

# Valutazione della Ricerca: Manuale di Sopravvivenza per l'Informatico

GIANSALVATORE MECCA

Dipartimento di Matematica e Informatica  
Università della Basilicata

Versione 1.2 – Revisione del 7 aprile 2010

## SOMMARIO

*Questo documento discute metodologie di valutazione della produzione scientifica di ricercatori che appartengono al settore informatico. Il suo obiettivo è fornire un contributo metodologico alla discussione sulla valutazione della ricerca, mettendo in luce alcune particolarità nelle pratiche adottate dagli informatici per la pubblicazione dei loro risultati scientifici. Queste particolarità, peraltro ampiamente documentate nella letteratura internazionale, rendono i criteri di valutazione della ricerca tipicamente adottati in altri settori delle scienze fortemente inadeguati se applicati alla valutazione della ricerca informatica. Oltre a questo contributo relativo al metodo, il documento si spinge fino a discutere alcuni algoritmi specifici per la valutazione della produzione scientifica dei singoli e delle strutture, allo scopo di illustrare più concretamente come sia possibile procedere alla valutazione in accordo con le idee discusse nel documento.*

## 1 Introduzione

E' certamente sempre terreno scivoloso quello che riguarda la valutazione della ricerca. Altrettanto delicato è il compito del ricercatore che debba illustrare le particolarità della propria ricerca a colleghi di altre discipline. Eppure, per quanto sfortunata, la necessità di argomentare queste particolarità è oramai riconosciuta come inevitabile per gli informatici. Per citare una recentissima pubblicazione sull'argomento [Meyer et al, 2009]: “*In many countries, computer scientists must repeatedly explain the specificity of their disciplines to colleagues from other areas, for example in hiring and promotion committees*”.

La definizione di modalità corrette per la valutazione della qualità della ricerca in un settore strategico come l'informatica e l'ingegneria informatica è stata presto identificata come una necessità, per evitare di penalizzare ricercatori promettenti, o di ripartire le risorse in modi che non riflettono il corretto valore scientifico delle proposte. Non a caso, già nel 1994, la National Science Foundation (NFS) americana commissionò al Computer Science and Telecommunications Board (CSTB) uno studio [CSTB, 1994] finalizzato a studiare le peculiarità della ricerca sperimentale nel settore dell'informatica e dell'ingegneria informatica, e di conseguenza a definire linee guida per la valutazione dei ricercatori in questi settori, con la seguente motivazione: “*The NFS request was motivated by the concern that experimental computer scientists and engineers may face special challenges in being evaluated and in creating appropriate research environments*” [CSTB, 1994]. Questo studio iniziale è stato seguito negli Stati Uniti da iniziative successive, fino al documento promosso dalla Computing Research Association nel 1999 [Patterson et al, 1999], che ribadisce e rinforza le idee sostenute negli studi precedenti, e che secondo molti ha profondamente influenzato le pratiche di valutazione della ricerca americane. Più recentemente, un documento del genere è stato prodotto anche in Europa [Meyer et al, 2009], su iniziativa di Informatics Europe<sup>1</sup>, l'associazione dei Dipartimenti Europei di Computer Science.

Tutti questi studi raggiungono conclusioni sostanzialmente unanimi.

L'obiettivo di questo documento è quello di fornire un contributo alle discussioni sulla valutazione della ricerca sintetizzando le principali idee di questi documenti e discutendo alcune proposte concrete. Il documento è organizzato in tre parti. Nella prima parte, quella principale, vengono riassunti gli elementi

<sup>1</sup> <http://www.informatics-europe.org>

principali della ricerca in informatica ed ingegneria informatica che rendono impossibile l'applicazione di criteri tradizionalmente applicati in altre discipline scientifiche, ed in particolare:

- il ruolo centrale svolto dalle pubblicazioni in atti di convegni internazionali;
- la scarsa rappresentatività degli indici ISI per la valutazione dell'impatto delle pubblicazioni.

Questa parte è impostata come un'ampia rassegna della letteratura internazionale per evitare l'equivoco – purtroppo spesso collegato a queste operazioni – di voler sostenere teorie strumentali o punti di vista puramente personali. In quest'ottica, ha obiettivi simili ad altri documenti del genere, come i già citati [CSTB, 1994] [Patterson et al, 1999] [Meyer et al, 2009], dal momento che introduce principi generali, elementi di conoscenza e linee guida che possono essere incorporati in molti modi nei processi di valutazione.

La seconda e la terza parte, che hanno entrambe un taglio ed obiettivi diversi, sono volutamente contenute nelle appendici a questo documento, in modo da snellirne l'organizzazione e semplificarne la lettura. Nella seconda parte, contenuta nell'Appendice I., il documento assume un taglio più pragmatico, entrando nel merito di alcuni modelli diffusi per la valutazione dell'attività della ricerca, con l'obiettivo di proporre correttivi coerenti con i principi discussi nella prima parte. La terza parte, infine, contenuta nell'Appendice II., discute alcuni commenti ricevuti dopo la lettura di versioni preliminari di questo documento.

Il documento è orientato ai non informatici che si trovano a valutare l'attività di ricerca degli informatici, ma anche agli informatici stessi, ai quali vuole fornire alcuni punti di riferimento per meglio argomentare le peculiarità delle proprie ricerche di fronte a modelli di valutazione chiaramente ispirati da altre discipline. Ha inoltre la velleità di avviare una discussione più strutturata su questo tema all'interno della comunità scientifica informatica nazionale.

La filosofia ispiratrice di tutto il documento è quella secondo cui la valutazione della ricerca è una straordinaria opportunità, ma corre il rischio di diventare controproducente se condotta in modo inadeguato e prescindendo da un'analisi seria delle specificità delle aree.

## **2 L'Importanza delle Conferenze Internazionali**

Un elemento centrale di tutti gli studi sull'argomento citati in questo lavoro è la constatazione che per un informatico gli atti di convegni sono spesso sedi di pubblicazione preferibili rispetto alle riviste, come è possibile facilmente indurre dalle seguenti citazioni:

*“The experimental computer scientist and engineer (ECSE) research community depends heavily on conferences to communicate new knowledge, and conferences are widely regarded as the preferred medium for maximizing the intellectual impact of ECSE research. However, the tenure and promotion process at many universities does not give conference presentations and publications a weight appropriate to their significance in the field, preferring instead publications in archival journals.”* [CSTB, 1994]

*“For experimentalists conference publication is preferred to journal publication, and the premier conferences are typically more selective than the premier journals.”* [Patterson et al, 1999]

*“In the Computer Science publication culture, prestigious conferences are a favorite tool for presenting original research – unlike disciplines where the prestige goes to journals and conferences are for raw initial results. [...] It is important not to use journals as the only yardstick for computer scientists.”* [Meyer et al, 2009]

Per sgombrare immediatamente il campo da equivoci, è opportuno rimarcare che le pubblicazioni negli atti dei principali convegni internazionali del settore informatico sono cosa ben diversa dalle cosiddette “comunicazioni a convegno” diffuse in altri settori. Ci sono infatti almeno tre elementi che distinguono il processo di pubblicazione di tutti i principali convegni informatici da quello dei congressi di altri settori:

- il processo di revisione è estremamente strutturato, e prevede almeno tre revisioni (in alcuni casi di più);
- i tassi di accettazione sono tipicamente molto bassi, spesso tra il 10% e il 20%, decisamente più bassi di quelli delle principali riviste del settore;
- il lavoro sottomesso e la successiva pubblicazione sono soggetti ad un limite di pagine, che però è relativamente alto (può arrivare fino a 10-15 pagine in doppia colonna, ovvero l'equivalente di circa 25-40 pagine nel formato a singola colonna e 30-50 pagine nel formato delle Lecture Notes della Springer).

