC può essere collegata all'hash dell'allegato. Un dossier in conservazione può includere attestazioni su versioni puntuali. Un report di audit verso un fornitore critico può essere sigillato al momento della ricezione. Una configurazione di sicurezza può essere attestata prima di una modifica.

Nel settore bancario, gli usi più rilevanti riguardano log e audit trail ad alta frequenza, catena di evidenza per la segnalazione degli incidenti, report di terze parti ICT, configurazioni critiche, registri di subfornitura, snapshot di contratti, esiti di test di resilienza e prove relative a sistemi AI. La funzione è sempre la stessa: rendere verificabile un artefatto senza trasferire il suo contenuto a un ulteriore osservatore esterno.

## 5.9 CertiSigma innestato negli agenti AI

L'integrazione con agenti AI è una delle applicazioni più rilevanti del modello. Un agente AI non produce solo output. Produce eventi: ricezione di un input, selezione di una policy, chiamata a un tool, interrogazione di una knowledge base, generazione di un output, richiesta di approvazione umana, scrittura di un ticket, modifica di un record, produzione di un report. Ciascuno di questi eventi può essere rappresentato da un manifest minimo e attestato crittograficamente.

CertiSigma può essere innestato a livello SDK, API o middleware nel ciclo operativo dell'agente. Il sistema cliente genera localmente un event manifest con i dati necessari al livello probatorio scelto: identificativo dell'agente, versione del modello, hash del prompt template, hash della policy attiva, elenco o hash dei tool disponibili, hash dell'input o del suo contenitore, hash dell'output, riferimento alla knowledge base, eventuale approvazione umana, timestamp locale, ambiente di esecuzione, identificativo del flusso di lavoro. Il contenuto sensibile può restare nel perimetro del cliente. Verso CertiSigma viene inviato il digest del manifest o una radice aggregata.

Questo non rende l'agente “spiegabile” in senso pieno. Rende l'agente forensicamente ricostruibile. In caso di audit, incidente o contestazione, il cliente può esibire gli artefatti originali conservati internamente e dimostrare che corrispondono alle impronte attestate. La prova non dipende dalla buona fede del log applicativo o dalla memoria del fornitore AI. Dipende da una catena crittografica esterna al sistema che ha generato l'evento.

Il modello è particolarmente utile nei processi regolati. In un flusso AML, si può attestare quale versione dell'agente ha classificato un alert e quale supervisione umana è intervenuta. In un flusso di lavoro di credito, si può attestare quale dataset di policy era attivo al momento di una raccomandazione. In un help desk bancario, si può provare che un agente customer-facing aveva determinati limiti operativi. In un incidente cyber, si può dimostrare quando un agente di incident response ha prodotto una timeline, quali fonti ha consultato e quali output sono stati approvati.

## 5.10 Census e la tracciabilità crittografica dei set di addestramento AI

Census svolge un ruolo complementare a CertiSigma. Se CertiSigma produce attestazione esterna su impronte e radici aggregate, Census opera nel perimetro locale del cliente come strumento di inventario, hashing, manifestazione e versionamento degli artefatti. Nel contesto AI, questa funzione diventa centrale per i set di addestramento e per tutto ciò che alimenta il ciclo di vita del modello.

Un dataset AI non è un oggetto statico. È una sequenza di versioni: corpus grezzo, selezione iniziale, esclusioni, normalizzazione, deduplicazione, annotazione, segmentazione, generazione di embedding, split tra training, validation e test, eventuale sintesi di dati artificiali, fine-tuning, valutazione, rilascio. Ogni fase produce artefatti. Senza un inventario crittografico, la domanda “quali dati hanno contribuito a questo modello?” resta dipendente da tabelle interne, esportazioni manuali o dichiarazioni difficili da verificare.

Census può calcolare localmente gli hash dei file, costruire manifest strutturati, registrare percorsi, metadati tecnici, versioni, dipendenze, esclusioni, trasformazioni e relazioni tra dataset. Il cliente può poi attestare verso CertiSigma il digest del manifest o la root Merkle del corpus, mantenendo i dati nel proprio ambiente. Il servizio esterno non riceve i file di addestramento, non vede il contenuto dei documenti e non interpreta il significato dei dati. Riceve

solo ciò che serve a provare che una determinata configurazione informativa esisteva in quel momento.

