I Dati parte 2

Linguaggi di marcatura, Firma Digitale, Blockchain

Giovanni Sartor

Giuseppe Contissa

I linguaggi di marcatura

I linguaggi per la marcatura (*markup language*) consentono di individuare le componenti di un documento testuale racchiudendole

- all'interno di etichette (*tag*) riconoscibili automaticamente, e inserire ulteriori informazioni, dette metadati (*metadata*) all'interno
- del documento, distinguendole rispetto al suo contenuto testuale.

Tipi di markup

Markup proprietari e non proprietari:

- i markup *proprietari* sono definiti e controllati da uno o più produttori di software (anche nel caso che essi non possano vantare diritti di proprietà intellettuale sullo standard);
- i markup *non proprietari* (pubblici) sono invece definiti e controllati da enti imparziali (di solito affiliati a un'organizzazione internazionale che si occupa di standardizzazione).

Markup leggibili e markup non leggibili:

- se il markup è *leggibile*, una persona potrà senza notevole sforzo cogliere le indicazioni di marcatura e individuare le porzioni di testo cui si riferiscono;
- se il markup è *non leggibile*, invece ciò sarà impossibile, o richiederà comunque un notevole sforzo.

Tipi di markup

Markup procedurale e dichiarativo

- il markup *procedurale* consiste nell'inserire nel documento istruzioni per la presentazione del testo (ingrandire i caratteri, allineare il paragrafo a destra o a sinistra, usare il corsivo, ecc.);
- il markup *dichiarativo* (detto anche semantico) consiste nel caratterizzare le porzioni del testo a seconda della loro funzione nel testo stesso, senza specificare come esse debbano essere presentate o elaborate (ad esempio, nell'indicare che una porzione di un testo normativo è una rubrica, un comma, un riferimento normativo, ecc.).

Linguaggi e metalinguaggi

- i *linguaggi di marcatura* specificano le etichette e i comandi da utilizzare nella marcatura di un testo;
- i *metalinguaggi di marcatura* consentono invece al loro utilizzatore di definire il proprio linguaggio di marcatura, cioè, di definire etichette (più in generale, modelli documentali) che corrispondano ai propri bisogni.

HTML e XML

HTML (*HyperText Markup Language* - linguaggio per la marcatura dell'ipertesto) è il linguaggio normalmente utilizzato nella preparazione di pagine web (e quindi nella predisposizione di testi destinati a essere visualizzati in rete). XML (*eXtensible Markup Language* - linguaggio estensibile per la marcatura) è utilizzato per caratterizzare le componenti strutturali e semantiche dei documenti.

- mentre HTML è uno strumento per il markup procedurale, in cui si specifica direttamente la presentazione grafica del testo, XML consente di realizzare markup dichiarativi, nei quali si caratterizza invece la funzione semantica degli elementi testuali;
- mentre HTML è un linguaggio di marcatura (la sua sintassi specifica un determinato insieme di etichette da introdurre nel testo da marcare), XML è un metalinguaggio (la sua sintassi specifica come l'utilizzatore possa creare le proprie etichette, che poi inserirà nei testi da marcare).

Esempio di testo in HTML

<html>

<head>

<title>Legge 22 aprile 1941, n. 633</title>

</head>

<body>

<big>Protezione del diritto d’autore e di altri diritti connessi al suo esercizio</big>

(G.U. n.166 del 16 luglio 1941)</ br>

</ br> TITOLO I</ br>

Disposizioni sul diritto d’autore

CAPO I</ br>

Opere protette

Esempio di testo in HTML

Art. 1

Sono protette ai sensi di questa legge le opere dell’ingegno carattere creativo che appartengono alla letteratura, alla musi alle arti figurative, all’architettura, al teatro e alla cinema qualunque ne sia il modo o la forma di espressione.

Sono altres&iacute protetti i programmi per elaboratore come opere letterarie ai sensi della convenzione di Berna sulla prot delle opere letterarie e artistiche ratificata e resa esecutiva legge 20 giugno 1978, n. 399 , nonch&eacute le banche di dati che per la scelta o la disposizi del materiale costituiscono una creazione intellettuale dell’autore

…

</body> </html>

Esempio di testo in XML

... <intestazione>

<tipoDoc>Legge</tipoDoc>

<DataDoc norm=19410422> 22 aprile 1941</DataDoc>

n. <numDoc>146 </numDoc>

<titoloDoc>Protezione del diritto d’autore e di altri diritti connessi al suo esercizio </titoloDoc>

</intestazione>

...

<articolato>

<articolo id="art1">

<num>Art. 1.</num>

<rubrica>Opere protette</rubrica>

<comma id="art1-com1"> <num>1.</num>

<corpo>Sono protette ai sensi di questa legge le

Esempio di testo in XML

opere dell’ingegno di carattere creativo che appartengono alla letteratura, alla musica, alle arti figurative, all’architettur teatro ed alla cinematografia, qualunque ne sia il modo o la forma di espressione. </corpo>

</comma>

<comma id="art1-com2">

<num>2.</num>

<corpo> Sono altresì protetti i programmi per elaboratore come opere letterarie ai sensi della convenzione di Berna sulla protezione delle opere letterarie e artistiche ratificata e resa esecutiva con legge 20 giugno 1978, n. 399, nonch&eacute le banche di dati che per la scelta o la disposizione del materiale costituiscono una creazione intellettuale dell’autore

</corpo>

</comma>

DTD e XML-Schema

DTD (*document type definition* - definizione del tipo di documento): definizione degli elementi da usare nella marcatura di un testo e del modo in cui possono essere combinati. XML-Schema: nuovo linguaggio, più complesso, per la stessa funzione.