Per esemplificare meglio questi aspetti, di seguito vengono riportati tassi di accettazione di alcuni prestigiosi convegni internazionali<sup>2</sup>:

- International Conference on Software Engineering (ICSE): 13%
- Principles of Programming Languages (POPL): 18%
- ACM SIGMOD International Conference on the Management of Data: 14%

E' interessante il caso della conferenza SIGMOD, che negli ultimi anni organizza il proprio processo di revisione in modo molto articolato per migliorare la selezione dei contributi:

- in una prima fase, ciascun articolo viene revisionato da tre membri del comitato di indirizzo del convegno; gli articoli vengono assegnati ai membri del comitato scientifico sulla base dei loro interessi di ricerca e delle loro competenze specifiche;
- il risultato della prima valutazione viene inviato agli autori, che hanno la possibilità di replicare ai commenti dei revisori fornendo un “feedback”; il feedback viene inoltrato ai revisori; in questa fase, in caso di giudizi contrastanti tra i revisori o di segnalazioni da parte degli autori nel loro feedback, un revisore aggiuntivo può essere assegnato al lavoro;
- al termine di questa fase, sulla base delle valutazioni dei revisori vengono selezionati i lavori da presentare al convegno e pubblicare negli atti del convegno; questi lavori, però, sono sottoposti ad una ulteriore verifica, chiamata “*repeatability and workability evaluation*” [Manegold et al, 2009]; la repeatability track ha come obiettivo quello di verificare la solidità e la ripetibilità delle valutazioni sperimentali contenute nel lavoro accettato; a questo scopo viene nominato un secondo comitato di revisori; gli autori sono invitati a fornire copia dei propri sistemi software e dei dati sperimentali ai revisori; ogni lavoro viene assegnato a due revisori, che cercano di riprodurre i risultati contenuti nell'articolo, ed esprimono una valutazione che viene resa pubblica.

Alla luce di queste informazioni, dovrebbe risultare più comprensibile il fatto che queste sedi di pubblicazioni vengano tradizionalmente considerate privilegiate dagli informatici: “*The reason conference publication is preferred to journal publication, at least for experimentalists, is the shorter time to print (7 months vs 1-2 years), the opportunity to describe the work before one's peers at a public presentation, and the more complete level of review (up to 4-5 evaluations per paper compared to 2-3 for archival journals). Publication in the prestige conferences is inferior to the prestige journals only in having significant page limitations and little time to polish the paper. In those dimensions that count most, conferences are superior.*” [Patterson et al, 1999].

A questo proposito, è interessante analizzare i risultati di uno studio condotto nel 2005 sulle principali riviste e conferenze del settore delle basi di dati (*databases*) [Rahm and Thor, 2005]. Lo studio, tra l'altro, utilizza Google Scholar per generare una classifica delle pubblicazioni in atti di convegni più citate in assoluto tra il 1994 e il 2003. Delle prime cinque pubblicazioni della classifica, solo due hanno successivamente dato origine ad una versione estesa su rivista. In uno dei due casi, la

<sup>2</sup> Si tratta di valori medi calcolati considerando le edizioni dal 2005 al 2007. Fonti: [SE Conferences], [DB Acceptance Rates].

versione su rivista è stata pubblicata 4 anni dopo l'articolo originale presentato alla conferenza. E' tra l'altro interessante il fatto che questo studio mostra che gli articoli pubblicati nelle principali conferenze relative alle basi di dati ricevono un numero di citazioni che è ordini di grandezza superiore a quello delle principali riviste del settore. Rilevante appare anche il fatto che le cinque pubblicazioni su rivista più citate nel periodo in esame sono tutte rassegne della letteratura (*survey*), oppure rassegne sui risultati di progetti di ricerca oramai scientificamente maturi, tipologia di lavoro per cui le riviste appaiono effettivamente sedi di pubblicazione più adeguate.

E' importante enfatizzare ulteriormente il fatto che questa singolarità nella cultura scientifica degli informatici è spesso vissuta in modo conflittuale dagli stessi informatici. Negli ultimi anni è cresciuta la consapevolezza di alcune distorsioni di questa pratica scientifica [Crowcroft, 2009], e del prezzo che queste comportano. Questo fatto ha giustificato una serie di interventi recenti [Vardi, 2009] [Reed, 2009] [Fortnow, 2009] secondo i quali sarebbe opportuno riconsiderare le pratiche scientifiche nel settore informatico, ed una serie di iniziative concrete (es: [Snodgrass, 2003] [Bernstein et al, 2005] [Jagadish, 2008]) per avvicinare il meccanismo di pubblicazione a quello di discipline più tradizionali. Una delle più significative è la creazione della rivista "Proceedings of the VLDB Endowment" [PVLDB], che ha l'obiettivo di riavvicinare la pubblicazione in atti di congressi a quella più tradizionale su rivista nel settore delle basi di dati. E' troppo presto per giudicare l'efficacia di queste iniziative<sup>3</sup> e capire quanto questo movimento di idee riuscirà ad influire realmente sulle consuetudini del settore. In ogni caso, rimarrà per anni il problema di valutare l'attività scientifica condotta secondo gli schemi attuali.

Per concludere, vale la pena di notare che spesso, nel discorso, viene fatto riferimento a ricercatori di ambito sperimentale. Le considerazioni in questione valgono in effetti anche per gli informatici teorici, ma con una minore enfasi: "*Though conference publication is highly regarded in the theoretical community, there is a long tradition of completing, revising, and extending conference papers for submission and publication in archival journals.*" [Patterson et al, 1999].

Più in generale, è opportuno rimarcare esplicitamente che l'obiettivo del documento non è quello di disconoscere il valore delle riviste come sedi di pubblicazione per gli informatici: esistono riviste in tutti i sottosectori dell'informatica che rappresentano certamente sedi di pubblicazione di primissimo livello. Alla luce della discussione fino a questo punto, dovrebbe però essere chiaro che, al fianco delle riviste, nelle attività di valutazione della ricerca devono essere tenute in debita considerazione anche le pubblicazioni in atti di conferenze di pari livello.

### **3 La Valutazione della Qualità e gli Indici Bibliometrici**

#### **3.1 Quantità e Impatto**

E' importante enfatizzare il fatto che le considerazioni della sezione precedente valgono certamente per le conferenze di primo livello del settore, ma progressivamente meno per le conferenze di livello inferiore. E queste ultime sono decisamente la maggioranza.

Negli ultimi anni si è assistito ad una notevolissima proliferazione degli eventi scientifici di argomento informatico (convegni, workshop, simposi) [Zhuang et al, 2007]. Questa proliferazione, com'è comprensibile, è stata spesso nemica della qualità, generando molte cosiddette conferenze "write-only", i cui contenuti rivestono un interesse scientifico marginale. In alcuni casi (fortunatamente pochi), addirittura nullo [SCIgen]. In misura minore, considerazioni analoghe valgono anche per alcune riviste, soprattutto per via dei nuovi meccanismi di pubblicazione elettronica che ne semplificano la creazione.

<sup>3</sup> E' sintomatico il fatto che il CINECA, nella sua base di dati delle pubblicazioni scientifiche, si rifiuta di considerare Proceedings of the VLDB Endowment come una rivista, e piuttosto la considera come una serie monografica o come un libro.

