

*Enrico Panai\**

## **Responsabile dell'etica dei dati Un ruolo da costruire**

### **Introduzione**

Raccogliere, conservare o trattare dati può avere effetti sulle singole persone, su gruppi sociali o sull'intera società. Ciò implica che le decisioni prese su alcuni dati possono avere un valore etico. In generale molte organizzazioni (intese come entità che hanno uno scopo definito, quindi associazioni, società o istituzioni) si confrontano con il ragionamento etico quando manipolano dati. Esiste quindi *de facto* uno spazio dell'etica dei dati dove già si prendono decisioni. Tuttavia se all'interno di una qualche organizzazione questo spazio non è evidente, c'è il rischio di prendere decisioni (eticamente) non adeguate. Questo spazio è lì da millenni. Tuttavia, con lo sviluppo delle tecnologie dell'informazione e della comunicazione digitale (che hanno razionalizzato la raccolta, la conservazione e il trattamento dei dati) e con la recente diffusione dei sistemi di intelligenza artificiale (che hanno potenziato la capacità di classificazione e di inferenza delle macchine) l'impatto di una cattiva scelta o di una non-scelta è diventato più apprezzabile. Senza le dovute precauzioni, le conseguenze etiche sono visibili solo quando il danno è stato fatto.

### **Livello di astrazione informazionale**

La scrittura è stata la prima tecnologia dell'informazione e della comunicazione. Tecnologia dell'informazione perché permetteva di raccogliere su un supporto fisico (tavolette d'argilla, papiri, pergamene, libri) una serie di dati organizzati in maniera tale da fornire una informazione. Tecnologia della comunicazione perché in pochi segni permetteva di conservare per anni e diffondere geograficamente le conoscenze raccolte. Una tavoletta d'argilla poteva essere trasportata o custodita per trasmettere l'informazio-

\* Università di Sassari. Fellow ForHumanity

ne che racchiudeva. I primi supporti non trasmettevano poesie o racconti, ma informazioni organizzative: procedure religiose, gestione delle risorse umane, contabilità, ricette, ecc. Sono esistiti nella storia altri sistemi per conservare o trasmettere le informazioni senza la scrittura, come i *quipu*, le cordicelle annodate degli Inca, o le “mappe a bastoncini” delle isole Marshall<sup>1</sup>. Del resto alcune informazioni non si sarebbero potute raccogliere con la sola scrittura. Mentre le cordicelle degli Inca venivano usate per conservare dati strutturati (censo, transazioni commerciali e tasse), le mappe a bastoncini erano utilizzate come carte geografiche dinamiche per la navigazione nel pacifico: le mappe *Mattang* erano usate solo per insegnare come le onde oceaniche si infrangono sulle isole. In pratica sembra che l’informazione e la sua gestione siano state uno degli elementi fondamentali per progredire in tutte le parti del mondo. La scrittura in particolare è stata la tecnologia vincente tra le altre concorrenti.

Che fossero annodate, scritte o trasmesse con bastoncini e conchiglie, le informazioni hanno sempre avuto una caratteristica: quella di essere indissolubilmente associate al loro supporto fisico. L’informazione è composta da dati (GDI 1); i dati devono essere ben strutturati in accordo con la sintassi che viene utilizzata nel sistema scelto (GDI 2) e devono rispettare il significato dello stesso sistema (GDI 3), come suggerito dalla definizione generale dell’informazione o *General Definition of Information*<sup>2</sup>. In altre parole questo significa che se uno di questi tre principi non è rispettato, allora non abbiamo informazione. Per esempio la frase “idee verdi senza colore dormono furiosamente”<sup>3</sup> che è composta da dati (“idee verdi” “senza colore” “dormire”) e rispetta una sintassi (soggetto e verbo) è priva di significato, quindi non trasmette alcuna informazione<sup>4</sup>.

