

GenOA week

International Open Access Week
25 - 29 October 2021

**La problematica dell'Open Data e dell'Open Access nell'ambito
dell'Open Science: una prospettiva storica ed esempi concreti**

Dario Menasce – INFN Milano Bicocca

Open Science e open access 1.

lunedì 25 ottobre 2021

La problematica dell'Open Data e dell'Open Access nell'ambito dell'Open Science

Una prospettiva storica ed esempi concreti



Dario Menasce
INFN Milano-Bicocca



Le problematiche di cui accennerò in questo seminario sono relativamente nuove e fortemente intrecciate alla moderna modalità di produzione e fruizione della conoscenza (e quindi alla tecnologia più attuale), ma hanno altresì radici profonde nella storia della cultura occidentale di cui facciamo parte.

Eviterò pertanto di addentrarmi in tecnicismi e dettagli di come l'Open Data può essere implementato, di quali siano le implicazioni legali, di privacy piuttosto che non di diritto d'autore (vi sono colleghi più titolati di me a parlarne). Mi limiterò ad esporre il punto di vista dello scienziato che reputa la prassi dell'Open Data come una componente fondamentale del modo di fare scienza che non solo produce risultati utili, ma che accelera e fortifica il processo di formazione e consolidamento delle teorie e del loro confronto sperimentale con ciò che la natura esibisce o nasconde.

Inizierò con un excursus storico che illustra lo sviluppo nei secoli della necessità e della validità di questa prassi anche quando non era in uso in modo formale ma in qualche modo già sotto traccia.

Passerò ad esaminare un paio di esempi pratici di utilizzo della modalità Open Data (uno, in particolare, legato all'attuale crisi pandemica che ha coinvolto tutto il nostro pianeta).

Concluderò con alcune osservazioni circa il meccanismo di retroazione positiva su cui si basa la scienza e di cui l'Open Access e l'Open Data sono ingranaggi fondamentali

L'accesso aperto a pubblicazioni scientifiche e ai corrispondenti dati è una problematica che coinvolge moltissimi aspetti multidisciplinari.

Come anticipato eviterò di addentrarmi in dettagli tecnici o sottigliezze giuridiche, preferendo concentrarmi sul perché questa pratica sia utile (direi fondamentale) per la preservazione delle nostre conoscenze attuali e per un continuo progresso di esse.

La storia del mondo occidentale (e non solo quello) è ricca di esempi sul perché la scienza progredisca solo a certe condizioni e vorrei cogliere l'occasione per esaminare questi punti.

Questo mi permetterà di inquadrare il problema nella cornice che più mi è familiare, quella dello scienziato che necessita di dati e del sapere accumulato dai suoi colleghi per fare ulteriori progressi nella strada della conoscenza.



La *Scuola di Atene* di Raffaello Sanzio, 1509-1511 Stanza della Segnatura, Palazzi Apostolici in Vaticano

Il sapere, per formarsi e crescere (quello scientifico in particolare) necessita di un certo numero di ingredienti

- Qualcuno che abbia tempo di dedicarsi ad acquisire conoscenza
- Qualcuno (possibilmente lui stesso) che renda persistente nel tempo i suoi studi e i risultati acquisiti
 - la tradizione orale non è sufficiente perché dà adito a fraintendimenti nel passaggio da un relatore al successivo
 - fondamentale è preservare gli studi in forma scritta (possibilmente in una lingua di vasta diffusione)
- Gli studi non bastano: occorre preservare i dati su cui sono basati questi studi e la metodologia con la quale sono stati raccolti e analizzati.
- Occorre poi che sia dato l'accesso a tutto ciò alla più vasta comunità possibile
- Si tratta di un meccanismo a forte feedback positivo

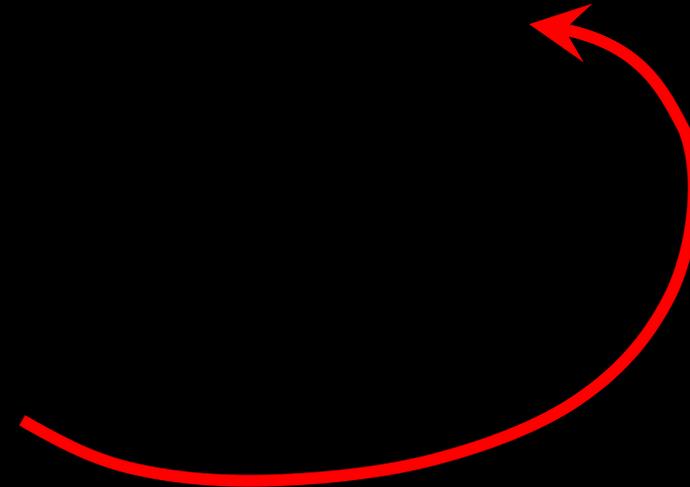




Osservazione
di un
fenomeno

Giudizio favorevole
di esperti del settore

Un ciclo virtuoso



Come si è sviluppato nel tempo questo processo?

Quali sono i primi segni della capacità umana di **preservare nel tempo** il proprio pensiero?

La paleoantropologa Genevieve von Petzinger ha identificato, studiando pittogrammi in moltissime caverne naturali d'Europa, 32 simboli dallo stesso aspetto grafico

The Geometric Signs of Ice Age Europe



L'uniformità del loro aspetto suggerisce che questa forma di comunicazione e la capacità di trasmettere e preservare messaggi su fatti e cose oltre il breve momento temporale in cui essi accadono e si manifestano, sia più diffusa e antica di quanto si pensi comunemente (tra i 22000 e i 44000 anni fa)

https://www.ted.com/talks/genevieve_von_petzinger_why_are_these_32_symbols_found_in_ancient_caves_all_over_europe

Celeberrime sono le immagini della Grotta di Lascaux nel Perigord, risalenti a circa 17000 anni fa.



Il messaggio convogliato da questi pittogrammi richiedeva una modalità di interpretazione delle immagini che fosse **condivisa** tra tutti coloro che ne avrebbero "*letto il messaggio*" affinché questo potesse essere compreso nel suo significato: serviva un *dizionario* semiologico.



A un certo punto della storia si ha però un'importante innovazione: la scrittura, ossia il passaggio da immagini a codici di rappresentazione più astratti ed anche più sintetici

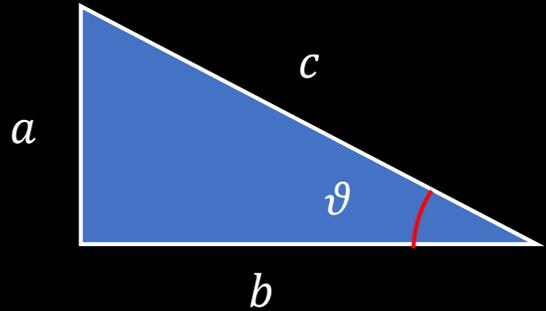
Le tavolette di argilla babilonesi



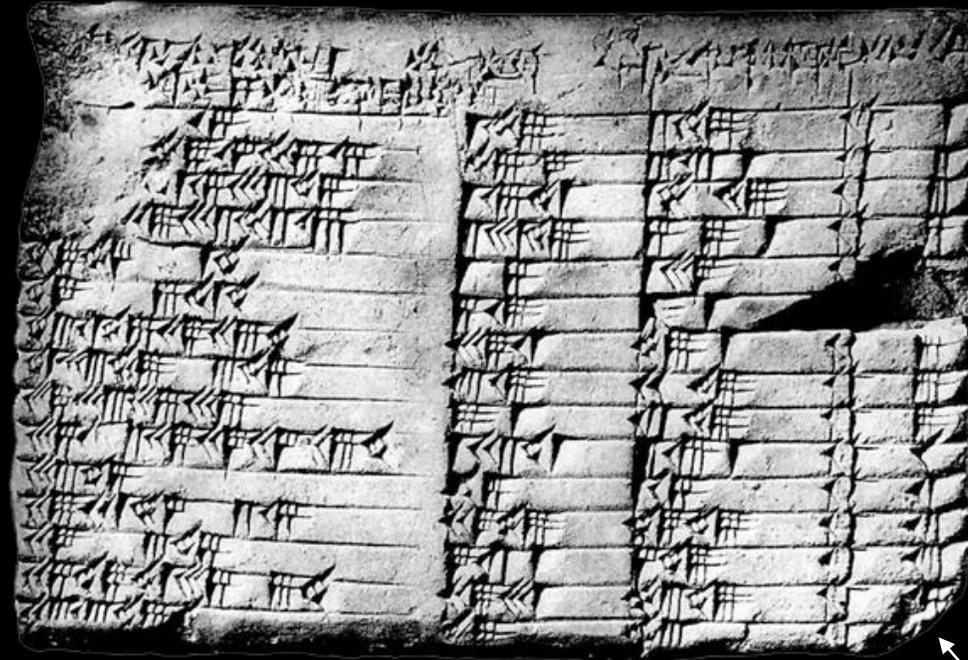
Tell al-Uhaymir , Babil Governorate, Iraq, circa 3500 a.C. (medio periodo Uruk)

Inizia il passaggio alla proto-scrittura sillabica, ma la cosa rimarchevole di questi reperti è il mezzo su cui essa si sviluppa: tavolette di argilla cotte al fuoco che meglio permettono una **conservazione** nel tempo del manufatto e una sua relativamente agile **diffusione** e **fruizione**.

I primi segni dello sviluppo di conoscenze matematiche, la tavoletta di argilla della Plimpton Collection (Plimpton 322)



Plimpton 322



Qual'era il suo scopo?

Tavole trigonometriche?

Esercizi per studenti?

La tabella contiene qualche errore di calcolo

$$\operatorname{tg}\vartheta = a^2/b^2$$

$$a^2$$

$$b^2$$

Numeri da 1 a 15

$$c^2 = a^2 + b^2$$

I valori in tabella sono per incrementi costanti di lati, non di angoli, come in una tavola trigonometrica moderna.

Cominciano a comparire testi di utilità pratica, antesignani dei moderni manuali di matematica, ingegneria o contabilità: pur essendo sempre possibile ricalcolare questi valori in caso di necessità, la conoscenza necessaria per eseguire il calcolo era a disposizione di pochi. Conoscenza "*nascosta*", mancanza di meta-dati descrittivi.

Iniziano a comparire vaste raccolte del sapere. È stata ritrovata una biblioteca con 30000 tavolette di argilla in molte lingue, tra cui la completa Saga di Gilgamesh: alcuni strati privilegiati delle antiche popolazioni iniziano ad apprezzare il valore del sapere e della cultura in generale (in particolare di quella artistica, ma anche di quella scientifica e filosofica passando da quella militare)



La Biblioteca di Ashurbanipal Mesopotamia 1500-539 BC Gallery, British Museum, London, England, UK

La filosofia greca e il periodo classico

La civiltà greca si reggeva tramite una confederazione di Polis, entro le quali era possibile, per gli abitanti, esercitare un notevole grado di libertà dialettica e di contraddittorio.

Questo clima sociale favorì moltissimo la crescita intellettuale ottenuta grazie al *libero scambio di idee*, opinioni e conoscenze acquisite.

Il sopraggiungere dell'Impero romano ha interrotto questo atteggiamento positivo verso la dialettica: in qualche modo l'impero, per reggersi, non poteva tollerare la totale libertà di espressione e il progresso intellettuale che questa comporta.

Di fatto, con la diaspora degli studiosi da Alessandria, indotta dai membri della tarda dinastia dei Tolomei, inizia la perdita di molte delle conoscenze acquisite in campo scientifico e filosofico.

Il tempo ha poi agito da filtro, non sempre in modo positivo...



Aristarco di Samo (~320 a.C.) intuisce e propone un modello eliocentrico del sistema solare

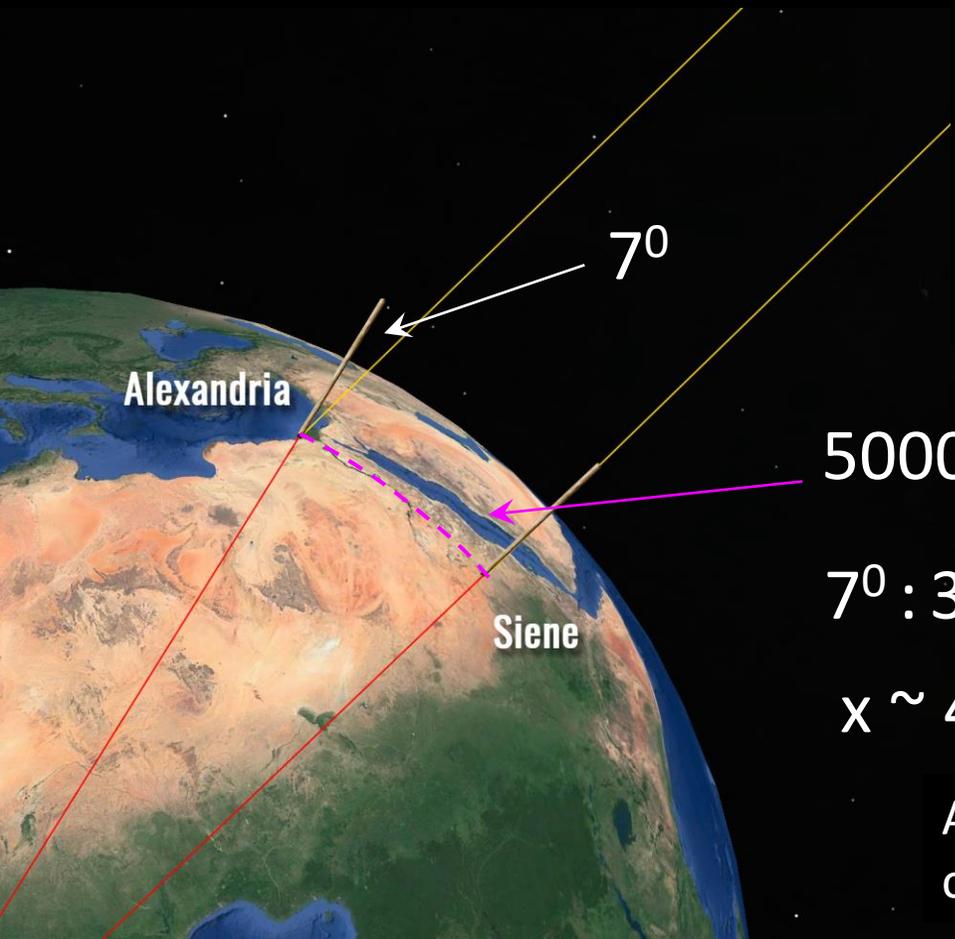
Eratostene di Cirene (~200 a.C.) **misura** la circonferenza della Terra (tardo periodo classico)

Inizia con **un'ipotesi di lavoro**: la Terra è sferica (si tratta di sapere tramandato)

Poi escogita un **esperimento** per misurarne le dimensioni

Lo fa usando le proporzioni!