Per un training set, questo significa poter dimostrare che una specifica versione del corpus era composta da determinati artefatti. Per un validation set, significa poter collegare i risultati di valutazione al dataset effettivamente usato. Per un test set, significa proteggere l'integrità di un benchmark interno. Per un embedding store, significa attestare la versione della base informativa interrogabile da un agente. Per un fine-tuning, significa congelare il perimetro informativo prima dell'addestramento e collegarlo al modello risultante.

CertiSigma e Census non certificano la qualità, la liceità, la completezza, la neutralità o la rappresentatività del dataset. Queste valutazioni restano in capo al titolare, al fornitore AI, agli auditor, ai DPO, ai responsabili del rischio e agli organismi competenti. Il contributo è diverso: fornire una base verificabile per discutere quelle valutazioni. Senza prova dell'esistenza e dell'integrità del dataset, ogni valutazione successiva poggia su un fondamento debole.

## 5.11 Catena di evidenza per modelli, dataset e agenti

L'unione tra Census e CertiSigma consente di costruire una catena di evidenza lungo l'intero ciclo AI. Census fotografa localmente artefatti, dataset e trasformazioni. CertiSigma attesta impronte, manifest e radici. Il cliente conserva contenuti, chiavi, significato organizzativo e documentazione sostanziale. Un auditor può verificare la corrispondenza tra contenuto esibito e impronta attestata. Un'autorità può ricostruire la sequenza temporale. Un fornitore può dimostrare che una versione era in uso prima di una contestazione. Un cliente può provare che un output è stato generato da una configurazione specifica.

Questa architettura è particolarmente coerente con il principio “prova fuori, contenuto dentro, significato sotto controllo del cliente”. La prova esce perché deve essere verificabile da terzi. Il contenuto resta dentro perché può essere sensibile, personale, bancario, proprietario o coperto da segreto. Il significato resta sotto controllo del cliente perché solo il cliente conosce il contesto, la policy, il dataset, il contratto o il processo operativo a cui l'impronta si riferisce.

## 6. Scenari applicativi

### 6.1 Catena di evidenza degli incidenti DORA

In un incidente ICT maggiore, la banca deve ricostruire la sequenza temporale: rilevamento, classificazione, escalation, contenimento, comunicazioni, report intermedi, analisi finale, remediation. Ogni passaggio produce artefatti: log, screenshot, esportazioni da SIEM, ticket, email, report, configurazioni, liste di indicatori, snapshot di sistemi. CertiSigma può attestare digest e manifest di questi artefatti mentre l'incidente è in corso, riducendo il rischio che la catena probatoria venga contestata successivamente.

L'attestazione content-blind è utile perché durante un incidente i contenuti possono essere estremamente sensibili: vulnerabilità, credenziali, clienti impattati, architettura interna, indirizzi IP, indicatori non pubblici. Non è necessario trasferire tutto a un terzo per provare che l'evidenza esisteva in quella forma in quel momento.

### 6.2 Fornitori ICT critici e configurazioni

DORA richiede controllo dei fornitori ICT, incluse clausole, audit, subappalto, localizzazione, livelli di servizio, exit strategy e resilienza. In questo contesto, CertiSigma può attestare versioni di contratti, report di audit, dichiarazioni di subfornitura, prove di test, configurazioni condivise, certificazioni e comunicazioni critiche. Census può inventariare localmente pacchetti documentali e creare manifest verificabili.

Il beneficio non è solo probatorio. È organizzativo. Un gruppo bancario con molte entità può creare una memoria verificabile delle versioni approvate e degli stati di conformità, senza centralizzare necessariamente ogni contenuto sensibile in un repository unico accessibile a troppe funzioni.

