

***Il processo di valutazione VQR
negli Atenei italiani.
Punti di forza e di debolezza dell'attuale modello e
proposte di miglioramento***

**a cura dei
Gruppi di lavoro
in seno alla Commissione Ricerca CRUI**

Ottobre 2022

Il documento che segue è frutto del lavoro di riflessione e approfondimento condotto all'interno della Commissione Ricerca e Terza Missione della Conferenza dei Rettori delle Università Italiane, coordinata dal prof. Claudio Pettinari, Rettore dell'Università degli studi di Camerino.

L'elaborazione è a cura di gruppi di lavoro composti da Delegati e Prorettori degli atenei per le tematiche oggetto di approfondimento (la composizione è riportata alla fine del documento) ed è stato condiviso con i Delegati e Rettori di tutti gli atenei.

Indice

Introduzione	3
Parte I – Peer Review e GEV	4
Parte II – ISPD e Dipartimenti di Eccellenza	8
Parte III – Terza missione	12
Parte IV – Reforming Research Assessment	19
Conclusioni	23
Composizione dei Gruppi di lavoro	25

Introduzione

L'efficacia di un impianto di valutazione passa inevitabilmente attraverso la tempestività e la stabilità dei suoi meccanismi, la trasparenza metodologica e la chiarezza dei risultati.

Con riferimento alla **tempestività e alla stabilità dell'impianto di valutazione**, gli Atenei ritengono essenziale assicurare forme di garanzia in merito alla continuità dei meccanismi di valutazione, affinché gli sforzi progressivi possano non disperdersi e sia possibile attuare un indirizzo e una programmazione delle attività nel medio/lungo termine, come peraltro richiesto dai meccanismi di accreditamento delle sedi (AVA3, nello specifico). Diventa, quindi, essenziale che il Ministero dell'Università e della Ricerca definisca immediatamente le Linee Guida per il prossimo esercizio della VQR 2020-2024, auspicabilmente in continuità rispetto alla VQR3 2015-2019, in modo che l'ANVUR possa adottare tempestivamente i provvedimenti di sua competenza. Se, al contrario, vi fosse l'intenzione di modificare (anche solo in parte) le regole del gioco, sarebbe ancora più importante una condivisione immediata con gli Atenei nelle forme opportune. Diversamente, si rischierebbe una vanificazione del percorso che molti Atenei hanno già intrapreso per prepararsi al prossimo esercizio valutativo.

In merito alla **trasparenza metodologica e alla chiarezza dei risultati**, i criteri di valutazione applicati dai GEV sono stati esplicitati nel bando in modo spesso generale. Il rapporto di fiducia fra gli Atenei e l'organismo valutatore rappresenta il presupposto fondamentale del successo di qualsiasi esercizio di valutazione, anche in un'ottica di indirizzo, programmazione e miglioramento. Questo rapporto di fiducia si fonda sulla chiarezza dei metodi di valutazione, da comunicarsi ex-ante, e sulla trasparenza nell'esposizione dei risultati, che devono essere strettamente riconducibili all'applicazione dei criteri di giudizio.

Nei prossimi mesi, il sistema italiano di valutazione della ricerca dovrà, inoltre, affrontare la sfida rappresentata dalle **nuove linee guida promulgate dalla Commissione Europea** (e già accolte favorevolmente da LERU e da EUA) per la valutazione della ricerca condotta dalle strutture e dai singoli docenti/ricercatori. La Commissione chiede agli Atenei europei di allinearsi alle nuove indicazioni e di renderle operative nell'arco di 1-2 anni, con implicazioni importanti sia per i criteri attualmente utilizzati nella VQR che per i concorsi universitari alla base dei reclutamenti e delle progressioni di carriera del personale docente.

Parte I – Peer Review e GEV

Premessa

L'obiettivo della Valutazione della Qualità della Ricerca (VQR) è valutare la qualità della ricerca realizzata dalle Università e dagli Enti di Ricerca. Gli indicatori che da essa scaturiscono contribuiscono a determinare l'entità della quota premiale del Fondo di Finanziamento Ordinario del sistema universitario e l'indicatore ISPD di ciascun dipartimento valutato. La VQR si prefigge di fornire indicazioni chiare alle stesse Istituzioni affinché si possa:

- migliorare la qualità complessiva della ricerca italiana;
- stimolare l'adozione di politiche virtuose di investimento e programmazione;
- promuovere un utilizzo efficiente delle risorse finanziarie destinate al reclutamento al fine di premiare le aree già forti e al contempo rafforzare le aree più deboli in ciascuna Istituzione.

Ne deriva l'assoluta necessità di un sistema che sia al contempo trasparente e replicabile e soprattutto basato su indicatori definiti a priori, precocemente, e pertanto monitorabili nel tempo dalle istituzioni.

