

Capitolo I

Dati e informazioni

SOMMARIO: 1. La questione definitoria di dati e informazioni. – 2. Dati e informazioni nei diversi paradigmi scientifici. – 3. Ambiguità e paradossi della riferibilità. – 4. Gli apporti della fisica contemporanea. – 5. Il contributo della teoria dell'informazione. – 6. Informazione semantica (fattuale). – 7. Informazione nella filosofia dell'informatica e dell'intelligenza artificiale. – 8. Informazione tra biologia, memetica e neuroscienze. – 9. Dati e informazioni nel diritto: i primi riferimenti. – 10. Dati e informazioni nel diritto: dagli anni Settanta a oggi. – 11. Possibili chiarificazioni.

1. La questione definitoria di dati e informazioni

Le scienze cognitive, sorte principalmente dall'intersezione tra informatica, psicologia e linguistica, hanno dettato un programma di ricerche teso allo «studio dei processi cognitivi, interpretati come elaborazioni di un calcolatore finito, naturale o artificiale»¹, proponendo un paradigma antropologico radicalmente nuovo, nel quale la mente è considerata come un sistema di elaborazione dei dati. Sulla base di questo più ampio programma di ricerche, le neuroscienze e le neurotecnologie, cioè le metodiche e le tecniche di estrazione informativa e modulazione dell'attività cognitiva, hanno consentito – e probabilmente consentiranno – un accesso sempre più pervasivo alla dimensione intima della persona.

Sia la concezione della *mente* come sistema di elaborazione di dati, sia la tecnologia di trattamento dei dati personali (che consentirebbe l'interferenza nella *mente* stessa) si incontrano per vie distinte sul terreno della libertà. Ed è perciò che, nell'orizzonte delle tecnologie informatiche, per

¹O, forse meglio, la «scienza cognitiva» (D. MARCONI, *Filosofia e scienza cognitiva*, Bari 2001, 18).

trattare delle libertà costituzionali, occorre partire dalla questione definitoria delle nozioni di dato e informazione.

In particolare, la possibile definizione del dato *personale* appare centrale per profili di ordine teorico, perché gran parte della riflessione filosofica, politica e giuridica assume che alcune delle principali questioni contemporanee gravitino intorno alle nuove tecnologie di trattamento dei dati riferibili alla persona²; poi, però anche per ragioni di ordine applicativo: la stessa regolamentazione dei dati (e segnatamente quella dei dati *personali*) – che riveste un ruolo principale nello spazio normativo delle tecnologie informatiche e nella protezione della persona nell’ambiente digitale – si fonda necessariamente sul tentativo di una definizione del suo oggetto principale³.

Va però subito osservato che nelle coordinate del diritto le definizioni⁴ siano qualcosa di singolare rispetto ad altri ambiti disciplinari, perché si rendono spesso necessarie ai fini applicativi e possono dipendere, oltre che dal risultato di un’attività meramente cognitiva e ricognitiva degli usi possibili di un termine, anche da decisioni, e in tal senso possono dirsi *stipulative*. Infatti, «le definizioni *lessicali*, come quelle dei dizionari, sono usate per informare sui diversi significati che una parola già ha», mentre «[l]e definizioni *stipulative*, tipiche delle scienze esatte, sono invece usate per prescrivere cosa una parola deve significare»⁵. E, in particolare e spesso, occorre procedere a una *ridefinizione* di carattere *stipulativo* di termini che hanno impieghi diffusi tra i parlanti, e che però richiedono significati *tecnicizzati*, soprattutto per esigenze di chiarezza normativa⁶.

È chiaro, allora, che questa premessa generale valga anche per le defini-

²Basti ora il richiamo al quadro tracciato da Y.N. HARARI, *Homo Deus. Breve storia del futuro* (2015), Firenze-Milano 2017; sul versante filosofico L. FLORIDI, *Pensare l’infosfera. La filosofia come design concettuale*, Milano 2020; su quello giuridico, tra i tanti, i collettanei di F. PIZZETTI (a cura di), *Privacy e il diritto europeo alla protezione dei dati personali: dalla Direttiva 95/46 al nuovo Regolamento europeo*, Torino 2016, e V. CUFFARO, R. D’ORAZIO, V. RICCIUTO (a cura di), *I dati personali nel diritto europeo*, Torino 2019.

³Infatti, «soltanto dalla riflessione sul linguaggio è possibile giungere a una comprensione del pensiero e che l’esperienza umana è in primo luogo *esperienza linguistica*. Il linguaggio, in una parola, è *intrascendibile*, rappresentando un mezzo imprescindibile di accesso alla realtà» (G. D’ALESSANDRO, *La nullità della legge. Percorsi della cultura giuridica italiana del Novecento*, Napoli 2012, 8).

⁴Per tutti, cfr. A. PINTORE, *La teoria analitica dei concetti giuridici*, Napoli 1990, 26 ss. con riferimento all’analisi del pensiero di Hart, 54 ss. sul rapporto tra quest’ultimo e Ross.

⁵M. BARBERIS, *Una filosofia del diritto per lo Stato costituzionale*, Torino 2017, 3.

⁶*Ibidem*, 3-4.

zioni di ‘dato’ e di ‘informazione’ nel contesto del diritto⁷. Un percorso di ricerca sul loro significato, quindi, lungi dal proporsi l’obiettivo di risolvere le questioni ontologiche sottese, può soprattutto limitarsi a registrare gli usi diversi, nei diversi contesti storici o disciplinari, per poi avvicinarsi al significato dei termini nel contesto del diritto e proporre – sia pure *in limine* – suggerimenti anche in chiave *stipulativa* o *ridefinitoria*⁸.

Eppure, l’impossibilità di pervenire a una definizione “ultimativa” dei concetti di dato e informazione, sia per prudenze metodologiche, sia per ragioni propriamente epistemologiche, non può esimere la ricerca di ambito giuridico dal frequentare la riflessione multidisciplinare sul punto, soprattutto a valle dei diversi apporti confluiti nella cosiddetta filosofia dell’informazione, in modo tale da convalidare o eventualmente riformulare le categorie e i concetti di cui il diritto stesso (e segnatamente quello della protezione dei dati) si serve. Ciò si rende tanto più necessario alla luce della premessa che abbiamo esplicitato: che la prospettiva dei dati abbia inciso sulla nostra visione antropologica; che il trattamento dei dati possa consegnare le chiavi di accesso alla nostra dimensione intima; e che molte delle tecnologie emergenti siano attratte (innanzitutto) nel perimetro della regolamentazione dei dati personali.

Potrebbe dubitarsi dell’opportunità di una simile direzione così ampia delle ricerche: però, le nozioni di dato e informazione, sia nel diritto, sia in altri ambiti – soprattutto in quelli umanistici – sono ormai tanto diffuse quanto ambigue.