Accumulare pubblicazioni di limitato interesse scientifico in queste sedi minori non è un'impresa particolarmente ardua. E' quindi convinzione di molti autorevoli ricercatori che valutare la ricerca debba significare valutarne *l'impatto* sulla comunità scientifica, e non esclusivamente la quantità [Parnas, 2007]. *“Measures should not address volume but impact. Publication counts only assess activity. Giving them any other value encourages 'write-only' journals, speakers-only conferences, and stakhanovist research profiles favoring quantity over quality.”* [Meyer et al, 2009]. In altri termini, *“the fundamental basis for academic achievement is the impact of one's ideas and scholarship on the field”* [Patterson et al, 1999].

Ovviamente quantità e impatto sono necessariamente collegate: un ricercatore che scriva molti articoli ha maggiori probabilità di avere impatto sul settore di uno che ne scrive meno, purchè gli articoli che il primo scrive siano articoli di valore. Viceversa, però, un ricercatore che scriva moltissimi lavori di scarso valore scientifico ha certamente meno impatto di un altro che ne scriva relativamente pochi ma che portino contributi riconosciuti. Più in generale, è possibile dire che i metodi quantitativi hanno come obbiettivo quello di stimare la “quantità di impatto” collegata alla produzione scientifica del ricercatore nel periodo di riferimento, e quindi fatalmente sono destinati a tenere in considerazione entrambi gli aspetti.

Valutare l'impatto della produzione scientifica in termini di avanzamento delle conoscenze e dello stato dell'arte è però un obiettivo ambizioso. La strategia considerata più credibile per perseguire questo obiettivo è certamente quella del peer-reviewing [Patterson et al, 1999]. Purtroppo, questa non è priva di difetti collegati alla soggettività delle valutazioni dei revisori [Meyer et al, 2009]<sup>4</sup>. Peraltro, non è sempre concretamente fattibile né economicamente perseguibile, soprattutto nel contesto della valutazione periodica dell'attività scientifica degli individui a livello locale (risulta, viceversa, decisamente più appropriata in termini di rapporto costi-benefici per la valutazione delle strutture a livello nazionale [CIVR, 2007]).

Nelle valutazioni ripetute dei singoli con frequenza relativamente alta svolte a livello locale è quindi spesso necessario adottare approcci quantitativi, certamente più economici. Questi approcci necessariamente adottano vari livelli di approssimazione nella valutazione della qualità della ricerca.

Un primo livello di approssimazione, considerato di solito accettabile, è quello di valutare l'impatto delle pubblicazioni scientifiche sulla base della reputazione delle relative sedi di pubblicazione, siano esse riviste o conferenze. Questa è certamente un'approssimazione, dal momento che dipende fortemente dalle scelte dell'autore al momento della sottomissione e dalla sua percezione personale del valore delle sedi di pubblicazione. Ma certo è ragionevole l'idea che le pubblicazioni di maggior impatto siano pubblicate in sedi prestigiose.

### 3.2 Gli Indici ISI

E' decisamente più delicato, invece, il secondo livello di approssimazione, ovvero quello che si introduce al momento in cui è necessario valutare il prestigio ed il valore scientifico delle sedi di pubblicazione. Nelle scienze tradizionali, questa valutazione si basa di solito su misure di carattere bibliometrico, e più specificamente sul cosiddetto *“Journal Impact Factor”* calcolato dal Thompson Institute for Scientific Information (ISI) nel suo Journal Citation Report.

L'impact factor, proposto da Eugene Garfield (un chimico) nel 1955 [Garfield, 1955], valuta l'impatto delle pubblicazioni di una rivista in un certo anno sulla base del numero medio di citazioni ricevute da articoli di quella rivista nei 2 (o 5) anni precedenti. Le citazioni vengono calcolate considerando esclusivamente le riviste indicizzate dall'ISI, la cosiddetta ISI Web of Science. La stessa ISI pubblica annualmente un ranking delle riviste in vari settori, sulla base del relativo impact factor.

---

<sup>4</sup> Si vedano anche, a questo proposito, le molte critiche sul sistema di selezione dei progetti da finanziare nell'ambito dei bandi PRIN italiani.

A questo riguardo, però, esistono in letteratura numerosissimi studi che sostengono la seguente tesi: *la ISI Web of Science e il relativo impact factor sono strumenti inadatti a valutare l'impatto delle pubblicazioni nel settore informatico*. E' letteralmente impossibile, per ragioni di spazio elencarli tutti, e quindi ci limiteremo ad alcuni dei più significativi: [Rahm and Thor, 2005] [Meho and Yang, 2007] [Kermarrec et al, 2007] [Meyer et al, 2009].

Ci sono numerose ragioni che supportano questa tesi. Queste ragioni sono discusse in dettaglio nei documenti citati, che è facilmente possibile reperire sulla rete. In estrema sintesi, diremo che i criteri secondo cui l'ISI include le sedi di pubblicazione nella sua base di dati sono completamente disallineati con le pratiche di pubblicazione degli informatici: alcune riviste di buon livello non sono indicizzate, altre di livello decisamente inferiore sì; gli atti di molti dei principali congressi non compaiono per nulla, mentre atti di convegni minori sono arbitrariamente inclusi nel database.

L'inadeguatezza della copertura delle sedi di pubblicazioni da parte della base di dati ISI è dimostrata dalla "copertura interna" [Moed and Visser, 2005] delle citazioni contenute nella base di dati, ovvero la percentuale di citazioni di lavori contenuti nella base di dati che si riferiscono a lavori anch'essi contenuti nella base di dati; per la fisica e per la chimica questo indice supera l'80%; per l'informatica si attesta attorno al 38%.

Una citazione decisamente efficace sull'argomento è la seguente: *"The (ISI) database has little understanding of what constitutes computer science. The 50 most cited references in computer science according to ISI include entries as 'Chemometrics in food science', from a journal called 'Chemometrics and Intelligent Laboratory Systems', a topic and a publication entirely alien to computer science. This is not an isolated example; most of the entries on the list, even those which are related to computer science, are not recognizable to computer scientists as milestone contributions. The cruelest comparison is with the list of most cited computer science works on the CiteSeer<sup>5</sup> site, devoted to computer science; while imperfect like any such selection, the CiteSeer lists many articles and books familiar to computer scientists. It has not a single entry in common with the ISI list."* [Meyer et al, 2009].

Studi recenti [Rahm and Thor, 2005] [Meho and Yang, 2007] [Kermarrec et al, 2007] mostrano in modo consistente come Google Scholar<sup>6</sup>, per quanto esso stesso non alieno da difetti, rappresenti una alternativa decisamente preferibile alla base di dati ISI, sia per ampiezza e copertura delle citazioni contenute, che per qualità dei risultati ottenibili. Peraltro, adottando Google Scholar come base di dati per l'analisi delle citazioni è possibile vedere come gli atti di convegni di primo livello abbiano impact factor decisamente superiori rispetto a quelli delle riviste di primo livello del settore [Rahm and Thor, 2005].

Vale la pena di notare che critiche rispetto all'adozione degli indici ISI per la valutazione della ricerca sono state mosse anche da ricercatori di altri settori. Si veda ad esempio [Figà Talamanca, 2000].

L'insistita applicazione delle misure bibliometriche dell'ISI ha spesso spinto i ricercatori dei settori penalizzati a rifiutare del tutto gli approcci quantitativi alla valutazione della ricerca. Questo è probabilmente un atteggiamento eccessivo, soprattutto in un'ottica pragmatica; secondo molti, le valutazioni quantitative sono destinate a rimanere parte integrante del nostro sistema: *"it is unrealistic to discard the idea of numeric criteria entirely. Even if one ignores the political context (the need for universities to show tangible results, and the attraction of numbers), it is not clear that the suggested alternatives are always better"* [Meyer et al, 2009]. E' quindi più sensato cercare di intervenire per smussarne i difetti (si veda anche, a questo proposito, [CUN, 2010]).