Contesto

La VQR3 ha previsto una serie di novità importanti rispetto agli esercizi precedenti:

- Definizione di un numero massimo di prodotti attesi ($3 \times N$) calcolato sull'integrale del numero di accreditati nei dipartimenti (N), ma con la possibilità che ciascun singolo ricercatore possa conferire un numero di prodotti variabile fra 0 e 4;
- Un punteggio in trentesimi (somma di 3 punteggi in decimi per ciascuno dei 3 criteri richiamati dal bando) determinante per la collocazione dei prodotti nelle categorie in accordo ad una tabella di transcodifica definita da ANVUR;
- Esenzioni opzionali che si applicano al numero di prodotti da conferire per la struttura (e non necessariamente al singolo ricercatore che ne avrebbe diritto) e che quindi consentono lo sviluppo di una strategia di Istituzione e dipartimentale;
- Peer review estesa a tutti i prodotti conferiti e "informata" da indicatori per le aree bibliometriche;
- Indicazione per ciascun GEV di popolare le 5 categorie di merito con percentuali minime e massime definite a priori;
- Sorteggio dei GEV in un insieme di candidati in possesso di specifici requisiti;
- Attenzione ancora maggiore ai prodotti che presentavano un numero elevato di autocitazioni.

Punti di Forza

Nella sua attuale configurazione la VQR ha posto, più che in passato, i Dipartimenti al centro del processo di valutazione. In particolare attraverso:

- Numero variabile di prodotti associati al singolo ricercatore pur in presenza di un numero complessivo di prodotti richiesto al dipartimento pari a $3 \times N$, con N numero di accreditati per l'esercizio della VQR;
- Possibilità per più dipartimenti della stessa Istituzione di presentare lo stesso prodotto in base al numero di coautori;
- Possibilità che alcune tipologie di prodotti possano avere valenza "doppia" rispetto alle altre, possibilità che era presente già nella VQR2 e che è stata opportunamente confermata nella VQR3;
- Definizione di ulteriori sotto-aree per tener conto di differenze e specificità all'interno di un'unica area CUN.

Tutto questo ha determinato una interessante presa di coscienza di tutti i ricercatori, dei Dipartimenti e dell'intero Ateneo, aumentando il livello di responsabilizzazione rispetto a queste tematiche, aspetto già di per sé molto positivo.

Punti di debolezza

All'esito dei risultati della VQR3 sono emersi i seguenti punti critici:

- l'indicazione sulla distribuzione dei prodotti nelle categorie di giudizio introdotta nel Bando (art. 7 c. 11) ha indotto di fatto alcuni GEV ad effettuare una valutazione comparativa dei prodotti presentati. Si fa notare inoltre che l'indicazione della distribuzione dei prodotti è stata presumibilmente costruita sulla base della distribuzione dell'intera popolazione mondiale di prodotti; i prodotti conferiti dalle istituzioni, che costituiscono un sottoinsieme dei prodotti totali, sono stati inevitabilmente scelti tra i prodotti migliori, spostando significativamente la distribuzione corrispondente verso le categorie più alte (A e B); le due popolazioni statistiche non sono, pertanto, confrontabili e il tentativo di ricondurre la seconda (prodotti conferiti) alla prima (popolazione globale) costituisce una forzatura;
- per le aree bibliometriche pure e per quelle non bibliometriche ma informate (si vedano ad esempio le 13a e 13b) la valutazione dei revisori ha avuto in molti casi un peso eccessivo rispetto a quello della collocazione bibliometrica del prodotto;
- per le aree non bibliometriche, la pluralità degli approcci insita nella peer review pura potrebbe aver amplificato il peso della discrezionalità dei membri GEV sia nella valutazione sia nella scelta dei revisori esterni;
- i criteri di valutazione applicati dai GEV, esplicitati nel bando in modo alquanto generale, non risultano trasparenti in relazione all'utilizzo dell'informazione bibliometrica;
- i GEV hanno agito in assenza di conflitto di interesse, per come definito attualmente. Tuttavia, la VQR3 ha altresì messo in luce ulteriori possibili situazioni di conflitto che dovrebbero essere considerate;
- le novità introdotte, pur apprezzabili, sono state presentate con una tempistica che non ha consentito alle istituzioni di tenerne conto nelle dinamiche accademiche sulle quali si prefiggono di incidere;
- alcuni membri GEV si sono trovati a valutare un numero di prodotti molto elevato (in area 06, più di 600 prodotti per singolo valutatore, da analizzare in soli 45 giorni), con il rischio di non poter garantire per ciascuno di essi l'approfondimento necessario per una corretta valutazione;
- il singolo ricercatore ha avuto a disposizione il valore numerico dei giudizi (in scala 1-10) e la categoria finale di appartenenza risultante per i soli prodotti che ha presentato, senza alcuna motivazione; ciò non fornisce alcuna opportunità di miglioramento. Analogamente, ai fini di una autovalutazione volta al miglioramento, si ritiene controproducente che gli Atenei non possano conoscere le valutazioni, e le relative motivazioni, attribuite ai prodotti che gli Atenei stessi hanno conferito, sentiti i Dipartimenti (Bando VQR, art. 6, comma 1);
- si è chiesto al revisore/valutatore di esprimere una votazione con scala 1-10 e si è poi tradotta questa scala lineare (3-30 per 3 valutatori) in una scala non lineare (A, B, C, D ed E corrispondenti a 1, 0,8, 0,4, 0,1, 0): questo modo di procedere produce risultati che, dal punto di vista matematico, sono incongruenti, perché una differenza minima nella scala 3-30 può tradursi nell'attribuzione di una classe differente, con variazione del punteggio sintetico complessivo che oscilla tra il 20% (nel passaggio da A a B, cioè da 1 a 0,8) e addirittura il 50% (nel passaggio da B a C, cioè da 0,8 a 0,4);
- il bando non chiariva se l'indicatore dimensionale di una istituzione sarebbe stato definito in base al numero di prodotti attesi (al lordo delle esenzioni) o al numero di prodotti conferiti (al netto delle esenzioni);