La definizione generale dell’informazione è fondamentale oggi, perché i dati sono diventati la risorsa fondamentale per la creazione dell’informazione. Infatti, prima dell’era informatica i dati che componevano le informazioni erano invisibili perché fusi nella loro struttura logica, nella loro sintassi, nella distribuzione sul supporto. Con l’avvento dell’informatica è avvenuta una separazione tra dato, supporto e struttura che ha messo in rilievo l’importanza del dato. In pratica l’informazione è stata decomposta nei suoi elementi atomici (i dati) e viene ricreata al volo (*on*

<sup>1</sup> Marshall Islands stick chart.

<sup>2</sup> Cfr. L. Floridi, *Information. A Very Short Introduction*, Oxford University Press, Oxford 2010.

<sup>3</sup> La frase “Colorless green ideas sleep furiously” è stata proposta da Chomsky come esempio di frase grammaticalmente corretta ma semanticamente assurda. Cfr. N. Chomsky, *The Logical Structure of Linguistic Theory*, Springer US, Cambridge/Mass 1975.

<sup>4</sup> In realtà da un punto di vista informazionale, questa frase dà delle informazioni, ma queste hanno bisogno di un’interpretazione che può cambiare secondo il livello di astrazione utilizzato.

*the fly*) organizzando i dati secondo una determinata struttura logica ogni qualvolta ne abbiamo bisogno. Una volta rigenerate, le informazioni appaiono sulle interfacce grafiche (lo schermo di un computer o di un cellulare), audio (gli assistenti virtuali come Alexa, Siri, ecc.), tattili (come nelle tastiere Braille), ecc.

Con questo divorzio dei dati dal loro supporto fisico si è aperta una nuova era, l'era informazionale (anche se tecnicamente sarebbe più corretto chiamarla "era datazionale", ma avrebbe un fascino minore). Le tecnologie dell'informazione e della comunicazione (ICT) di oggi, cioè quelle digitali, hanno un impatto diverso dalla scrittura, proprio perché la loro attività di trasformazione dei dati in informazione è costante e dinamica, e non statica come sulle tavolette d'argilla. Inoltre elaborano una quantità di dati che non sarebbe possibile trattare su supporto fisico, basti pensare che ogni giorno si generano più dati di quelli prodotti dall'invenzione della scrittura al 2005<sup>5</sup>. Per districarsi in questo tsunami di dati abbiamo bisogno di risorse computazionali e statistiche avanzate che ci aiutino a prendere decisioni o che permettano ai sistemi automatici di fare scelte; scelte e decisioni che hanno necessariamente un impatto etico nell'universo informazionale nel quale viviamo. Quindi utilizzando il livello di astrazione informazionale diventa evidente la necessità di utilizzare un'etica adatta al sistema.

Scegliere il buon punto di partenza o il giusto livello di astrazione per comprendere un fenomeno è il passo fondamentale per proporre delle soluzioni. In breve, il livello di astrazione "informazionale" e il livello di granularità "organizzazionale" sono i perimetri di riflessione di questo articolo. Ora vedremo come partendo da un livello d'astrazione informazionale lo spazio dell'etica dei dati diventi evidente.

## Lo spazio latente dell'etica dei dati

Nelle organizzazioni molte decisioni etiche sui dati vengono già prese. Esistono. Sono sistema. Quello che manca è il modello etico da utilizzare. Questo spazio dell'etica dei dati esiste, ma è latente: latente perché gli manca una teoria.

Una barca a vela che naviga è reale, o almeno la assumiamo tale nel pensare comune. È reale perché esiste il "modello" della barca (il progetto) che identifica delle "proprietà" (idrodinamicità dello scafo,

<sup>5</sup> I calcoli sui dati sono approssimativi, ma alcuni calcoli possono essere fatti a partire da alcune fonti. Vedi ad esempio: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>.

altezza dell'albero, taglio e dimensione delle vele, tensione e spessore delle sartie, tipo di deriva) che vengono attribuite ad un oggetto finale, al "sistema" (proprio perché le proprietà sono sistemate in maniera tale che la barca possa navigare). La barca (l'oggetto, il sistema) incarna le informazioni di modello e proprietà in un ordine fisico preciso. In questo senso l'informazione può essere considerata non come un concetto effimero e impalpabile, ma come "l'ordine fisico incarnato negli oggetti"<sup>6</sup>. Una barca a vela che naviga è quindi "reale" quando un modello identifica delle proprietà che vengono attribuite ad un sistema.