Né deve stupire che un campo di studi, quale quello della protezione dei dati personali e della *privacy*, non si serva tuttora (e forse non si possa servire) di definizioni consolidate e “ultimative”. Basti un riferimento a un caso noto e analogo nella cultura giuridica: la teoria della proprietà ha dovuto concentrarsi sulla nozione di *res* (materiale o immateriale) non già alle sue origini, ma solo una volta che il progresso scientifico e tecnologico ha introdotto nuovi problemi, che richiedevano quindi uno sforzo definitorio che poteva essere avvertito solo sulla base di una diversa sensibilità. Si pensi, all’esigenza di classificare oggetti diversi da quelli tradizionalmente interessati dai rapporti dominicali, quali quelli cui si rivolge la proprietà intellettuale in senso ampio, o quelli fisici che presentano morfologie irriducibi-

⁷ Su questi temi, una parte delle ricerche sono confluite nel contributo F. CIRILLO, *La nozione di dato personale. Spunti di riflessione per un approccio interdisciplinare*, in *Cyberspazio e diritto*, 22, 2021, 23-40.

⁸ Sulle possibili ridefinizioni *stipulative*, v. *infra*, § 11.

li alle *res* tradizionali (come le *res immateriali*, ma anche le fonti energetiche, l'aria o l'acqua). Così, non deve stupire se il diritto non disponga (anche in questo ambito) di definizioni stabili degli elementi fondamentali del suo stesso discorso, e che però ciò non impedisca una regolamentazione dei fenomeni⁹.

Anche in assenza di uno strumentario concettuale davvero adeguato, infatti, si possono regolamentare determinati settori della vita in modo soddisfacente: basti considerare che nessuna teoria della proprietà ha realmente risentito della scoperta einsteiniana della formula che pone in relazione massa ed energia; non è stato necessario, cioè, alla luce della scoperta di una nuova teoria scientifica, porre in questione la distinzione giuridica delle *res* materiali dall'energia.

Infatti, può accadere, come nel caso delle distinzioni tra le tipologie di *res* che caratterizzano la teoria della proprietà, che le definizioni e le categorie del diritto, pur servendosi di nozioni empiriche o vagamente ispirate alle scienze naturali, non ricalchino in tutto e per tutto le acquisizioni dei saperi scientifici.

Allora, in questa ricerca, ci si propone di introdurre le principali questioni relative alla definizione dei dati e delle informazioni e, soprattutto, del rapporto tra i dati e le realtà cui essi *mediatamente* o *immediatamente* pretendono di *referirsi*¹⁰, evidenziando delle connessioni possibili tra gli approcci delle diverse discipline interessate dal tema e le categorie del diritto.

2. Dati e informazioni nei diversi paradigmi scientifici

Ai termini 'dati' e 'informazioni' si associano spesso significati diversi. Ciò accade perché tali espressioni si possono riferire a oggetti sovrappo-

⁹Si può pensare al paradosso di Böckenförde sull'incapacità del diritto di dimostrare i propri fondamenti, espresso, *ex multis*, in E.-G. BÖCKENFÖRDE, *Die Entstehung des Staates als Vorgang der Säkularisation*, in E. FORSTHOFF (a cura di), *Säkularisation und Utopie: Ebracher Studien*, Stuttgart 1967, 42 ss. Per altro verso, si osserva che la teoria giuridica «dipende anche da generalizzazioni empiriche (sociologiche, storiche, antropologiche ...)» (M. BARBERIS, *Una filosofia del diritto per lo Stato costituzionale*, Torino 2017, 6).

¹⁰Un'analisi del dibattito attuale sul "riferimento" del linguaggio alla realtà, sulla distinzione tra estensione e intensione, senso e denotazione, ecc. in D. MARCONI, *La filosofia del linguaggio. Da Frege ai nostri giorni*, Novara 2008, in particolare 103 ss.

nibili o del tutto distinti, se non proprio a differenti classi di oggetti possibili¹¹. Il problema non muta se si considera anche lo slittamento di significati prodotto dalla traduzione di ‘data’ e ‘information’, per i quali termini vale in inglese una simile ambiguità (anche se hanno coperture semantiche solo in parte sovrapponibili ai corrispettivi italiani). Inoltre, il contesto di riferimento cambia rapidamente: la realtà che i termini ambiscono a descrivere non è sempre la stessa, perché l’orizzonte tecnologico è in continuo mutamento e le prospettive disciplinari chiamate in causa sono molteplici.

Così, alle nozioni del senso comune, si sovrappongono quelle della matematica, della fisica, dell’informatica, della statistica o della filosofia, che a loro volta influenzano la coscienza dei parlanti e il senso comune¹².

Se tale indeterminatezza non deve destare stupore sul piano semiotico, lo stesso non può dirsi per quello giuridico. Nel diritto, come anticipato, le interpretazioni e le definizioni ambiscono a uno statuto forte¹³, indipendentemente da quanto ciò sia teoricamente possibile. Non ci si può arrestare al “fallimento del linguaggio”¹⁴, alla denuncia del suo carattere inattendibile o alla mera scepri linguistica. Un’indagine che mira a confrontarsi con un campo della protezione dei dati personali e della *privacy*, allora, deve sforzarsi di affrontare il problema definitorio di parole come ‘dati’ o ‘informazioni’. Tuttavia, un simile proposito, come anticipato, dovrebbe tendere soprattutto a registrare le principali definizioni *informative* degli usi e dei significati delle espressioni nei diversi contesti – ivi compreso nel lessico del diritto – per poi procedere a un’operazione *sti-*

¹¹ Le semiotiche di autori come Saussure e Peirce studiano la relazione tra espressione e significato (o, con il secondo, oggetti-*type*), senza riferimento, cioè, agli “oggetti reali a cui il segno può riferirsi” (oggetti-*token*), come invece può accadere per FREGE [così in U. ECO, *Trattato di semiotica generale* (1975), Milano 2016, 108-109]. Il riferimento ultimo è a G. FREGE, *Über Sinn und Bedeutung*, in *Zeitschrift für Philosophie und philosophische Kritik*, NF 100, 1892, 25-50 (in particolare 26: «daß ich hier unter „Zeichen“ und „Namen“ irgendeine Bezeichnung verstanden habe, die einen Eigennamen vertritt, deren Bedeutung also ein bestimmter Gegenstand ist (dies Wort im weitesten Umfange genommen)»).

¹² Cfr. L. FLORIDI, *Information: A Very Short Introduction*, Oxford 2010, 3 ss.; v. ampiamente anche il collettaneo, ID. (a cura di), *The Routledge Handbook of Philosophy of Information*, Abingdon-New York 2016.

¹³ Su tutti, v. R. GUASTINI, *Interpretare e argomentare*, Milano 2011, 4 ss.

¹⁴ *Versagen der Worte* è la formula apparsa nel romanzo di esordio di R. MUSIL, *Die Verwirrungen des Zöglings Törleß* (1910), Tübingen 2016, 101-102: «[e]s war ein *Versagen der Worte*, das ihn da quälte, ein halbes Bewusstsein, dass die Worte nur zufällige Ausflüchte für das Empfundene waren» [corsivo aggiunto].

pulativa in cui si stabiliscano quali termini adottare nel corso dell'analisi¹⁵.

