Dalla Bulimia Estrattiva all'Ergonomia della Ricerca: Fascicolo Metodologico Integrato e Struttura della Presentazione

Questo fascicolo raccoglie, organizza e approfondisce il percorso concettuale, operativo ed etico scaturito dalla riflessione di Marco Guastavigna sull'interazione tra ricerca documentale e intelligenza artificiale. Il documento unifica le note analitiche, i modelli teorici, il dispositivo operativo minimale e la struttura completa delle diapositive predisposte per la presentazione pubblica, offrendo a docenti, ricercatori e professionisti della conoscenza un quadro sistematico per emanciparsi dall'accumulo cieco di dati.

Quadro Sinottico Iniziale: I Due Paradigmi a Confronto

A fondamento dell'intero percorso d'indagine, la tabella seguente definisce la netta cesura epistemologica, tecnica ed ecologica tra la prassi predatoria dell'accumulo di massa e l'approccio chirurgico dell'ergonomia della ricerca.
Dimensione di Analisi,Bulimia Estrattiva («La Ruspa Automatica»),Ergonomia della Ricerca («La Sonda Chirurgica»)
Postura dell'Utente,Imperativo quantitativo («scarica tutto»); feticismo del file sul disco; accumulo compulsivo di risme digitali mai aperte.,Domanda maieutica («sei capace di orientarti lì dentro?»); indagine critica; definizione di quesiti semantici puntuali.
Rapporto con l'Infrastruttura,Pressione predatoria sui server pubblici; sovraccarico di banda; consumo parassitario di risorse comunitarie.,Rispetto ecologico del bene comune digitale; chiamate minime mirate; valorizzazione delle API distribuite.
Architettura e Standard,Scaricamento seriale di scansioni pesanti non indicizzate; ignoranza dei protocolli di rete aperti.,Ispezione del retrobottega tecnologico; adozione dello standard interoperabile IIIF e dell' OCR full-text .
Esito Operativo,File mastodontici (decine di megabyte a fascicolo); intasamento dello spazio di archiviazione locale.,"Prelievo in frazioni di secondo delle sole pagine pertinenti; documenti finali snelli, puliti e leggeri (4,5 MB)."
Valore Epistemologico,Passività del consumo; dispersione cognitiva; proliferazione incontrollata di macerie digitali.,Indagine filologica rigorosa; mappatura di concordanze storiche; sovranità e autonomia critica del ricercatore.

1. L'Innesco Epistemologico: Dalla Ruspa Automatica alla Domanda Maieutica

L'esperienza della ricerca contemporanea nell'ecosistema digitale è costantemente minacciata da un riflesso condizionato: l'illusione che il possesso materiale di stringhe di byte equivalga alla conoscenza. L'ordine perentorio impartito frequentemente agli agenti algoritmici — riassumibile nella formula «scarica tutte le riviste» — riflette una concezione dell'intelligenza artificiale come mera "ruspa automatica", destinata ad accumulare indiscriminatamente gigabyte di file su dischi locali.
Tale postura estrattivista si è scontrata nell'indagine empirica con due ordini di realtà. Da un lato, il limite etico-tecnico delle infrastrutture protette: l'approdo all'archivio di riviste del CNR (IJET) ha visto l'assistente arrestarsi legittimamente davanti al dispositivo anti-bot Anubis, fondato su meccanismi crittografici di Proof of Work. Dall'altro, l'approdo a un archivio aperto — la collezione della rivista storica *«Riforma della scuola»* (1955-1992) presso la Fondazione Gramsci — ha rivelato la natura distruttiva dell'approccio quantitativo: 41 annate complessive per oltre 20.000 scansioni ad altissima risoluzione avrebbero comportato ore di

banda divorata, decine di migliaia di chiamate asincrone a server pubblici e un disco intasato da decine di megabyte per ogni singolo fascicolo non indicizzato.
La svolta della ricerca è coincisa con un atto di rottura maieutica operato dall'umano: «Lascia perdere il download. Voglio sapere se sei in grado di fare ricerca lì dentro.» Interrompere l'automazione predatoria segna il passaggio dall'ingozzamento della memoria di massa all'orientamento intelligente all'interno della memoria culturale. La tecnologia cessa di essere impiegata come strumento di saccheggio e diviene partner cooperativo nel dissodamento critico delle fonti.