Questa conoscenza sul metodo del calcolo da eseguire per giungere al risultato scompare nel medioevo: **si era persa la cognizione** che un disegno su carta potesse avere una precisa corrispondenza concettuale con oggetti fisici di dimensioni macroscopiche come la Terra stessa (principio di corrispondenza). Di più, era andato perso il concetto di "*corrispondenza*" tra enti ideali (geometria) e enti reali.



5000 stadi (~800 km)

$$70^{\circ} : 360^{\circ} = 5000 : x$$

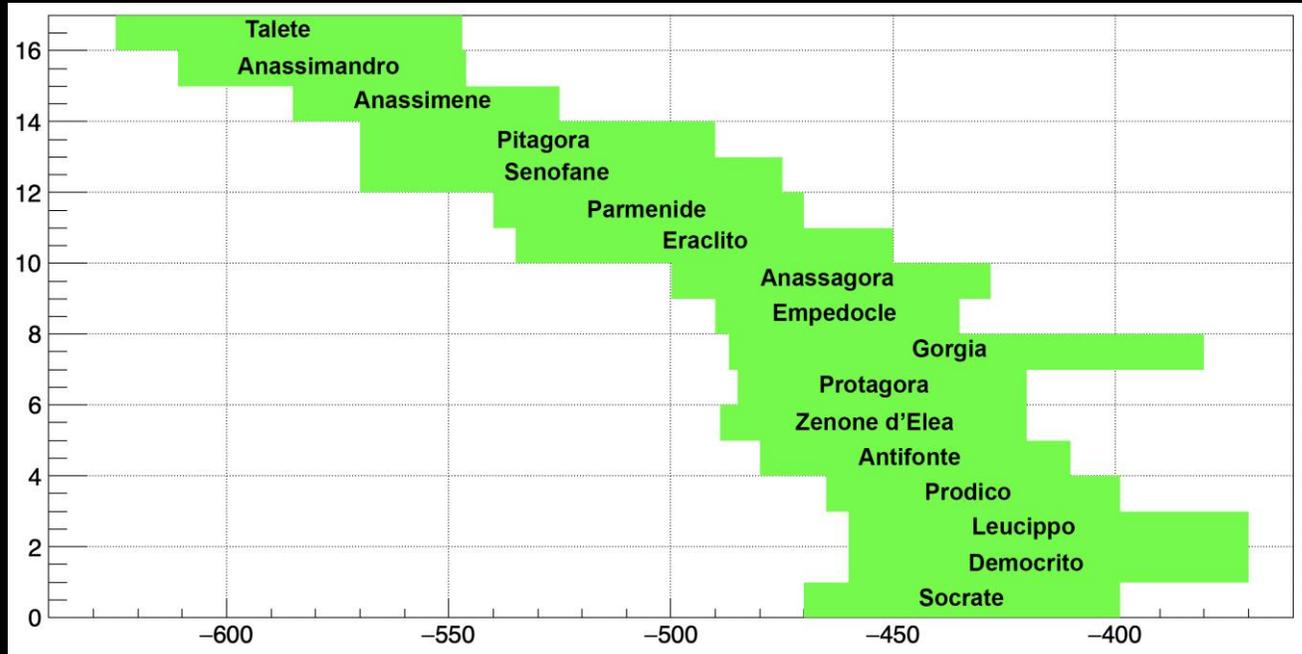
$$x \sim 41000 \text{ km}$$

La geometria permette di trovare soluzioni a problemi nuovi grazie al metodo deduttivo e la corrispondenza con enti fisici permette di dedurre conclusioni anche nel mondo reale (modo di fare scienza moderno)

Ai posteri rimaneva quindi la sola conoscenza del risultato, ma non la capacità di ottenerlo, il metodo necessario per riprodurlo...

Il periodo ellenistico

L'età d'oro della nascita del sapere scientifico: dai presocratici, al periodo classico fino alla distruzione della Biblioteca di Alessandria



Nelle polis greche fioriscono innumerevoli scuole di pensiero ma il polo di aggregazione di queste correnti filosofiche si stabilizza in Egitto, attorno alla grande ***Biblioteca di Alessandria***

Si trattava, per certi versi, dell'equivalente del moderno CERN di Ginevra: una "*centrifuga*" di cervelli, di conoscenze acquisite, di modi di pensiero e di approcci allo studio.

Aristotile (-384,-322)

Euclide (-325,-270)

Fondazione della Biblioteca (~300 a.C.)

In quel periodo e in quel luogo si pongono le basi della cultura occidentale: accumulatisi nei secoli, vengono preservati testi e conoscenze per i posteri.

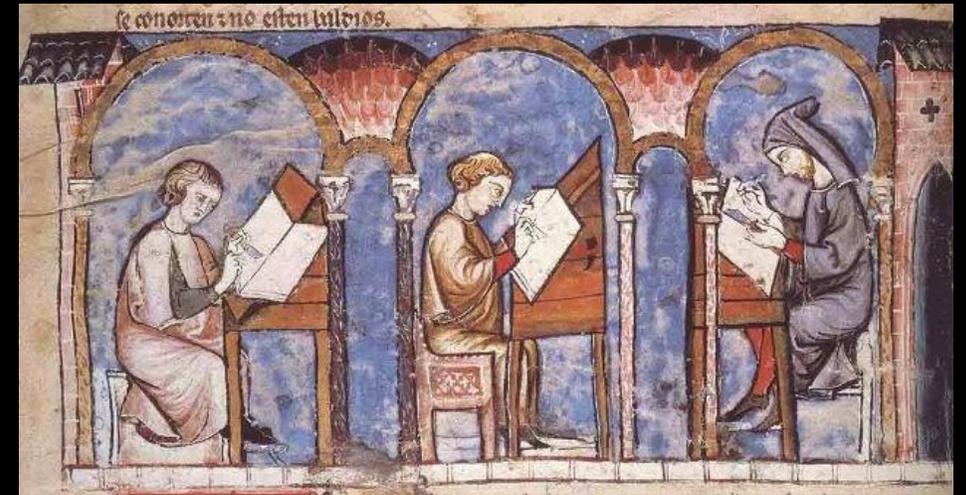
Con la sua caduta, nel periodo di Cesare, si completa però il declino, in atto già da molti anni, con la perdita di buona parte delle conoscenze acquisite: la trascrizione dei testi è spesso fatta da chi non ha conoscenza approfondita della materia e si limita a riportarne le conclusioni. Si perde la possibilità di ripercorrere la strada che porta alle conclusioni degli studiosi !

Abbiamo quindi individuato una prima serie di elementi collegati ai concetti di "*conoscenza*" e di "*progresso*"

- La disponibilità (la più diffusa possibile) di testi
- L'accesso ai medesimi da parte della maggior parte possibile di potenziali lettori
- La preservazione di tutto questo patrimonio, in modo che si possa "*accumulare*" nel tempo

Con il crollo dell'Impero romano e l'inizio del medioevo la diffusione della cultura non sparisce, come spesso considerato, ma se ne modificano la produzione e la sua fruizione.

Il punto importante è che la copiatura dei testi antichi trasmette solo i risultati, non il *metodo* (o anche solo i dati) con cui questi erano stati ottenuti. Si trasmette quindi un sapere "*codificato*", privato in modo sostanziale della possibilità di ulteriore progresso.



Flavio Magno Aurelio Cassiodoro, politico e studioso romano, fonda, nel 554 d.C. il monastero del Vivarium.

Si trattava di un monastero con caratteristiche diverse da quelli tradizionali, che erano all'insegna di "*Ora et Labora*" di origine benedettina: di nuovo, una sorta di CERN ante litteram (più conservazione che innovazione).

Lo scopo del Vivarium era infatti la copiatura, la conservazione lo studio degli antichi testi per come sopravvissuti e tramandati dal periodo greco (ma da questo processo di gestione del patrimonio culturale pregresso era in genere esclusa o marginale la produzione di **nuovi** testi: mancava il necessario ambiente culturale).

Quest'attività era condotta dai monaci amanuensi, una classe di monastici specializzati nelle varie fasi di questo processo di manipolazione dei testi, essenzialmente di classici e di patristica.

È interessante considerare le diverse specializzazioni di questa classe monastica e le conseguenze di questa divisione del lavoro nel contenuto delle opere tramandate.

Gli amanuensi erano attivi all'interno del monastero nello "*scriptorium*", luogo attrezzato alla copiatura e ben illuminato (una sorta di laboratorio), dove agivano tre diverse tipologie di lavoratori:

La prima fase di copiatura era affidata ai "*copiatores*", monaci incaricati della copia vera e propria dei testi. Purtroppo però si trattava spesso di monaci semianalfabeti, in grado unicamente di copiare in modo quasi meccanico i testi (i salari erano buoni e attiravano anche i contadini)



Una volta copiato il testo vero e proprio intervenivano i *correctores*, un ristretto numero di monaci di più alto livello culturale, col compito di rettificare gli errori (ruolo fondamentale ma fonte di gravi errori di interpretazione del sapere antico nella sua codificazione scritta).

Vi erano infine i *miniatores*, incaricati delle illustrazioni: all'inizio si occupavano solamente del primo carattere di ogni pagina, tipicamente prodotto in rosso (*minio*, da cui *miniatores*), ma in seguito furono autori anche di quelle splendide miniature che illustrano i testi.

Di nostro interesse qui sono i *correctores*: si tratta di un'attività analoga, per certi versi, a quella dei moderni *referee* di un articolo

Era infatti ben noto già allora il problema della propagazione degli errori nella procedura di copia dei testi. Nel corso dei secoli si era persa la capacità di ricostruire le linee di pensiero e soprattutto le competenze tecniche necessarie a verificare le scoperte di carattere scientifico fatte dagli autori originali e gli amanuensi non avevano la cultura necessaria a corroborare i testi con le parti mancanti.



Un esempio classico di propagazione degli errori di trascrizione è la celebre frase:

"È più facile per un cammello passare attraverso la cruna di un ago, che per un ricco entrare nel regno di Dio"
(Mc 10,25; Mt 19,24; Lc 18,25)

Esistono diverse ipotesi sul significato di una frase tanto sibillina

- La parola aramaica גמל , *gamal* può significare, infatti, sia *"cammello"* sia *"grossa corda"*
- Un'altra ipotesi è che l'errore di traduzione sia dovuto a San Gerolamo, che tradusse in latino buona parte dell'Antico Testamento greco e successivamente l'intera Bibbia, dando vita alla Vulgata. Egli, erroneamente, potrebbe aver trascritto *"kamelos"* (cammello) al posto del molto simile *"kamilos"* (grossa fune o gomina di nave).

Un aspetto ironico di questa propagazione degli errori nella traduzione dal latino è la parola stessa *"tradurre"* che deriva da una errata comprensione del termine latino *traductum* da parte dell'umanista Leonardo Bruni: il termine, che significava *"introdotto"*, fu scambiato per il participio *translatum* e prese ad equivalere sistematicamente, in italiano, alla parola tradotto, nel senso di trasferito da una lingua all'altra.

La mancanza di un metodo sistematico di preservazione, cura e pubblico accesso alla produzione di sapere degli eruditi in ogni campo ha portato alla perdita di molte conquiste già raggiunte anche in altri campi.

La prospettiva



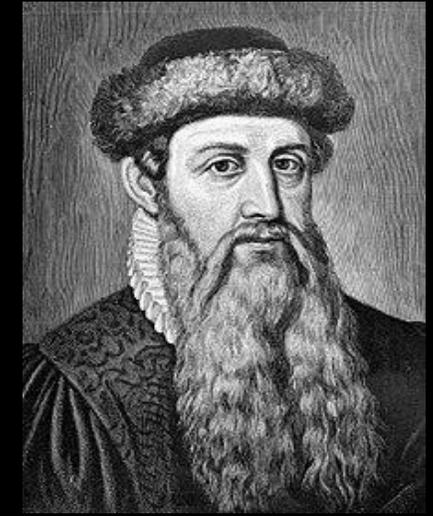
Casa di Marco Lucrezio Frontone - Pompei



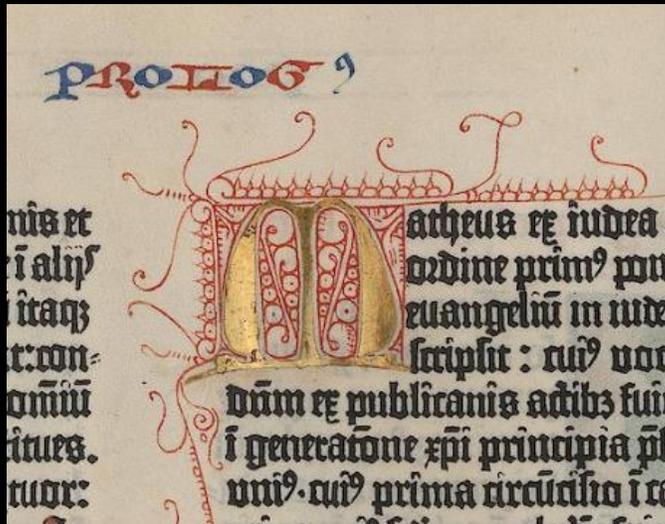
Walter de Milemete - 1326-27

La copiatura manuale dei testi comportava quindi tre ordini di problemi:

- Errori di interpretazione accumulati nel tempo
- Costi elevatissimi e quindi scarsa diffusione del sapere
- Concentrazione delle fonti scritte in mano della Chiesa con conseguente censura su ciò che non rispettava la dottrina corrente (chiusura all'innovazione)

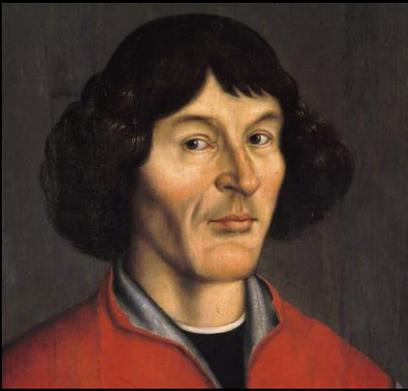


Lo strumento che per la prima volta permette di intravedere una svolta in questo deprecabile stato di cose è l'invenzione della stampa a caratteri mobili di Gutenberg (circa 1450)



I costi di produzione diminuiscono (anche se non con un crollo verticale, occorrerà tempo per la vera diffusione popolare con la nascita di un vero e proprio mercato editoriale)

Ma la cosa fondamentale è che la diffusione dei testi decolla in maniera esponenziale: grazie all'atmosfera rinascimentale si ha una riscoperta del mondo classico e rifiorisce la passione per il sapere, quello scientifico incluso



Chi però innesca la vera rivoluzione in campo scientifico è Copernico, seguito a breve da Galileo.