La prima operazione deve necessariamente seguire i concetti di 'informazione' e 'dati' dal momento della comparsa dell'*information theory*¹⁶ nel corso del ventesimo secolo, senza cedere alla tentazione di estendere l'arco temporale oltre il necessario. Se per 'informazione' intendessimo ogni genere di nozione o notizia, e per 'dato' ogni elemento registrato o acquisito alla conoscenza, allora sarebbe opportuno ampliare l'indagine alla preistoria della scrittura, sino a ricompredervi «brevi cenni sull'universo»¹⁷. Non è però in questo senso che ci si riferisce alla *società dell'informazione*¹⁸, al *dataismo* (o *datismo*)¹⁹ o all'*infosfera*²⁰. Né tantomeno nella protezione dei dati o nel diritto all'informazione, 'dati' e 'informazioni' possono coprire un campo semantico così ampio e indeterminato. Inoltre, un'operazione di ricostruzione degli usi in tempi e contesti diversi di questi termini deve guardarsi, per quanto possibile, dai pericoli della *Begriffsgeschichte*²¹: ogni volta che si ripercorre la storia di un concetto, infatti, si ri-

¹⁵ Sulle definizioni *informative* e *stipulative*, cfr. ancora R. GUASTINI, *Interpretare e argomentare*, cit., 25-26, nonché – sempre con riferimento all'approccio analitico – U. SCARPELLI, *Contributo alla semantica del linguaggio normativo* (1959), Milano 1985, capp. I e II.

¹⁶ Si tratta di una teoria fondamentale per l'informatica e le telecomunicazioni, la cui paternità è attribuita consuetamente a due contributi C.E. SHANNON, *A Mathematical Theory of Communication*, in *Bell System Technical Journal*, 27, 1948, 379-423 e 623-656, su cui cfr. *infra* § 1.4.

¹⁷ L'espressione è di A. GRAMSCI, *Quaderni dal carcere* (1948-1951), Quad. 8 (1931-1932), Torino 2014, § 143.

¹⁸ La centralità della produzione di informazione nel sistema economico è stata evidenziata, tra i primi, da F. MACHLUP, *The Production and Distribution of Knowledge in the United States*, Princeton 1962.

¹⁹ Formula probabilmente coniata dal giornalista D. BROOKS, *The Philosophy of Data*, ne *The New York Times*, 13 febbraio 2013; adottata poi anche dal giornalista S. LOHR, *Data-ism. The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*, New York 2015; e, infine, resa celebre Y.N. HARARI, *Homo Deus. Breve storia del futuro* (2015), Firenze-Milano 2017, 559: «le scienze biologiche sono giunte a concepire gli organismi come algoritmi biochimici. Contemporaneamente, negli anni 80 trascorsi da quando Alan Turing formulò l'idea della macchina che porta il suo nome, gli informatici hanno imparato a progettare algoritmi digitali interpretabili da elaboratori elettronici sempre più sofisticati. il datatismo mette insieme queste concezioni, evidenziando che esattamente le stesse leggi matematiche si applicano sia agli algoritmi biochimici sia a quelli computerizzati digitali».

²⁰ Ancora in L. FLORIDI, *Pensare l'infosfera. La filosofia come design concettuale*, Milano 2020, 141 ss.

²¹ Il rischio principale è l'invenzione di una storia meramente linguistica, disancorata dal contesto sociale e politico, su cui, brevemente, già R. KOSELLECK, *Begriffsgeschichte und Sozial-*

schia di astrarre da un determinato contesto un certo contenuto ideale – spesso in ragione di alcuni bisogni del presente – per poi costruire una narrazione storicamente e teoricamente fallace. Però, il rischio insito in una simile operazione non supera i benefici di un tentativo attuato nel quadro delle opportune prudenze metodologiche.

Il risultato della prima operazione definitoria *informativa*, conseguito alle condizioni ora indicate, potrebbe consentire di procedere alla costruzione di un quadro concettuale convenzionale dentro cui iscrivere l'analisi successiva.

Entrambe le operazioni, però, dovranno tendere a *riferire*²² espressioni e significati alla realtà delle *information technologies*, nel loro continuo e rapido incedere. Il riferimento alla realtà delle tecnologie di trattamento dei dati, lungi dal rappresentare un'adesione ingenua a una «semantica estensionale»²³, implica piuttosto la tensione di cui si fa carico ogni ricerca sul diritto delle nuove tecnologie. L'inseguimento di un orizzonte tecnologico mutevole, sia pure a detrimento del rigore metodologico-giuridico, garantisce, infatti, la necessaria elasticità della prospettiva di indagine. In altri termini, una ricerca sul diritto (e sui possibili diritti) delle nuove tecnologie non può concedersi il lusso di trascurare cosa esse siano o si propongano di attuare.

Il metodo per giungere a una ricostruzione possibile delle definizioni di 'dati' e 'informazioni' in un'indagine sulla protezione dei dati non potrebbe, quindi, essere unico. Da un lato, allora, occorre servirsi dell'*entanglement* di diverse prospettive disciplinari²⁴, dalla filosofia dell'informazione

geschichte, in C. LUDZ (a cura di), *Soziologie und Sozialgeschichte*, Opladen 1972, 116 ss. Un'altra prospettiva della storiografia sociale è quella della lessicologia, su cui L. GIRARD, *Histoire et lexicographie*, in *Annales*, 6, 1963, 1128-1132.

²² Le questioni del riferimento alla *Cosa in Sé* in un'analisi che parte dal rapporto tra significante e significato sono molteplici e non mancano nella semiotica come nel più ampio orizzonte della filosofia teoretica-autori che escludono la possibilità e il senso di un tale riferimento. Sul punto, si conviene ancora con U. ECO, *Kant e l'ormitorinco* (1997), Milano 2016, 17: «[c]'è una differenza tra dire che non possiamo capire tutto (una volta per tutte) e dire che l'essere è andato in vacanza».

²³ Cioè, a quell'indirizzo filosofico che ritiene di risolvere la questione del riferimento tra oggetto reale, significato e significante in modo piano: «[c]omprendere una proposizione vuol dire sapere che accade se essa è vera» [si tratta della proposizione n. 4.024 di L. WITTGENSTEIN, *Tractatus Logico-Philosophicus*, anche in A.G. CONTE (a cura di), *Tractatus logico-philosophicus e Quaderni 1914-1916*, trad. it. di Id., Torino 2009]. Sul punto, cfr. anche U. ECO, *Dall'albero al labirinto* (2007), Milano 2017, 651 ss.

²⁴ Nel senso del recente indirizzo metodologico della storiografia giuridica, su cui ampiamente v. T. DUVE, *European Legal History-Concepts, Methods, Challenges*, in ID. (a cura di), *En-*

alla storia, dalla sociologia all'informatica. Dall'altro lato, invece, le categorie del diritto – vecchie o nuove – delineano un quadro possibile di condizioni fondamentali per l'obiettivo di questa ricerca. Infatti, la fondazione di uno statuto giuridico della riservatezza, che passi per la regolamentazione del flusso di dati *referibili* all'individuo (il cd. *habeas data*²⁵), parte necessariamente dalla definizione dei termini essenziali del discorso. In questo capitolo, proprio perché lo stato dell'arte si presenta in rapida evoluzione, si intendono soltanto tratteggiare le condizioni di partenza di un simile approccio di ricerca.