### 6.3 AI governance e model lifecycle

Nel ciclo di vita del modello, Census può fotografare dataset, versioni, split, trasformazioni, report di valutazione e artefatti di rilascio. CertiSigma può attestare le impronte dei manifest e collegare temporalmente ogni fase. Questo permette di dimostrare che un modello è stato addestrato o valutato rispetto a un determinato perimetro informativo, che un benchmark interno non è stato modificato dopo i risultati, che un fine-tuning si riferisce a una versione

congelata del corpus, che un modello rilasciato corrisponde a un determinato pacchetto di evidenze.

Questa funzione è utile anche per fornitori AI che lavorano con clienti regolati. Il fornitore può mantenere i dati nei propri sistemi, ma offrire al cliente un pacchetto di evidenza verificabile. Il cliente può sottoporlo ad audit senza ricevere ogni dettaglio proprietario, se il contratto prevede livelli differenziati di divulgazione. L'evidenza crittografica diventa un ponte tra trasparenza e protezione del know-how.

## 6.4 Agenti AI in esercizio

Per gli agenti AI in produzione, l'obiettivo è attestare eventi selezionati, non ogni token generato. La progettazione deve essere proporzionata. Un agente a basso rischio può attestare solo versioni di configurazione e release. Un agente che interagisce con clienti può attestare sessioni, output finali e approvazioni. Un agente inserito in processi regolati può attestare anche tool call, policy attive, knowledge base e interventi umani.

Il livello di dettaglio deve essere deciso dal titolare del rischio. CertiSigma fornisce il meccanismo di attestazione. Census e gli strumenti locali producono manifest e impronte. La governance decide cosa è rilevante, per quanto tempo conservarlo, chi può accedervi e in quali condizioni può essere esibito.

## 6.5 Contenzioso, audit e ricostruzione forense

Il valore di una prova emerge spesso mesi o anni dopo. In un contenzioso, un audit o una verifica forense, il problema è dimostrare che un certo artefatto è quello originale o che una certa versione esisteva prima di un evento. La catena CertiSigma-Census consente di presentare contenuti conservati internamente e collegarli a impronte attestate esternamente. La verifica può essere svolta senza dipendere dalla memoria del sistema applicativo originario.

Questo è particolarmente importante quando il sistema originario è stato compromesso, migrato, dismesso o aggiornato. Un log conservato nello stesso ambiente dell'incidente può essere contestato. Un digest attestato esternamente prima o durante l'evento ha una forza diversa, perché rende più difficile alterare retroattivamente la storia senza lasciare tracce.

## 7. Conclusioni

Il percorso sviluppato in questa relazione mostra che DORA non produce solo un nuovo regime di conformità ICT. Produce una domanda strutturale di evidenza. Le banche devono conoscere il proprio perimetro tecnologico, governare fornitori e subfornitori, classificare incidenti, testare resilienza, documentare decisioni, dimostrare configurazioni e ricostruire eventi. Questa domanda incontra una realtà operativa frammentata, stratificata e sempre più attraversata da sistemi AI.

L'intelligenza artificiale agentica amplia il problema. Un agente non è solo un software che restituisce un output. È un sistema operativo capace di comporre azioni, consultare basi dati, invocare strumenti e produrre artefatti. Per governarlo servono log, ma i log da soli non bastano. Servono evidenze verificabili, temporalmente ordinate e separabili dai contenuti sensibili. Lo stesso vale per i dataset: la governance AI richiede di provare la genealogia del corpus, le versioni, le trasformazioni e la relazione tra dati, modello e output.

Le attestazioni temporali tradizionali risolvono una parte del problema, ma possono introdurne un altro: l'osservabilità laterale dei flussi. Una TSA esterna può non vedere il contenuto, ma può osservare pattern operativi, volumi, frequenze e correlazioni. Per soggetti bancari e regolati, questi metadati possono essere sensibili. La risposta deve quindi separare verificabilità e osservabilità.

CertiSigma e Census si collocano in questo spazio. Lavorano su impronte, non su contenuti. Distribuiscono il trust su T0, T1 e T2. Usano batching e Merkle root per ridurre la visibilità verso la TSA esterna. Mantengono una coerenza giurisdizionale svizzera ed europea. Dichiarano con prudenza ciò che sono e ciò che non sono. Insieme, costituiscono uno strato di evidenza per documenti, log, configurazioni, fornitori, agenti AI e dataset.