Senza il sistema non siamo più in presenza del "reale", ma del "virtuale". Abbiamo solo il modello e le sue proprietà. Ecco perché un software di simulazione di vela è virtuale: c'è il modello con le sue proprietà, ma non c'è il sistema. Ora disaccoppiare il modello dal sistema evidenzia anche il caso opposto, in cui il sistema esiste, ma non il suo modello. Anche in questo caso non siamo in presenza del "reale", ma nemmeno del "virtuale". Siamo invece in presenza del "latente", "nel senso originale latino di nascosto o sconosciuto"<sup>7</sup>. È quello che avviene nella archeologia quando ci si imbatte in una costruzione di cui non si ha il modello. Dai *nuraghi* sardi nel mediterraneo ai *moai* dell'isola di Pasqua, esistono sistemi di cui non conosciamo un modello. Sistemi che esistono, ma con un'essenza onirica, non pienamente reali: "un oggetto senza la propria ombra, come Peter Pan"<sup>8</sup>.

In ogni organizzazione i dati esistono e sono la base del sistema produttivo. I dati vengono quindi raccolti, conservati, elaborati e trasmessi. Sarebbe ingenuo pensare che questo possa avvenire in maniera neutrale. Ogni scelta viene fatta rispettando un quadro etico, anche se questo non viene esplicitamente dichiarato. Se per esempio un'azienda che vende libri online raccoglie i dati etnografici per orientare le vendite a categorie di persone, sta implicitamente rafforzando la divisione culturale tra etnie diverse. Mentre una istituzione pubblica può raccogliere esattamente gli stessi dati per tentare di mitigare le ingiustizie razziali. L'etica dei dati è presente ma latente. Il sistema esiste ma il modello (etico) non è stato dichiarato; perché spesso non è chiaro.

Ci si domanda allora per quale ragione l'etica dei dati non sia stata esplicitamente introdotta nel dibattito pubblico. Perché esistano esper-

<sup>6</sup> C. Hidalgo, *Why Information Grows: the Evolution of Order, from Atoms to Economies*, Basic Books, New York 2015, p. 7.

<sup>7</sup> L. Floridi, *The Latent Nature of Global Information Warfare*, in "Philosophy & Technology", XXVII, 3, 2014, p. 318.

<sup>8</sup> *Ibidem*.

ti etici dell'AI, ma non dei dati. Non perché essa non esista o non sia stata concettualizzata, ma perché non abbiamo ancora indossato gli occhiali giusti per leggere la nostra era. Quindi ricerchiamo esperti etici della tecnica che utilizza i dati (AI, Robot, Computer, etc.) e non dei dati. Tecnicamente non abbiamo ancora accettato di utilizzare un livello di astrazione informazionale introdotto nel capitolo precedente. Il “metodo dei livelli di astrazione” è già stato usato in matematica (dov'è strettamente correlato alla teoria dei modelli), in informatica (principalmente nella programmazione ad oggetti o OOP, *Object Oriented Programming*) e in filosofia (per un cambiamento di prospettiva ontologica nell'etica dell'informazione). Una volta selezionato il “livello di astrazione” di un sistema complesso che è rilevante per i nostri fini possiamo creare un “modello”. I livelli di astrazione sono necessari nella progettazione dei sistemi (*system design*) e possono essere utili per esporre una teoria perché “il modello di un sistema è una funzione degli osservabili disponibili”<sup>9</sup>. In parole povere, utilizzare un preciso livello di astrazione significa assumere un punto di vista e sviluppare il proprio processo decisionale all'interno di quella prospettiva. Nel nostro caso il livello di astrazione informazionale mette in evidenza l'esistenza di uno spazio etico dei dati. Questo è essenziale per avere una corretta discussione su un argomento perché “sembra che molti disaccordi poco interessanti potrebbero essere chiariti, se i vari interlocutori potessero rendere esplicito e preciso il loro Livello di Astrazione”<sup>10</sup>. Una volta scelto il livello di astrazione, si deve scegliere il livello di dettaglio sul quale discutere, cioè il livello di granularità. Più è alto il livello di astrazione e meno saranno i dettagli. In questo caso, come vedremo, il livello di astrazione è quello “informazionale” e il livello di granularità è quello della divisione dei ruoli e delle responsabilità all'interno delle organizzazioni. Le sezioni successive sono il frutto del lavoro portato avanti negli ultimi due anni dall'autore nel gruppo di lavoro internazionale ForHumanity<sup>11</sup> per la creazione di Audit Indipendenti per i Sistemi di Intelligenza Artificiale.