In quest'ottica, nella valutazione della ricerca informatica possiamo certamente affermare che risulta inaccettabile la seguente assunzione (piuttosto frequente nella pratica), secondo cui una rivista

---

<sup>5</sup> <http://citeseer.ist.psu.edu/>, dati rilevati nel febbraio 2008

<sup>6</sup> <http://scholar.google.com>

indicizzata dall'ISI è una sede di pubblicazione di maggior prestigio ed impatto di una rivista non indicizzata dall'ISI, o di una conferenza.

Altrettanto inaccettabile è l'assunzione secondo la quale le misure di impatto basate su indici ISI possono essere utilizzate per il confronto dell'attività di ricerca in aree diverse (es: informatica, matematica, chimica). Ci sono numerosissimi elementi che differenziano in modo significativo le pratiche di pubblicazione di aree diverse. Un esempio è il numero di coautori (oltre che l'ordine dei nomi, un elemento considerato di minore importanza nel settore informatico). Un'analisi del numero medio di autori di articoli pubblicati tra il 2007 e il 2008 [Meyer et al, 2009] fornisce i seguenti dati, da cui si evincono eterogeneità macroscopiche:

- Nature: massimo numero di autori: 22, numero medio: 7,3;
- American Mathematical Monthly: massimo 6, medio: 2
- International Conference on Principles of Programming Languages (POPL): massimo: 7, media 2,7 (analogamente per la conferenza OOPSLA)

Un altro esempio è la variazione, spesso significativa, nel numero medio di pagine dei lavori pubblicati su riviste di settori diversi. Alla luce di questi elementi, emerge l'assoluta necessità di interpretare e correggere i dati quantitativi prima del confronto tra le aree, tenendone in considerazione le differenze.

Ugualmente preoccupante è la tendenza a valutare la qualità delle pubblicazioni sulla base del numero di citazioni ricevute dalla pubblicazione stessa, come prefigurato nel decreto legge per la valutazione della ricerca degli atenei italiani tra il 2004 e il 2008 [CIVR 2004-2008]. Per quanto questo possa rappresentare effettivamente un modo per stimare l'impatto di una specifica pubblicazione, resta vero che l'attendibilità di queste valutazioni è strettamente collegata all'attendibilità della base di citazioni adottata.

Per ulteriori elementi di riflessione sui limiti degli indici bibliometrici e dell'impact factor nella valutazione della ricerca si veda [Adler et al, 2008].

### **3.3 Ranking e Classificazioni**

Dovendo procedere ad una stima dell'impatto delle sedi di pubblicazione informatiche, siano esse riviste o conferenze, è quindi indispensabile adottare classificazioni ("ranking") più solidi e credibili scientificamente rispetto al ranking ISI.

E' diffusa la pratica di classificare le sedi di pubblicazione in fasce, con una tipica organizzazione di questo genere:

- fascia A+/A: riviste o conferenze di eccellenza/di primo livello scientifico;
- fascia B: riviste o conferenze di buon livello scientifico
- fascia C: riviste o conferenze di livello scientifico accettabile
- fascia D: riviste o conferenze di limitato interesse scientifico

A queste può essere aggiunta una ulteriore fascia (NC), composta dalle riviste e conferenze non classificate.

Utilizzando gli indici ISI, limitatamente alle riviste indicizzate, è naturale adottare il seguente algoritmo di classificazione: considerata la classifica rispetto all'impact factor (ISI Journal Ranking), viene selezionato il primo quartile come fascia A (il 25% delle riviste con impact factor più elevato), il secondo quartile come fascia B e così via. Questo tipo di classificazione ha una sua coerenza interna, limitatamente alle riviste contenute nella base di dati ISI, ma è ovviamente insoddisfacente per il settore informatico.

Sono state sviluppate quindi numerose ulteriori classificazioni sia delle riviste che delle conferenze informatiche, ordinariamente messe a punto da appositi comitati di valutazione, e quindi non basate su indici bibliometrici. Di seguito ne vengono elencate alcune:

- il GRIN (Gruppo Nazionale di Informatica) italiano, ha prodotto tra il 2003 e il 2004 una classificazione sia delle riviste del settore, che delle principali conferenze [GRIN]; dopo la prima pubblicazione, la classificazione non è stata ulteriormente aggiornata, e quindi risente di un certo livello di obsolescenza;
- l'associazione CORE Australiana (Computer Research and Education Association of Australasia) pubblica da anni una classificazione molto dettagliata delle riviste e delle conferenze informatiche [CORE]; la classificazione viene aggiornata quasi annualmente; attualmente sono disponibili i dati della classificazione 2008 per le riviste, e di quella 2009 per le conferenze;
- l'Università di Singapore pubblica un ranking delle conferenze in vari settori dell'informatica [U. Singapore Conf Listing];
- l'Università di Hong Kong ha promosso un sito che offre servizi di ranking delle riviste sulla base di vari indici bibliometrici [Journal Ranking].

Alcuni singoli e/o istituzioni hanno intrapreso l'iniziativa di classificare le conferenze o le riviste del settore informatico<sup>7</sup>, oppure quelle di sottosectori specifici<sup>8</sup>, come le basi di dati o l'ingegneria del software. Ovviamente l'attendibilità di queste classificazioni è pari a quella del soggetto classificatore, ed alla solidità della metodologia adottata. E' cruciale, infatti, diffidare di classificazioni di scarsa autorevolezza che possono essere volutamente manipolate [Ranking Expose].

Adottare una classificazione accurata, tanto delle riviste, tanto delle conferenze informatiche, è quindi un problema centrale della valutazione. Ci sono svariate strategie che è possibile adottare per avere una classificazione di qualità, ognuna con i propri pregi e i propri difetti:

- la prima è ovviamente quella di procedere ad una nuova classificazione manualmente, con l'ausilio di esperti, e mantenerla periodicamente per incorporare le nuove sedi di pubblicazione e tenere in considerazione le variazioni nello status di quelle già classificate; questo approccio ha però due limiti che lo rendono spesso inapplicabile; per cominciare è decisamente costoso in termini di tempo e risorse; inoltre, corre il rischio di essere contaminato da negoziazioni di carattere politico tra i sottosectori, ognuno dei quali percepisce le proprie sedi di pubblicazione come particolarmente rilevanti;
- la seconda strategia consiste nell'adottare, come già discusso in precedenza, una delle classificazioni esistenti, come ad esempio quella del CORE; questo approccio è decisamente più economico, ma finisce per rispecchiare la visione del settore dell'organizzazione che ha compilato la classificazione;
- un terzo approccio, che cerca di mantenere i vantaggi dei precedenti eliminandone per quanto possibile i difetti, è quello di costruire una nuova classificazione sulla base delle varie classificazioni esistenti; in sintesi, questo approccio consiste nell'attribuire un valore numerico di reputazione ad una sede di pubblicazione sulla base della sua considerazione nelle classificazioni di partenza (valori massimi per la classe A e minimi per la classe D ed NC); un modo per farlo è

---

7 Alcuni esempi sono riportati di seguito:

<http://webdocs.cs.ualberta.ca/~zaiane/htmldocs/ConfRanking.html>

[http://www.cs.albany.edu/~ashwin/Conf\\_rank.html](http://www.cs.albany.edu/~ashwin/Conf_rank.html)

<http://www3.ntu.edu.sg/home/assourav/crank.htm>

Si tratta per lo più di dati derivati da una classificazione delle conferenze informatiche adottate dall'Università di Singapore per una sua valutazione dell'attività di ricerca svolta nel 1999 [Long et al, 1999].