- nel DM che ha istituito la VQR3 era previsto di rendere disponibili i risultati dell'esercizio di valutazione unicamente in forma aggregata per istituzione per aree scientifiche (presumibilmente coincidenti con quelle CUN) e non per i singoli prodotti di cui i ricercatori sono co-autori. È stato il bando ANVUR a stabilire che questo dovesse essere fatto. Peraltro, questi risultati sono stati resi disponibili al singolo in una forma estremamente distorta del concetto di VQR come di un esercizio che valuta le istituzioni e non i singoli ricercatori. Infatti, sono stati resi visibili solo i prodotti che l'istituzione ha assegnato (seguendo una logica di ottimizzazione dei risultati per la struttura) al ricercatore e non come era accaduto in passato di tutti i prodotti presentati dalla istituzione di cui il ricercatore stesso è co-autore. Questo ha generato in molti colleghi la percezione che questa sia stata una valutazione personale.

Proposte

- La stabilità dell'impianto di valutazione è essenziale per non dare indicazioni contrastanti al sistema e non disperdere gli sforzi fatti dalle istituzioni per migliorare i loro profili di qualità. ANVUR, per quanto di sua competenza, dovrebbe fornire il prima possibile adeguate garanzie sulla stabilità dei meccanismi VQR e dei criteri dei GEV nel prossimo esercizio e segnalare all'interlocutore politico l'opportunità di non stravolgere il disegno. Una comunicazione tempestiva è ancora più importante qualora il MUR e/o ANVUR decidessero di cambiare alcune regole rispetto alla precedente VQR; in tal caso, le modifiche dovrebbero essere ragionevoli e condivise con gli Atenei. Idealmente si potrebbe tendere a quanto avviene nel Regno Unito per il Research Excellence Framework (REF) in cui criteri di valutazione sono resi noti tre anni prima del conferimento dei prodotti. Questo peraltro consentirebbe di creare le interfacce per il processo di valutazione con notevole anticipo e quindi rendere complessivamente più rapido il processo VQR;
- al fine di affiancare alla valutazione dei prodotti migliori delle strutture un'indicazione sulla loro qualità media, si propone che tutte le Istituzioni presentino per ognuno dei ricercatori accreditati un numero di prodotti variabile da 1 a 4, anziché da 0 a 4 come nella VQR3 (così che ciascun ricercatore presenti almeno un prodotto), fermo restando il numero complessivo di prodotti da conferire pari al triplo del numero dei ricercatori accreditati;
- considerata la composizione del GEV e la mole delle attività di peer review sarebbe auspicabile avere almeno un membro GEV per ciascun SSD; si dovrebbe inoltre prevedere il ricorso ad un'adeguata platea di revisori esterni, opportunamente formati allo scopo e (qualora praticabile per l'area di riferimento) possibilmente stranieri. In ogni caso la scelta dei revisori esterni dovrebbe essere affidata a criteri che assicurino sufficienti garanzie di trasparenza e imparzialità attingendo preferenzialmente alla lista di revisori disponibili in possesso di elevata qualificazione (criteri partecipazione al bando GEV);
- non dovrebbero essere date indicazioni sulla distribuzione desiderata dei prodotti nelle categorie alla fine del processo di valutazione, peraltro definite sulla base di popolazioni statistiche non comparabili (distribuzione mondiale vs prodotti conferiti dall'istituzione), dato che impongono una valutazione comparativa dei prodotti (e la VQR non è questo);
- nei settori bibliometrici il ricorso al parere di revisori potrebbe essere adottato per tutti e soli quei prodotti che ricadono nelle seguenti fattispecie:
 - prodotti che pur vantando una eccellente collocazione editoriale non hanno un numero adeguato di citazioni (al netto delle autocitazioni e rapportato all'anno di pubblicazione) tale da far ritenere il prodotto assolutamente eccellente e rilevante;
 - prodotti che vantano un notevole numero di citazioni (al netto delle autocitazioni e rapportato all'anno di pubblicazione) pur in presenza di una sede editoriale non eccellente;

- prodotti che hanno un numero di autocitazioni (da parte di tutti i coautori) la cui percentuale sul totale delle citazioni sia maggiore o uguale al 50% del totale delle citazioni
- prodotti evidentemente interdisciplinari il cui impatto nelle singole discipline appaia di differente qualità;
- prodotti pubblicati su riviste ritenute predatorie o considerate poco trasparenti nei processi di peer-review;
- nei settori non bibliometrici si auspica una maggiore considerazione della collocazione editoriale rispetto al ricorso della peer review esterna. Pertanto, analogamente a quanto proposto per il contesto bibliometrico, anche per quello non bibliometrico si potrebbe prevedere la non adozione del parere dei revisori esterni per pubblicazioni di riconosciuto prestigio a livello sia internazionale sia nazionale quali:
 - riviste di Classe A rilevanti ai fini dell'Abilitazione Scientifica Nazionale per il Settore Concorsuale di riferimento dell'autore;
 - riviste di fascia A del ranking delle Associazioni scientifiche di riferimento internazionale e nazionale;
 - monografie accreditate dalle Associazioni scientifiche di riferimento, laddove sia stata effettuata una revisione fra pari con esperti anonimi.