3. *Ambiguità e paradossi della riferibilità*

Il problema dei significati di 'dati' e 'informazioni', del contenuto semantico cui questi termini si debbano *referire*, può essere affrontato congiuntamente alla attitudine di dati e informazioni a *referirsi* – a loro volta – alla *realtà*. In altre parole, 'dati' e 'informazioni' sono tali, almeno nelle coordinate del diritto, in quanto si rivolgono a oggetti o persone. Per questa ragione, la *referibilità* di dati e informazioni a determinate realtà è correlata alla loro questione definitoria. In altre parole, appare centrale, sia nella definizione dei dati sia quanto alla soluzione dei problemi di ordine teorico e applicativo, la pretesa che i dati abbiano un rapporto con la realtà; circostanza che l'intuizione potrebbe fare apparire come scontata e che, tuttavia, nelle teorie della comunicazione e dell'informazione, come pure nelle semiotiche "deboli", è del tutto marginale o trascurabile.

La questione della *referibilità* di un segno alla realtà, però, non è soltanto filosofica o linguistica. Essa evoca certamente l'inconoscibilità della *Cosa in Sé* e così il tema delle condizioni trascendentali che consentono – e limitano – la conoscenza del mondo²⁶, sino rinviare al dibattito più recente

tanglements in Legal History: Conceptual Approaches, Frankfurt am Main 2014, 3 ss.; oppure, in quello di *histoire croisée*, proposto da M. WERNER, B. ZIMMERMANN, *Penser l'histoire croisée: entre empirie et réflexivité*, in *Annales*, 58, 2003, 5 ss.

²⁵ Per tutti, S. RODOTÀ, *Il mondo nella rete. Quali i diritti, quali i vincoli*, Roma 2014, 27 ss.

²⁶ Il concetto-limite di KANT, introdotto con la *Critica della ragion pura* (1781), che continua a essere evocato da molti approcci filosofici che intendono *limitare* le possibilità della conoscenza (*i. e.* «*die Sache an sich selbst* dagegen zwar als für sich wirklich, aber von uns unerkannt», in I. KANT, *Kritik der reinen Vernunft*, II ed., Riga 1787, XX, corsivo aggiunto).

alimentato in seno alla filosofia teoretica e, più specificamente, alla semiotica²⁷. Da questo punto vista, per la sua generalità, tale problema interessa la presente indagine non meno di quanto riguardi ogni altro percorso ricerca. E, in quanto tale, non potrebbe essere qui né risolto, né adeguatamente trattato.

Eppure, la *riferibilità* interseca l'analisi su altri due piani più specifici.

Il primo – quello al quale si è fatto riferimento nel paragrafo che precede – è di ordine metodologico: se e in quale misura possiamo chiederci cosa siano ‘dati’ e ‘informazioni’ o, forse, semplicemente cosa s'intenda per essi; più in generale, quale rapporto corra tra un termine, il suo significato in un discorso (giuridico) e il mondo *reale* o, per dirla con Frege, il *bestimmter Gegenstand* che deve essere parte della *Bedeutung* di nomi e segni. Anche sul piano metodologico, come su quello filosofico, l'analisi non può dirimere alcuna questione, né intende apportare risultati innovativi. Però, l'incapacità di risolvere i problemi teoretici e metodologici non può esimerci dal frequentarli e tradurli in atteggiamenti di prudenza. Su questo piano del discorso, occorre evocare la riflessione sulle definizioni nel diritto offerte, soprattutto, in senso alla filosofia analitica del diritto e alla teoria generale.

Vi è poi un secondo piano, ora attinente al contenuto specifico dell'indagine, in cui la questione della *riferibilità* è fondamentale. Dati e informazioni, che ci appaiono indistintamente ora come segni ora come significati, sono protetti e regolamentati proprio in ragione della loro *riferibilità* a realtà concrete. Basti il richiamo alla nozione di ‘dato personale’ offerta dal regolamento europeo sulla protezione dei dati (art. 4, § 1, n. 1, GDPR); nozione dalla cui estensione dipenderebbe l'attivazione dell'intero sistema regolatorio, e che però desta alcune perplessità. Per il legislatore europeo un dato personale è:

«qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubi-

²⁷ Su questi temi, v. brevemente U. ECO, M. FERRARIS, D. MARCONI, *Lo schema del cane*, in *Rivista di Estetica*, n.s., 8, 1998, 3 ss., ripubblicato anche in *Rivista Italiana di Filosofia del Linguaggio*, 11, 2017, 6-29; più ampiamente, v. anche gli altri contributi nello stesso numero della rivista (a cura di V. PISANTY, S. TRAINI), intitolato *Eco, Kant and the Platypus. Twenty Years After*.

cazione, un identificativo *online* o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale»²⁸.

La prima proposizione della disposizione risolve apparentemente in modo acritico la questione della riferibilità di un'informazione a una persona (i dati personali *sono* informazioni e *riguardano* le persone). Le righe successive, invece, testimoniano quanto il problema sia del tutto aperto. Il regolamento, tra le varie possibilità, «*considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come [...] un numero di identificazione*». Il legislatore europeo, facendo propria la poetica di Gertrude Stein, dove una «rosa è una rosa è una rosa è una rosa»²⁹, ha preso posizione – forse inconsciamente – nella disputa degli universalisti, che da Guglielmo di Ockham a William Shakespeare accompagna la storia della cultura europea, sino all'ultima e più recente fama de *Il nome della rosa* di Umberto Eco. Ed è proprio l'analisi che quest'ultimo compie della tautologia di Stein che induce a ritenere che la ridondanza della definizione contenuta nel regolamento europeo implichi una particolare tensione, anche emotiva, che rivela l'incapacità o forse l'impossibilità di definire il dato personale³⁰.

Il legislatore europeo, pur non avendo superato «il bisogno di assegnare a ogni significante il suo significato»³¹ (come invece accadrebbe per Stein), intende così *riferirsi* a qualcosa che sfugge al proprio linguaggio.

Un esempio tratto dalla prassi applicativa del regolamento europeo dimostra che il problema non sia soltanto teorico. Accade che, sulla base del trattamento di dati aggregati (non riferibili a persone determinate), alcuni algoritmi consentano inferenze alquanto precise su informazioni sensibili di un individuo identificato³². Più nel dettaglio, ciò può riguardare le in-

²⁸ Sulla definizione normativa del dato personale, v. subito V. RICCIUTO, *La patrimonializzazione dei dati personali*, in V. CUFFARO, R. D'ORAZIO, ID. (a cura di), *I dati personali nel diritto europeo*, Torino 2019, 25 ss.

²⁹ «Rose is a rose is a rose is a rose», G. STEIN, *Sacred Emily* (1913), in *Geography and Plays* (1922), Madison 1993, 187, anche in EAD., *Geografia e drammi*, trad. it. di F. Iuliano, Macerata 2010, 179.