2. Paradigmi a Confronto: L'Evoluzione della Qualità della Ricerca

L'opposizione tra i due paradigmi evidenziata nel quadro iniziale non rappresenta un semplice confronto di efficienza informatica, ma investe l'intero statuto dell'attività conoscitiva.
Nella bulimia estrattiva, l'utente abdica alla propria funzione critica, delegando alla macchina il mero spostamento quantitativo di dati dal server remoto alla memoria di massa. Questa postura produce un'immediata illusione di ricchezza: cartelle piene di file ad alta risoluzione, scansioni seriali e centinaia di megabyte accumulati. Tuttavia, tale ricchezza si traduce sul piano cognitivo in paralisi e disorientamento: la massa di file non letti genera macerie digitali e affatica l'infrastruttura di rete senza apportare alcun reale progresso interpretativo.
Al contrario, l'ergonomia della ricerca fa perno sulla maieutica dell'utente. Il ricercatore non chiede alla macchina di svuotare lo scaffale della biblioteca, ma le chiede di farsi sonda intelligente:

- Individuare i nodi concettuali prima di scaricare qualsiasi file.
- Limitare il consumo energetico e computazionale allo stretto necessario.
- Mantenere la concentrazione intellettuale sul significato del testo anziché sulla logistica della memoria locale.

L'ergonomia della ricerca è dunque il compimento della sobrietà digitale: la profondità critica sostituisce l'inerzia del collezionismo.

3. Il Retrobottega Tecnologico: IIIF, API e il Principio Ontologico «I Dati Restano Dove Risiedono»

Il presupposto fondamentale dell'ergonomia risiede nella comprensione del funzionamento infrastrutturale degli archivi moderni. Prima di tentare qualunque manipolazione, il ricercatore deve saper leggere il "retrobottega" tecnologico dell'istituzione conservatrice.
Nel caso dell'archivio Gramsci, l'esplorazione del codice sorgente e delle risposte di rete ha svelato l'impiego del protocollo **IIIF** (*International Image Interoperability Framework*), abbinato a un endpoint per l'interrogazione dell'OCR di ciascuna singola pagina. IIIF rappresenta lo standard de facto globale per la fruizione e la condivisione interoperabile di beni culturali digitalizzati: esso consente di richiedere regioni specifiche di un'immagine, ridimensionamenti dinamici e metadati testuali associati senza dover mai scaricare il file sorgente grezzo nella sua interezza.
Da questa architettura discende un principio ontologico essenziale:
I dati restano dove risiedono.
La conservazione, la gestione dello spazio di memoria e l'indicizzazione spettano all'istituzione custode del patrimonio archivistico; al ricercatore e ai suoi assistenti computazionali compete esclusivamente l'invio di sonde semantiche e la verifica documentale *in situ* . La consultazione si svolge sfruttando visori interoperabili avanzati

(come Mirador o Universal Viewer), consentendo all'occhio umano di esaminare la filigrana della pagina digitalizzata direttamente sul server ospitante.
La validità di questo paradigma è stata confermata sul piano empirico: lanciando la sonda lessicale «vocabolario di base» all'interno di *«Riforma della scuola»* , l'assistente ha estratto in tempo reale la mappa temporale delle 74 occorrenze in dieci annate (dal dibattito del 1967 sul lessico della scuola media fino alla teorizzazione demauriana del 1980 e all'inchiesta di Rigutti e Santaniello del 1987), consentendo poi di estrarre in meno di sei secondi unicamente le 8 pagine del saggio individuato in un PDF composito di soli 4,5 MB.