<!ELEMENT Procura (#PCDATA | Parte | ElezioneDomicilio | AutenticazioneFirma | DataAtto)*>

<!ELEMENT ElezioneDomicilio (#PCDATA)>

<!ELEMENT AutenticazioneFirma (#PCDATA | DataAutenticazioneFirma)*>

<!ELEMENT DataAutenticazioneFirma (#PCDATA)>

<!ATTLIST DataAutenticazioneFirma Tipo CDATA #FIXED "date">

DTD per l’elemento procura

<!ELEMENT Parte (#PCDATA | CodiceFiscale | CognomeODenominazione | Nome | Titolo | Residenza | RecapitoTelefonico | LuogoDiNascita | DataDiNascita | Avvocato)*>

DTD per l'elemento procura

```
<Parte>
    <CodiceFiscale>MSSMR1960CODFISC</CodiceFiscale>
     <CognomeODenominazione>Rossi Mario</CognomeODenominazione>
    <Titolo>Dottor</Titolo>
    <Residenza>Bologna</Residenza>
    <RecapitoTelefonico>+39-06-11111111</RecapitoTelefonico>
     <LuogoDiNascita>Milano</LuogoDiNascita>
     <DataDiNascita>1/1/1960</DataDiNascita>
    <Avvocato>
    <CodiceFiscale>BNCLC1955CODFISC</CodiceFiscale>
    <CognomeODenominazione>Bianchi Luisa </CognomeODenominazione>
     </Avvocato>
</Parte>
```

XML e document models

Modelli di documento: struttura comune a tutti i documenti di una certa classe

- Modelli *descrittivi*: ricoprono tutti (o quasi) i testi della classe (e.g. tutte le leggi presenti e passate), così da consentire la marcatura di ciascuno di quei testi.
- modelli *prescrittivi*: indicano requisiti strutturali e sintattici che i testi di una certa classe dovrebbero idealmente possedere (per essere più facilmente ritrovati, compresi, e gestiti, anche con l'aiuto dell'informatica). I modelli prescrittivi possono essere imposto mediante il software per la redazione dei documenti.

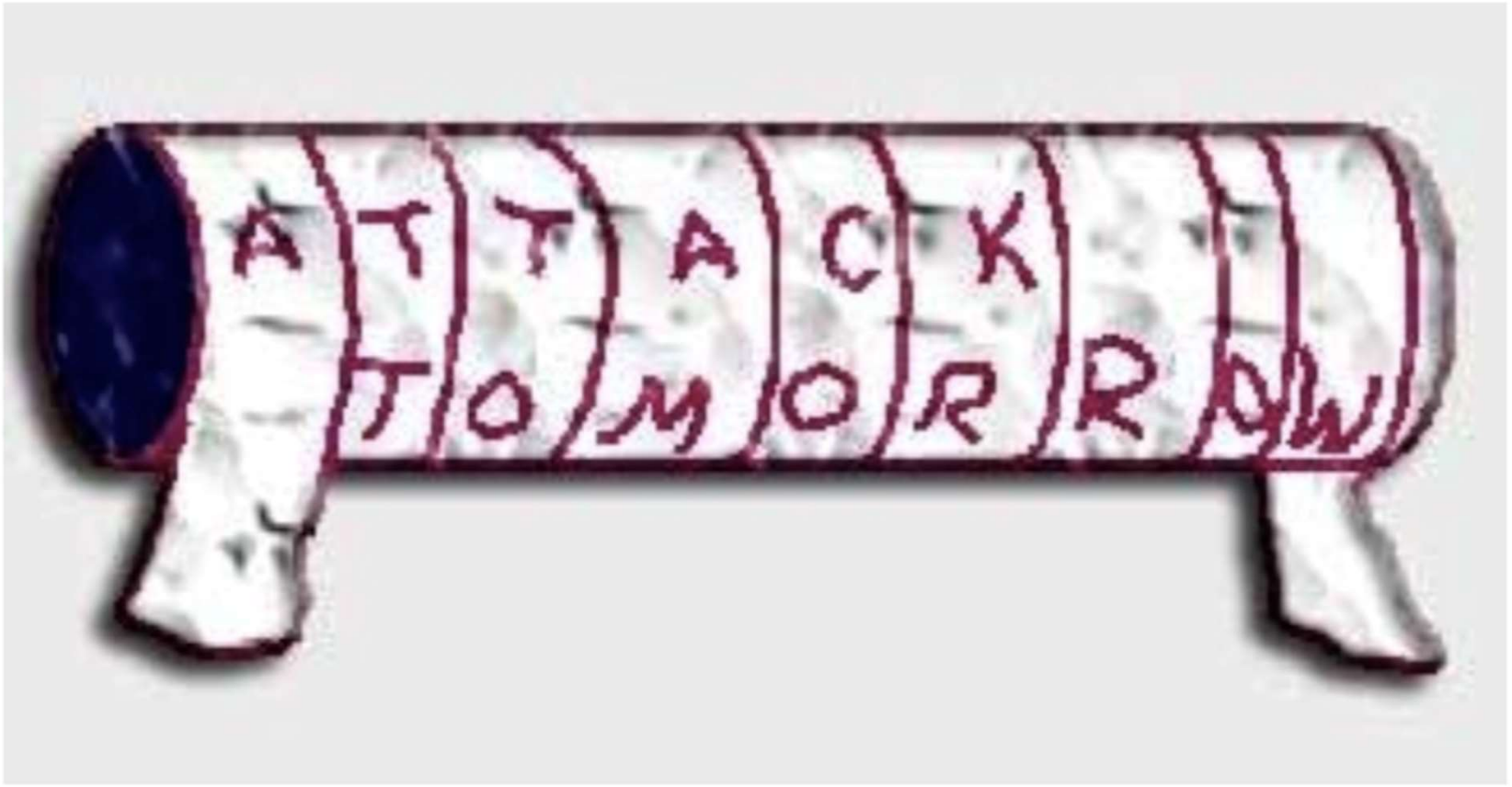
Crittografia

Consente di trasformare un testo leggibile (testo in chiaro) in un testo non più comprensibile al lettore (testo cifrato) e di risalire dal testo cifrato al testo in chiaro:

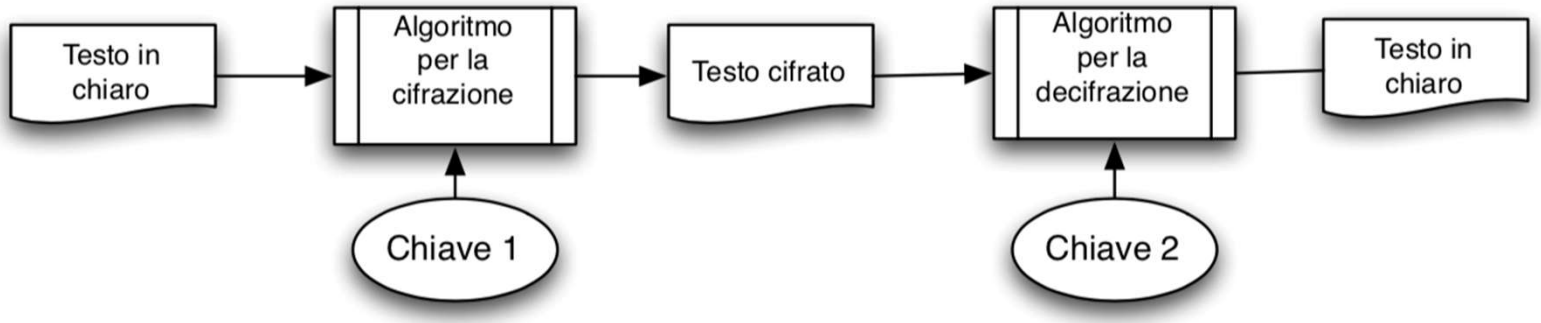
- garantisce la riservatezza (confidenzialità) delle comunicazioni
- consente di verificare l'integrità e autenticità dei messaggi

La crittografia di Cesare

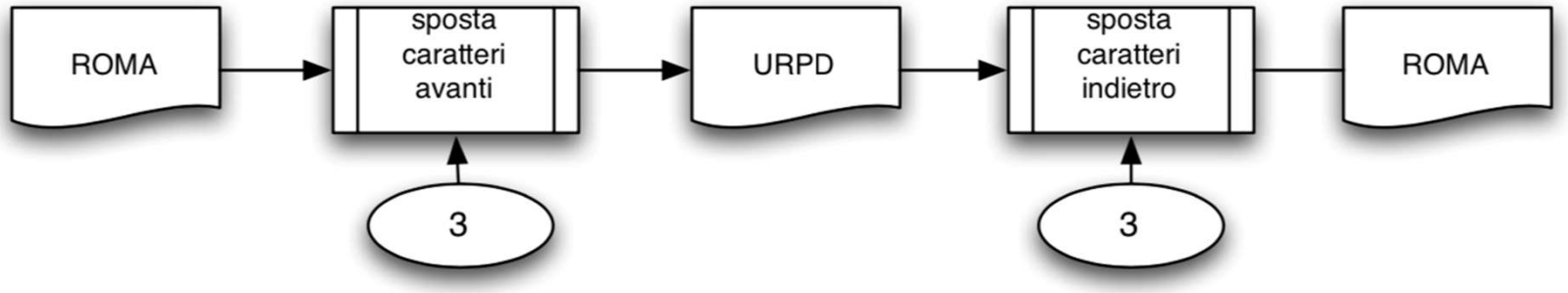
La crittografia degli spartani



La crittografia e l'informatica



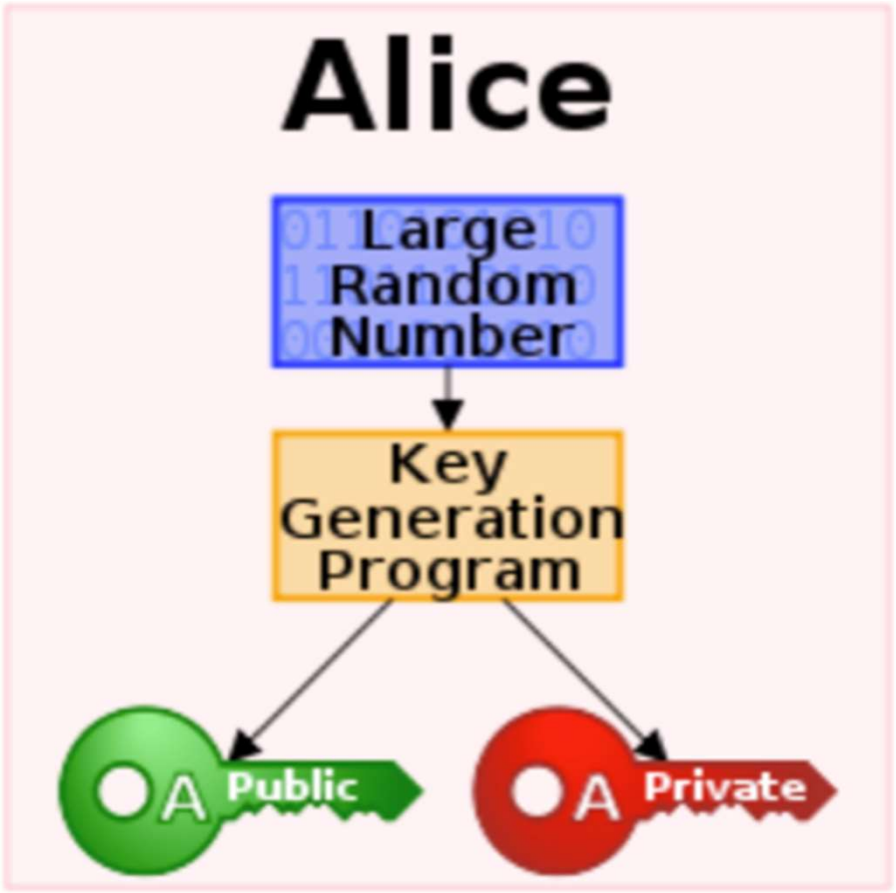
La crittografia di Cesare



Crittografia simmetrica e asimmetrica

- i sistemi a chiave simmetrica, che usano la stessa chiave sia per la cifrazione sia per la decifrazione del messaggio
- i sistemi a chiave asimmetrica, che usano due diverse chiavi, tra loro complementari, per effettuare con una la cifrazione e con l'altra la decifrazione.