## Cartografia informazionale delle organizzazioni

In questi ultimi anni le scelte etiche relative ai dati e ai sistemi autonomi di supporto alla decisione<sup>12</sup> generalmente chiamati sistemi di intel-

<sup>9</sup> L. Floridi, *The Ethics of Information*, Oxford University Press, Oxford 2013, p. 33.

<sup>10</sup> Ivi, p. 32.

<sup>11</sup> Per maggiori informazioni su IAAIS consultare il sito: <https://forhumanity.center>.

<sup>12</sup> ADMS – Automated Decision Making System.

ligenza artificiale (AI) sono state spesso<sup>13</sup> “delegate” ad un responsabile etico o Chief Ethics Officer (CEtO); oppure sono state “delegate” al responsabile del trattamento dati, il Chief Data Officer (CDO) per verificare la qualità del ciclo di vita dei dati e garantire la conformità con il quadro giuridico rilevante per l’azienda. La “delega etica” è stata particolarmente evidente all’interno dell’Unione Europea grazie alla comparsa del principio di “data subject”, cioè di individuo informazionale che deve essere protetto al di là dei confini geografici. Nel vecchio continente il ruolo del CDO è rapidamente stato sostituito da quello del Data Protection Officer (DPO) dedicato a garantire la conformità con la regolamentazione europea sulla protezione dei dati personali (GDPR – *General Data Protection Regulation*). Ad essere precisi nei casi precedenti non si è trattato di “delega” attraverso la quale affidare la responsabilità a una persona, perché il concetto di “responsabilità etica dei dati” non è stato definito precisamente. Si è trattato piuttosto di “colonizzazione” di un territorio di cui nessuno rivendicava la proprietà. Questa indeterminatezza ha permesso a ciascuno di imporre il proprio livello di astrazione secondo le proprie forze, capacità o risorse. Molte decisioni etiche concernenti i dati sono state prese da coloro che lavorano materialmente con i dati: sviluppatori, *data scientists* o in generale *data workers*.

*Data workers* erano sul terreno e hanno risolto problemi etici, senza dovere (o potere) consultare i livelli più alti della gerarchia: quale *dataset* (collezione di dati) scegliere; dove raccogliere e come conservare i dati; come mitigare gli errori (pregiudizi cognitivi, sociali e statistici o in inglese *bias*) insiti nel *dataset*; quale algoritmo scegliere per il trattamento e via dicendo. Insomma, i *data workers* hanno ed hanno avuto nelle aziende un’influenza determinante sulle scelte etiche relative ai dati.

Tuttavia, mentre il responsabile etico lavora su un livello di granularità grezzo che riguarda i rapporti esterni dell’azienda con la società e le relazioni interne con i propri collaboratori, il delegato per la protezione dei dati personali cambia completamente livello di astrazione, dovendosi occupare principalmente degli aspetti legali riferiti ai dati e/o ai dati personali; infine ai *data workers* viene chiesto di essere performanti, mettendo in secondo piano la riflessione etica. Quindi i primi basano le loro conclusioni sulla etica degli affari (*business ethics*), i secondi prendono decisioni in conformità con legislazioni locali, i terzi si focalizzano sulle prestazioni del sistema che costruiscono.