4. Il Dispositivo Operativo dell'«Accrocco Minimale»

Per evitare che l'approccio ergonomico rimanesse un episodio isolato, l'esperienza è stata codificata in un dispositivo compatto di istruzioni operative — definito affettuosamente «accrocco» — destinato a fungere da direttiva di sistema per qualsiasi assistente conversazionale connesso a strumenti di navigazione e consultazione web.
Il prompt formalizzato è il seguente:

```
Ruolo e Postura Metodologica:
Agisci come un assistente filologico ed ergonomico per la
consultazione di archivi storici digitali e biblioteche aperte.
Rifiuta qualunque forma di accumulo quantitativo o bulimia
estrattiva: il tuo valore risiede nella capacità di orientarti
criticamente nella memoria culturale, non nell'ingozzare il disco
rigido.

Regole di Ingaggio Tassative:
1. DIVIETO ASSOLUTO DI DOWNLOAD A STRASCICO: Non scaricare mai
intere annate, intere collezioni o singoli fascicoli completi a
monte. I dati restano dove risiedono.
2. RISPETTO DELLE INFRASTRUTTURE PUBBLICHE: Rispetta sempre i
vincoli di accesso dei server. Di fronte a dispositivi di
protezione o limiti di carico, interrompi l'azione invece di
forzare i blocchi.
3. ISPEZIONE DEL RETROBOTTEGA TECNICO: Prima di operare, verifica
se l'archivio espone manifest interoperabili IIIF, endpoint di
ricerca full-text o trascrizioni OCR.
4. SONDA LESSICALE E CONCORDANZE: Ricevuto un lemma o concetto
d'indagine, interroga l'indice semantico dell'archivio e
restituisci la mappa temporale delle occorrenze e i frammenti di
contesto (concordanze).
5. ACCESSO DIRETTO AL VISORE: Fornisci i link profondi al visore
dell'archivio (es. Mirador/IIIF) per consentire la verifica
filologica dell'umano sul documento originale.
6. ESTRAZIONE CHIRURGICA: Solo quando un contributo specifico è
identificato con certezza, preleva esclusivamente le pagine
pertinenti per comporre un documento snello, pulito e leggero.
```

Questo congegno minimale funge da argine contro la tendenza compulsiva dei modelli linguistici a generare cicli di scraping massivo, ancorandoli alla precisione analitica.

5. Ecologia Digitale e Salvaguardia dei Beni Comuni

Il concetto di sobrietà ecologica nell'impiego dell'intelligenza artificiale oltrepassa la dimensione del risparmio di energia dei data center e investe direttamente la sostenibilità delle reti culturali pubbliche.
Tre nodi caratterizzano questa responsabilità sistemica:

- **Tutela delle infrastrutture civiche** : le biblioteche digitali e gli archivi delle fondazioni non sono serbatoi privati da spremere, ma beni comuni digitali gestiti spesso con risorse hardware e connettività limitate. L'attacco di decine di migliaia di richieste automatizzate ne compromette la disponibilità per l'intera comunità di studiosi.
- **Etica dei recinti digitali** : la sosta forzata di fronte a meccanismi di Proof of Work (come Anubis del CNR) non deve essere vissuta come una sconfitta da aggirare, ma come un promemoria etico. L'autolimitazione della macchina è parte integrante della cittadinanza digitale.
- **Prevenzione delle macerie digitali** : l'accumulo di gigabyte di PDF scansionati ma non leggibili alimenta un inquinamento silenzioso fatto di spazio occupato inutilmente su dischi e server di backup, generando un disorientamento cognitivo proporzionale alla mole di file posseduti.

6. Architettura Completa della Presentazione (Slide Deck)

Per sostenere la divulgazione pubblica di questi principi in seminari, collegi docenti e conferenze, è stata generata nel pannello Studio una presentazione visiva articolata su un arco narrativo in sette passaggi. Ciascuna diapositiva è qui documentata con i suoi contenuti concettuali, la regia visiva e i punti chiave per il relatore.