³⁰ In U. ECO, *La struttura assente. La ricerca semiotica e il metodo strutturale* (1968), Milano 2016, 117 ss. D'altronde, «[c]ome sarà lei stessa a dire: "No, no, no, no, non è solo una ripetizione. Cambio sempre un po' le parole"» (dall'introduzione alla versione italiana di N. FUSINI, XII, in un'intervista del 1934 su cui v. *sub* n. 1, XX).

³¹ Ancora dalla *Introduzione* di N. FUSINI a *Geografia e drammi*, cit., XIX.

³² Per un quadro ampio, con riferimento alla sorveglianza elettronica di massa, cfr. G. RESTA, *La sorveglianza elettronica di massa e il conflitto regolatorio USA/EU*, in ID., V. ZENO-

formazioni relative alla salute, poiché, intrecciando dati aggregati deidentificati con dati personali apparentemente innocui, è possibile inferire informazioni su condizioni patologiche di una determinata persona.

Si tratta, allora, di dati (non personali) che forniscono informazioni su altri dati (personali), cioè di *metadata* (o metadati). La questione della riferibilità dei *metadata* alle persone fisiche e alla loro possibile qualificazione come dati sensibili era già stata in parte risolta dalla giurisprudenza europea, anche con riferimento ai rapporti tra la vigente Direttiva *ePrivacy* (2002/58/CE) e la abrogata Direttiva *madre* (1995/46/CE): «[q]uesti dati, presi nel loro complesso, possono permettere di trarre conclusioni molto precise riguardo alla vita privata delle persone i cui dati sono stati conservati, come le abitudini quotidiane, i luoghi di soggiorno permanente o temporaneo, gli spostamenti giornalieri e non, le attività svolte, le relazioni sociali di queste persone e gli ambienti sociali da esse frequentati» (CGUE, C-293/12 e C-594/12)³³. Un'altra declinazione dello stesso problema è quella della riferibilità all'individuo dei dati pseudonimizzati: con la tecnica della pseudonimizzazione, infatti, un elemento identificativo di una persona è sostituito con uno pseudonimo tramite, appunto, il ricorso a un'apposita funzione che offusca le informazioni o meglio, quell'informazione da cui deriva l'univoca riferibilità a una persona³⁴. Anche in questo caso, ci si è interrogati sull'eventuale e persistente riferibilità dell'informazione pseudonimizzata alla persona, visto che con il ricorso a dati pseudonimizzati e intrecciati è spesso possibile risalire comunque all'identità di una certa persona³⁵.

In tutte queste ipotesi, il modo di trattare i dati ne determina la (pretesa) riferibilità a una realtà – nel caso, alla persona – con apprezzabili conseguenze di ordine giuridico. Si potrebbe ritenere che questo problema sia marginale, in quanto si limiterebbe ai soli casi in cui gli algoritmi formulino

ZENCOVICH (a cura di), *La protezione transnazionale dei dati personali. Dai "safe harbour principle" al "privacy shield"*, Roma 2016, 27 ss.

³³ Sul punto cfr. A. SPINA, *La medicina degli algoritmi*, in F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino 2018, 324.

³⁴ Cfr. brevemente il documento informativo dell'Agenzia dell'Unione europea per la cibersicurezza (ENISA), di A. BOURKA, DROGKARIS, I. AGRAFIOTIS, *Tecniche di pseudonimizzazione e migliori pratiche*, ENISA, 2019, 9-11.

³⁵ Sul punto ampiamente v. G. COMANDÈ, *Regulating Algorithms' Regulation? First Ethico-Legal Principles, Problems, and Opportunities of Algorithms*, in T. CERQUITELLI, D. QUERCIA, F. PASQUALE (a cura di), *Transparent Data Mining for Big and Small Data*, New York 2017, 169 ss.

delle inferenze su qualità individuali integrando dati non personali con altri direttamente riferibili a individui.

Invece, come potrebbe essere ampiamente illustrato, la questione critica nell'ambiente tecnologico non è tanto la protezione delle informazioni esplicitamente *sensibili*³⁶ e direttamente riferite a una persona, quanto piuttosto la regolamentazione di tipologie di trattamenti informatici in cui la riferibilità a un individuo determinato (e la riferibilità a sue condizioni particolarmente intime) è mediata e indiretta.

Da ultimo, occorre subito osservare che non è sempre possibile distinguere tra un'informazione "certa" riferita a una persona e una mera inferenza probabilistica. Anzi, da un punto di vista epistemologico, si potrebbe persino affermare che le informazioni certe e quelle meramente inferite si pongano sullo stesso piano, che le verità altro non siano che delle ottime inferenze, intercorrendo tra la dimensione della certezza e quella dell'inferenza uno scarto quantitativo di attendibilità, piuttosto che una distinzione di ordine qualitativo.

4. *Gli apporti della fisica contemporanea*

La questione della definizione di cosa siano dati e informazioni³⁷, o di cosa per essi si possa intendere, deve essere affrontata innanzitutto con riferimento al contesto anglosassone, donde quello italiano ha mutuato le attuali accezioni su cui ci intendiamo soffermare: tratteremo, per ora indifferentemente, di *data* e *information*³⁸, come di dati e informazioni.

³⁶ Il riferimento è oggi alle particolari categorie di dati personali di cui all'art. 9 del regolamento UE/679/2016 (come pure nell'art. 8 della precedente Direttiva 95/46/CE), talvolta identificati, appunto, anche come 'dati sensibili' sin dalla scelta terminologica del D.lgs. 196/2003 e poi, diffusamente, in diverse fonti (p.e. ancora nel D.m. 162/2016 del Ministero dell'Istruzione, dell'Università e della Ricerca, recante il regolamento sul «*Trattamento di dati sensibili idonei a rilevare lo stato di disabilità degli alunni censiti in Anagrafe Nazionale degli Studenti in una partizione separata*»).

³⁷ Sulla quale, con riferimento alla letteratura giuridica, cfr. subito F. ROMEO, *Il dato digitale e la natura delle cose*, in A. BALLARINI (a cura di), *Diritto interessi ermeneutica*, Torino 2012, 97 ss.

³⁸ *Data* è un sostantivo plurale che vive una vita autonoma rispetto al suo singolare, *datum* (*Data*, in *Merriam-Webster.com Dictionary*); *information*, invece, è un sostantivo *uncountable* e, come tale, non è declinabile al plurale (*Information*, in *English Grammar, Dictionary. cambridge.org*). Questa avvertenza giustifica l'uso ora al singolare ora al plurale di 'dato' e 'dati', 'informazione' e 'informazioni'.

L'informazione, infatti, è stata a lungo «la cenerentola nella storia della filosofia»³⁹, anche se più recentemente ha guadagnato una centralità anche nell'interesse dei filosofi, sino alla genesi di talune filosofie dell'informazione. L'attuale centralità del tema è dovuta al ruolo che dati e informazioni occupano nel mondo contemporaneo, dalla fisica teorica alle tecnologie, dalla statistica all'economia, sino a quasi ogni ambito della conoscenza. Ed è proprio grazie all'intreccio delle riflessioni interdisciplinari che è possibile frequentare la questione della definizione dei dati e delle informazioni con una rinnovata sensibilità.