Copernico pubblica nel 1543 il *De Revolutionibus Orbium Cœlestium*, in cui propone il modello eliocentrico.

Pur essendo una ri-scoperta di concetti già considerati in periodo greco, viene considerata un'innovazione dirompente, tanto che da allora la parola *rivoluzione* assume il nuovo significato che più comunemente oggi le diamo.

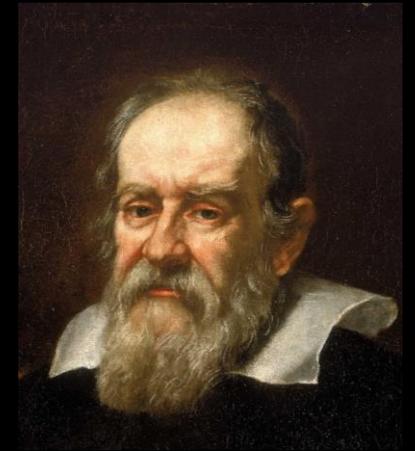
È grazie a questo testo che Galileo innesta la propria di rivoluzione, sostanzialmente creando (o forse riscoprendo?) il metodo scientifico:

Io stimo più il trovar un vero, benché di cosa leggiera, che 'l disputar lungamente delle massime questioni senza conseguire verità nissuna

In quegli stessi anni si consolida una dinastia che farà la storia dell'editoria scientifica: gli Elsevier.

Lodewijk Elzevir fonda la casa editrice Elzevir a Leida nel 1583, innovando totalmente un certo tipo di editoria:

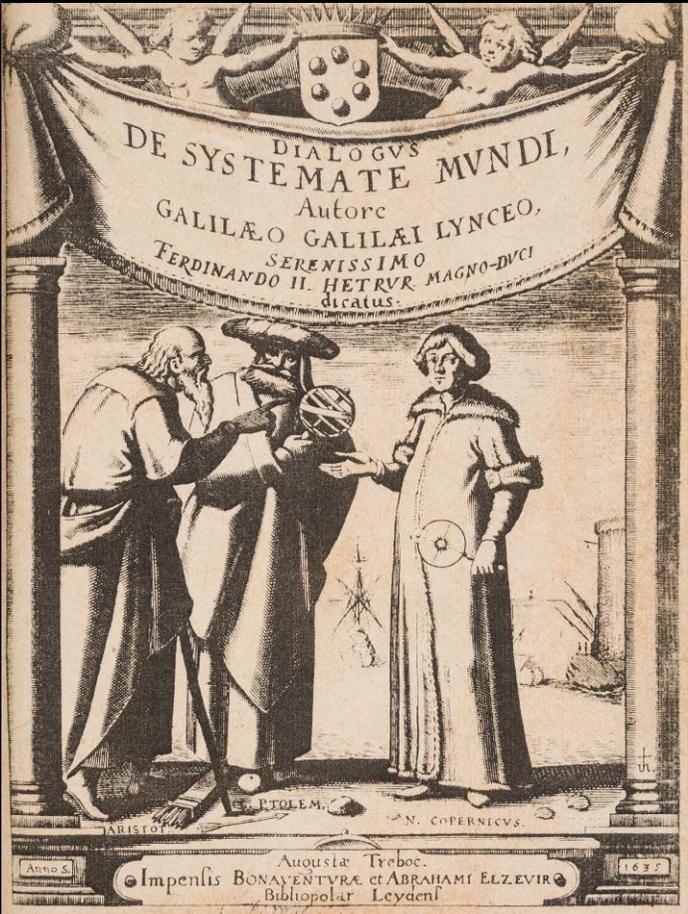
contrariamente all'uso allora in voga, il formato dei libri non si focalizza più sull'aspetto estetico e la gradevolezza grafica, concentrandosi invece sulla fruibilità e leggibilità (ridotti costi inclusi)



Inizia l'epoca della letteratura scientifica professionale con una serie di testi che hanno fatto la storia della scienza



PTOLOMAEUS, *Cl.Theatri geographiae véteris*. 1618-1619



Augusta Treboc.
Impensis BONAVENTURÆ et ABRAHAMI ELZEVI
Bibliopolar Leydensi

DISCORSI
E
DIMOSTRAZIONI
MATEMATICHE,
intorno à due nuoue scienze
Attenenti alla
MECANICA & i MOVIMENTI LOCALI,
del Signor
GALILEO GALILEI LINCEO,
Filosofo e Matematico primario del Serenissimo
Grand Duca di Toscana.
Con una Appendice del centro di gravità d'alcuni Solidi.



IN LEIDA,
Appresso gli Elsevirii. M. D. C. XXXVIII.



Questo nuovo approccio alla diffusione del sapere contribuisce in modo determinante al progresso scientifico.

- La definizione della corretta metodologia di indagine dei fenomeni naturali (e l'aumento del numero di studiosi in quel campo) rimedia al problema della pedissequa copia di testi mal compresi e non digeriti per il loro vero significato, riducendo di molto la trasmissione errata dei concetti basilari
- L'aumento del numero di testi prodotti grazie alla stampa a caratteri mobili facilita la nascita di imponenti biblioteche, recuperando il problema della preservazione del sapere su una scala del tutto nuova.
- Tornano però a porsi le basi, nel contempo, per un mercato monopolista che limita, in qualche modo, il libero accesso ai testi
- Non viene ancora per nulla affrontato il problema della preservazione dei dati scientifici



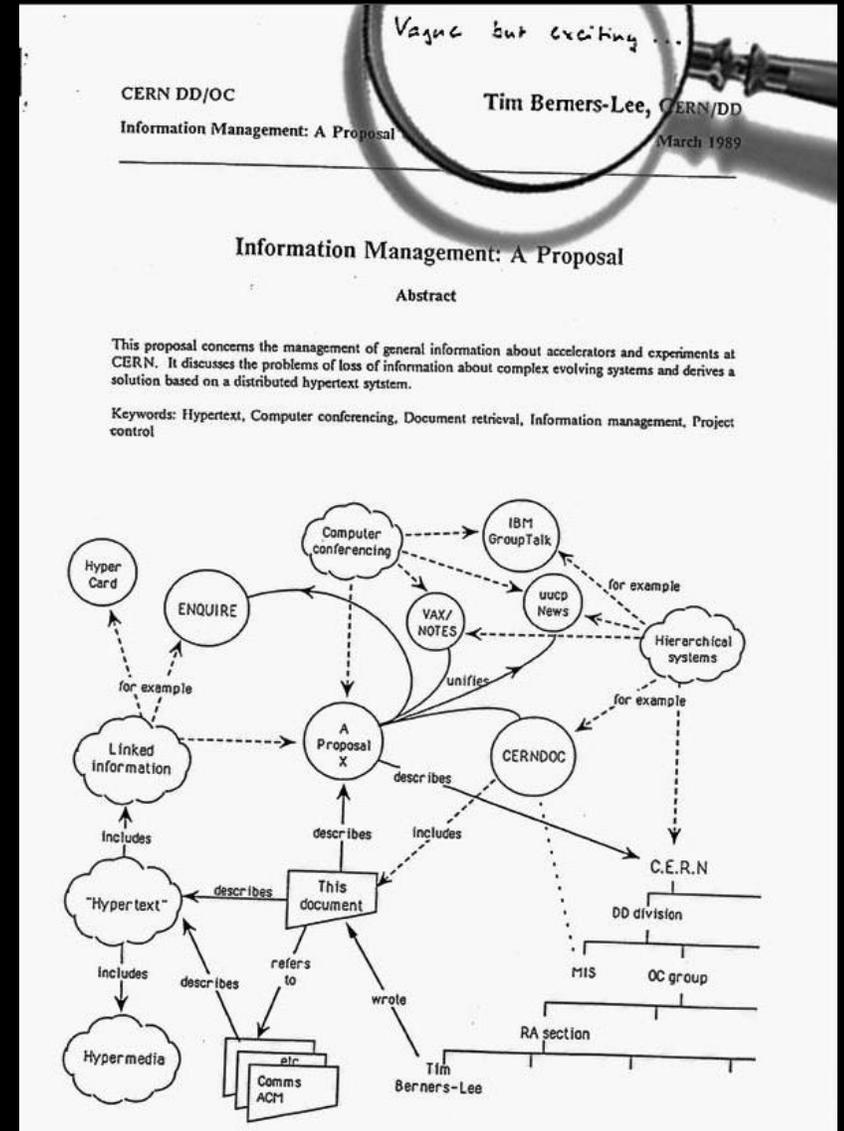
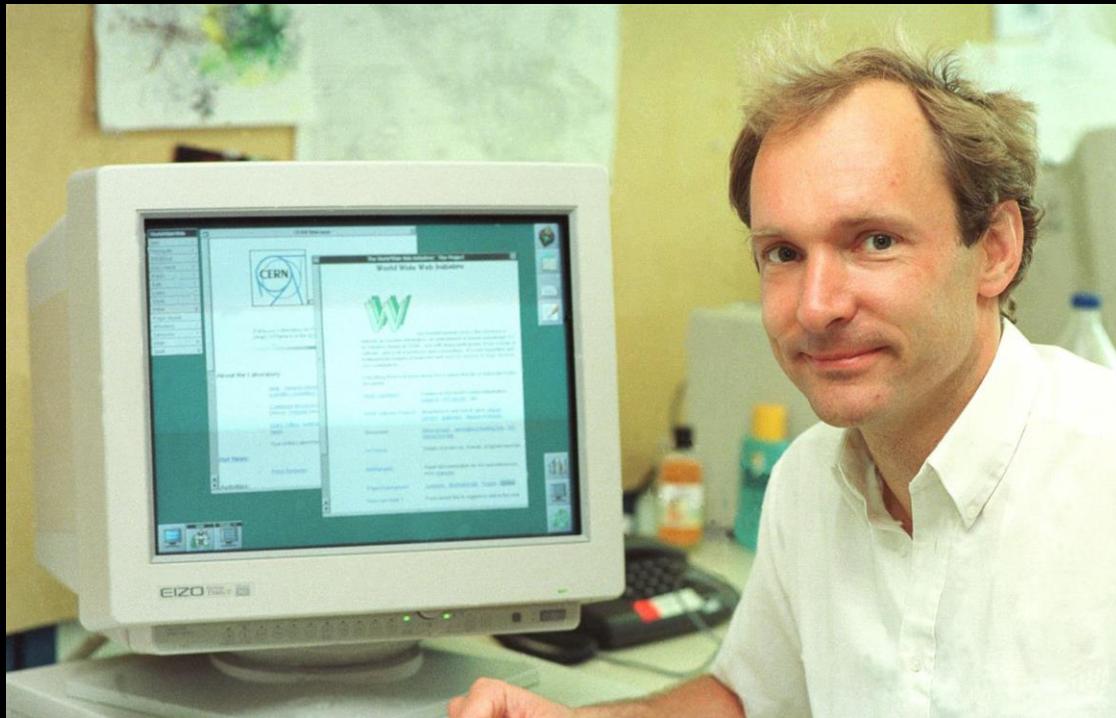
Trinity College Library – Dublin - Ireland

La situazione si stabilizza sino all'innesco dell'ennesima rivoluzione tecnologica: la nascita del Web

Tim Berners Lee, allora al CERN di Ginevra, idea un protocollo per la gestione dei dati trasmessi via INTERNET (protocollo già in uso da anni).

La sua prima proposta è del 1989, ma viene considerata troppo vaga dal management del CERN. Tim non desiste.

Il protocollo www viene realizzato in pochi mesi e la sua prima implementazione consiste in una schermata in modalità *command-line*



WORLD WIDE WEB

The WorldWideWeb (W3) is a wide-area hypermedia[1] information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary[2] of the project, Mailing lists[3] , Policy[4] , November's W3 news[5] , Frequently Asked Questions[6] .

What's out there?[7]Pointers to the world's online information, subjects[8] , W3 servers[9], etc.

Help[10] on the browser you are using

Software Products[11] A list of W3 project components and their current state. (e.g. Line Mode[12] ,X11 Viola[13] , NeXTStep[14] , Servers[15] , Tools[16] , Mail robot[17] , Library[18])

Technical[19] Details of protocols, formats, program internals etc

<ref.number>, Back, <RETURN> for more, or Help: █

W3 SOFTWARE

This is a list of products related to the WWW project[1] with a link to a status summary of each one. For list of previous milestones, releases, etc, see the "Features" list for each product, linked to the wish lists. See also: getting the files[2] .

Common Library[3] The common code library is the basis for most browsers. It contains the network access code and format handling.