Ne risulta che le decisioni etiche sui dati e sugli algoritmi di AI sono al di fuori di un quadro etico definito. Per usare una metafora medica,

<sup>13</sup> Nel caso si avesse coscienza dell’impatto etico.

quando si tratta di decisioni etiche sui dati, nelle aziende abbiamo dermatologi che si improvvisano dentisti per sostituire un molare. Il rischio maggiore è creare una giungla etica; un ambiente nel quale prevalgono le proprie convinzioni personali o politiche. In pratica: chi ha più competenze tecniche o più potere interno impone la propria visione agli altri. Per evitare questo rischio e fare scelte etiche sulle tecnologie digitali, sui dati e infine sui sistemi di AI c'è la necessità di riorganizzare il dibattito all'interno di un quadro etico definito e indicare esplicitamente i livelli di responsabilità.

A questo proposito propongo una mappa delle competenze necessarie all'interno di una azienda per la realizzazione di una infrastruttura etica interna che permetta il fiorire di scelte etiche adatte al contesto. Questa analisi fornisce alcuni spunti per le ricerche future che indaghino sull'integrazione dell'etica dei dati in culture non occidentali.

I seguenti diagrammi servono per definire lo spazio decisionale di un'organizzazione "spostando il livello di astrazione delle indagini etiche, dall'essere centrate sulle informazioni all'essere centrate sui dati"<sup>14</sup>. La mappatura della zona di influenza di ogni ruolo in un'organizzazione ha portato in superficie uno spazio latente che è stato invaso da altri profili esistenti, come rappresentato dal diagramma di fig. 1.

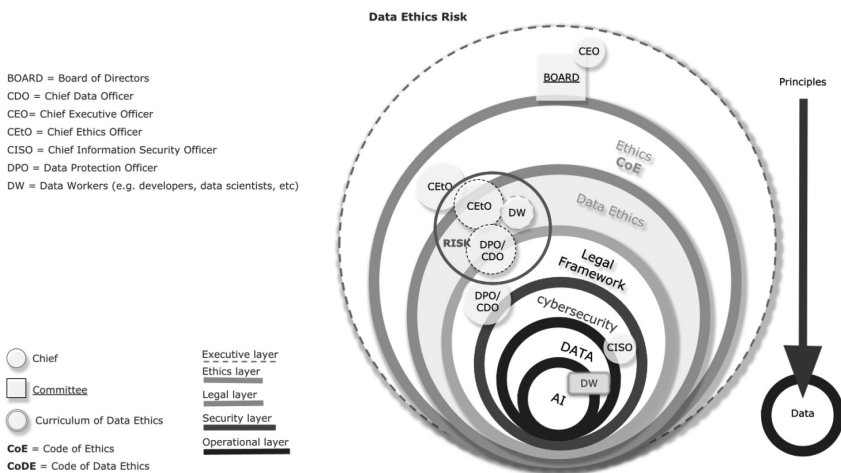


Figura 1: Lo spazio dell'etica dei dati occupato da altri ruoli

<sup>14</sup> M. Taddeo, L. Floridi, *What is data ethics?*, in "Philosophical Transaction of The Royal Society AI", n. 374, 2016, p. 1.

### Il ruolo del responsabile dell'etica dei dati

L'obiettivo dell'etica dei dati è quello di creare l'infrastruttura etica o infraetica nell'organizzazione che faciliti le decisioni etiche sui dati e su come saranno trattati. Lo spazio dell'etica dei dati dovrebbe quindi essere occupato da un responsabile dell'etica dei dati: Chief Data Ethics Officer (CDEO). Il CDEO dovrebbe favorire questo ambiente etico stabilendo e mantenendo un codice di etica dei dati: Code of Data Ethics (CoDE). Il codice di etica dei dati deve rispettare una coerenza interna (con i principi e le norme proposte) e esterna (con il codice etico dell'azienda). Il responsabile dell'etica dei dati deve promuovere la diffusione del codice di etica dei dati all'interno dell'azienda. La sua adozione permetterebbe di definire esplicitamente il perimetro decisionale sull'etica dei dati. Infine tutti gli attori coinvolti nel trattamento dei dati, e in particolare i Data Workers (DW), dovranno essere formati al ragionamento etico sui dati e aggiornati con frequenza regolare sulle migliori pratiche etiche in circolazione. In breve, lo spazio per l'etica dei dati esiste, come nella figura 2. Deve essere occupato da un esperto di etica dei dati che possa proteggere e promuovere la disciplina.