Per tutti gli altri prodotti, si suggerisce l'impiego del referaggio anonimo, che non può essere sostituito dall'applicazione di metriche e dati attraverso una valutazione automatica basata su indici e fattori;

- in ogni caso la variazione di più di una categoria rispetto all'indicazione bibliometrica (ad esempio, da A a C, oppure da C a A) dovrebbe essere ampiamente motivata secondo criteri codificati in modo preciso all'interno dei GEV;
- associare la dimensione di ciascuna struttura al numero di prodotti massimo da conferire al LORDO delle esenzioni collegando quindi la VQR alla sola qualità dei prodotti presentati;
- condividere a priori e limitare il numero delle Subject Category (SC) al fine di semplificare la collocazione dei prodotti e lasciare esclusivamente alla Istituzioni la facoltà di scegliere per ciascun prodotto la SC in cui si vuol essere valutati;
- basare la classificazione bibliometrica sulla scelta della banca dati/indice citazionale indicata dalla istituzione;
- i membri dei GEV devono godere di una riconosciuta autorevolezza presso la comunità scientifica della quale sono chiamati a gestire il processo di valutazione; oltre ad un adeguato curriculum scientifico-disciplinare, essi dovrebbero possedere una comprovata esperienza in attività di valutazione della ricerca.

Parte II - ISPD e Dipartimenti di Eccellenza

Contesto

L'obiettivo della valutazione della qualità della ricerca (VQR) e in particolare del finanziamento dei Dipartimenti di Eccellenza dovrebbe essere migliorare la qualità complessiva della ricerca italiana, incentivando le strutture a mettere in campo politiche virtuose di programmazione, autovalutazione, utilizzo delle risorse finanziarie e reclutamento volte a questo scopo. In questo breve documento si intendono articolare punti di forza, punti di debolezza e proposte di cambiamento, rispetto a quattro dimensioni strutturali che caratterizzano il progetto di finanziamento dei 180 Dipartimenti di Eccellenza:

- indice ISPD;
- formazione delle graduatorie dei dipartimenti;
- distribuzione dei finanziamenti;
- procedura di presentazione dei progetti.

Aspetti specifici di osservazione

1. Calcolo ISPD, trasparenza e riproducibilità

Punti di forza

I dipartimenti universitari sono fortemente diseguali in termini di numerosità, di settori di ricerca rappresentati, di tipologia di ricerca (con valutazione bibliometrica e no). Pertanto, la definizione di un indicatore standardizzato rappresenta un elemento di potenziale chiarezza e utilità nel collocare le performances di ricerca della singola struttura entro l'eterogeneo panorama nazionale.

Punti di debolezza

Un processo di fatto non completamente trasparente, in quanto basato su indicatori il cui calcolo non è replicabile per l'indisponibilità dei dati in ingresso, non rispetta l'irrinunciabile criterio di riproducibilità né gli standard della comunità scientifica. È stato altresì da più esperti rilevato che l'indicatore ISPD definito e utilizzato in questo esercizio VQR3 ha scarsa robustezza statistica.

Il complesso di queste criticità rischia di minare la fiducia nelle procedure di valutazione stesse, demotivando e deresponsabilizzando i ricercatori, le strutture e le istituzioni rispetto all'impegno nella ricerca e all'obiettivo del miglioramento.

Ai fini di valutazione della ricerca scientifica e nell'ottica di un approccio basato sul metodo scientifico, dovrebbero essere noti i dati di input, oltre agli algoritmi e ai risultati.

Proposte

Si ritiene opportuno che ANVUR renda noti (anche anonimizzando i SSD o i SC) per ogni singola struttura e istituzione i dettagli dei loro specifici valori necessari per comprendere il risultato ottenuto. Questo risulterebbe utile anche ai fini di valutazione ex-post delle scelte compiute nella selezione dei prodotti conferiti (responsabilità delle strutture) e nella composizione dei dipartimenti; ciò è tanto più necessario dati gli effetti di sensibilità dell'indicatore e la sua scarsa robustezza statistica a fronte dei problemi della valutazione dei prodotti come evidenziati nella Parte I di questo documento (si vedano in particolare i Punti di debolezza 1 e 9).

In particolare, si ritiene opportuno che ANVUR:

- renda note: (i) le distribuzioni reali dei voti dei diversi SSD e/o delle aggregazioni di SSD utilizzati; (ii) per ogni dipartimento, i voti nei settori con numero di prodotti presentati sufficientemente numeroso da rendere impossibile l'associazione della valutazione al singolo docente/ricercatore;
- pubblici il codice mediante il quale viene calcolato ISPD, inclusa la parte che determina le aree preminenti;
- chiarisca dettagliatamente nelle note metodologiche il metodo di standardizzazione per i casi di SSD a numerosità sotto-soglia e il significato delle aree che risultano sopra e sotto media nella classifica;
- calcoli il valore di ISPD non in modo discretizzato al semi-intero, ma includendo almeno 3 cifre decimali.

2. Formazione delle graduatorie dei dipartimenti

Punti di forza

L'affiancamento della valutazione del progetto alla graduatoria fornita dall'ISPD è apprezzabile, in quanto evita che il finanziamento venga assegnato in maniera del tutto automatica sulla base di un indicatore che presenta diverse criticità.