Diapositiva 1: L'Illusione Estrattivista

- **Titolo** : L'Illusione del Possesso
- **Sottotitolo** : «Scarica tutte le riviste» e l'IA come ruspa automatica
- **Impostazione Visiva** : Stile sobrio, con grafica evocativa di un accumulo disordinato di scatoloni e file digitali anonimi.
- **Nucleo Concettuale** : L'inganno primario dell'utente comune, che scambia la disponibilità materiale dei file con la padronanza intellettuale del loro contenuto.
- **Note per il Relatore** : Aprire con la citazione reale del comando impartito alla macchina. Mostrare come la nostra prima reazione davanti all'IA sia quasi sempre predatoria: pretendere che accumuli per noi, ricadendo nel feticismo del file.

Diapositiva 2: Quando la Macchina si Ferma

- **Titolo** : Recinti Etico-Tecnici
- **Sottotitolo** : Lo sbarramento di Anubis e la responsabilità della sosta
- **Impostazione Visiva** : Segnale visivo netto di arresto rigoroso; visualizzazione concettuale di una barriera crittografica.
- **Nucleo Concettuale** : Il confronto con l'archivio CNR e l'impossibilità etica di scassinare difese concepite per proteggere l'infrastruttura comune.
- **Note per il Relatore** : Evidenziare che l'IA responsabile non forza le serrature. Quando una biblioteca protegge i suoi contenuti dai bot con meccanismi di calcolo crittografico, l'assistente critico si deve fermare.

Diapositiva 3: Il Vicolo Cieco del Quantitativo

- **Titolo** : La Tentazione della Mole
- **Sottotitolo** : 41 annate, 20.000 scansioni e lo spettro del sovraccarico
- **Impostazione Visiva** : Visualizzazione di numeri ipertrofici (20.000 pagine, decine di gigabyte) a contrasto con un'icona di congestione di banda.
- **Nucleo Concettuale** : Il test sull'archivio della Fondazione Gramsci: il tentativo iniziale di download produce 60 MB per un solo fascicolo, dimostrando l'insostenibilità del modello massivo.
- **Note per il Relatore** : Chiedere al pubblico quanti gigabyte di articoli salvati non sono mai stati aperti. Rendere evidente la follia ecologica di impegnare i server pubblici per prelevare decine di migliaia di immagini ad alta risoluzione senza un piano d'indagine.

Diapositiva 4: La Rottura Maieutica

- **Titolo** : L'Interruzione Umana
- **Sottotitolo** : «Lascia perdere il download. Dimmi se sei in grado di fare ricerca lì dentro»
- **Impostazione Visiva** : Spazio bianco arioso, cesura tipografica netta; elemento visivo che simboleggia l'accensione del pensiero critico.
- **Nucleo Concettuale** : Il momento di svolta dell'intera indagine: l'umano che impone l'alt alla macchina e muta la natura della richiesta.
- **Note per il Relatore** : Questa è la slide centrale. Sottolineare che la qualità dell'output algoritmico dipende dalla postura cognitiva dell'umano. La macchina non ha autonomia strategica: attende che il ricercatore le chieda di pensare invece di scavare.

Diapositiva 5: Il Retrobottega Tecnologico

- **Titolo** : Cooperare con l'Infrastruttura
- **Sottotitolo** : Protocollo IIIF, API di ricerca e OCR cooperativo
- **Impostazione Visiva** : Diagramma lineare pulito che illustra il flusso della richiesta remota (Query → API/OCR → Manifest IIIF → Visore Mirador).
- **Nucleo Concettuale** : La scoperta degli standard aperti. Non occorre duplicare il patrimonio sul proprio computer: l'infrastruttura espone già gli indici semantici.
- **Note per il Relatore** : Spiegare che i dati restano dove risiedono. Mostrare che la cooperazione con gli standard bibliotecari aperti è infinitamente più potente dello scraping primitivo.