La formula più sintetica è: prova fuori, contenuto dentro, significato sotto controllo del cliente. Questa non è solo una scelta tecnica. È una risposta architetturale alla tensione che attraversa la regolazione europea: documentare senza esporre, provare senza centralizzare, rendere verificabile senza trasformare l'evidenza in sorveglianza.

In un mercato in cui DORA, AI Act, GDPR, eIDAS 2.0, NIS2 e discipline settoriali continueranno a sovrapporsi, il valore non sarà soltanto avere sistemi conformi. Sarà avere sistemi capaci di produrre prove proporzionate, minimizzate,

verificabili e durature. CertiSigma e Census possono essere letti come infrastrutture per questa fase: non sostituiscono governance, audit o valutazioni legali, ma forniscono il terreno crittografico su cui quelle valutazioni possono poggiare con maggiore solidità.

## Appendice A · Esempi di manifest attestabili

Gli esempi che seguono hanno funzione illustrativa. Non costituiscono template normativi, né individuano campi obbligatori. Il contenuto effettivo, i campi e la granularità devono essere definiti in base al rischio, al settore, al contratto, alla base giuridica, alle policy interne e agli obblighi applicabili.

### A.1 Manifest per dataset AI

- Identificativo interno del dataset e versione.
- Data di congelamento della versione.
- Algoritmo di hashing usato.
- Elenco dei file o radice Merkle dell'elenco.
- Numero di artefatti inclusi ed esclusi.
- Regole di preprocessing applicate.
- Riferimento a training, validation o test split.
- Hash del report di qualità o valutazione.
- Eventuale HMAC o salting per elementi a bassa entropia.
- Digest finale attestato verso CertiSigma.

### A.2 Manifest per agente AI

- Identificativo dell'agente e versione del pacchetto applicativo.
- Versione del modello o del provider model endpoint.
- Hash del system prompt o della policy attiva.
- Hash della configurazione degli strumenti disponibili.
- Hash della knowledge base o riferimento alla root del corpus Census.
- Hash dell'input o del contenitore dell'input.
- Hash dell'output o dell'artefatto prodotto.
- Identificativo dell'approvazione umana, se prevista.
- Timestamp locale e timestamp applicativo T0 CertiSigma.
- Digest finale o root Merkle del ciclo evento.

### A.3 Manifest per catena di evidenza dell'incidente

- Identificativo dell'incidente e fase del processo.
- Artefatti inclusi nel pacchetto probatorio.

- Hash dei log o degli export tecnici.
- Hash dei ticket, report e comunicazioni rilevanti.
- Riferimento alla classificazione dell'incidente.
- Riferimento alla notifica o relazione inviata all'autorità.
- Hash delle evidenze di remediation.
- Root Merkle del pacchetto complessivo.
- Attestazione T0, T1 e T2 quando disponibili.

## Appendice B · Riferimenti normativi e tecnici essenziali

- Regolamento (UE) 2022/2554, Digital Operational Resilience Act.
- Regolamento delegato (UE) 2025/301, RTS sui contenuti e termini per initial notification, intermediate report e final report degli incidenti ICT maggiori.
- Regolamento di esecuzione (UE) 2024/2956, ITS sui template standard del Registro delle Informazioni DORA.
- Regolamento (UE) 910/2014 e successive modifiche, eIDAS.
- Regolamento (UE) 2024/1689, Artificial Intelligence Act.
- Direttiva (UE) 2022/2555, NIS2.
- RFC 3161, Internet X.509 Public Key Infrastructure Time-Stamp Protocol.
- RFC 6962, Certificate Transparency, per il riferimento alla struttura Merkle e alla domain separation.
- OpenTimestamps, protocollo di timestamping distribuito su blockchain Bitcoin.
- Trusted Lists europee dei prestatori di servizi fiduciari qualificati.

https://certisigma.ch/legal/imprint