Punti di debolezza

I risultati, e il modo in cui sono stati resi noti, evidenziano ancora di più alcune criticità insite nella definizione di ISPD. Di seguito elenchiamo le criticità più significative.

Eccessiva polarizzazione verso valori molto elevati (o molto bassi). Il numero di dipartimenti che ha riportato una valutazione pari a 100 (e presumibilmente pari a 0) pare eccessivo. Questa polarizzazione ha diversi effetti negativi: a) amplifica la differenza di punteggio tra i dipartimenti che si trovano in cima alla graduatoria e quelli che, pur ammessi a presentare domanda, hanno ottenuto punteggi intermedi, rendendo di fatto quasi impossibile recuperare il gap attraverso il progetto; b) appiattisce le differenze tra i dipartimenti migliori, ponendoli tutti sullo stesso piano; c) genera disaffezione o rigetto vero e proprio nei molti dipartimenti che si sono visti assegnare un punteggio pari a 0.

Instabilità di ISPD. È stato osservato, e verificato anche tramite simulazioni, che lo spostamento di una percentuale anche minima di lavori da una classe all'altra può comportare variazioni molto significative del valore di ISPD, minando ancora una volta la fiducia dei dipartimenti nel meccanismo di valutazione e nella sua efficacia, demotivando i ricercatori.

Proposte

Utilizzo diretto dei percentili del voto standardizzato di dipartimento al posto di ISPD. In questo modo il numero di dipartimenti con voto pari a 100 sarebbe esattamente pari all'1%, così come quelli con voto pari a 99, 98 ecc., con una classifica più graduale e una minore instabilità.

Quantizzazione più fine di ISPD per evitare di avere una percentuale troppo elevata di dipartimenti con punteggio pari a 100.

In assenza di cambiamenti nel calcolo dell'indicatore, il peso relativo di ISPD e progetto per la selezione dei Dipartimenti di Eccellenza potrebbe essere riconsiderato.

3. Distribuzione dei fondi dell'Eccellenza

Punti di forza

Il vincolo sul numero dei dipartimenti ammissibili per ateneo pone un limite rispetto al rischio di finanziamento concentrato su poche istituzioni.

Punti di debolezza

I dipartimenti già finanziati nella tornata 2018-23 hanno avuto un obiettivo vantaggio per una parte del periodo considerato da VQR3, rispetto ai dipartimenti non finanziati, vantaggio che non è considerato nella selezione.

L'assenza di criteri di selezione legati ai finanziamenti precedenti aumenta il rischio di contribuire ingiustificatamente a perpetuare disuguaglianze tra atenei e tra dipartimenti di uno stesso ateneo. Quale esempio, si citano qui le disuguaglianze territoriali. Mentre il numero di dipartimenti statali nelle regioni del Sud, nelle regioni di transizione e Isole (Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna) è pari al 31,86% dei dipartimenti statali (253/794), il numero di dipartimenti delle regioni del Sud, di transizione e delle isole nella graduatoria per l'Eccellenza 2023-27 è pari al 14,86% (52/350).

Talvolta l'investimento finanziario sul singolo dipartimento può risultare sovradimensionato rispetto agli obiettivi di rafforzamento e ampliamento della ricerca e della sua reputazione internazionale.

Proposte

- Tenere in considerazione i finanziamenti già ottenuti dai dipartimenti per la stessa misura, ad esempio valutando un eventuale 'handicap' per un successivo finanziamento di dipartimenti già finanziati, o viceversa un vantaggio per i Dipartimenti non finanziati, oppure prevedendo graduatorie distinte
- Riconsiderare il numero totale dei dipartimenti finanziati: si propone di aumentarne il numero, anche intervenendo sull'entità del finanziamento per singolo dipartimento, pur parametrandola ai diversi costi della ricerca.

4. Procedura di presentazione dei progetti

Punti di forza

La struttura del template da compilare per la presentazione del progetto, in cui in ogni sezione vengono specificati gli obiettivi e le linee guida per valorizzare la sezione, rende il processo di stesura più semplice; inoltre la specifica dettagliata nel Decreto Ministeriale dei criteri di valutazione permette un'autovalutazione in fase di stesura molto utile per migliorare il progetto. La strutturazione del budget già su base quinquennale permette una pianificazione ragionata e una successiva implementazione già ben strutturata nel tempo.

Punti di debolezza

La concomitanza del periodo di apertura della procedura per la presentazione dei progetti con il periodo della fase di chiusura della tornata precedente impegna sui due fronti i dipartimenti.

La comunicazione preliminare del risultato della selezione interna, sia per il numero di dipartimenti sia per il singolo dipartimento "campione", rischia di essere un'operazione cieca. La scelta di vincolare la visibilità delle schede progetto alla comunicazione dei dipartimenti da candidare e del campione di Istituzione risulta immotivata e genera notevoli complessità organizzative. Il vincolo

della comunicazione preliminare ha impedito alle istituzioni di valutare internamente, ai fini della scelta del dipartimento campione, anche i progetti da sottoporre.

Risulta inoltre che, su richiesta, alcuni atenei abbiano potuto accedere alla compilazione dei progetti senza preliminare comunicazione del risultato della selezione interna per il dipartimento campione, mentre altri sono rimasti all'oscuro di questa possibilità.