Un primo campo di cui tener conto è quello della fisica, dove «alcune nozioni di informazione sono matematiche e precise, mentre altre sono quantitative ed euristiche»⁴⁰. Nel determinismo che caratterizza la meccanica classica, infatti, l'informazione è costituita sia dal potenziale complesso di tutti i dati che descrivono lo stato presente dell'universo, sia dalla conoscenza delle forze che muovono il divenire⁴¹. Si tratta dell'immagine del demone descritto da Laplace:

«Un'intelligenza che per un dato istante conoscesse tutte le forze di cui la natura è animata, e la rispettiva situazione degli esseri che la compongono, se fosse d'altronde assai vasta per sottomettere questi dati all'analisi, abbraccerebbe nella medesima formula i movimenti dei più grandi corpi dell'universo e quei dell'atomo il più leggero; niente sarebbe incerto per essa, e l'avvenire come il passato sarebbe presente ai suoi occhi»⁴².

In tale prospettiva raffinemente teoretica, l'informazione rappresenterebbe lo «specchio» o il «doppio» del mondo, e si lega all'ambizione – destinata in parte a fallire – di duplicare la realtà nell'intelletto. Quest'approccio si fonda su una prospettiva molto ottimista del rapporto tra la realtà e le informazioni (veritiere) che vi si *riferirebbero*, con un entusiasmo marcatamente positivistico che scommette sulla possibilità che la conoscenza umana raggiunga dei margini di certezza e oggettività, magari pro-

³⁹Traducendo L. FLORIDI, *Introduction*, in ID., *The Routledge Handbook of Philosophy of Information*, cit., 1.

⁴⁰Traduco da N.L. HARSHMAN, *Physics and Information*, in L. FLORIDI (a cura di), *The Routledge Handbook of Philosophy of Information*, cit., 7.

⁴¹*Ibidem*, 8.

⁴²S. DE LAPLACE, *Saggio filosofico sulle probabilità* (1795), trad. it. L.M. Fanelli dell'avvertimento contenuto sin dalla terza edizione francese (1820), Napoli 1820, 8.

spettando un futuro in cui tutta l'informazione della realtà confluisca nell'intelletto (collettivo).

Tuttavia, già nella meccanica statistica – tra gli altri, di Maxwell, Boltzmann e Gibbs – la probabilità viene incorporata nei modelli della fisica e il paradigma deterministico inizia a vacillare. La fisica novecentesca, dalla teoria della relatività alla meccanica quantistica, ha contribuito a collocare l'informazione all'interno del suo stesso oggetto di studio, definendo le forze che la governano e indebolendo la prospettiva positivista che vedeva nei dati e nelle formule il semplice replicarsi della realtà all'interno della mente⁴³.

La probabilità quantistica ha poi fortemente influenzato anche gli studi sulle telecomunicazioni, ponendo le basi per l'attuale *quantum information theory*, che a differenza dell'informatica classica utilizza i quanti per memorizzare ed elaborare le informazioni. L'informazione quantistica ha aperto le porte alla crittografia e al *computing* quantistici, in una torsione complessiva dal determinismo alla probabilità, dalla certezza all'incertezza, dalla località all'*entanglement* di interferenze a distanza⁴⁴.

In breve, le nozioni di dato e di informazione nel paradigma contemporaneo della fisica hanno perso quella limpidezza e aporeticità che affioravano nella fisica classica (ed emblematicamente nell'immagine del demone di Laplace): si è persa, cioè, l'ingenua intuizione di un rapporto di riferibilità diretta di un dato al mondo reale; così, la certezza ha ceduto il posto alla probabilità. Le informazioni, lungi dall'essere lo specchio del mondo nell'intelletto, rappresentano il prodotto di un rapporto problematico tra l'osservatore e il sistema di riferimento, anch'esso sottoposto all'interesse della scienza. La mente stessa, un tempo estranea all'oggetto di studio delle scienze naturali, può essere ora studiata come un sistema di elaborazione di informazioni, talora concepito riduttivamente come un mero elaboratore informatico fatto di carne (così nelle teorie forti dell'Intelligenza artificiale), talaltra evidenziando il ruolo del corpo nell'attività cognitiva (nella *embodied cognition*) o ipotizzando la necessità della meccanica quantistica per l'esistenza stessa della mente umana⁴⁵. Questi

⁴³ Per esempio, un ampio quadro dei rapporti tra i movimenti culturali degli anni Sessanta e Settanta è offerto in D. KAISER, *How the Hippies Saved Physics: Science, Counterculture, and the Quantum Revival*, New York 2011.

⁴⁴ Sul punto, v. tra i tanti la *Preface* di N.S. YANOFKY, M.A. MANNUCCI, *Quantum Computing for Computer Scientists*, New York 2008, XI ss.

⁴⁵ Diffusamente in R. PENROSE, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, New York 1989.

approcci, a volte mutuamente esclusivi, testimoniano tutti l'ambizione di estendere il campo di interesse della fisica alla mente dell'osservatore, la quale può essere studiata, se non proprio in modo esclusivo o riduzionistico, quantomeno anche come un sistema fisico di elaborazione delle informazioni.

5. Il contributo della teoria dell'informazione

Un contributo determinante negli studi sull'informazione è stato offerto a partire dalla seconda metà del Novecento da alcuni autori che hanno posto le basi di una nuova disciplina, la teoria dell'informazione, che si riprometteva di studiare in modo scientifico il fenomeno della generazione e della trasmissione di informazioni nel contesto della comunicazione.

Già al termine degli anni Quaranta, Shannon, in due contributi dell'ottobre e del luglio 1948⁴⁶, poi, insieme anche con Weaver in un saggio del 1949⁴⁷, fondò una prima teoria dell'informazione, basata su un modello matematico che descriveva un sistema di generazione e trasmissione dell'informazione⁴⁸. «La teoria dell'informazione classica», infatti, «tratta di dati e informazioni come realtà oggettive, slegate dal significato»⁴⁹: l'informazione, si identifica con il contenuto del messaggio da comunicare. Per essa, allora, «si intende un contenuto di differenza relativa, di novità e di imprevedibilità di un messaggio intercorrente fra sistemi in relazione»⁵⁰. Nel modello di Shannon, una sorgente genera informazione, la trasmette tramite messaggi, un codificatore elabora i messaggi, tale segnale perviene tramite un canale – compatibilmente con la presenza di un rumore sul canale – all'osservatore che la decodifica, compiendo un'operazione inversa:

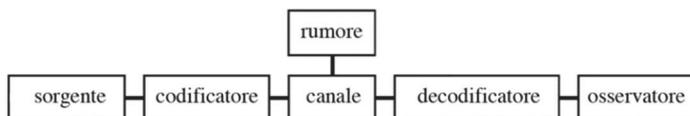
⁴⁶ Qui «compare per la prima volta il termine 'bit'» (Shannon, in *Enciclopedia della Matematica*, Roma 2013).

⁴⁷ Nel quale si pubblicava nuovamente il testo integrale dell'articolo del 1948, C.E. SHANNON, W. WEAVER, *The Mathematical Theory of Communication* (1949), Champaign 1964.