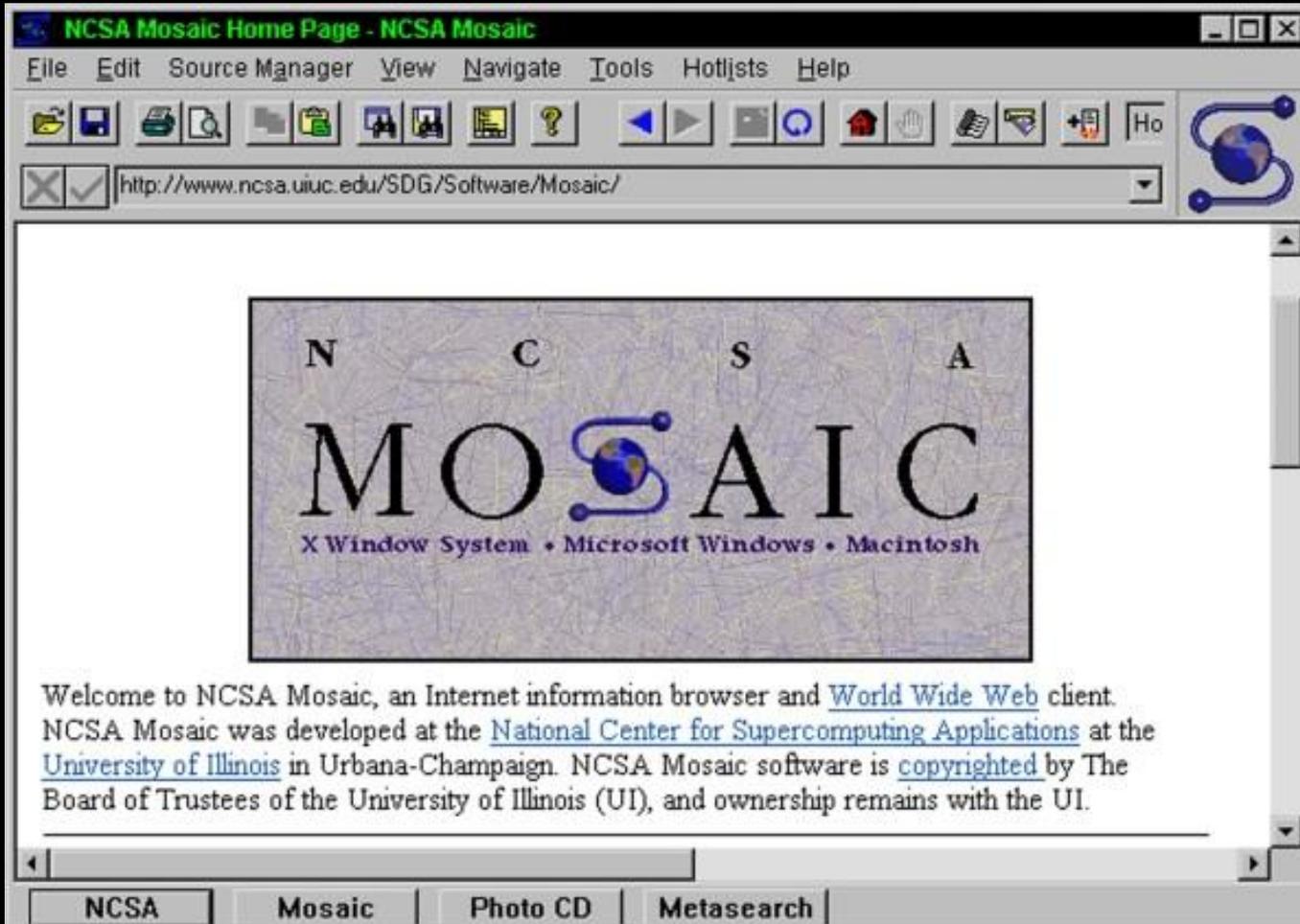
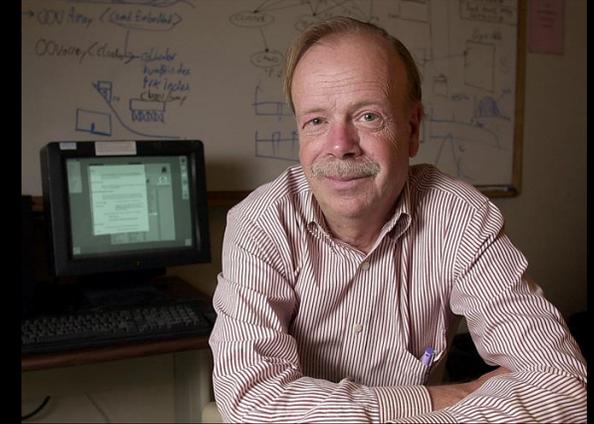
CLIENTS

LINE MODE BROWSER[4] THIS PROGRAM GIVES W3 READERSHIP TO ANYONE WITH A DUMB TERMINAL. NOT AS FLASHY AS A WINDOWS IMPLEMENTATION, IT COVERS A WIDE CLASS OF USERS WHO DO STILL NOT HAVE WINDOW FACILITIES, AND IS A GENERAL PURPOSE INFORMATION RETRIEVAL TOOL.

CURSES BROWSER[5] ASSUMES A CHARACTER-GRID TERMINAL WITH CURSOR ADDRESSING, AND PROVIDES A FULL-SCREEN INTERFACE TO THE WEB.

Verso la fine del 1990, dopo un viaggio a Fermilab, prende accordi con l'Università di Urbana nell'Illinois per la realizzazione del primo **browser**, Mosaic.

Paul Kunz dello Stanford Linear Accelerator Center si entusiasma del progetto dopo una visita al CERN e installa un primo server alla biblioteca scientifica del laboratorio.



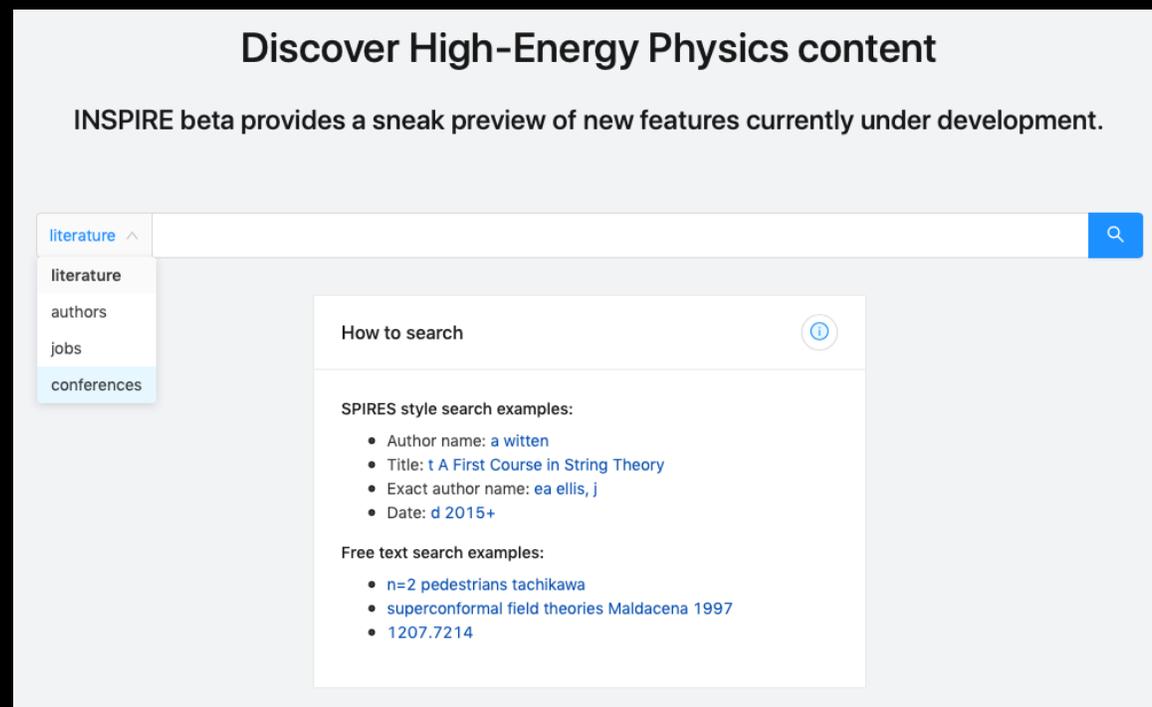
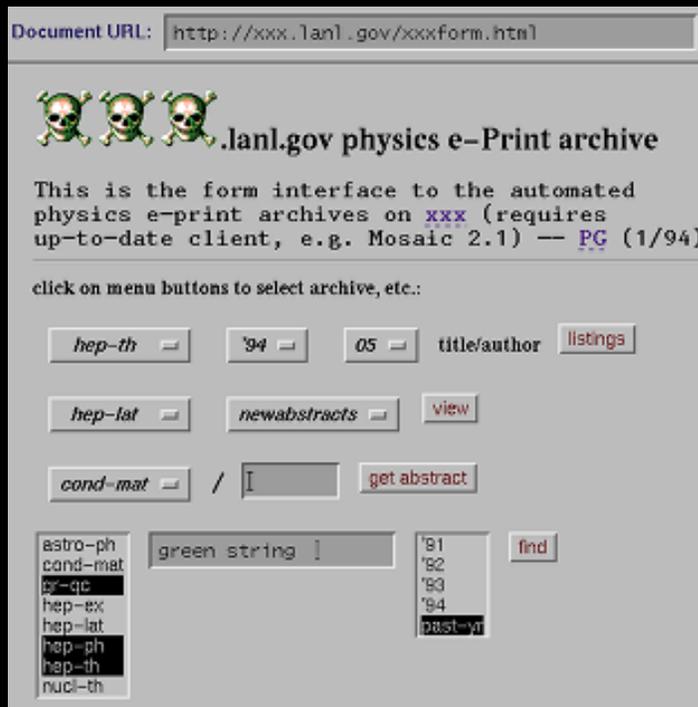
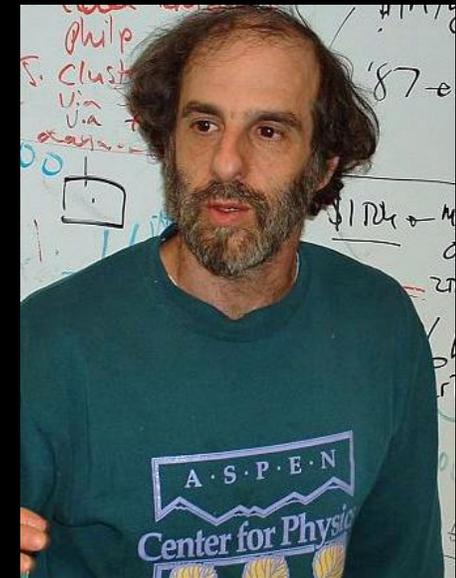
La bibliotecaria Louise Addis coglie appieno le possibilità date dal nuovo protocollo e crea il primo programma che può essere eseguito in remoto: un accesso alle risorse editoriali della biblioteca.

Al progetto contribuiscono lo specialista di software George Crane e il fisico delle particelle Tony Johnson, una sinergia tra le due discipline che si è mostrata molto fruttifera nei decenni



Il vero salto di qualità, per quanto riguarda gli articoli scientifici, lo fa però Paul Ginsparg, dei laboratori di Los Alamos (sempre fisica)

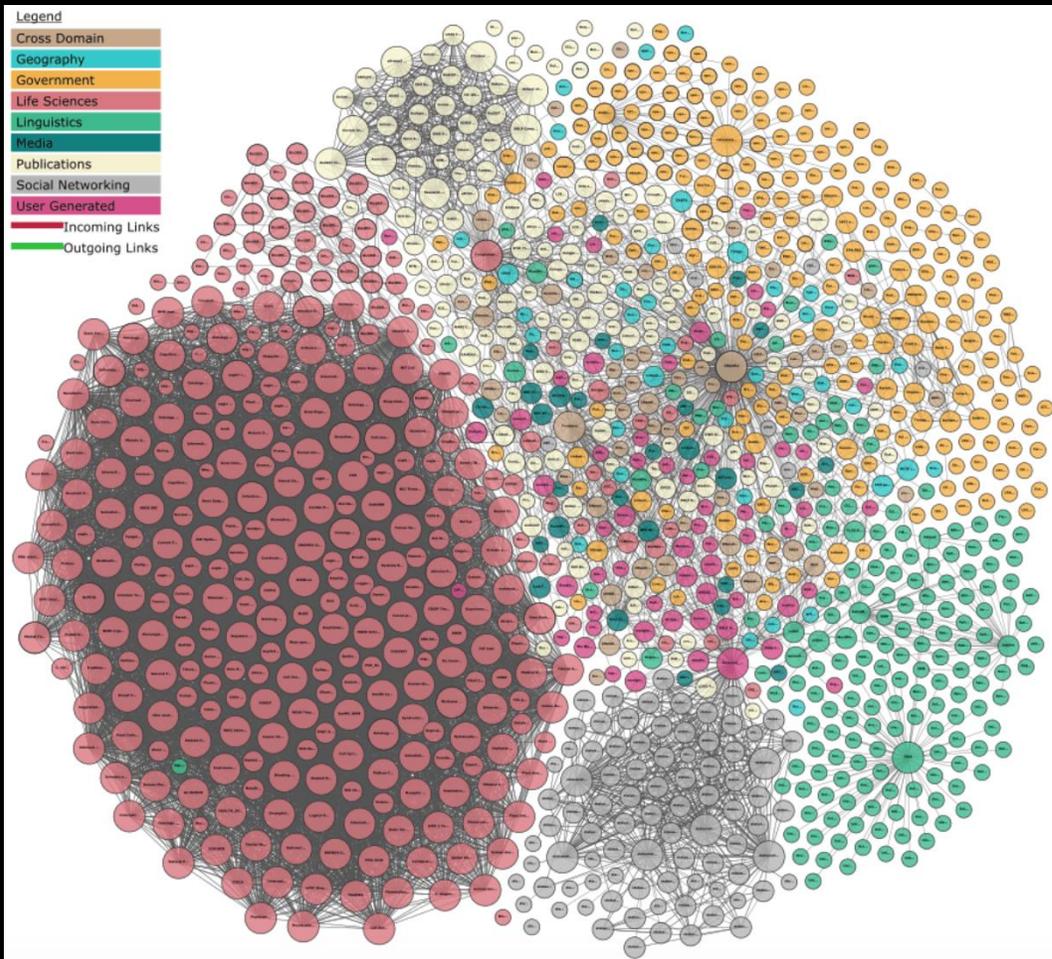
Crea **arXiv**, la prima biblioteca virtuale di pre-print (versioni non ancora giudicate da un collegio di referee ma ispirate a criteri di openess): subito SLAC si attrezza con INSPIRE-HEP per fornire analogo servizio per le pubblicazioni del laboratorio.



Nasce in quel momento la possibilità pratica, per lo meno dal punto di vista tecnologico, di un modello di Open Access nel quale le parole chiave di FAIR (Findability, Accessibility, Interoperability, and Reusability) possono cominciare a trovare una concreta e valida applicazione (R. Stallman, movimento Open Source).

L'uso sempre più diffuso del servizio di *preprint* inizia a modificare l'atteggiamento degli studiosi circa le pubblicazioni: diventa parte delle aspettative comuni il poter accedere al sapere prodotto dai colleghi della propria disciplina e di quelle contigue senza la necessità di recarsi fisicamente in una biblioteca (con tutti i problemi che questo comporta).

Il vero punto di forza di questa soluzione tecnologica è stato però un altro: la possibilità di reperire materiale tramite ricerche guidate (il *web semantico*).