Proposte

Evitare che l'apertura della procedura per la presentazione dei progetti si sovrapponga alla fase di chiusura della tornata precedente. La presentazione dei nuovi progetti dovrebbe seguire la chiusura. Inoltre, l'apertura della procedura di compilazione non dovrebbe essere legata all'indicazione da parte degli atenei del dipartimento sottoposto alla I fase: l'identificazione del dipartimento da mandare alla prima fase di valutazione sia posticipata alla fase finale di presentazione dei progetti. Le istituzioni e le strutture devono essere messe tutte nelle stesse condizioni, anche dal punto di vista delle regole e possibilità di intervento nelle piattaforme dedicate.

Parte III – Terza missione

Contesto

La valutazione della Terza missione nel processo VQR 2015-2019, basata sull'impatto dei casi studio, ha favorito un ripensamento critico da parte degli Atenei in merito al ruolo di questa missione, spesso giudicata ancillare rispetto alla Didattica ed alla Ricerca o comunque marginale.

Ciò ha determinato un'interessante presa di coscienza da parte di tutti i soggetti valutati, incrementando il livello di responsabilizzazione degli Atenei e dei Dipartimenti in merito al ruolo svolto nel proprio contesto di riferimento. Superando la visione tradizionale della Terza missione come processo unidirezionale di trasferimento tecnologico dagli Atenei verso gli altri attori territoriali, si è affermata l'idea che le attività di collaborazione esterna delle Università consistano in processi di apprendimento reciproco con le istituzioni pubbliche, le organizzazioni sociali e le imprese. Tali processi sono volti a favorire la co-produzione di innovazione e conoscenza per la riduzione delle disuguaglianze e lo sviluppo sostenibile dei territori.

Grazie a questa nuova visione, che ha ispirato anche il processo VQR, gli Atenei sono sempre più consapevoli di avere la possibilità e la responsabilità di contribuire al soddisfacimento delle esigenze delle proprie comunità di riferimento. La condivisione di risorse e conoscenze che ne deriva può svolgere un ruolo fondamentale in una prospettiva di giustizia sociale e ambientale e di maggiore apertura internazionale dei sistemi locali.

Questo rinnovamento concettuale della Terza missione ha un impatto anche sui suoi rapporti con la Didattica e la Ricerca, pur nella salvaguardia del valore fondamentale dell'autonomia di tali attività universitarie. La Didattica, infatti, viene stimolata dal continuo dialogo degli Atenei con gli altri attori sociali: da un lato, l'offerta formativa viene spesso efficacemente orientata dalle mutevoli esigenze esterne; dall'altro, i docenti - opportunamente indirizzati - possono trarre dal dialogo sociale spunti di riflessione e indicazioni da trasferire agli studenti. Dal suo canto, la Ricerca è fortemente stimolata dalle attività di collaborazione sociale degli Atenei, come sta emergendo anche dai vari progetti connessi al Piano Nazionale di Ripresa e Resilienza.

La Terza missione si configura oggi come un insieme di valori e di iniziative che stanno cambiando tutte le attività degli Atenei, amplificandone l'impatto sulla società. In questo senso, appare opportuno rimetterne in discussione la denominazione, superando la sua forma generica di numero ordinale che suggerisce l'idea di una missione residuale rispetto alla Didattica e alla Ricerca, rendendone esplicito il contenuto e sottolineando il carattere reciproco dei processi di interscambio di conoscenze in cui si attua la Terza missione.

Aspetti specifici di osservazione

1. Bando VQR (definizione di impatto, nomenclature utilizzate, numerosità dei casi studio attesi, ecc.)

Punti di forza

La valutazione basata su casi studio fa emergere il radicamento delle Università sul territorio, responsabilizza gli Atenei e permette di valorizzare le peculiarità di determinate azioni.

La definizione di impatto viene opportunamente declinata per i diversi campi di azione.

È stato concesso un margine di libertà nella scelta degli indicatori da proporre per illustrare l'impatto dell'azione di Terza missione.

La definizione di impatto è ampia ed articolata. Essa consente di abbracciare attività non solo di trasferimento tecnologico in senso stretto, ma anche iniziative di impegno sociale, economico ed istituzionale in senso lato.

Si abilita la Terza missione nella sua accezione ampia, in grado anche di coinvolgere attivamente le discipline delle scienze umane e sociali.

Le nomenclature sono ampie ed articolate.

Punti di debolezza

L'indicatore IRAS4 è stato calcolato rispetto al numero di prodotti della ricerca conferiti che abbiano ottenuto un giudizio almeno pari a "rilevanza sufficiente" (DM n. 289 del 25/03/2021), mentre il numero di casi studio attesi è stato limitato alla metà del numero di Dipartimenti. Si tratta di una incoerenza che può aver determinato distorsioni, in quanto il numero di casi studio presentati da ciascun Ateneo dipende esclusivamente da scelte organizzative dell'Ateneo stesso (numero dei Dipartimenti).

Gli indicatori di impatto (sociale, culturale, economico) suggeriti per ogni singolo campo di azione talvolta appaiono più come indicatori di realizzazione o di risultato che come veri e propri indicatori di impatto. Tra l'altro, sarebbe opportuno considerare che, molto spesso, si è trattato di azioni con impatti di più lunga durata rispetto al periodo di esercizio e che la persistenza degli impatti costituisce una tematica assai rilevante.