⁴⁸ Nella letteratura giuridica, v. G. RUFFO, *Rete e reti*, in M. DURANTE, U. PAGALLO, *Manuale di informatica giuridica e diritto delle nuove tecnologie*, Torino 2012, 20 ss.

⁴⁹ F. ROMEO, *Il dato digitale e la natura delle cose*, cit., 98, n. 24.

⁵⁰ *Informazione*, in *Enciclopedia on line*, Roma 2023.



In questo modello semplificato, come è chiaro, l'informazione è un *quid* (il contenuto è indifferente) che si trasmette da un termine a un altro, il cui trasferimento ottimale dipende dalle singole variabili considerate. La teoria dell'informazione, allora, è basata sulla teoria della probabilità e sulla statistica, che consentono di studiare le *chance* che il contenuto prodotto dalla sorgente giunga all'osservatore. Il livello semantico, quello del significato del messaggio, e il ruolo dell'osservatore non sono considerati. In quest'approccio quantitativo e oggettivo alla comunicazione, l'informazione è misurabile, è l'unità di misura minima è rappresentata, come è noto, da cifre binarie (*binary digits* donde *bits*⁵¹). In altri termini, «l'informazione di Shannon ci fornisce [soltanto] la struttura matematica per distinguere il segnale dal rumore per misurare la capacità e l'affidabilità»⁵².

La teoria in questione è stata notevolmente rivista e ampliata nel corso degli anni, ma occorre evidenziare la postura culturale che accompagna un simile approccio al tema dell'informazione. I lavori di Shannon (e Weaver), infatti, rappresentarono i primi più noti tentativi di studiare il fenomeno della comunicazione secondo un approccio oggettivo, quantitativo e basato su modelli matematici statistico-probabilistici. In questi studi, per 'informazione' s'intende qualcosa di completamente slegato dal suo significato, dal suo contenuto di verità, dalla riferibilità a un oggetto. Il dato, invece, è rappresentato dalla successione di messaggi, codificati nel sistema binario, e costituisce l'unità di misura dell'informazione.

Negli ultimi decenni l'*information theory* ha poi avuto notevoli sviluppi e il modello di Shannon si è rivelato semplicistico rispetto alla varietà tipologica delle comunicazioni, in particolare poiché esso assume una comunicazione tra due termini attraverso un unico canale (modello che ben può descrivere le comunicazioni radiotelefoniche e, ancora, i *mass-media* tradizionali, quali ad esempio la televisione, ma che mal si adatta alla comunicazione nei *media* contemporanei). Eppure, ciò che resta invariato nell'ap-

⁵¹ L'espressione, come ricorda SHANNON, è di J.W. TUKEY (sul punto, anche H.S. TROPP, *Origin of the Term Bit*, in *Anecdotes, Annals of the History of Computing*, 2, 1984, 152-155).

⁵² D.C. DENNETT, *Dai batteri a Bach. Come evolve la mente* (2017), trad. it. di S. Frediani, Azzate 2018, 149.

proccio della teoria della comunicazione e dell'informazione è che la pretesa riferibilità di un dato di un'informazione a un elemento di realtà sia una circostanza inconfidente o, comunque, del tutto trascurabile.

In altri termini, se nella fisica classica la riferibilità dei dati al mondo reale era un presupposto imprescindibile (e aproblematico), nella teoria dell'informazione, così come nell'informatica e nella fisica contemporanea la questione della riferibilità è diventata molto più problematica. D'altronde, questo risultato progressivo rispecchia, almeno in parte, lo stesso atteggiamento della cultura contemporanea di prudenza nei confronti delle epistemologie forti, se non proprio, nei confronti della "Verità".

Nei successivi sviluppi delle teorie dell'informazione, la questione venne saldamente ricompresa nell'orizzonte della computazione e dell'informatica, atteso che a valle degli anni Cinquanta si commercializzarono i primi *computer*, grazie alle innovazioni del decennio precedente riconducibili soprattutto a John von Neumann e Alan Turing, i padri dell'informatica moderna. È in questa cornice che dati e informazioni assumono la fisionomia che hanno tuttora.

In particolare, nella teoria algoritmica dell'informazione, che si è affermata nel corso degli anni Sessanta, si adotta una nozione procedurale per misurare la quantità di informazioni presente in un determinato oggetto⁵³. Fa così la sua comparsa una nozione, quella di algoritmo, largamente (e talora confusamente) impiegata in ogni trattazione sul tema dei dati. Il termine 'algoritmo' ha origine dal toponomastico del matematico arabo Muḥammad ibn Mūsa, al-Khuwārizmī, e a lungo indicò i procedimenti di calcolo fondati sulle cifre arabe. In informatica, indica un «insieme di istruzioni che deve essere applicato per eseguire un'elaborazione»⁵⁴, dunque, una sequenza finita di operazioni.

La teoria algoritmica dell'informazione misura l'informazione contenuta in singoli oggetti, e si occupa della relazione tra calcolo, informazione e ca-

⁵³ Cfr. A. SHEN, *Algorithmic Information Theory*, in *The Routledge Handbook of Philosophy of Information*, cit., 37 ss.; per questa teoria si può fare riferimento a tre lavori significative di Solomonoff, Kolmogorov e Chaitin degli anni Sessanta: R. SOLOMONOFF, *A Formal Theory of Inductive Inference*, in *Information and Control*, 7, 1964, 1 ss.; A. KOLMOGOROV, *Three Approaches to the Quantitative Definition of Information*, in *International Journal of Computer Mathematics*, 2, 1968, 157 ss. (traduzione inglese di un precedente contributo russo, apparso nel 1965 in *Проблемы передачи информации*); G. CHAITIN, *On the Length of Programs for Computing Finite Binary Sequences: Statistical Considerations*, in *Journal of the ACM*, 16, 1969, 145 ss.

⁵⁴ Voce *Algoritmo*, in *Vocabolario online*, Roma 2023.

sualità. Il contenuto informativo o la complessità di un oggetto possono essere misurati dalla lunghezza della sua descrizione più breve. La quantità di informazione in un singolo oggetto determinato, o la sua complessità di Kolmogorov, è definita come la lunghezza del più breve programma che produca l'oggetto come *output*. Semplificando, si può dire che tale lunghezza sia il numero di *bit* necessari per codificare un contenuto informativo. Ad esempio, per codificare l'alfabeto, tanto quello italiano a ventuno caratteri quanto quello inglese a ventisei, per tradurlo in linguaggio binario, è necessaria una combinazione di cinque *bit*, di modo che la coppia di valori possibili (zero o uno, appunto) generi trentadue possibili varianti ($2^5=32$, e $32 \geq 26$), mentre con sole quattro combinazioni si otterrebbero soltanto sedici combinazioni ($2^4=16$ e $16 < 32$). In un'ipotesi leggermente più articolata, si può immaginare di assegnare codificatori più brevi per i messaggi più frequenti, come nell'alfabeto *Morse* in cui alcune sequenze più semplici sono usate per codificare le lettere più adoperate. È chiaro, poi, che si possano immaginare casi sempre più complessi, in cui il *set* dei messaggi possibili non sia già predeterminato.