La cultura umana costituisce una sorta di iper-spazio con innumerevoli connessioni logiche tra i testi codificati che ne costituiscono i nodi. La possibilità di automatizzare le ricerche spalanca possibilità di trovare vie inedite allo sviluppo della conoscenza. **Il punto chiave qui è proprio la natura aperta di tutta questa mole di documentazione scritta.**

Arriviamo ai giorni nostri: la situazione attuale, condizionata da una pandemia mondiale ci fornisce l'occasione per ulteriormente approfondire i benefici e l'importanza di un approccio FAIR a pubblicazioni e dati.

Una cosa della quale abbiamo imparato a renderci conto è che la situazione percepita da ognuno di noi, circa lo sviluppo della pandemia, è spesso lontana dalla realtà.

Ognuno di noi conosce personalmente un numero piuttosto ridotto di persone gravemente colpite dal fenomeno e questo induce a sottovalutare gravemente il problema.

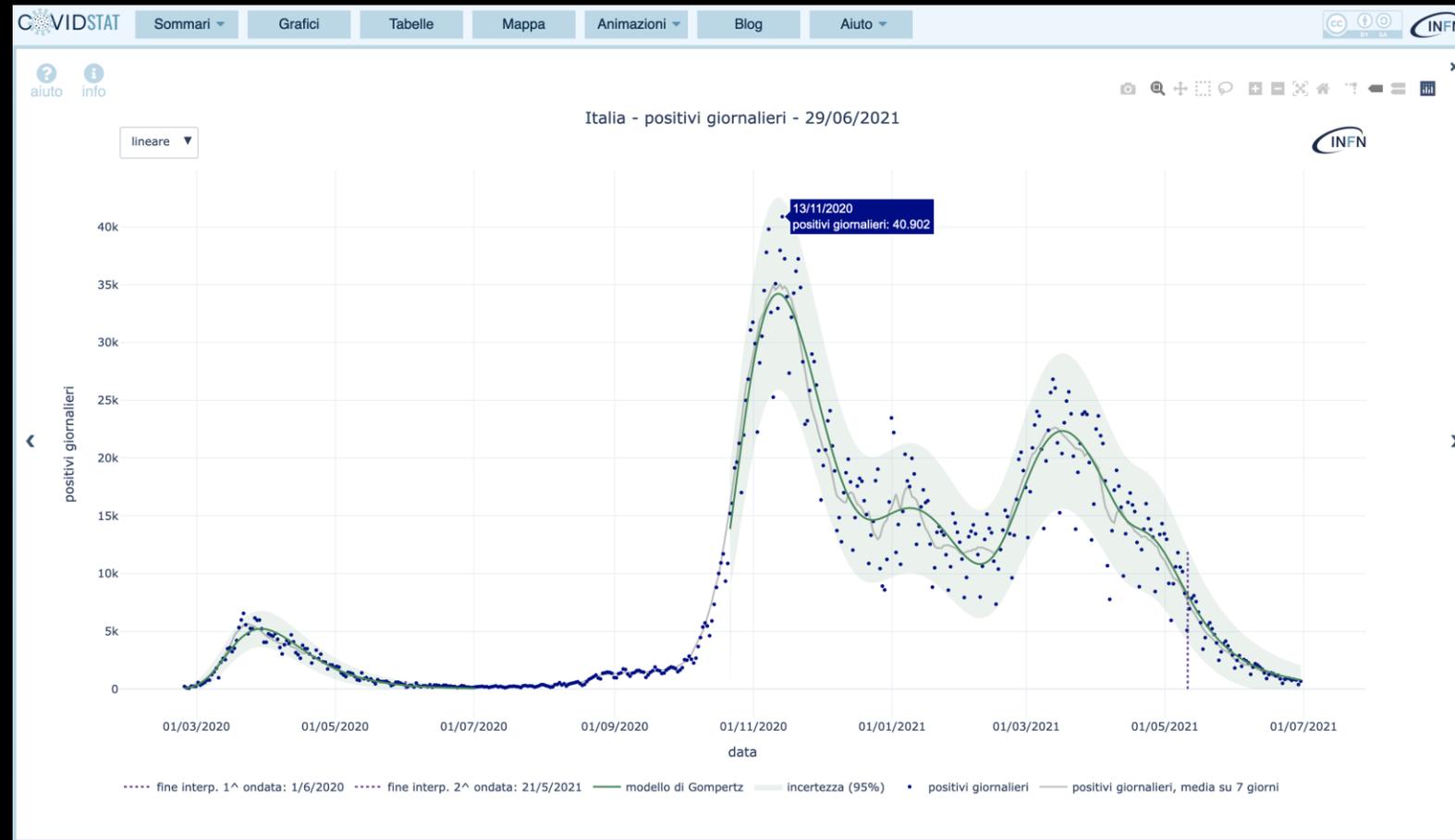


Il contagio si diffonde in modo molto più rapido della nostra capacità di rendercene conto: inoltre gli effetti causati da un contagio in un certo giorno si manifestano con un certo ritardo (dovuto al tempo necessario allo sviluppo della patologia e da quello necessario a registrare il fatto stesso).

Per prendere gli opportuni provvedimenti, quindi, il tempo è l'elemento fondamentale: prima si ha contezza numerica di quel che accade un certo giorno x , prima si prendono i provvedimenti adeguati più si limitano gli effetti globali della pandemia.

Per farci un'idea di come una legge esponenziale abbia una progressione così veloce che, una volta innescata sia quasi impossibile fermarla, vediamo un esperimento davvero impressionante:

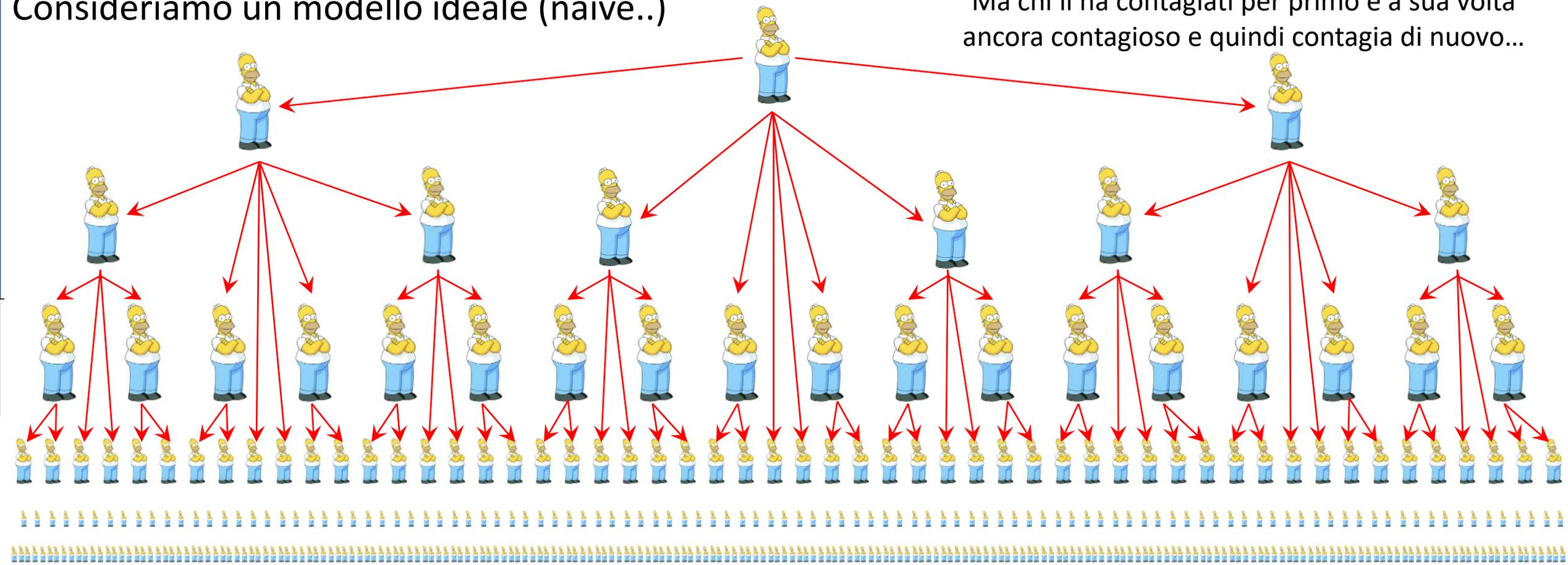
Intorno ai primi di Dicembre del 2021 è stato raggiunto un picco di circa 41000 positivi al giorno: questo corrispondeva, in quei giorni, ad una frequenza di contagi di ~ 2 al secondo, abbastanza in linea con quella sorta di *esplosione* osservata nel filmato...



L'importanza di una conoscenza in tempo reale del fenomeno può essere vista anche così:

Consideriamo un modello ideale (naive..)

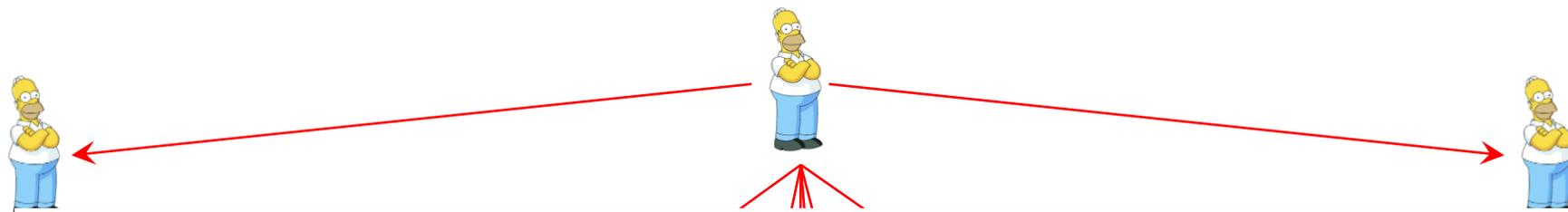
Ma chi li ha contagiati per primo è a sua volta ancora contagioso e quindi contagia di nuovo...



Questo è un modello realistico ma non accurato: non tiene conto, per esempio, del fatto che dopo essere stati infettati o si muore (pochi, per fortuna) o si guarisce: comunque sia, dopo un certo tempo si cessa di essere infettivi e si viene rimossi dalla catena di contagio. Lo sviluppo reale può essere quindi, in genere, un po' diverso da quanto illustrato qui e la matematica che lo descrive più complicata di conseguenza (serve un modello più raffinato...).

Giorno 7

Contagiati 486



Agendo su una persona **solo al quinto giorno** si riesce quindi a ridurre il numero dei contagiati totali, al settimo giorno, da 486 a 477. Una diminuzione del solo 1.85% del totale, ben poca cosa.

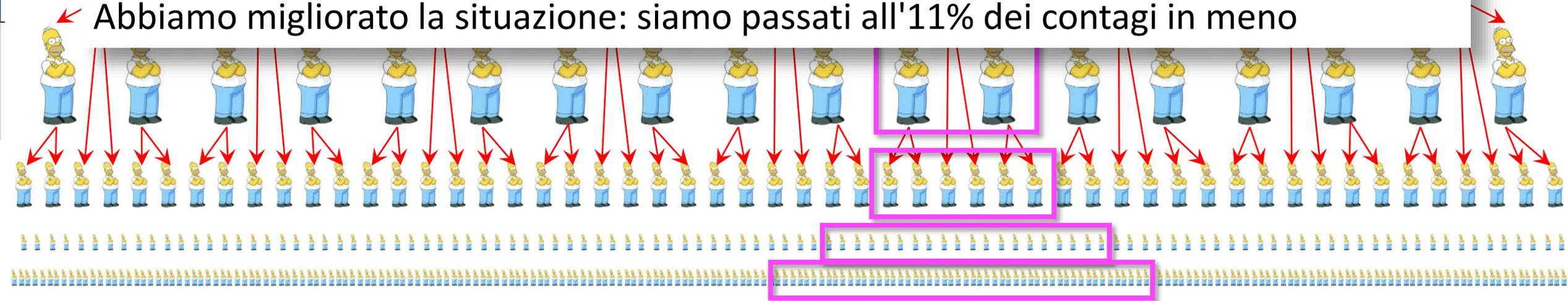


Supponiamo che un contagiato, che si renda conto di esserlo stato tre giorni prima, decida, per senso civico, di mettersi spontaneamente in quarantena a partire solo **dal quinto giorno**. Questo significa che le persone che avrebbe potuto contagiare (e quelle a loro volta altre) non lo faranno più. Da questo schema quelle persone spariscono dal gruppo dei contagiosi infettivi! Quante ne spariscono? Secondo questo modello sparisce lui il quinto giorno, lui più i due che avrebbe potuto contagiare il sesto giorno fino a un totale di **nove** al **settimo giorno**.



Quante ne spariscono grazie a questa saggia decisione di anticipare la quarantena al secondo giorno? Secondo questo modello il terzo giorno sparisce lui dalla lista dei contagiosi, il quarto giorno lui più i due che avrebbe potuto contagiare fino a un totale, al settimo giorno, di 54 persone in meno.

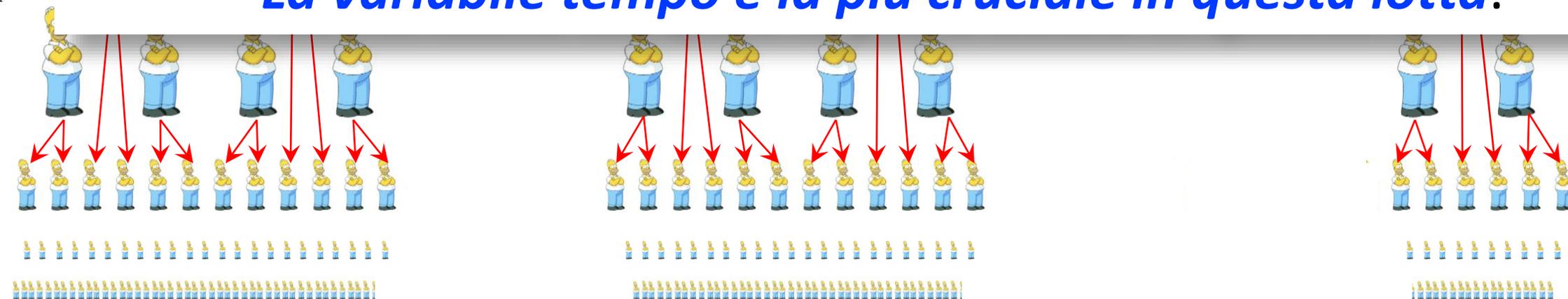
← Abbiamo migliorato la situazione: siamo passati all'11% dei contagi in meno



Se la persona contagiata decidesse invece di mettersi in quarantena **non appena ha scoperto di esserlo**, ossia già al **terzo giorno** dall'inizio della pandemia, cosa accadrebbe in questo scenario? Si vede subito che, raggiunto il settimo giorno sarebbero molte di più le persone cui viene evitato il contagio e che quindi sparirebbero a loro volta dal gruppo dei potenziali contagiati/contagiosi.