Proposte

Si suggerisce la predisposizione di linee guida contenenti una serie di definizioni relative a indicatori di realizzazione (output), indicatori di risultato (outcome), indicatori di impatto (outreach, impact), ecc., nonché volte alla scelta coerente degli indicatori medesimi. In proposito, si ritiene fondatamente che il quadro d'insieme derivante dalla passata VQR 2015-19 dovrebbe consentire ad ANVUR di predisporre le auspiccate linee guida in modo compiuto ed opportunamente rappresentativo.

Si potrebbero riportare nelle linee guida alcune "buone pratiche" di indicatori (e relativi aspetti informativi) come possibili esempi per ciascuna tipologia o classe di casi studio conferiti, lasciando tuttavia libertà alle istituzioni di individuare gli indicatori più adatti per evidenziare l'impatto dei singoli casi studio.

Si potrebbe ampliare la numerosità dei casi studio richiesti, magari limitandola in ciascun ambito della tassonomia proposta, affinché gli Atenei possano attuare una scelta articolata di merito. Tale proposta dovrebbe essere connessa alla classe dimensionale degli Atenei.

Il numero di casi studio dovrebbe essere parametrato non al solo numero di Dipartimenti, ma anche alla numerosità di afferenti, intesi come docenti, personale tecnico amministrativo e bibliotecario. Si tratta di una questione rilevante anche dal punto di vista concettuale, perché permetterebbe agli Atenei di dare la giusta enfasi alla Terza missione. In altri termini, la struttura organizzativa dell'Ateneo non può essere l'unica variabile che determina il numero di casi attesi.

Si suggerisce di valutare una modalità idonea a valorizzare l'importanza dell'interdisciplinarietà, così da incoraggiare azioni trasversali di Ateneo. In tale prospettiva, si potrebbe ad esempio considerare anche l'opportunità di una logica di premialità valutativa. Questa potrebbe valere per possibili casi studio che abbiano comportato il coinvolgimento sinergico ed interdisciplinare di aree/docenti afferenti a più di un Dipartimento del medesimo Ateneo. Al riguardo, si potrebbe altresì chiedere di prevedere anche indicatori in grado di tener conto di tale aspetto dell'interdisciplinarietà.

Come ultimo, si potrebbe valutare la possibilità di riproporre casi studio in differenti VQR che hanno un impatto consecutivo nel tempo.

2. Tassonomia dei campi d'azione

Punti di forza

I campi d'azione disponibili sono stati ampi e definiti in modo abbastanza chiaro pur con alcune inevitabili sovrapposizioni. In aggiunta, agli Atenei è stata lasciata libertà nella scelta e una doppia opzione. Ciò ha consentito agli Atenei di mostrare i propri orientamenti di Terza missione in campi differenti, scegliendo tra quelli presenti in una tassonomia molto ampia.

La strutturazione per campi di azione permette la valorizzazione della multidisciplinarietà e preserva la specificità degli ambiti disciplinari, in quanto alcuni campi di azione sono logicamente correlati a determinati contesti.

Punti di debolezza

La scelta di richiedere la presentazione dei casi studio strutturata per campi d'azione potrebbe aver portato a difficoltà e distorsioni nella valutazione. Infatti, alcuni campi di azione erano talmente ampi che ne comprendevano idealmente anche altri (es. Agenda 2030, trasferimento tecnologico, incubatori, ecc.), con conseguente potenziale difficoltà da parte degli Atenei di una collocazione univoca delle proprie attività.

Il collegamento dei casi studio a un campo di azione principale ed eventualmente ad altri secondari non appare chiaro e ben definito.

Proposte

Sarebbe opportuna una più chiara definizione dei campi di azione, eventualmente suddividendo le categorie in sotto categorie, soprattutto con riferimento ai campi più trasversali e di più nuova istituzione. A fronte di ciò, si potrebbe comunque prevedere una categoria residuale per "Altre attività di Terza missione, non riconducibili alle precedenti".

Sarebbe utile offrire orientamenti più precisi nella definizione degli indicatori di impatto per ogni "item" della tassonomia.

Per favorire una ampia e diversificata presentazione dei casi di studio, si potrebbe incoraggiare la presentazione di casi di diversa tipologia.

Si dovrebbe garantire la massima omogeneità possibile dei criteri adottati per le valutazioni dei casi collocati nei diversi campi.

3. Composizione del GEV interdisciplinare (numerosità, selezione, criteri, ecc.)

Punti di forza

I membri GEV erano sufficientemente numerosi e provenienti da diversi Atenei, ambiti geografici e aree culturali, così da garantire la multidisciplinarietà e la capacità di cogliere correttamente aspetti diversi dell'impatto.

La selezione della composizione attraverso bando pubblico e "chiamata per chiara fama" è convincente.

Punti di debolezza

Non si può parlare di GEV 'interdisciplinare' quanto piuttosto di GEV 'multidisciplinare' in quanto la creazione di un linguaggio comune richiede un adeguato percorso di apprendimento.

Tra i criteri di selezione andrebbe maggiormente considerato anche l'aver ricoperto ruoli di governance in tema di Terza missione e l'esperienza scientifica sui temi della interdisciplinarietà e della Terza missione.

La stragrande maggioranza dei docenti selezionati era italiana con sede lavorativa in Italia. Questo potrebbe aver aumentato il rischio di conflitto di interesse del GEV rispetto ai soggetti valutati.