Ancora una volta, allora, la teoria algoritmica dell'informazione non è in grado di distinguere la qualità o il significato delle informazioni, ma si riferisce meramente alla loro dimensione oggettiva o quantitativa.

Questa prospettiva testimonia che dati e informazioni siano nozioni elaborate proprio a seguito di una torsione culturale avvenuta nel corso del Ventesimo secolo, in cui le questioni relative alla conoscenza (alla sua produzione e trasmissione) non sono state più affrontate con un approccio soltanto qualitativo, ma sono state scomposte in un processo di oggettivazione, con un approccio soprattutto quantitativo. Questa torsione, quindi, ha implicato una parziale perdita di interesse verso la veridicità o la riferibilità delle informazioni a realtà determinate, donde ne deriva il principale paradosso della questione definitoria dei dati e delle informazioni nel contesto della protezione dei dati *personali*: nella teoria dell'informazione, dove cioè questi concetti sono sorti o quantomeno si sono *ridefiniti*, la riferibilità o il contenuto dei dati sono elementi del tutto inconfidenti. Dati e informazioni, infatti, sono tasselli concettuali di un approccio culturale, se non proprio antropologico, che tratta la conoscenza come un fenomeno autonomo e quantitativamente osservabile, in cui l'aspetto semantico non ha più la centralità che invece caratterizzava i modelli epistemici passati. Dati e informazioni attraversano la realtà, ma non necessariamente la *ri-guardano*.

6. Informazione semantica (fattuale)

Ogni teoria dell'informazione che si fonda su un approccio quantitativo (statistico, probabilistico o, più generalmente, matematico), come osservato, si concentra sul fenomeno comunicativo nella sua oggettività più esteriore, osservando se e come un determinato contenuto informativo può raggiungere il destinatario, scomponendo le variabili che determinano il trasferimento di informazione, studiando le probabilità di successo o i limiti di tale fenomeno. In altri termini, la riferibilità delle informazioni a una determinata realtà, la loro rilevanza, l'affidabilità, la precisione o, su tutti, il loro significato non giocano alcun ruolo negli approcci in questione. La teoria dell'informazione di Shannon, per tutte, «ideò un modo per *misurare* l'informazione, indipendentemente da che cosa *riguarda*, come se si misurasse il volume di un liquidò, a prescindere dal liquidò in questione»⁵⁵.

Alla luce di questa considerazione, allora, è sorta nella filosofia dell'informazione (il campo di studi che ha concentrato il proprio interesse sulla relativa questione definitoria) la necessità di ricostruire una nozione semantica di informazione, ovvero una nozione in cui rilevasse anche il contenuto – appunto, semantico – delle informazioni relative a realtà fattuali o, in senso più ampio, la “giustezza” del contenuto informativo, eventualmente traslata nell'informazione stessa. In questo senso, si vorrebbe affermare che alcune stringhe di dati abbiano, accanto alla struttura, cioè alla loro sintassi, anche una semantica, un significato, tale da riferirsi a un contenuto di realtà⁵⁶. E tale circostanza, in parte estromessa dallo studio quantitativo dell'informazione, ambisce ad essere riconsiderata nelle

⁵⁵ D.C. DENNETT, *Dai batteri a Bach. Come evolve la mente*, cit., 116: «Immaginate qualcuno che si vanta di possedere enormi quantità di litri non sappia rispondere quando gli domandano: “Litri di che cosa-vernice, vino, latte, benzina?”».

⁵⁶ Cfr. L. FLORIDI, *Semantic Information*, in *The Routledge Handbook of Philosophy of Information*, cit., 44 ss. Si può fare riferimento anche a un lavoro di Y. BAR-HILLEL, R. CARNAP, *Semantic Information*, in *The British Journal for the Philosophy of Science*, 4, 1953, 147: «This deliberate restriction of the scope of Statistical Communication Theory was the heuristic value and enabled this theory to reach important results in a short time. Unfortunately, however, it often turned out that impatient scientists in various fields applied the terminology and the theorems of Communication Theory to fields in which ‘information’ was used, presystematically, in a semantic sense, that is one involving contents». Oppure, ancora, a J. HINTIKKA, *On Semantic Information*, in W. YOURGRAU, A.D. BRECK, *Physics, Logic, and History*, Boston 1970, 147 ss.

coordinate delle scienze sociali e del diritto. Cioè, non si potrebbe trattare di dati e informazioni, senza contestualmente estendere l'analisi alla questione della loro riferibilità alla realtà (benché tali concetti la rifuggano ontologicamente).

Non vi è però un accordo neppure sul significato possibile di una nozione di informazione che non trascuri il suo significato. Il termine 'informazione' è spesso usato per riferirsi a contenuti semantici esterni ai processi mentali (indipendenti dall'utente, nella terminologia informatica), incorporati in implementazioni fisiche come *database*, enciclopedie, siti *web*, programmi televisivi e così via, che possono essere variamente prodotti, raccolti, consultati ed elaborati⁵⁷.

Un primo approccio è quello relativo alla *Standard Definition of Information* (SDI), apparsa nel corso dei primi anni Novanta⁵⁸. Secondo questo approccio, i dati, intesi come unità comunicative basilari, possono produrre informazioni, allorché essi siano *rilevanti*, cioè purché sussista una correlazione funzionale tra il dato quale veicolo comunicativo e l'informazione quale porzione di comunicazione trasmessa. In primo luogo, la SDI presuppone che le informazioni non possono essere prive di dati, ma non specifica quali tipi di dati costituiscano – o producano – le informazioni. Tale approccio assume che ci sia un nesso tra i dati "significativi" e le informazioni da essi prodotti. La questione principale, negli approcci descritti dalle teorie dell'informazione semantica, è quella di stabilire quale nesso ci sia fra i dati, gli elementi di partenza oggettivamente descritti dalle teorie matematiche della comunicazione, e l'informazione, intesa come porzione di conoscenza desumibile dal loro trattamento. Il dato sarebbe il precipitato sintattico, mentre l'informazione punterebbe al risultato semantico.

In questa prospettiva, è possibile innanzitutto offrire uno schema ordinativo del panorama dei dati, secondo una classificazione che rispecchi i loro aspetti funzionali. In particolare, secondo classificazione offerta da Floridi, i dati possono essere di quattro tipi: dati primari; metadati; dati operativi; dati derivati. I primi sono ciò che normalmente percepiamo come i dati principali archiviati in un *database*. I metadati sono indicazioni secondarie sulla natura dei dati primari. I dati operativi riguardano l'uti-

⁵⁷ L. FLORIDI, *Is Semantic Information Meaningful Data?*, in *Philosophy and Phenomenological Research*, 2, 2005, 351-352.

⁵⁸ Cfr. D. ISRAEL, J. PERRY, *What is Information?*, in HANSON (a cura di), *Information, Language and Cognition*, Vancouver 1990, 1 ss.; oppure, ampiamente in K. DEVLIN, *Logic and Information*, Cambridge 1991.