8 Esempi: <http://www-cse.ucsd.edu/fileadmin/cse/attach/database.doc>  
<http://people.engr.ncsu.edu/txie/seconferences.htm>

calcolare la reputazione complessiva della conferenza o della rivista sommando i suoi valori di reputazione nelle classificazioni di partenza; infine, utilizzare questi valori per costruire una nuova classifica e quindi una nuova classificazione sulla base dei percentili.

## 4 Il Resto del Documento

Le sezioni precedenti discutono i principi che dovrebbero ispirare la valutazione della ricerca degli informatici. Vale la pena di notare che simili indicazioni emergono anche dal recente documento del CUN sugli “Indicatori di Attività Scientifica e di Ricerca” [CUN, 2009], che, con riferimento al settore Informatica dell'Area 01, dichiara chiaramente inapplicabili indici bibliometrici come l'impact factor ISI e sancisce la pari dignità di trattamento tra le pubblicazioni su rivista e quelle in atti di congressi internazionali.

Per dare maggiore concretezza alla discussione sui sistemi di valutazione, ed evitare che questo documento si limiti ad enunciare esclusivamente principi e linee guida, sono state previste due appendici. L'Appendice I. prende in esame alcuni modelli specifici utilizzati per la valutazione della ricerca. Commenti, discussioni e questioni aperte sono contenuti nell'Appendice II.

### Ringraziamenti

Di seguito è riportato l'elenco dei colleghi informatici che hanno contribuito con le loro osservazioni a migliorare questo documento: Ugo Erra, Paolo Papotti, Carlo Sartiani, Giuseppe Scanniello. Li ringrazio sentitamente. Ringrazio inoltre i colleghi matematici Marica De Bonis, Elvira Di Nardo, Onofrio Di Vincenzo, Martin Funk e Maria Grazia Russo per i commenti e le discussioni sulla valutazione della ricerca nel settore matematico.

## Bibliografia

[ADLER ET AL, 2008] R. Adler, J. Ewing, P. Taylor, *Citation Statistics*, A Report from the International Mathematica Union, in cooperation with the International Council of Industrial and Applied Mathematics and the Institute of Mathematical Statistics, 2008  
<http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>

[BERNSTEIN ET AL, 2005] P. A. Bernstein, D. De Witt, A. Heuer, Z. Ives, C. S. Jensen, H. Meyer, M. T. Ozsu, R. T. Snodgrass, K.-Y. Whang, and Widom, J. , Database Publication Practices, *Proceedings of the 31st VLDB Conference*, 1241-1245, 2005

[CIVR 2004-2008] *D. M. 19/03/2010 n. 8 - Linee Guida VQR 2004-2008*,  
<http://attiministeriali.miur.it/anno-2010/marzo/dm-19032010-n-8.aspx>

[CIVR, 2007] Comitato di Indirizzo per la Valutazione della Ricerca, *Valutazione Triennale della Ricerca 2001-2003*, 2007

[CORE] *Computing Research and Education Association of Australasia*, <http://www.core.edu.au/>

[CROWCROFT, 2009] , *Scaling Internet Research Publication Processes to Internet Scale*, Communications of the ACM, 52 (1), 27-30, 2009

[CSTB, 1994] Computer Science and Telecommunications Board, *Academic Careers for Experimental Computer Scientists and Engineers*, 1994 [http://www.nap.edu/openbook.php?record\\_id=2236](http://www.nap.edu/openbook.php?record_id=2236)

[CUN, 2009] *CUN - Indicatori di Attività Scientifica e di Ricerca*,  
[http://www.cun.it/media/100062/indicatori\\_completo.pdf2009](http://www.cun.it/media/100062/indicatori_completo.pdf2009)

[CUN, 2010] *CUN - Parere sulla Valutazione della Ricerca (25/03/2010)*,  
[http://www.cun.it/media/104777/ps\\_2010\\_03\\_25\\_002.pdf](http://www.cun.it/media/104777/ps_2010_03_25_002.pdf)

[DB ACCEPTANCE RATES] *Acceptance Rates of Major Database Conferences*,  
<http://wwwhome.cs.utwente.nl/~apers/rates.html>

[DIT] , *Documento Indicatori DIT*, Department of Information and Communication Technology - University of Trento, 2005 (Documentazione Interna)

[EGGHE, 2006] L. Egghe, *Theory and Practice of the g-index*, *Scientometrics*, 69 (1), 131-152, 2006

[FIGÀ TALAMANCA, 2000] A. Figà Talamanca, *L'Impact Factor nella Valutazione della Ricerca e nello Sviluppo dell'Editoria Scientifica* , *IV Seminario su Sistema Informativo Nazionale Per la Matematica* , -, 2000 <http://siba2.unile.it/sinm/4sinm/interventi/fig-talam.htm>

[FORTNOW, 2009] L. Fortnow, *Time for Computer Science to Grow-Up*, *Communications of the ACM*, 58 (8), 33-35, 2009

[GARFIELD, 1955] E. Garfield, *Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas*, *Science*, 122 (3159), 108-111, 1955

[GRIN] *GRIN - Valutazione della Ricerca*, <http://www.di.unipi.it/grin/ricerca.html>

[HIRSCH 2005] J. E. Hirsch, *An Index to Quantify an Individual's Scientific Research Output*, , 2005 <http://arxiv.org/abs/physics/0508025> arXiv:physics/0508025

[JAGADISH, 2008] H. V. Jagadish, *The Conference Reviewing Crisis and a Proposed Solution*, *ACM Sigmod Record*, 37 (3), 40-45, 2008

[JOURNAL RANKING] *Red Jasper's Journal Ranking.com* , <http://www.journal-ranking.com>

[KELCHTERMANS, 2009] S. Kelchtermans, R. Veugeliers, *The Great Divide in Scientific Productivity. Why the Average Scientist Does Not Exist*, Katholieke Universiteit Leuven, 2009 <http://ideas.repec.org/p/ner/leuven/urnhdl123456789-213759.html>

[KERMARREC ET AL, 2007] A.-M. Kermarrec, E. Faou, J.-P. Merlet, P. Robert, L. Segoufin, *What do Bibliometric Indicators Measure ?* , 2007 (INRIA Evaluation Committee Analysis Document)

[LONG ET AL, 1999] P. M. Long, T. K. Lee, J. Jaffar, *Benchmarking Research Performance in Department of Computer Science*, School of Computing, National University of Singapore, 1999 <http://www.comp.nus.edu.sg/~tankl/bench.html>

[MANEGOLD ET AL, 2009] S. Manegold, I. Manolescu, L. Afanasiev, J. Feng, G. Gou, M. Hadjieleftheriou, S. Harizopoulos, P. Kalnis, K. Karanasos, D. Laurent, M. Lupu, N. Onose, C. Rè, V. Sans, P. Senellart, T. Wu, D. Shasha, *Repeatability and Workability Evaluation of SIGMOD 2009*, *SIGMOD Record*, 38 (3), 40-43, 2009

[MEHO AND YANG, 2007] L. I. Meho, K. Yang, *Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science versus Scopus and Google Scholar*, *Journal of the American Society for Information Science and Technology*, 58 (13), 2105-2125, 2007

[MEYER ET AL, 2009] B. Meyer, C. Choppy, J. Staunstrup, J. van Leeuwen, *Research Evaluation for Computer Science*, *Communications of the ACM*, 52 (4), 31-34, 2009 [http://www.informatics-europe.org/ECSS08/papers/Research\\_evaluation\\_CACM.pdf](http://www.informatics-europe.org/ECSS08/papers/Research_evaluation_CACM.pdf)