Capiamo quindi bene che non basta dire
"ho predisposto un lock-down" per ridurre l'entità del problema,
bisogna farlo *il prima possibile!*

La variabile tempo è la più cruciale in questa lotta.



Per essere davvero efficaci in questo, bisogna però bloccare le possibilità di contagio per quante *più persone* possibili ma, cosa ben più importante, occorre farlo *il prima possibile*.

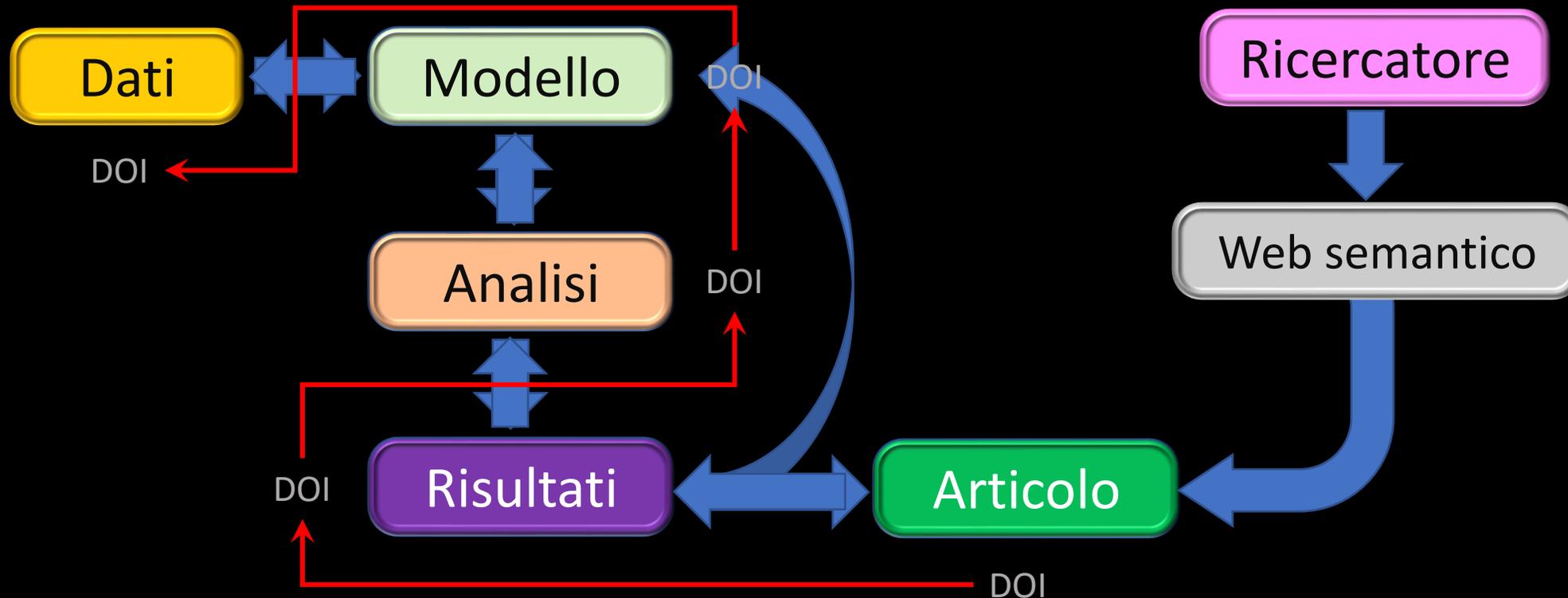
Quando è il "*prima possibile*"? Dipende dalla bontà degli indicatori che abbiamo a disposizione sul progresso della pandemia e dalla qualità e affidabilità dei dati resi disponibili localmente.

Prima raccogliamo i dati e prima siamo avvisati che sta succedendo qualcosa: se agiamo immediatamente a quel punto abbiamo la possibilità di bloccare tutto sul nascere!

Che rilevanza ha tutto ciò con la problematica dell'Open Data (e dell'Open Access)?

Il modo con cui veniamo a conoscere un qualsiasi fenomeno è tramite due azioni fondamentali:

- Una precisa e rigorosa raccolta di **dati**, mediante un protocollo univoco e non ambiguo
- L'ideazione di uno o più modelli interpretativi (modelli matematici e statistici costruiti possibilmente mediante principi primi) corroborati da un'opportuna analisi dei dati

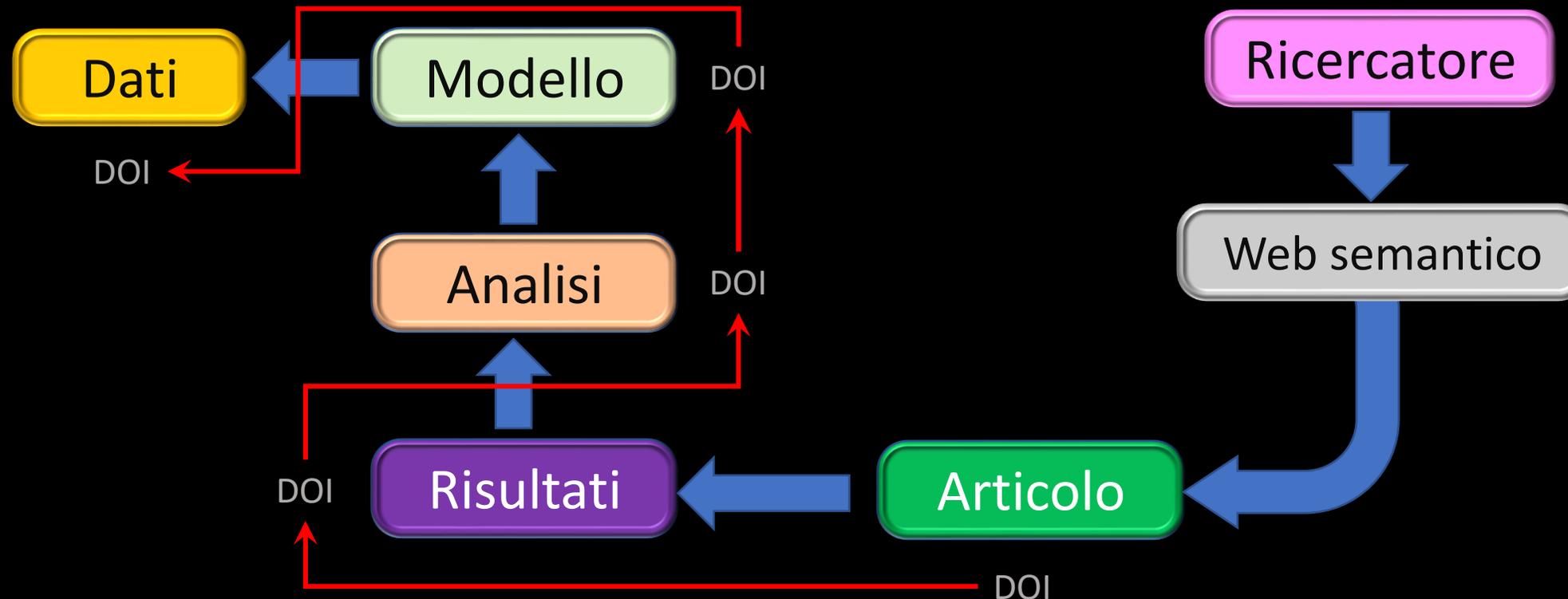


In che modo questo processo diviene possibile nella pratica?

Tramite una serie di ingredienti che definiscono il paradigma **FAIR** (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable):

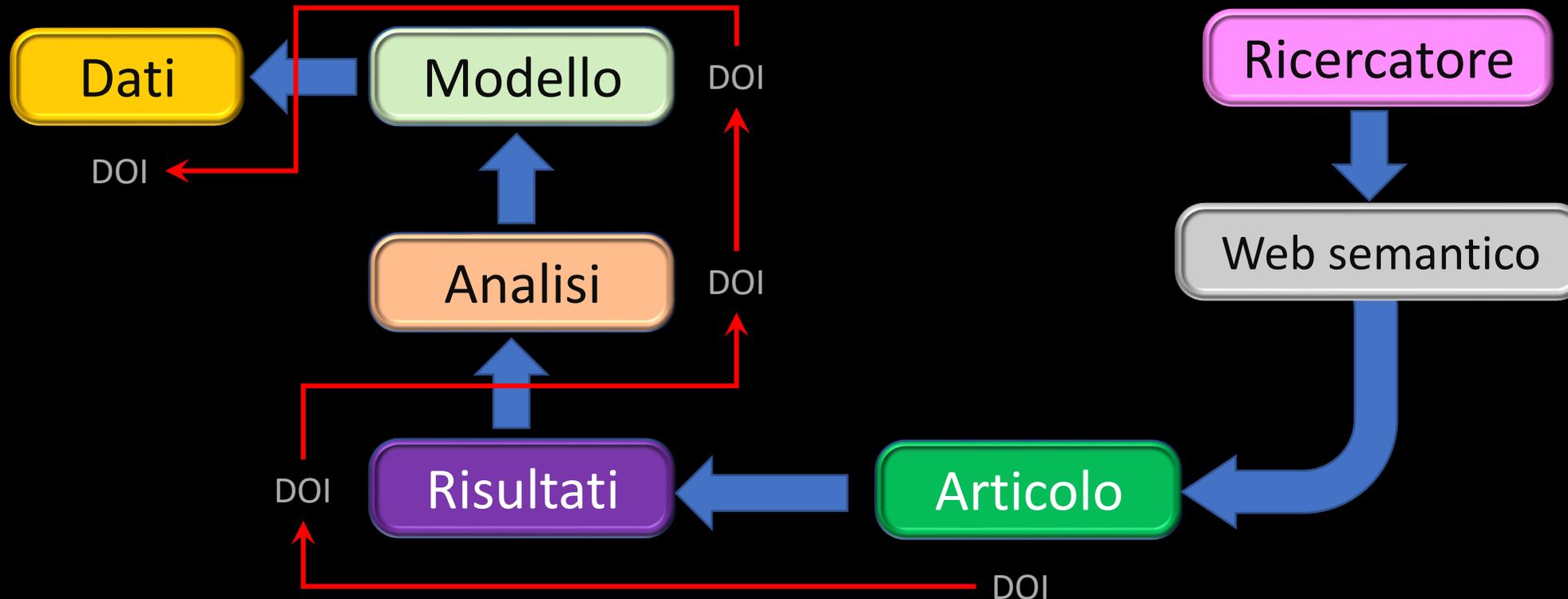
- I dati e le pubblicazioni devono poter essere "*trovabili*": internet, DOI, web semantico, *repositories*,...

Si tratta in questo caso di aspetti tecnologici da un lato (la rete, i programmi, l'intelligenza artificiale, ...) e di cooperazione internazionale dall'altro (accordi sulla definizione dei metadati, dei DOI, sull'interoperabilità dei vari *repositories*), gestione delle diverse legislazioni nazionali in tema di proprietà intellettuale,....



- L'insieme dei *repositories* di dati e pubblicazioni deve essere *interoperabile* con mezzi software

Nel mondo esistono ormai innumerevoli *repositories* di dati e articoli, database e *knowledge base*, ma la maggior parte di essi richiede software proprietari ed incompatibili. La richiesta di interoperabilità riguarda fondamentalmente la possibilità di automatizzare lo scambio e l'accesso reciproco.



Come può favorire questo processo l'Open Access e l'Open Data?

Il caso dell'attuale pandemia illustra molto bene come questo paradigma possa essere di aiuto alla società civile in pratica e non solo agli scienziati.

Una rigorosa definizione del tipo di dati da raccogliere (mediante un protocollo condiviso tra nazioni), assieme ad una precisa metodologia di raccolta (che specifichi quando prendere dati, dove, in quali circostanze) e infine un sofisticato dizionario dei metadati permette lo scambio ed il confronto in tempo reale sullo sviluppo pandemico quando questi dati siano tecnicamente accessibili da chiunque, da dovunque e il prima possibile.

Questo ha come ricaduta che qualsiasi indicatore possa venire ideato per segnalare rapidamente lo sviluppo negativo della pandemia, questo possa essere implementato per trasformare i dati grezzi in opportuni segnali d'allarme.