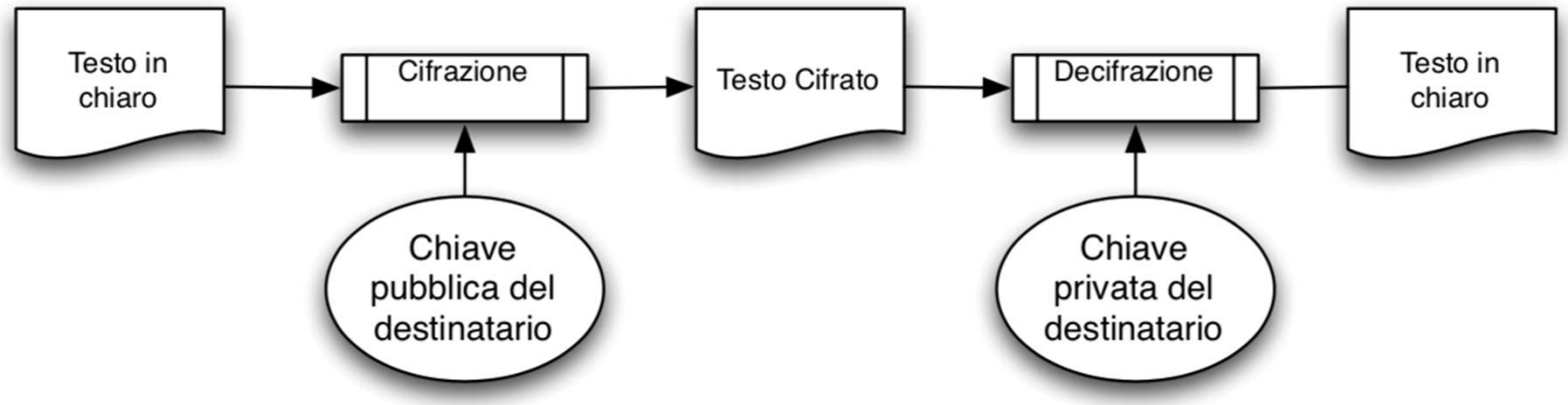
Chiavi asimmetriche



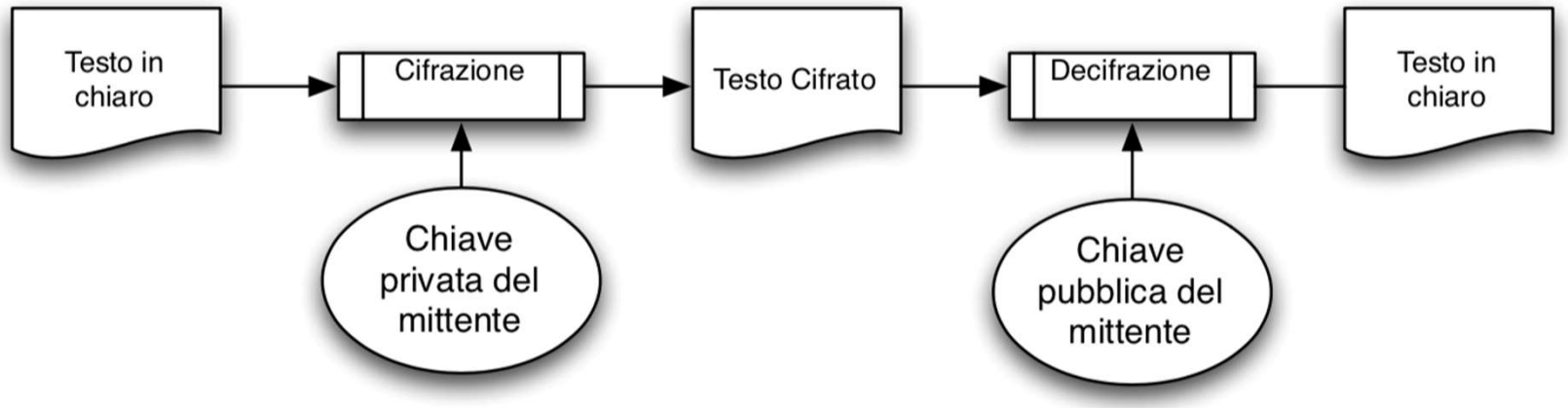
Sistemi a chiave pubblica

- ciascuno dei soggetti che si scambiano informazioni ha la titolarità di una diversa coppia di chiavi
- una chiave della coppia, la *chiave privata* rimane segreta, conosciuta solo al titolare,
- l'altra chiave della coppia la *chiave pubblica* è resa pubblicamente accessibile