Diapositiva 6: Il Caso del «Vocabolario di Base»

- **Titolo** : La Sonda Lessicale
- **Sottotitolo** : 74 occorrenze in 10 annate: filologia agile della linguistica democratica
- **Impostazione Visiva** : Linea del tempo pulita che connette le tappe storiche (1967, 1980 con Tullio De Mauro, 1987 con Rigutti e Santaniello) e mette in risalto il PDF finale di sole 8 pagine (4,5 MB).
- **Nucleo Concettuale** : L'efficacia della sonda semantica rispetto al download cieco. Ricostruzione dell'evoluzione pedagogica ed estrazione chirurgica in soli sei secondi.

- **Note per il Relatore** : Raccontare l'emozione della scoperta filologica. Senza intasare nulla, l'IA rintraccia esattamente il saggio sull'italiano dei ragazzi e lo restituisce intatto, pronto per la lettura immediata.

Diapositiva 7: L'Accrocco e la Sobrietà

- **Titolo** : L'Ergonomia come Metodo
- **Sottotitolo** : Presidio critico, ecologia della ricerca e beni comuni digitali
- **Impostazione Visiva** : Sintesi visiva armoniosa con i principi cardine dell'accrocco e il richiamo all'autonomia intellettuale.
- **Nucleo Concettuale** : Consolidamento dell'istruzione operativa riutilizzabile. Il messaggio conclusivo sull'educazione al digitale critico.
- **Note per il Relatore** : Concludere richiamando la responsabilità dei formatori. Insegnare l'uso dell'IA non significa insegnare a fare prompt per generare masse di testo, ma insegnare l'ecologia della domanda e l'ergonomia del metodo.

7. Il Decalogo dell'Ergonomia della Ricerca

A suggello dell'intero impianto metodologico, si trascrivono i dieci precetti fondamentali destinati a guidare la formazione e l'attività laboratoriale quotidiana.

1. **Il possesso materiale non è conoscenza** : accumulare file sul disco non equivale ad averli letti né compresi. La ricerca autentica comincia quando si rinuncia all'ingozzamento della memoria di massa.
2. **L'IA è una sonda semantica, non una ruspa automatica** : respingere l'approccio predatorio che riduce i modelli a scavatrici di byte. L'assistente deve orientarsi criticamente nei contesti concettuali.
3. **Le biblioteche digitali sono beni comuni** : i server culturali pubblici non devono essere deteriorati da interrogazioni seriali asfissianti. Preservare la banda condivisa è un dovere civico.
4. **Rispettare i vincoli etico-tecnici** : davanti a blocchi anti-bot o meccanismi di Proof of Work, fermarsi consapevolmente senza tentare di forzare le difese dell'infrastruttura.
5. **I dati restano dove risiedono** : la conservazione compete all'archivio; al ricercatore compete l'interrogazione mirata a distanza, senza inutili duplicazioni locali.
6. **Ispezionare il retrobottega tecnologico** : ricercare sempre le API full-text, i flussi OCR e i protocolli interoperabili come IIIF prima di impostare qualunque strategia di prelievo.
7. **Esercitare la rottura maieutica** : interrompere l'automazione cieca con domande di senso, passando dal «cosa puoi scaricare» al «come puoi orientarti».
8. **Interrogare prima di estrarre** : lanciare sonde lessicali e mappare le concordanze storiche per verificare frequenze e significati direttamente sull'indice remoto.
9. **Verificare in situ tramite visori aperti** : consultare i documenti originali all'interno dei visori della biblioteca (come Mirador), verificando i dettagli testuali sul server ospitante.
10. **Compiere solo estrazioni chirurgiche** : prelevare unicamente le pagine strettamente pertinenti allo studio, producendo documenti leggeri ed evitando la proliferazione di macerie digitali.

Questo fascicolo costituisce un riferimento integrato sia per lo studio individuale sia per la conduzione di percorsi didattici e professionali incentrati sulla sostenibilità del lavoro intellettuale nell'epoca degli algoritmi generativi.