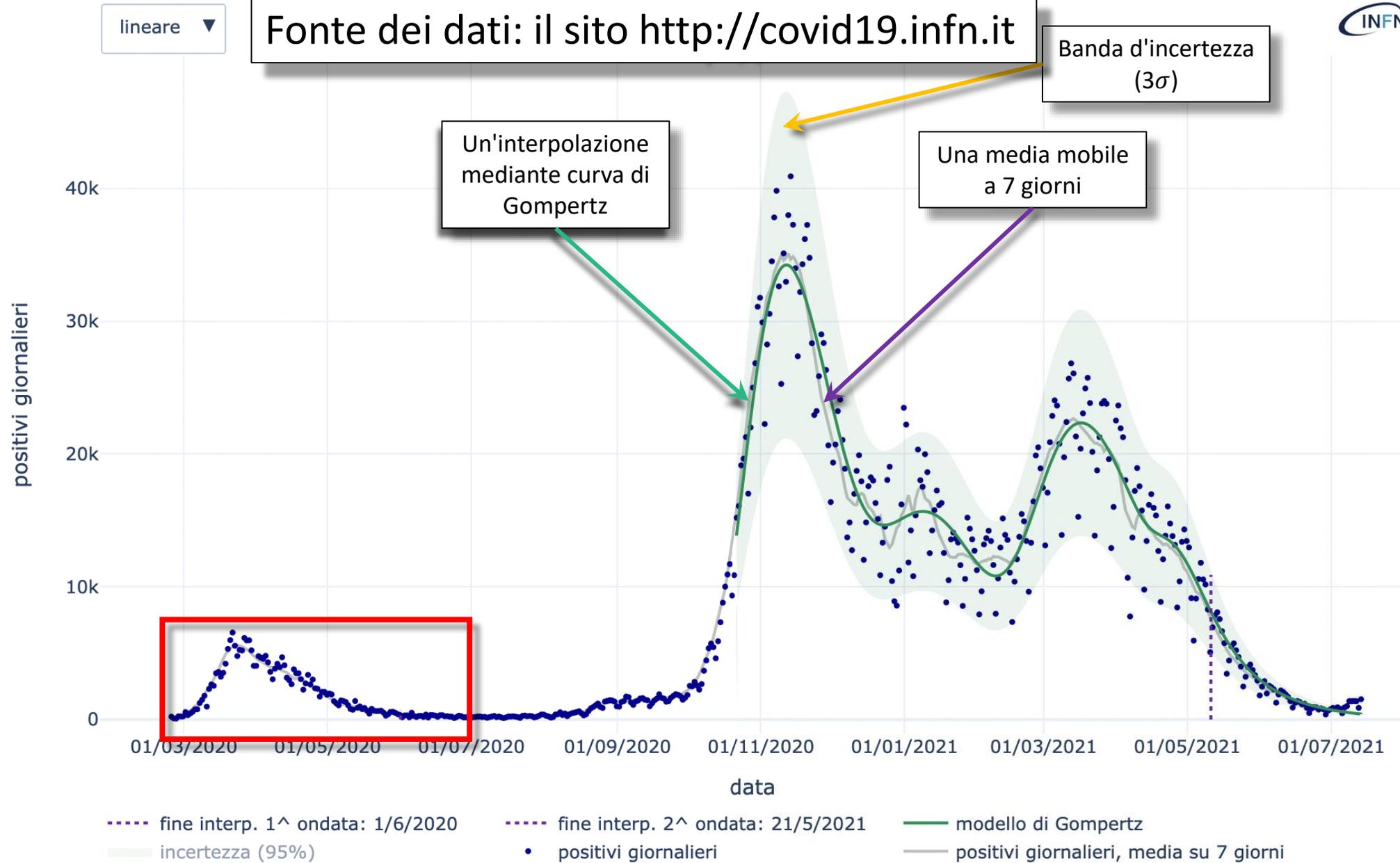
La *conditio sine qua non* per un meccanismo di questo genere è l'insieme delle condizioni precedentemente menzionate.

Esplicitiamo tutto questo mediante un preciso esempio pratico:

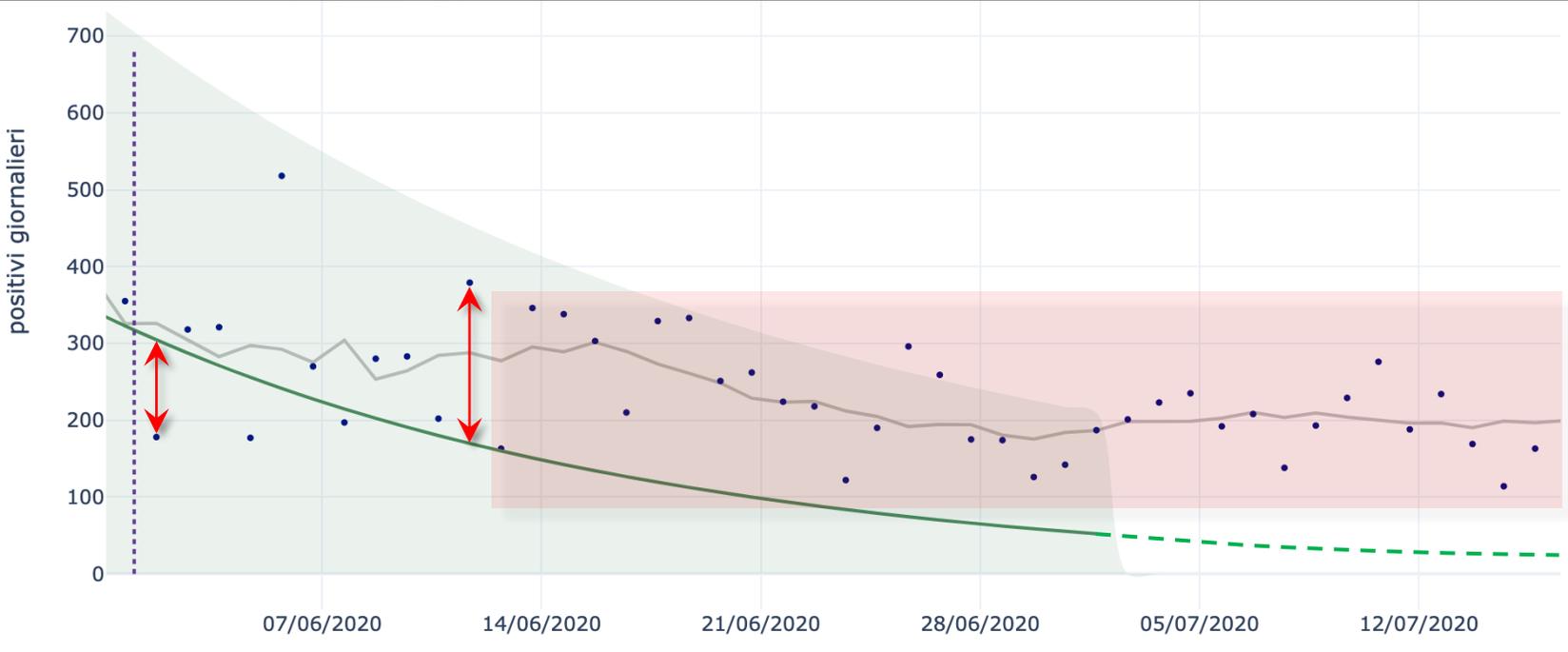
un potente ed efficace indicatore della ripresa della pandemia dopo un periodo di apparente cessazione di nuovi casi di infezione

Italia - positivi giornalieri - 13/07/2021

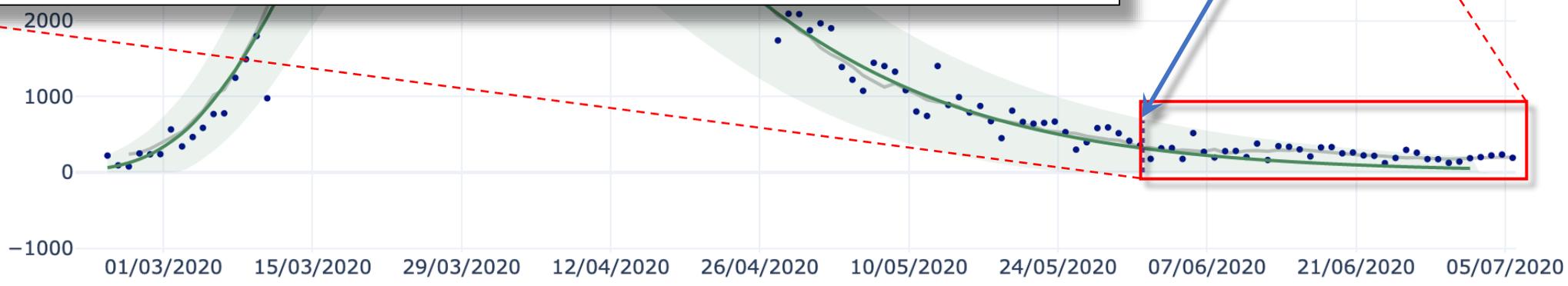
Fonte dei dati: il sito <http://covid19.infn.it>



Italia - positivi giornalieri - 13/07/2021

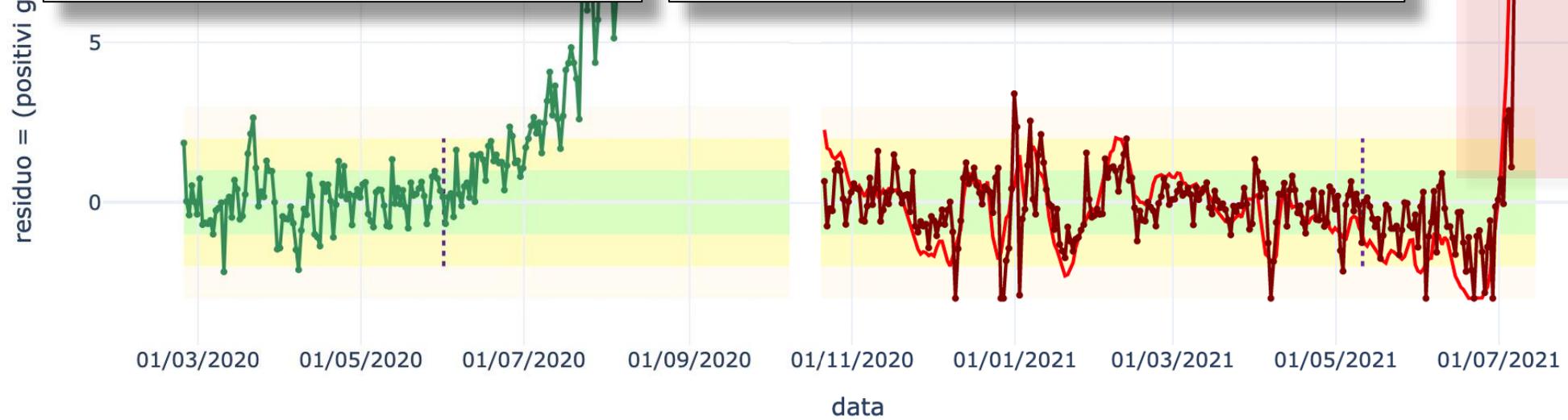
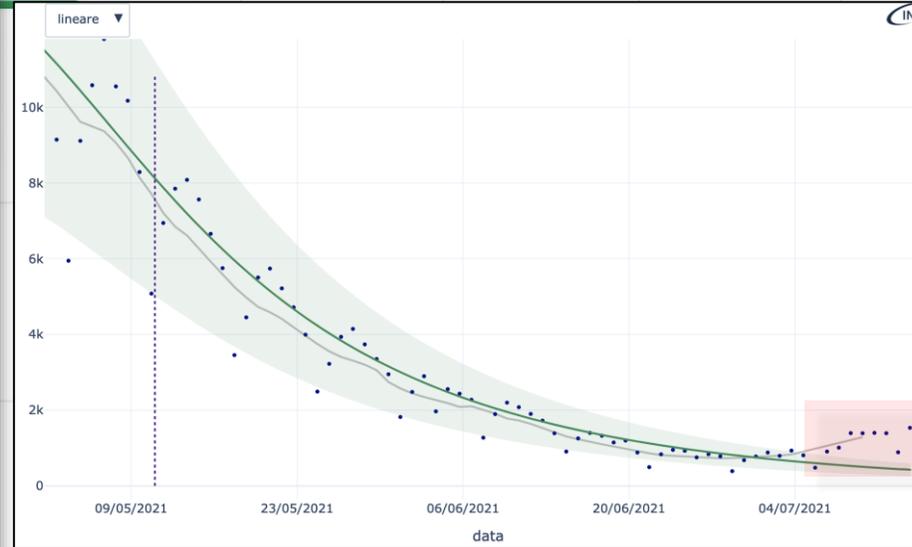
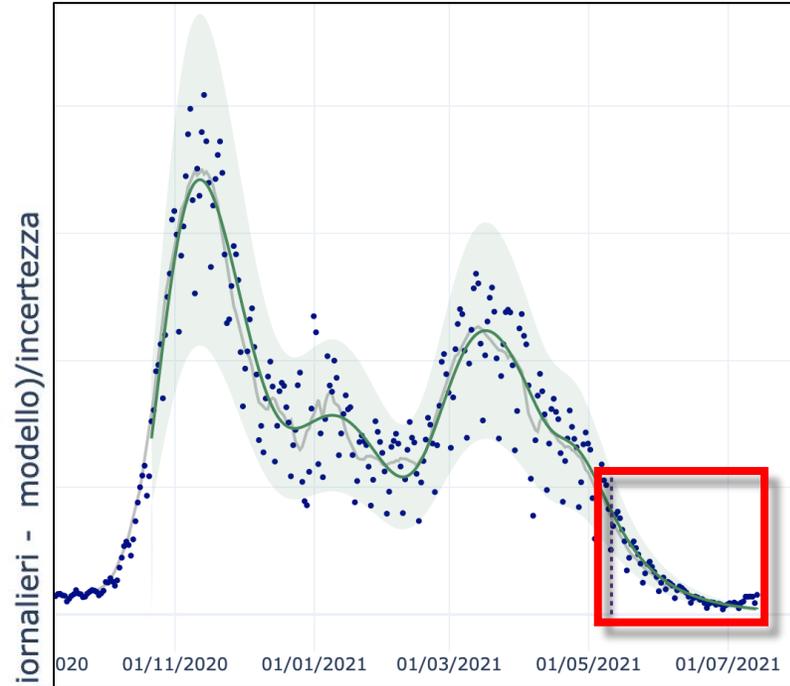


Ultimo punto utilizzato per l'interpolazione



- fine interp. 1^ ondata: 1/6/2020
- fine interp. 2^ ondata: 21/5/2021
- modello di Gompertz
- incertezza (95%)
- positivi giornalieri
- positivi giornalieri, media su 7 giorni

Italia - positivi giornalieri - residui - 13/07/2021



Questo tipo di indicatore dello stato di progresso della pandemia è ovviamente uno strumento importante, per ottenere il quale occorrono tutti gli strumenti dell'Open Science.

In primis un sito web nel quale gli utenti possono cercare le informazioni di loro interesse mediante un'interfaccia che li aiuti nella navigazione:

<http://covid19.infn.it>

Di per sé un sito non può qualificarsi come un repository di Open Access e neppure di Open Data: gli mancano tutta una serie di attributi che rispondono agli standard internazionali per rispondere ai principi FAIR (metadati, persistenza nel tempo dei prodotti,...)

Allo scopo di rispondere a questi requisiti, al sito è stato affiancato un repository INFN che queste caratteristiche, invece, le possiede tutte:

<https://www.openaccessrepository.it/>

The screenshot shows the COVIDSTAT INFN website dashboard. At the top, there is a navigation menu with options: Sommari, Grafici, Tabelle, Mappa, Animazioni, Blog, and Aiuto. The main header features the COVIDSTAT INFN logo and the text 'Aggiornato al 4/7/2021'. Below this, there are several key statistics presented in a grid:

Nuovi casi	Decessi	Terapia Intensiva	Ospedalizzati	Tamponi	Tasso di positività	Indice R_t	Tempo di raddoppio	Persone vaccinate
+808 -13.30% (su ieri) Totale 4.263.317	+12 -45.45% (su ieri) Totale 127.649	-7 -3.55% Totale 197	-37 -2.37% Totale 1.561	+141.640 -37.91% (su ieri) Totale 72.340.685	1% 1% (ieri) positivi / casi testati	0.68 0.83 ↔ 0.56 CovidStat (95% conf.)	-13 giorni → dimezzamento → raddoppio	18.714.761 Dosi somministrate: 53.385.493 4/7/2021 - 15:07:00

Below the statistics, there are five main navigation buttons: Grafici, Tabelle, Mappa, and Sommario, each accompanied by a small thumbnail image. At the bottom, there is a section titled 'Ultimi post dal blog' with three links: 'Indicatori di ripartenze (parte seconda)', 'Indicatori di ripartenze (parte prima)', and 'Attenzione!!!!'.