[MOED AND VISSER, 2005] H. F. Moed, M. S. Visser, *Developing Bibliometric Indicators of Research Performance in Computer Science: an Exploratory Study*, *Proceedings of the Conference of the International Society for Scientometrics and Infometrics*, 275-279, 2005

[PARNAS, 2007] D. Parnas, *Stop the Numbers Game - Counting Papers Slows the Rate of Scientific Progress*, *Communications of the ACM*, 50 (11), 19-21, 2007

[PATTERSON ET AL, 1999] D. Patterson, L. Snyder, J. Ullman, *Computing Research Association - Best Practices Memo: Evaluating Computer Scientists and Engineers for Promotion and Tenure*, *Computing Research News*, (September), A-B, 1999 [http://www.cra.org/resources/bp-view/evaluating\\_computer\\_scientists\\_and\\_engineers\\_for\\_promotion\\_and\\_tenure/](http://www.cra.org/resources/bp-view/evaluating_computer_scientists_and_engineers_for_promotion_and_tenure/)

[PoP] *Publish or Perish*, <http://www.harzing.com/pop.htm>

[PVLDB] *Proceedings of the VLDB Endowment*, <http://www.vldb.org/PVLDB>

[RAHM AND THOR, 2005] E. Rahm and A. Thor, *Citation Analysis of Database Publications*, SIGMOD Record, 34 (4), 48-53, 2005

[RANKING EXPOSE] *Avoid Anonymous Online Conference Ranking Sites*, <http://rankingexpose.com/>

[REED, 2009] D. Reed, *Publishing Quarks: Considering our Culture*, Computing Research News, 21 (2), 3, 2009

[SCI GEN] *SCIgen - An Automatic CS Paper Generator*, <http://pdos.csail.mit.edu/scigen>

[SE CONFERENCES] *Software Engineering Conferences*, <http://people.engr.ncsu.edu/txie/seconferences.htm>

[SNODGRASS, 2003] R. T. Snodgrass, *Journal Relevance*, SIGMOD Record, 32 (3), 11-15, 2003

[U. SINGAPORE CONF LISTING] *University of Singapore - Conference Listing*, [http://www.cais.ntu.edu.sg/content/research/conference\\_list.jsp](http://www.cais.ntu.edu.sg/content/research/conference_list.jsp)

[VARDI, 2009] M. Vardi, *Conferences vs. Journals in Computing Research*, Communications of the ACM, 52 (5), 5, 2009

[ZHUANG ET AL, 2007] Z. Zhuang, E. Elmacioglu, D. Lee, C. Lee Giles, *Measuring Conference Quality by Mining Program Committee Characteristics*, *Proceedings of the International Conference on Digital Libraries*, 225-234, 2007

## I. Appendice: Alcuni Modelli Esemplificativi

In questa appendice vengono discussi due diversi modelli applicabili per la valutazione della ricerca, allo scopo di metterne in luce le criticità alla luce dei principi enunciati nella prima parte, e suggerire interventi migliorativi.

I modelli discussi si basano tutti su alcuni presupposti comuni:

- sono modelli orientati alla classificazione della ricerca dei singoli o delle strutture, di solito utilizzati in valutazioni periodiche in cui è indispensabile adottare criteri quantitativi, oggettivi e rapidamente applicabili; si esclude a priori, in queste occasioni, di adottare tecniche di peer-reviewing; casi tipici sono la ripartizione di risorse di Ateneo a livello locale (fondi per il sostegno alla ricerca, borse di dottorato, assegni di ricerca);
- possono prendere in considerazione prodotti della ricerca di vario genere (brevetti, revisioni, commenti, ecc.); per semplicità, in questa discussione ci concentreremo esclusivamente su tre tipologie di prodotti: articoli su rivista, articoli su atti di convegni con processo di revisione, e libri;
- si basano sul presupposto di attribuire a ciascun ricercatore o struttura un punteggio che stimi la qualità della ricerca svolta; il punteggio è di norma ottenuto combinando i punteggi attribuiti a ciascuno dei prodotti della ricerca nel periodo di riferimento.

Nell'ambito di questo modello comune, i due metodi si differenziano sensibilmente nelle modalità secondo cui attribuiscono i punteggi alle singole pubblicazioni e nella strategia per calcolare il punteggio complessivo per la produzione scientifica del ricercatore.

### Il Modello ISI-non ISI

Chiamerò “modello ISI-non ISI” qualsiasi modello per la valutazione della produzione scientifica basato sul presupposto secondo cui le pubblicazioni di primo livello sono essenzialmente quelle ospitate da riviste indicizzate dall'ISI.

Questo modello utilizza una classificazione dei prodotti basata su questo presupposto per attribuire i punteggi alle pubblicazioni. Un esempio potrebbe essere il seguente:

- categoria A+: libro o monografia a contenuto scientifico e diffusione internazionale (punti P1)

- categoria A: rivista internazionale ISI (punti P2)
- categoria B: rivista internazionale non-ISI o capitolo di libro (punti P3)
- categoria C: rivista nazionale o conferenza internazionale con processo di revisione (punti P4)

Tipicamente i punteggi sono organizzati in modo tale per cui  $P1 > P2 > P3 > P4$ . Il punteggio complessivo per la produzione scientifica del ricercatore o della struttura è ottenuto sommando i punteggi dei singoli prodotti, calcolati secondo la classificazione riportata sopra.

Alla luce dei principi e delle linee guida discussi nella prima parte di questo documento, è possibile muovere una serie di critiche che riguardano difetti piuttosto evidenti. Un critica di carattere generale è rappresentato dal fatto che il metodo misura prevalentemente la quantità della produzione scientifica, e molto meno la qualità di questa produzione. La strategia per la valutazione dell'impatto delle pubblicazioni è infatti relativamente poco accurata, e piuttosto basata su una filosofia secondo cui esistono pubblicazioni di maggior valore (le riviste ISI) e pubblicazioni di minor valore (quelle non ISI). Il valore viene riconosciuto sulla base della classificazione effettuata da un'ente esterno (l'ISI stesso). Ma è ovvio che l'indice ISI include tanto riviste di altissima selettività e valore scientifico, tanto riviste di valore decisamente inferiore, e quindi questo criterio è molto discutibile. A ben guardare, il metodo non dà una misura accurata nemmeno della quantità dell'attività scientifica, dal momento che ignora alcuni aspetti cruciali per la misurazione quantitativa, come per esempio il numero di autori dei singoli lavori.

Va detto, però, che a dispetto di questi difetti, il modello ha il pregio di essere di semplicissima applicazione e di facile condivisione nei contesti in cui sia necessaria una negoziazione politica sui criteri di valutazione. Di conseguenza viene applicato molto più frequente di quanto ci si aspetterebbe<sup>9</sup>.

Volendo mantenere lo spirito del metodo (senza necessariamente dividerlo), resta il fatto che si tratta di un modello del tutto inappropriato a valutare l'attività scientifica degli informatici. Per rendere il meccanismo di calcolo dei punteggi più adeguato a valutare l'attività degli informatici, sono ovviamente necessari almeno due interventi:

- considerare le conferenze oltre che le riviste come prodotti potenzialmente di primo livello;
- adottare una classificazione delle sedi di pubblicazione che non sia basata esclusivamente sull'appartenenza all'indice ISI.