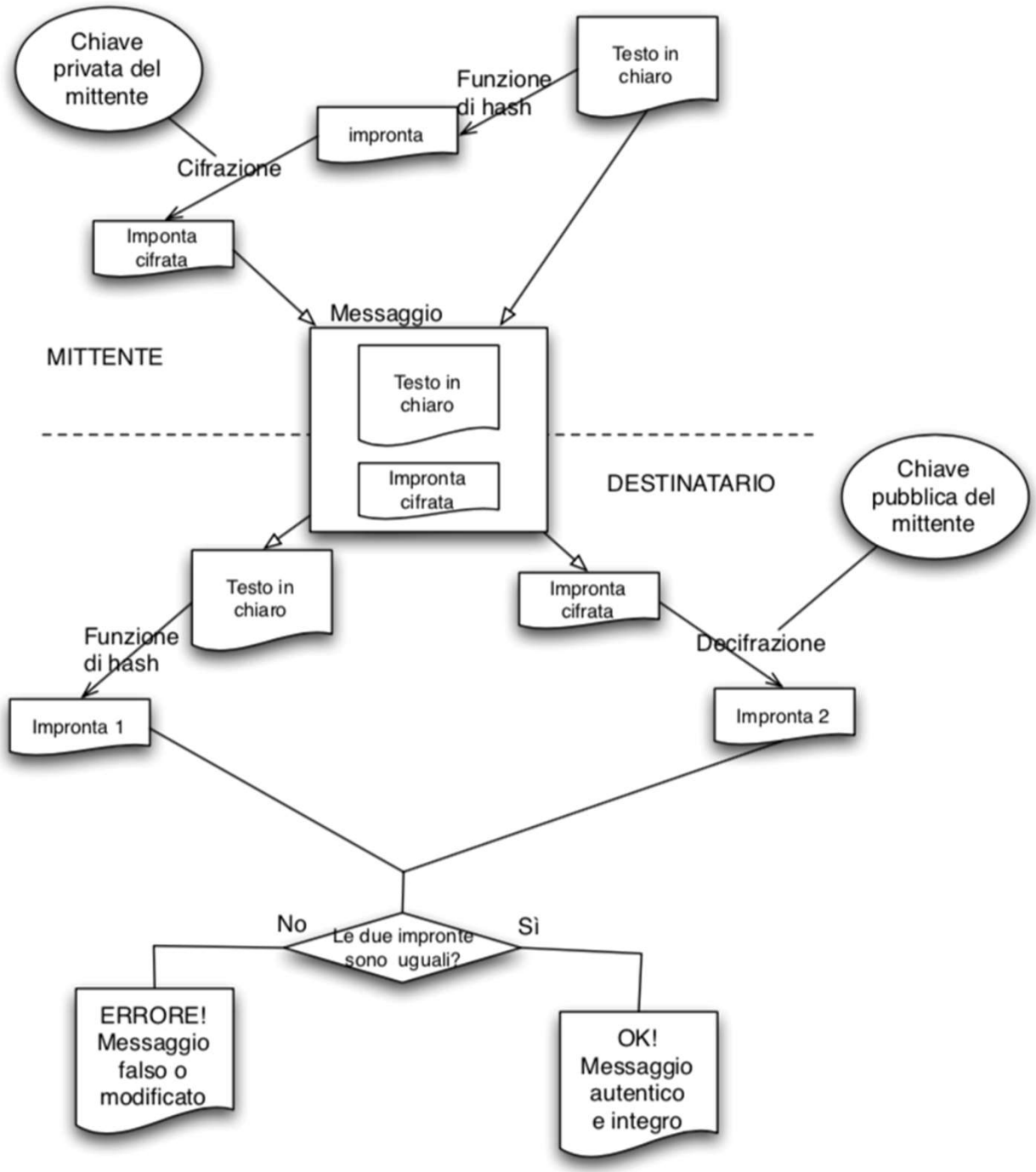
Crittografia per la confidenzialità



Crittografia per l'autenticità/integrità



Firma digitale



La procedura della firma digitale

- Il mittente crea l'impronta del testo in chiaro mediante la funzione di hash, che garantisce l'univocità dell'impronta.
- Il mittente cifra l'impronta usando la propria chiave privata. Ottiene così la cosiddetta firma digitale del testo in chiaro, che viene inviata al destinatario assieme al testo in chiaro.
- Il mittente spedisce testo e impronta.
- Il destinatario applica al testo in chiaro ricevuto la funzione di hash estraendone una nuova impronta.
- Egli decifra l'impronta cifrata ricevuta con il messaggio, mediante la chiave pubblica del mittente.
- Infine, il destinatario confronta l'impronta estratta dal testo in chiaro con quella ottenuta attraverso la decifrazione.
- Se le due impronte sono identiche, ciò significa che il messaggio è stato codificato con la chiave privata del mittente, e non è stato modificato da alcuno. Sono quindi garantite la sua autenticità e integrità

Regolamento eIDAS

Electronic Identification Authentication and Signature, efficace dal 1 Luglio 2016

- ribadisce la distinzione tra firma elettronica semplice, qualificata e avanzata, e la corrispondente efficacia giuridica,
- Alle firme elettroniche “non possono essere negati gli effetti giuridici e l’ammissibilità come prova in procedimenti giudiziali per il solo motivo della sua forma elettronica o perché non soddisfa i requisiti per firme elettroniche qualificate” (Art. 25)
- Riconoscimento reciproco: “una firma elettronica qualificata basata su un certificato qualificato rilasciato in uno Stato membro è riconosciuta quale firma elettronica qualificata in tutti gli altri Stati membri”.

Firme elettroniche - Regolamento EIDAS I

Articolo 3
Definizioni
10) «firma elettronica», dati in forma elettronica, acclusi oppure connessi tramite associazione logica ad altri dati elettronici e utilizzati dal firmatario per firmare;
11) «firma elettronica avanzata», una firma elettronica che soddisfi i requisiti di cui all'articolo 26;
12) «firma elettronica qualificata», una firma elettronica avanzata creata da un dispositivo per la creazione di una firma elettronica qualificata e basata su un certificato qualificato per firme elettroniche;
14) «certificato di firma elettronica», un attestato elettronico che collega i dati di convalida di una firma elettronica a una persona fisica e conferma almeno il nome o lo pseudonimo di tale persona;

Firme elettroniche - Regolamento EIDAS II

15) «certificato qualificato di firma elettronica», un certificato di firma elettronica che è rilasciato da un prestatore di servizi fiduciari qualificato ed è conforme ai requisiti di cui all'allegato I;

Articolo 25 Effetti giuridici delle firme elettroniche

- 1. A una firma elettronica non possono essere negati gli effetti giuridici e l'ammissibilità come prova in procedimenti giudiziali per il solo motivo della sua forma elettronica o perché non soddisfa i requisiti per firme elettroniche qualificate.
- 2. Una firma elettronica qualificata ha effetti giuridici equivalenti a quelli di una firma autografa.
- 3. Una firma elettronica qualificata basata su un certificato qualificato rilasciato in uno Stato membro è riconosciuta quale firma elettronica qualificata in tutti gli altri Stati membri.