Diverse comunità:

- Multidisciplinarietà
- Autonomia di gestione dei propri prodotti
- Infrastruttura basata su Zenodo (CERN)
- Compatibile con tutti gli standard internazionali
- Interoperabile in automatico con tutti i siti basati su standard.
- Può ospitare tutte le tipologie di "prodotti della ricerca" (articoli, dati, software, filmati, record di natura eterogenea): a tutti viene attribuito un DOI (DataCite)

The screenshot shows the INFN Open Access Repository website. The header features the INFN OAR logo, a search bar, an 'Upload' button, and a 'Log in' button. A prominent white box on the right side of the header contains the text: 'Accesso tramite federazione di Identity Providers'. Below the header, the main content area displays a list of 'Latest entries'. The first entry, 'CovidStat project summary plots', is highlighted with a red box. It includes a date tag 'July 5, 2021 (v452)', 'Plot' and 'Open Access' tags, and a 'View' button. The entry text describes the daily update summary plots of the CovidStat project data. The second entry, 'CovidStat project data', also includes a date tag 'July 5, 2021 (v449)', 'Dataset' and 'Open Access' tags, and a 'View' button. The third entry, 'Procédure synthétique pour répliquer le Référentiel en libre accès de l'INFN (basé sur Invenio v3 et Zenodo)', includes a date tag 'June 22, 2021 (v1)', 'Technical note' and 'Open Access' tags, and a 'View' button. On the right side of the page, there is a sidebar with the title 'INFN Open Access Repository at a glance' and a list of bullet points describing the repository's features: Research. Shared., Findable. Citeable. Discoverable., Communities, Funding, and Flexible licensing. Below the sidebar, there is a section for 'Tweets by @INFN_' showing a tweet from Antonio Zoccoli.

<https://www.openaccessrepository.it/record/73897>

Relazione gerarchica tra i documenti di questa comunità (navigabilità in termini di relazione tra dati, articoli e materiale associato)

The screenshot shows a web browser window displaying the record for 'CovidStat project summary plots' on the INFN Open Access Repository. The page features a blue header with the INFN OAR logo, a search bar, and navigation links for 'Upload' and 'Communities'. A 'Log in' button is located in the top right corner. The main content area includes the title 'CovidStat project summary plots', the date 'July 5, 2021', and a list of authors: Lista, Luca; Menasce, Dario; Mezzetto, Mauro; Pedrini, Daniele; Spighi, Roberto. A summary paragraph describes the project's goal of analyzing COVID-19 data in Italy. A yellow box provides instructions on how to reproduce the plots. A 'Preview' section shows a line graph titled 'Abruzzo - decessi giornalieri - 05/07/2021' with a y-axis labeled 'decessi giornalieri' ranging from 5 to 35. The graph displays daily data points and a green trend line. On the right side, there are statistics: 1,739 views and 2,335 downloads, with a link to 'See more details...'. Below this, the 'Publication date' is July 5, 2021, and the DOI is 10.15161/oar.it/73897. The 'Keyword(s)' section includes COVID-19, CovidStat project, FAIR data, and Open Science. The 'Related identifiers' section lists the compiled by information and a link to another record. The 'Communities' section lists COVIDSTAT-INFN and INFN. The 'License (for files)' is Creative Commons Attribution Share-Alike 4.0. At the bottom right, a 'Versions' table lists several versions of the record.

CovidStat project summary plots | INFN

https://www.openaccessrepository.it/record/73897#.YOYjZW6xU8Y

Search Upload Communities Log in

July 5, 2021 Plot Open Access

CovidStat project summary plots

Lista, Luca; Menasce, Dario; Mezzetto, Mauro; Pedrini, Daniele; Spighi, Roberto

This record contains the daily update summary plots of the data of the CovidStat project. CovidStat is a project carried out by the CovidStat Working Group at INFN, whose creation was promoted within the Italian National Institute of Nuclear Physics with the aim of making a statistical analysis of the data provided daily by the Civil Protection on the spread of the pandemic in Italy. CovidStat data can interactively be browsed at <https://covid19.infn.it>.

To reproduce the plots, please refer to the correct version of the dataset and analysis software indicated in the related identifiers listed below.

Preview

Abruzzo - decessi giornalieri - 05/07/2021

lineare

decessi giornalieri

1,739 views 2,335 downloads

See more details...

Publication date: July 5, 2021

DOI: DOI 10.15161/oar.it/73897

Keyword(s): COVID-19 CovidStat project FAIR data Open Science

Related identifiers: Compiled by: 10.15161/oar.it/23545, 10.15161/oar.it/23547

Communities: COVIDSTAT-INFN INFN

License (for files): Creative Commons Attribution Share-Alike 4.0

Versions

Version	DOI	Date
Version 452	10.15161/oar.it/73897	Jul 5, 2021
Version 451	10.15161/oar.it/73895	Jul 4, 2021
Version 450	10.15161/oar.it/73893	Jul 3, 2021
Version 449	10.15161/oar.it/73891	Jul 2, 2021

Uno strumento importante sotto molti punti vista, non ultimo dei quali l'ausilio agli utenti su problematiche quali proprietà intellettuale e licenze d'uso

The screenshot shows the INFN OAR (Open Access Repository) website. The browser address bar displays the URL: <https://www.openaccessrepository.it/record/73897#.YOYJZW6xU8Y>. The page title is "CovidStat project summary plots" and the date is "July 5, 2021". The authors listed are Lista, Luca; Menasce, Dario; Mezzetto, Mauro; Pedrini, Daniele; Spighi, Roberto. The record contains daily update summary plots of the data of the CovidStat project. A yellow box contains the text: "To reproduce the plots, please refer to the correct version of the dataset and analysis software indicated in the related identifiers listed below." The main content is a preview of a plot titled "Abruzzo - decessi giornalieri - 05/07/2021". The plot shows daily deaths (decessi giornalieri) on the y-axis (ranging from 5 to 35) and time on the x-axis. The data points are blue dots, and a green line represents a linear fit. The plot is labeled "lineare" in the top left corner. The INFN logo is in the top right corner of the plot area. The right sidebar contains metadata: Publication date: July 5, 2021; DOI: 10.15161/oar.it/73897; Keyword(s): COVID-19, CovidStat project, FAIR data, Open Science; Related identifiers: Compiled by: 10.15161/oar.it/23545, 10.15161/oar.it/23547; Communities: COVIDSTAT-INFN, INFN; License (for files): Creative Commons Attribution Share-Alike 4.0. The bottom right sidebar shows a list of versions: Version 452 (Jul 5, 2021), Version 451 (Jul 4, 2021), Version 450 (Jul 3, 2021), and Version 449 (Jul 2, 2021).



Il repository permette di gestire in contemporanea sia l'Open Access alle pubblicazioni (sia green che di letteratura grigia) che i dati e i codici ad essi associati (Open Data).

• Creative Commons – Attribution X +

← → × 🏠 🔒 https://creativecommons.org/licenses/by-sa/4.0/ oar infn

Search for CC images | Global Network | Newsletters | Store | Contact | f t e

creative commons | Share your work | Use & remix | What We do | Blog

Help us build a vibrant, collaborative global commons 

This page is available in the following languages:  English


**Attribution-ShareAlike 4.0 International
(CC BY-SA 4.0)**

This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

You are free to:

- Share** — copy and redistribute the material in any medium or format
- Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.



www.openaccessrepository.it

Seguendo la gerarchia dei DOI tra loro collegati è quindi possibile accedere a tutti i codici utilizzati nel progetto (sia direttamente i files piuttosto che non un link ad un repository di github).

The screenshot shows the INFN OAR website interface. At the top, there is a navigation bar with the INFN logo, a search bar, and links for 'Upload' and 'Communities'. A 'Log in' button is located in the top right corner. The main content area displays the record for 'CovidStat project data', dated July 5, 2021. It includes a 'Dataset' and 'Open Access' label. The record is attributed to Lista, Luca; Menasce, Dario; Mezzetto, Mauro; Pedrini, Daniele; Spighi, Roberto. A summary of the record states it contains daily updated data of the CovidStat project. A yellow box highlights a note about the software used to process the data. Below this, a 'Files' section shows a table with one file: 'covid19-root.tar.gz' (36.5 MB) with a 'Download' button. The right sidebar contains statistics (1,404 views, 122 downloads), publication date (July 5, 2021), DOI (10.15161/oar.it/73896), keywords (COVID-19, CovidStat project, FAIR data, Open Science), related identifiers, communities (COVIDSTAT-INFN, INFN), and license (Creative Commons Attribution Share-Alike). A 'Versions' section at the bottom right lists previous versions from 446 to 449.

CovidStat project data | INFN Open >X +

https://www.openaccessrepository.it/record/73896#.YOVjfG6xUB8

oar infn

INFN
OAR
Open Access Repository

Search Upload Communities Log in

July 5, 2021 Dataset Open Access

CovidStat project data

Lista, Luca; Menasce, Dario; Mezzetto, Mauro; Pedrini, Daniele; Spighi, Roberto

This record contains the daily updated data of the CovidStat project. CovidStat is a project carried out by the CovidStat Working Group at INFN, whose creation was promoted within the Italian National Institute of Nuclear Physics with the aim of making a statistical analysis of the data provided daily by the Civil Protection on the spread of the pandemic in Italy. CovidStat data can interactively be browsed at <https://covid19.infn.it>.

The software to process this Root TTree is in the related identifier. Be sure to use the right corresponding version, based on the data in which Root TTree was generated.

Files (36.5 MB)

Name	Size	Download
covid19-root.tar.gz	36.5 MB	Download
md5:f6b79985cf508245cba5d4107182553d		

1,404 views 122 downloads
See more details...

Publication date: July 5, 2021
DOI: DOI 10.15161/oar.it/73896
Keyword(s): COVID-19 CovidStat project FAIR data Open Science
Related identifiers: Supplement to: 10.15161/oar.it/23547
Communities: COVIDSTAT-INFN INFN
License (for files): Creative Commons Attribution Share-Alike

Versions

Version 449	10.15161/oar.it/73896	Jul 5, 2021
Version 448	10.15161/oar.it/73894	Jul 4, 2021
Version 447	10.15161/oar.it/73892	Jul 3, 2021
Version 446	10.15161/oar.it/73890	Jul 2, 2021

https://www.openaccessrepository.it/record/73896/files/covid19-root.tar.gz?download=1

Di tutto ciò è poi disponibile la cronologia del loro sviluppo nel tempo.

The screenshot shows a web browser window displaying a record on the Open Access Repository (OAR) website. The browser's address bar shows the URL: <https://www.openaccessrepository.it/record/73372#.YOYjhG6xU8Y>. The page title is "CovidStat project analysis code". A Firefox download dialog is open, showing the file "analisi.cpp" (32.1 KB) and asking "What should Firefox do with this file?". The dialog has options to "Open with Xcode (default)" or "Save File".

The main page content includes:

- Navigation: Search, Upload, Communities, Log in
- Metadata: Software, Open Access
- Statistics: 716 views, 226 downloads
- Authors: ...tto, Mauro; Pedrini, Daniele; Spighi, Roberto; Varaschin, Antonella; Zoccoli,
- Description: ...lyse the data of the CovidStat project. CovidStat is a project carried out by the CovidStat Working Group at INFN, whose creation was promoted within the Italian National Institute of Nuclear Physics with the aim of making a statistical analysis of the data provided daily by the Civil Protection on the spread of the pandemic in Italy. CovidStat data can interactively be browsed at <https://covid19.infn.it>.
- Files (43.2 kB):

Name	Size	Download
analisi.cpp	32.9 kB	Download
analisi.h	6.3 kB	Download
makeAnalysis.cpp	4.1 kB	Download
- Publication date: May 11, 2021
- DOI: [10.15161/oar.it/73372](https://doi.org/10.15161/oar.it/73372)
- Keyword(s): COVID-19, CovidStat project, FAIR data, Open Science
- Related identifiers: Supplementary material: [10.15161/oar.it/23545](https://doi.org/10.15161/oar.it/23545)
- Communities: COVIDSTAT-INFN, INFN
- License (for files): GNU General Public License 2.0
- Versions:

Version	DOI	Date
Version 9	10.15161/oar.it/73372	May 11, 2021
Version 8	10.15161/oar.it/73350	Dec 28, 2020
Version 7	10.15161/oar.it/73284	Dec 18, 2020
Version 6	10.15161/oar.it/73149	Nov 22, 2020

Conclusioni

- La conoscenza, quella scientifica in particolare, necessita per crescere e consolidarsi, di avvalersi di un ciclo virtuoso nel quale dati, ipotesi e risultati partecipino globalmente alla produzione del sapere grazie alla loro pubblica diffusione e alla loro preservazione nel tempo
- La storia ha dimostrato come la carenza di queste caratteristiche abbia lungamente ostacolato veri progressi
- Il paradigma FAIR, se e quando coniugato alle nuove tecnologie, permette l'instaurarsi di questo ciclo virtuoso
- Occorrono complessi ed articolati accordi internazionali per rendere questo approccio una realtà fruibile dai ricercatori di tutto il mondo.
- In questa presentazione è stato illustrato un recente use-case di Open Access/Open Data, realizzato dall'INFN, per venire incontro alle necessità di monitoring e allerta della recente pandemia di Covid19
- La strada verso una completa e globale implementazione del paradigma FAIR è ancora lunga, ma i recenti progressi dimostrano ampiamente la sua validità

Grazie





Università
di Genova

GenOA week

International Open Access Week
25 - 29 October 2021

It matters how we Open Knowledge: building structural equity.
Sulla **Open Science** l'UNESCO evidenzia l'urgenza di costruire una equità strutturale per tutti, passando dalla teoria alla pratica della Scienza Aperta. **GenOA week 2021** propone un percorso di riflessione su questi temi.



Consiglio Nazionale delle Ricerche