Nel seguito supporremo di avere adottato una classificazione di questo tipo. Per rispettare la semplicità complessiva del modello e la sua natura "politica", è possibile pensare di adottare a questo scopo la classificazione delle riviste e delle conferenze effettuata dal GRIN. In alternativa, è possibile procedere secondo le linee guida per l'adozione di classificazioni discusse nella prima parte. Supporremo quindi di avere adottato una classificazione delle riviste e delle conferenze, basata su cinque fasce (A, B, C, D, NC). Coerentemente con l'impostazione del modello, secondo cui tutte le riviste ISI vengono considerate sedi di pubblicazione di primo livello (indipendentemente dal loro impact factor), e le riviste non-ISI sedi di pubblicazione di secondo livello, definiamo:

- *sede di pubblicazione informatica di primo livello* ogni rivista o conferenza di classe A, B o C secondo la classificazione adottata;
- *sede di pubblicazione informatica di secondo livello* ogni rivista o conferenza di classe D o NC secondo la classificazione adottata.

La classificazione dovrebbe essere poi emendata come segue:

---

<sup>9</sup> Varianti di questo modello sono state adottate in varie circostanze negli anni passati presso l'Università della Basilicata per la ripartizione delle risorse di Ateneo, a volte con interventi ad hoc per le Aree CUN 08 (Ingegneria Civile ed Architettura), 10 e 11 (Scienze Umane).

- categoria A+: libro o monografia a contenuto scientifico e diffusione internazionale (P1)
- categoria A: rivista internazionale ISI per i settori non-informatici, o **rivista informatica di primo livello**, o **conferenza informatica di primo livello** (punti P2)
- categoria B: rivista internazionale non-ISI per i settori non-informatici, **rivista informatica di secondo livello**, **conferenza informatica di secondo livello**, o capitolo di libro (P3)
- categoria C: rivista nazionale o conferenza internazionale **per i settori non informatici** con processo di revisione (punti P4)

Nonostante questo, resta un'ulteriore caratteristica estremamente discutibile di questo metodo. Riguarda la pretesa di confrontare punteggi ottenuti da ricercatori appartenenti ad aree completamente diverse. Da questo punto di vista sarebbe necessario, piuttosto che limitarsi a confrontare i punteggi ottenuti dalle somme, effettuare un'attività di normalizzazione. Anche questo aspetto sarà discusso nella prossima sezione.

## Il Modello Bologna

L'Osservatorio per la Ricerca dell'Università di Bologna<sup>10</sup> coordina da anni un'attività di valutazione della ricerca per i docenti e le strutture dell'Ateneo. Nel corso degli anni è stato messo a punto un modello di valutazione estremamente dettagliato, che si candida a rappresentare un punto di riferimento a livello nazionale. Nel seguito della sezione descriviamo rapidamente questo modello<sup>11</sup>.

Nel modello Bologna, l'*Indice di Produttività (IP)*<sup>12</sup> di un ricercatore in un periodo (di solito un triennio) si ottiene come somma dei valori di produttività assegnati ai lavori del ricercatore nel periodo. I lavori vengono classificati in 5 categorie AA, A, B, C, D. Ai lavori di ciascuna categoria viene assegnato un *valore di produttività*, come segue:

Classe	AA	A	B	C	D
Valore	1,5	1	0,6	0,3	0,05

Il valore di produttività viene poi moltiplicato per un *fattore correttivo (fc)* che dipende dal numero di autori, secondo la seguente formula:

$$fc = e^{-\frac{n-1}{12}}$$

La formula produce i seguenti valori:

Autori	1	2	3	4	5	6
fc	1	0,92	0,85	0,78	0,72	...

Prima del confronto con aree diverse, gli indici di produttività vengono sottoposti ad un'attività di normalizzazione, in modo da assorbire differenze macroscopiche tra le culture e le prassi di pubblicazione delle diverse aree. La procedura di normalizzazione è sofisticata, e ne omettiamo i dettagli per ragioni di spazio, ma, grossolanamente, corrisponde a calcolare una media pesata tra l'indice di produttività del ricercatore ed il valore normalizzato rispetto all'indice di produttività medio, secondo la seguente formula:

$$\frac{1}{3} IP + \frac{2}{3} \frac{IP}{IP_{medio\ area}}$$

<sup>10</sup> <http://www.unibo.it/Portale/Ricerca/Servizi+Docenti+Ricercatori/valutazione.htm>

<sup>11</sup> Si tratta della versione adottata per la valutazione del 2009.

<sup>12</sup> La dizione "indice di produttività" può risultare fuorviante, dal momento che il modello ha chiaramente l'ambizione di misurare la qualità e l'impatto dell'attività di ricerca oltre che la quantità; sarebbe stato forse più opportuno adottare dizioni alternative, come "indice di ricerca" o "indice di impatto".

Classificazione di prodotti e attività di ricerca per la sessione 2009 (prodotti: 2008; fonte: Anagrafe della Ricerca Unibo)				
Modello descrittivo	Tipi definiti		Tipi U	Tipi S
Libro	monografie e trattati scientifici in forma di libro (compresi i trattati, i commentari, i cataloghi critici, le traduzioni di testi letterari e le edizioni scientifiche con introduzioni aventi carattere di saggio, note o apparati, e simili)		AA-C	AA-C
	concordanze, indici, bibliografie, pubblicazione di fonti inedite		D-NC	D-NC
	traduzione		C-D	C-D
	edizione critica		AA-C	AA-C
Atto di convegno	convegni di rilevanza internazionale		A-C	C
	convegni di rilevanza nazionale		B-D	D
	riassunti (abstract)		NC	NC
Contributo specifico in volume	capitolo di libro		A-C	A-C
	voce di enciclopedia	di rilevante interesse scientifico a carattere espositivo	B-C D	B-C D
	prefazioni e brevi introduzioni		D	D
	traduzione		C-D	C-D
Pubblicazione in rivista	articolo	rivista con IF > del 66° percentile	A	A
		tra il 66° e il 33° percentile	B	B
		< del 33° percentile	C	C
	rivista senza IF (*)		A-D	C-D
	note a sentenza	di rilevante interesse scientifico a carattere espositivo	A-C D	A-C D
	recensioni		C-D	D-NC
	repliche, brevi interventi e simili		D	D-NC
traduzione		C-D	C-D	
Brevetto			C-D	C-D
	opera di rilevante interesse naz. o internaz in atlante o monografia		A-B	A-B

Figura 1: Modello Bologna - Classificazione dei Prodotti (riproduzione parziale)

La procedura è resa più complessa dalla necessità di rendere omogenei i due termini della somma, solo uno dei quali proviene da una normalizzazione. A questo scopo, entrambi i termini della somma vengono ulteriormente normalizzati, dividendo per la somma dei valori corrispondenti per tutti i partecipanti alla valutazione. La formula finale è la seguente:

$$IP_{Finale} = \frac{1}{3} \sum_{ateneo} \frac{IP}{IP_i} + \frac{2}{3} \frac{IP_{medioarea}}{\sum_{ateneo} \frac{IP}{IP_{medioarea}_i}}$$

L'attività di normalizzazione è un elemento centrale del metodo, che consente di assorbire le eterogeneità tra le aree discusse nel paragrafo precedente. E' altresì evidente che questa attività dipende fortemente dal presupposto secondo cui la distribuzione degli indici di produttività all'interno delle aree è statisticamente credibile. Questo richiede perlomeno che le aree abbiano una numerosità minima accettabile. In quest'ottica, nel caso in cui un'area sia relativamente piccola, può essere opportuno accorpala ad altre più numerose. Evidenze empiriche legate a valutazioni condotte negli anni successivi presso l'Università della Basilicata suggeriscono che è possibile pensare di accorpate (puramente allo scopo della normalizzazione) gli informatici con i matematici.