Firme elettroniche - Regolamento EIDAS III

Articolo 26
Requisiti di una firma elettronica avanzata
Una firma elettronica avanzata soddisfa i seguenti requisiti:
a) è connessa unicamente al firmatario;
b) è idonea a identificare il firmatario;
c) è creata mediante dati per la creazione di una firma elettronica che il firmatario può, con un elevato livello di sicurezza, utilizzare sotto il proprio esclusivo controllo; e
d) è collegata ai dati sottoscritti in modo da consentire l'identificazione di ogni successiva modifica di tali dati.

Art. 21CAD - Documento informatico sottoscritto con firma elettronica. I

- 2-bis). Salvo il caso di sottoscrizione autenticata, le scritture private di cui all'articolo 1350, primo comma, numeri da 1 a 12, del codice civile, se fatte con documento informatico, sono sottoscritte, a pena di nullità, con firma elettronica qualificata o con firma digitale. Gli atti di cui all'articolo 1350, numero 13), del codice civile redatti su documento informatico o formati attraverso procedimenti informatici sono sottoscritti, a pena di nullità, con firma elettronica avanzata, qualificata o digitale ovvero sono formati con le ulteriori modalità di cui all'articolo 20, comma 1-bis, primo periodo.
- 2-ter. Fatto salvo quanto previsto dal decreto legislativo 2 luglio 2010, n. 110, ogni altro atto pubblico redatto su documento informatico è sottoscritto dal pubblico ufficiale a pena di nullità con firma qualificata o digitale. Le parti, i fidefacenti, l'interprete e i testimoni sottoscrivono personalmente l'atto, in presenza del pubblico ufficiale, con firma avanzata, qualificata o digitale ovvero con firma autografa acquisita digitalmente e allegata agli atti.

Art. 24 CAD. Firme digitali

1. La firma digitale deve riferirsi in maniera univoca ad un solo soggetto ed al documento o all'insieme di documenti cui è apposta o associata.

2. L'apposizione di firma digitale integra e sostituisce l'apposizione di sigilli, punzoni, timbri, contrassegni e marchi di qualsiasi genere ad ogni fine previsto dalla normativa vigente.

3. Per la generazione della firma digitale deve adoperarsi un certificato qualificato che, al momento della sottoscrizione, non risulti scaduto di validità ovvero non risulti revocato o sospeso

I bitcoin

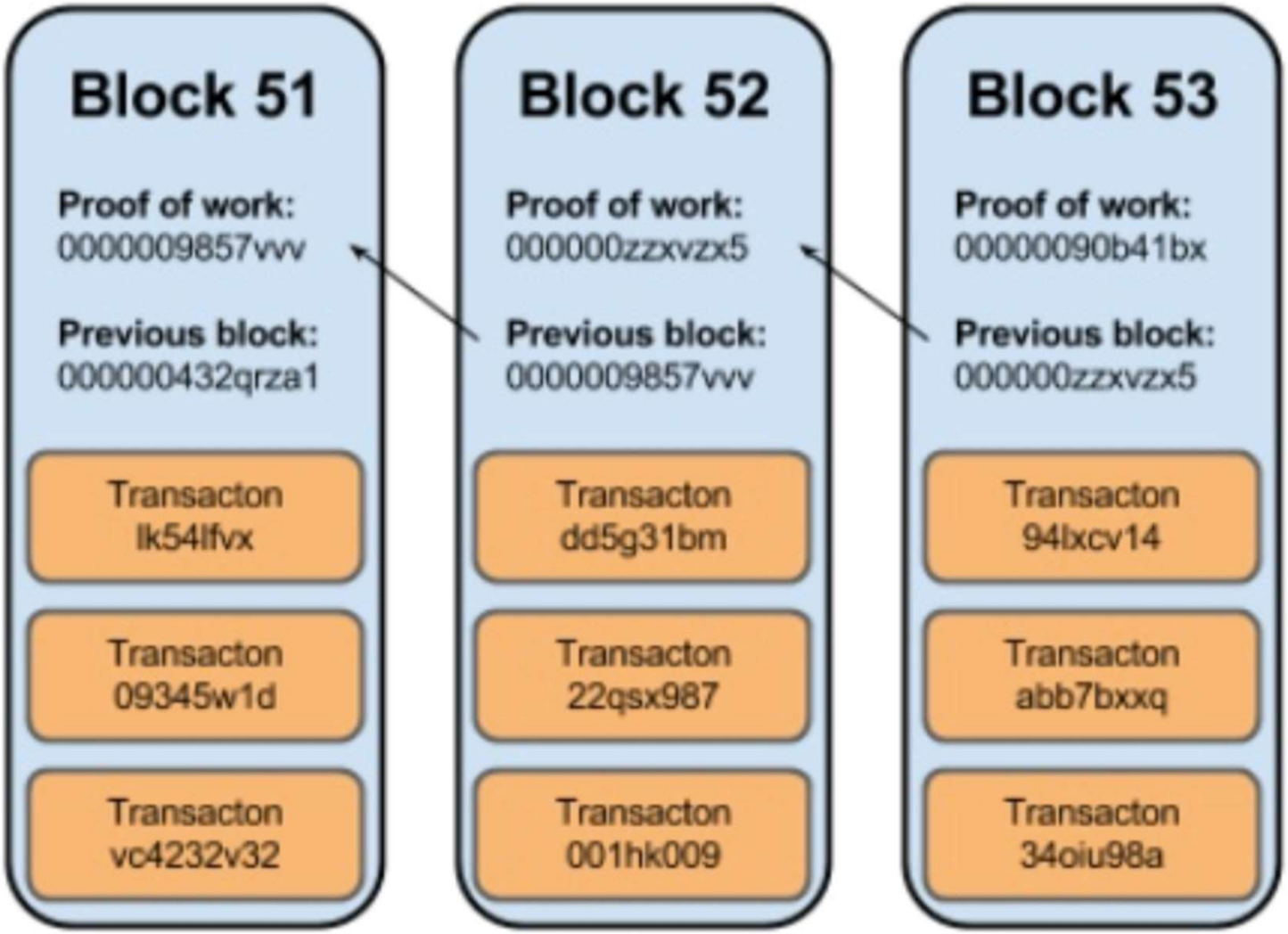
Che cosa sono

- moneta elettronica creata nel 2009, che ha conosciuto una rapida diffusione. Oggi è accettata da più di 100.000 venditori.