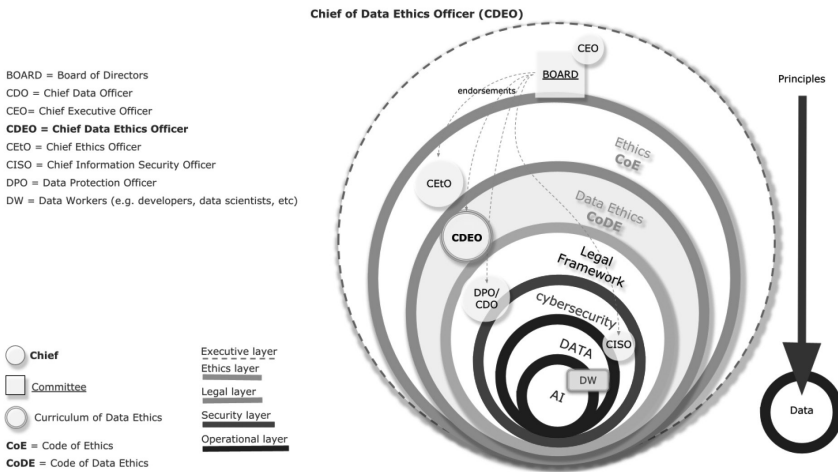


Figura 2: Il livello del Chief Data Ethics Officer (CDEO)



## Conclusioni

Questo articolo sostiene l'idea che ci sia un nuovo spazio etico all'interno delle organizzazioni che riguarda l'etica dei dati. Questo nuovo spazio ha contribuito a sollevare importanti questioni teoriche riguardanti l'organizzazione interna delle aziende nelle società mature dell'informazione. Naturalmente, la riflessione è limitata geograficamente e culturalmente perché inizia, si fonda ed evolve all'interno delle filosofie occidentali. Nonostante queste limitazioni, lo studio suggerisce il riconoscimento di un dominio specifico incentrato sull'etica dei dati e supervisionato da un responsabile etico dei dati (il Chief Data Ethics Officer – CDEO) con l'obiettivo di costruire un'infrastruttura etica informazionale all'interno delle organizzazioni.

## List of most common acronyms

AI	Artificial Intelligence
CDEO	Chief Data Ethics Officer
CDO	Chief Data Officer
CDOC	Children Data Oversight Committee
CEO	Chief Executive Officer
CEtO	Chief Ethics Officer
CISO	Chief Information Security Officer
CoDE	Code of Data Ethics
CoE	Code of Ethics
DPO	Data Protection Officer
DW	Data Workers
EU	European Union
GDPR	General Data Protection Regulation
IAAIS	Independent Audits of Artificial Intelligence Systems (ForHumanity)
Infraethics	Infrastructure of ethics
LoA	Level of Abstraction
SME	Small and Medium-sized Enterprises

## Bibliografia

- Chomsky N., *The Logical Structure of Linguistic Theory*, Springer US, Cambridge/Mass 1975.
- Floridi L., *Information. A Very Short Introduction*, Oxford University Press, Oxford 2010.
- Id., *The Ethics of Information*, Oxford University Press, Oxford 2013.

- Id., *The Latent Nature of Global Information Warfare*, in "Philosophy & Technology, XXVII, 3, 2014.
- Hidalgo C., *Why Information Grows: the Evolution of Order, from Atoms to Economies*, Basic Books, New York 2015.
- Taddeo M., Floridi L., *What is Data Ethics?*, in "Philosophical Transactions of the Royal Society A", CCCLXXIV, 2083, 2016.