L'attribuzione dei prodotti alle classi viene fatto secondo una griglia decisamente dettagliata, riportata in gran parte nella Figura 1. Si noti che la classe attribuita alle varie tipologie di prodotto è

distinta per il settore umanistico (“tipi U”) e quello tecnico-scientifico (“tipi S”). La classificazione avviene in parte in modo automatico, in parte sulla base di indicazioni dei Comitati d'Area, che, nell'ambito dei criteri fissati dall'Osservatorio della Ricerca, hanno il compito di coordinare le attività di valutazione relativamente ai vari settori scientifici. I Comitati d'Area possono intervenire sulla classificazione in vari modi; ad esempio, in alcune aree all'interno delle classi avviene una distinzione tra lavori brevi e lavori lunghi, sulla base di regole fissate dal Comitato d'Area relativo.

Il modello è chiaramente un modello elaborato e frutto di un'analisi approfondita del problema. E' apprezzabile l'impianto complessivo, e la cura nello studio degli aspetti diversi della valutazione. Ci sono però alcuni aspetti nei quali può essere ulteriormente emendato per renderlo maggiormente adeguato a valutare l'attività di ricerca degli informatici. Oltre alla possibilità di intervenire sui vari coefficienti, l'elemento centrale anche in questo caso pare quello della classificazione dei prodotti. In particolare:

- è molto discutibile il fatto di riservare alle pubblicazioni in atti di convegni internazionali del settore tecnico-scientifico la classe C; piuttosto, sarebbe opportuno consentire la classificazione di tutte le pubblicazioni in atti di convegni nelle classi A, B, C e D, a seconda del loro valore;
- è discutibile attribuire tout court le riviste ISI alle classi A, B, C esclusivamente sulla base dell'analisi del percentile nel loro ranking; peraltro, si tratta di una strategia di classificazione che si presta a moltissime varianti, tanto relativamente al valore dell'IF da considerare (quello dell'anno di pubblicazione ? quello dell'anno precedente ? quello medio dei due ?), tanto relativamente alla popolazione di riviste da considerare per il calcolo dei percentili (l'intero JCR ? le singole aree ? le aree allargate recentemente pubblicate dall'ISI ?). Un semplice esperimento di classificazione delle riviste ISI nei percentili sulla base degli impact factor rivela immediatamente come la classificazione di una stessa rivista sia spesso soggetta a fluttuazioni del tutto arbitrarie, che certo non rispecchiano cambiamenti significativi nel suo valore scientifico, ma piuttosto anomalie nella distribuzione degli articoli e delle citazioni nelle diverse annate.

In entrambi i casi si ripropone, di conseguenza, il problema della corretta classificazione delle conferenze e delle riviste informatiche. Si tenga presente che l'Osservatorio della Ricerca di Bologna effettua in ogni caso una classificazione per le riviste non incluse nell'indice ISI. Per adeguare maggiormente il metodo alla valutazione degli informatici, è indispensabile dotarsi di una classificazione di tutte le riviste e tutte le conferenze informatiche, secondo le linee discusse nella prima parte.

## **II. Appendice: Commenti e Questioni Aperte**

In questa sezione sono riportati alcuni commenti suscitati dalla lettura di versioni preliminari di questo documento. Ognuno di questi commenti rappresenta un aspetto interessante della valutazione che merita ulteriore discussione. La sezione è separata dal corpo principale del documento per evitare di appesantirne eccessivamente la lettura, ma anche perché, a differenza della prima parte, forzatamente contiene opinioni di carattere personale.

### **1. Valutare l'Anzianità e il Ruolo**

*“E' inappropriato che i sistemi quantitativi di valutazione della ricerca utilizzino gli stessi criteri per valutare ricercatori ed ordinari”*

Il commento tocca un aspetto molto interessante del processo di valutazione, e cioè la distinzione tra le misure e l'utilizzo che ne viene fatto. Questo documento si concentra per le ragioni discusse sui metodi quantitativi di valutazione della ricerca. E' estremamente difficile incorporare all'interno di un metodo quantitativo tutti gli aspetti correlati alla professione accademica. Questo vale, per esempio, per le caratteristiche del ruolo (per esempio il fatto che un giovane ricercatore ha spesso una minore

“inerzia” nella ricerca di un ordinario, così come il fatto che un ordinario è spesso soggetto a compiti didattici e organizzativi molto più gravosi rispetto a quelli di un ricercatore). Di conseguenza, qualunque sia il metodo adottato, per mantenerlo ragionevolmente semplice e sufficientemente rigoroso, è ragionevole pensare che la misura della qualità della didattica sia applicata a tutti i soggetti allo stesso modo, indipendentemente dal ruolo e dalle ulteriori attività ad esso collegate.

D'altro canto, le valutazioni della ricerca sono uno strumento finalizzato a definire strategie accademiche, di norma per la divisione delle risorse. E' indispensabile che gli aspetti non direttamente inclusi nella valutazione quantitativa vengano tenuti in considerazione al momento della definizione delle strategie e degli interventi. Per essere più concreti, è in questa fase che vanno previsti interventi perequativi che tengano in considerazione, per esempio, le differenze di anzianità. Paradossalmente, nulla impedisce che in questa fase un ricercatore che abbia una produzione promettente ma limitata per via della scarsa anzianità venga premiato, piuttosto che penalizzato dai risultati della valutazione.

Si noti che la definizione delle politiche basate sulle valutazioni della ricerca è un argomento molto studiato (si veda, ad esempio, [Kelchtermans, 2009]). Ci sono in Italia esempi di politiche di valutazione, come quella del Dipartimento di Informatica dell'Università di Trento [DIT], che tengono in considerazione sia i “fattori di merito” (per esempio la produzione scientifica), che “fattori di impedimento”, come il sottodimensionamento dei gruppi o i carichi didattici e organizzativi, in modo da definire politiche di ripartizione delle risorse più equilibrate.

In ogni caso, rimarcare con forza la distinzione tra i processi di raccolta dei dati e valutazione, e quelli di analisi e definizione delle strategie è indispensabile. Ancor più nell'attuale momento storico del sistema italiano, in cui il proliferare dell'offerta didattica impone sacrifici notevoli a molti.

## **2. Altri Utilizzi degli Indici Bibliometrici**

*“Sarebbe opportuno discutere anche altri indici bibliometrici, come l'indice h o l'indice g”*

Altri indici bibliometrici sono stati proposti negli anni per valutare l'impatto della produzione scientifica. Esempi noti sono *l'indice di Hirsch*, o *h-index* [Hirsch 2005], e le sue varianti, come il *g-index* [Egghe, 2006] (per una discussione più ampia delle ulteriori varianti, si veda [PoP]). In sintesi, l'indice di Hirsch di un ricercatore è il numero h delle sue pubblicazioni che negli anni hanno ricevuto almeno h citazioni. E' possibile intuire come questo indice sia pensato per valutare l'impatto complessivo di un ricercatore nel complesso della sua carriera. Per esempio, il CUN ha recentemente adottato l'h-index come parametro di riferimento per definire gli standard minimi di qualità scientifica che la carriera di un ricercatore dovrebbe rispettare per accedere ai diversi livelli concorsuali [CUN, 2009]. Di conseguenza, non viene ulteriormente discusso in questo documento, che si concentra esclusivamente sul problema della valutazione *periodica* della produzione scientifica dei singoli o delle strutture con riferimento a periodi limitati. E' ovvio però che misure come l'h-index o il g-index possono servire a collocare meglio la valutazione periodica della produzione scientifica del ricercatore nel più ampio contesto della sua carriera.

Resta ovviamente vera l'affermazione secondo cui la validità di questi indici è pari a quella della base di dati su cui vengono calcolati, e il principio secondo il quale è sempre opportuno interpretarne il significato, come ampiamente argomentato in [Adler et al, 2008].