Si basa sulla blockchain (catena di blocchi):

- Registro pubblico condiviso di cui esiste una copia in ciascun nodo della rete peer to peer che unisce gli utilizzatori di bitcoin.

Blockchain


Block 51
Proof of work:
0000009857vvv
Previous block:
000000432qrza1
Transacton
lk54lfvx
Transacton
09345w1d
Transacton
vc4232v32
Block 52
Proof of work:
000000zzxvzx5
Previous block:
0000009857vvv
Transacton
dd5g31bm
Transacton
22qsx987
Transacton
001hk009
Block 53
Proof of work:
00000090b41bx
Previous block:
000000zzxvzx5
Transacton
94lxcv14
Transacton
abb7bxxq
Transacton
34oiu98a

Come funzionano

Come si avvia una transazione

- Ogni soggetto è identificato da una coppia di chiavi
- Chi intende trasferire bitcoin predispone una transazione (un ordine di trasferimento) e la diffonde a tutti i nodi della rete

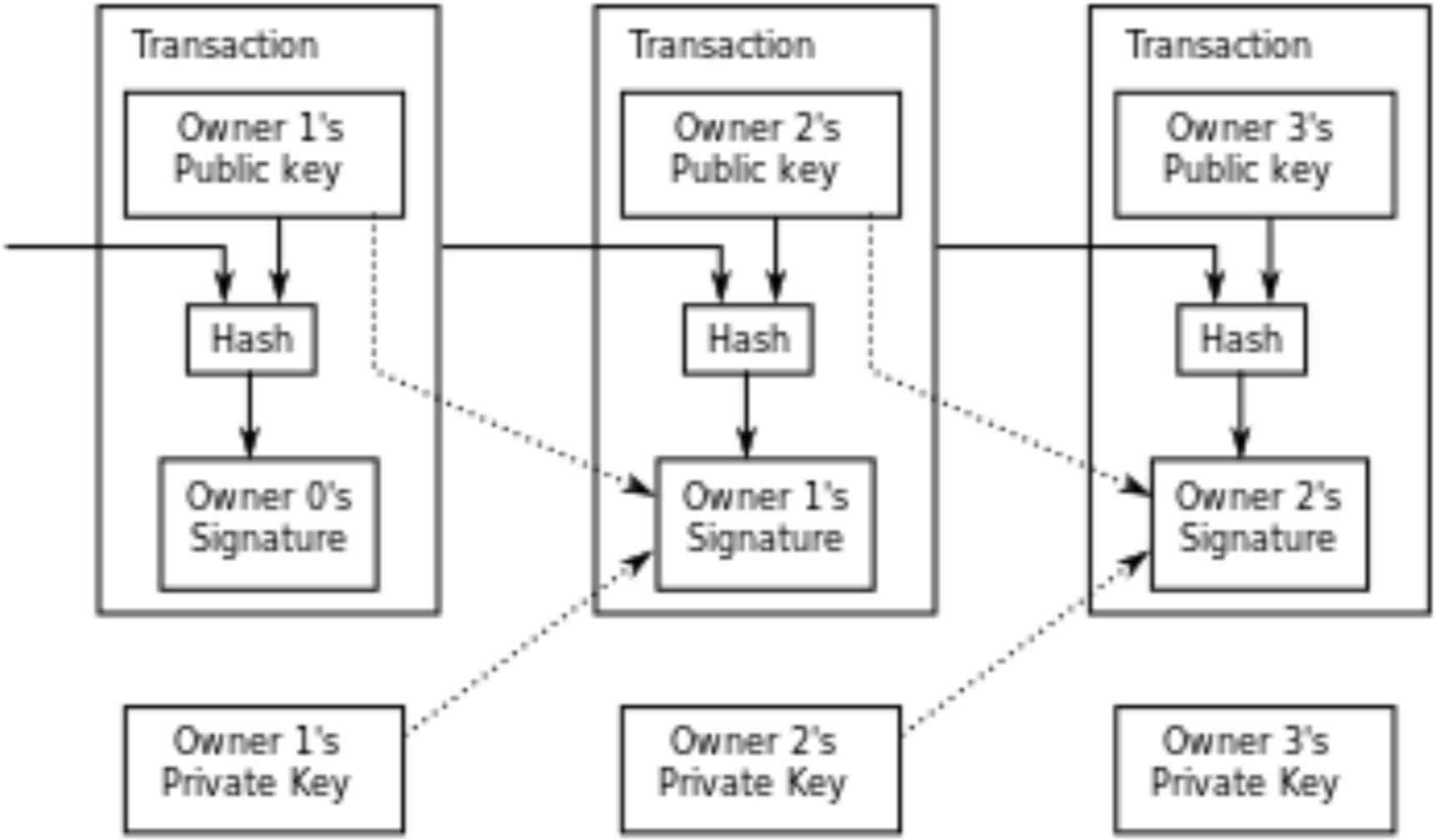
Come viene validata

- Un nodo valida la transazione (verifica identità e disponibilità dei fondi)
- Predispone gruppo di transazioni raccogliendole in un blocco
- Aggiunge al blocco un numero, sulla base del quale si può generare un'impronta del blocco con un certo numero di 0 iniziali (mining)
- Il nuovo blocco è validato dalla rete (accettato dagli altri nodi)

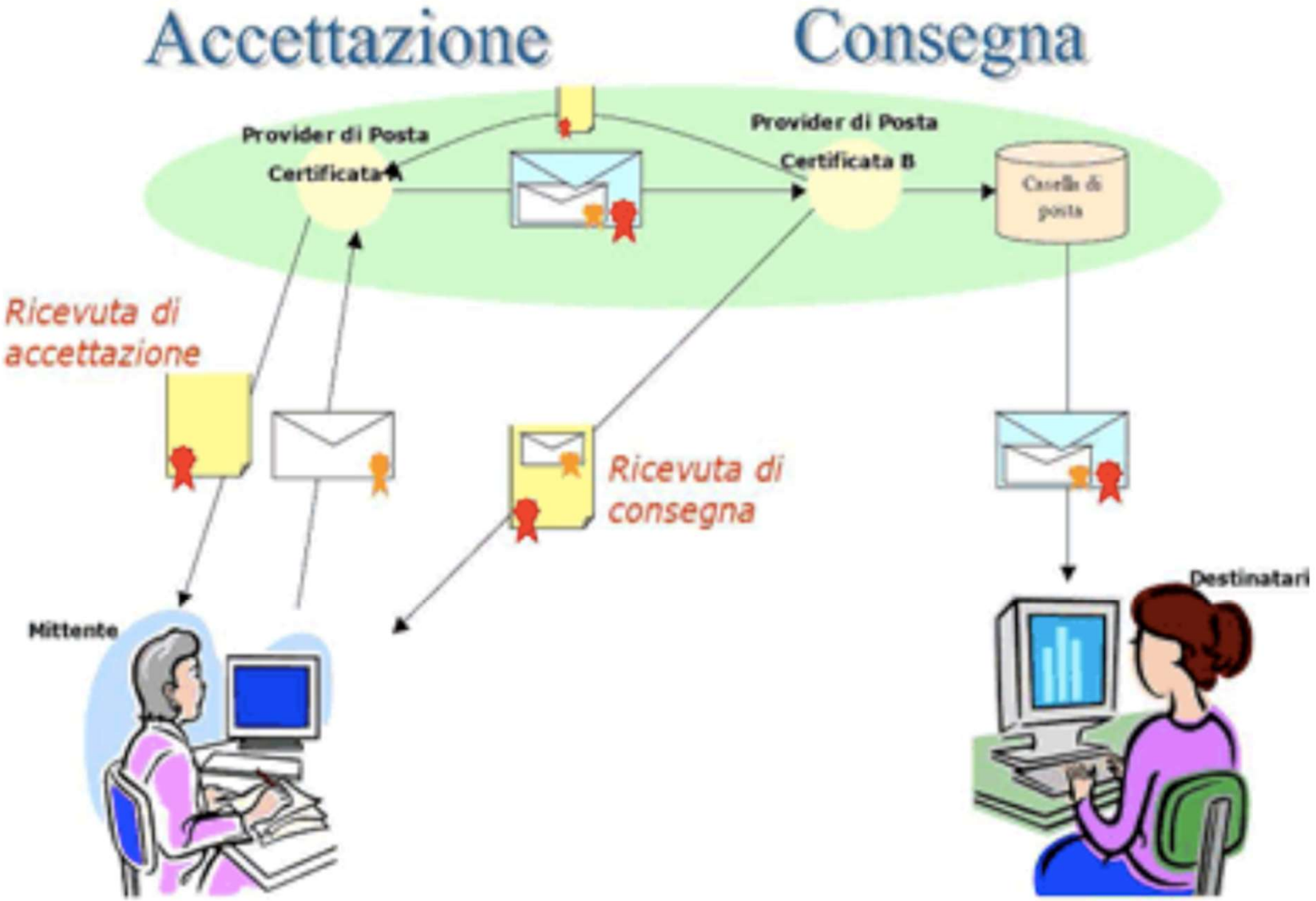
Usi di bitcoin e blockchain

- Scambi legali
- Scambi illegali nel darknet (Silkroad)
- Smart contracts (contratti che si autoeseguono)

Validazione di transazione



Posta certificata



Verso i Big Data

Quanto grandi?

- I dati analogici erano già nel 2013 solo il 7% dei dati totali, il 93% consistendo di dati digitali.
- Da circa uno zettabyte di dati digitalizzati disponibili nel 2009, siamo 8 zettabyte circa nel 2015, e passeremo a 35 nel 2020

Perchè?

- i dati sono sempre più raccolti automaticamente da ambienti virtuali e fisici
- il costo della memorizzazione dei dati è sceso rapidamente (più di 20 volte nel corso degli ultimi 15 anni
- strumenti di data analytics aumentano gli usi dei dati.

Che cos'è il fenomeno dei big data

il termine *big data* "fa riferimento a cose che si possono fare a una larga scala che non si possono fare ad una scala più piccola, per ottenere nuove comprensioni, o creare nuove forme di valore, in modi che cambiano i mercati, le organizzazioni e il rapporto tra cittadini e governi, e altro (Mayer-Schoberger e Cukier, 2013).

Data analytics

Data analytics: estrarre correlazioni probabilistiche inattese da grandi masse di dati

- tra dinamiche dei mercati e successive evoluzioni di domanda, offerta, e prezzi;
- tra la scelta di un certi “amici” e l’interesse a stabilire contatti con altre persone;
- tra curricula e capacità lavorative, tra caratteristiche personali e affidabilità creditizia;
- tra presenti situazioni del traffico e futuri ingorghi stradali;
- tra la frequenza di certe interrogazioni su Internet e la diffusione di epidemie o di tensioni politiche;
- tra gli stili di vita delle persone e certe loro patologie presenti o future;
- tra certi indizi, e futuri guasti o rotture di apparecchi e infrastrutture, ecc.

Big data: Opportunità e rischi

Opportunità:

- nuove prospettive alla ricerca scientifica, e
- scelte individuali, economiche, amministrative, e politiche sulla base di nuove e più ricche informazioni.

Rischi:

- Privacy: Tutti i dati diventano potenzialmente utili, oggi o in futuro (i dati come risorsa). Contrasto con i principi di finalità e minimizzazione.
- Potere e monopolio: Solo chi possiede i big data può sfruttare le tecnologie della data analytics