Proposte

I membri GEV devono godere di un profilo tale da ridurre al minimo possibili condizionamenti. Inoltre, il grado di "seniority" deve garantire una piena maturità scientifica, un'ampia e approfondita conoscenza del contesto e un riconosciuto prestigio a livello nazionale e internazionale.

Viene puntualizzata l'importanza che, per la valutazione di particolari aspetti dei casi studio, i GEV possano usufruire del referaggio di esperti riconosciuti sulla tematica oggetto del caso, anche esterni.

La redazione dei casi studio in lingua inglese, oltre che in italiano, potrebbe permettere l'inclusione di membri GEV stranieri.

Si suggerisce di costruire con tempestività la composizione del GEV e di sottoporre i membri a un percorso di apprendimento/addestramento non estemporaneo.

I componenti GEV potrebbero essere selezionati non tanto per ambito disciplinare, quanto per capacità di valutazione degli impatti che le azioni hanno sul territorio, oltre che per sensibilità alla Terza missione.

Sarebbe utile una ancora più ampia partecipazione dei diversi Atenei di diversa vocazione e dimensione tra i componenti del GEV.

4. Conferimento dei casi studio (campi richiesti, allegati, piattaforma di caricamento, lavoro interno agli Atenei, ecc.)

Punti di forza

È stata concessa libertà nell'inserire allegati e documentazione a supporto.

La limitatezza del numero di battute consentito per ciascun campo ha favorito la sintesi, seppur in una logica di esaustività, senza appesantire gli Atenei con lavori quali-quantitativi non necessari.

Punti di debolezza

La disomogeneità della documentazione a supporto dei casi non ha aiutato il lavoro di confronto e di valutazione.

La rigidità dei campi di azione previsti ha costretto gli Atenei a una narrativa non semplice.

La piattaforma non ha consentito l'utilizzo dell'editing che, invece, è uno strumento a servizio della chiarezza espositiva.

Proposte

Si consiglia l'utilizzo di una piattaforma che consenta l'uso di editing e immagini.

Si potrebbe articolare meglio la struttura della presentazione dei casi di studio con riferimento a: contesto di riferimento, cronologia, output e outcome, ecc.

Sarebbe necessario un repository perpetuo e condiviso per i casi studio, alla stregua di quanto concepito e impiegato per i prodotti della ricerca.

5. Processo di valutazione (criteri di valutazione, peso dei singoli criteri, punteggi, classi di merito, indicatori, ecc.)

Punti di forza

Il processo di valutazione condotto dai GEV ha assicurato l'anonimato e la mancanza di conflitto di interesse.

Punti di debolezza

Nella stesura dei casi studio, non sempre la rilevazione degli indicatori è stata semplice, soprattutto quando i dati erano difficilmente recuperabili.

Le classi di impatto e la modalità di attribuzione dei punteggi sembrano aver penalizzato le proposte che si trovavano ai limiti delle diverse soglie.

Il "range" di punteggio attribuito alla classe A sembra sia stato eccessivamente ridotto.

Proposte

Sarebbe opportuno far emergere le sinergie tra Terza missione, Didattica e Ricerca, ossia gli stimoli che dalla Terza missione sono arrivati per l'innovazione su Didattica e Ricerca (da certificare con indicatori comuni standard per tutti i casi).

Sarebbe alquanto utile consentire di valutare accuratamente l'impegno dell'Ateneo, anche chiarendo meglio il significato attribuito al termine "contributo della struttura proponente".

Si suggerisce un affinamento dei metodi di valutazione. In particolare, sarebbe auspicabile introdurre anche criteri volti a soppesare il contesto socio-economico dei diversi territori di diretto riferimento nei quali gli Atenei sono collocati.

Gli intervalli dei punteggi attribuiti alle diverse classi di merito dovrebbero essere resi maggiormente omogenei.

Sarebbe auspicabile l'attribuzione di un peso significativo alle attività di Terza missione presentanti anche risvolti di dimensione internazionale.

In aggiunta, potrebbe essere prestata una maggiore attenzione all'interdisciplinarietà dei casi studio.

6. Processo di restituzione (Giudizi del GEV, Rapporto ANVUR)

Punti di forza

Sono stati elaborati giudizi trasparenti, anche se non sempre esaustivi.

Punti di debolezza

Il livello delle valutazioni appare molto diversificato; in alcuni casi, i giudizi finali sono risultati più approfonditi, rendendo possibile comprendere le ragioni dei risultati, mentre in altri casi sono apparsi troppo sintetici.

Proposte

Laddove vengano segnalate mancanze relative alla numerosità/tipologia degli indicatori, sarebbe necessario suffragare i rilievi suggerendo possibili rimedi.

7. Impatto della valutazione (indicatori di risultato R4 e IRAS4, peso su FFO premiale, valorizzazione vs competizione, ecc.)

Punti di forza

È stato esplicitato il peso della valutazione su FFO.

Punti di debolezza

Possono esservi state distorsioni nell'indicatore IRAS4, basato sulla numerosità del personale dell'Ateneo, quando invece il numero di casi studio attesi è stato proporzionale al numero di Dipartimenti. Si potrebbe utilizzare per entrambi la numerosità del personale.

Il peso della valutazione su FFO premiale è risultato troppo scarso per assicurare nel futuro un impegno serio degli Atenei in questa fondamentale operazione di servizio alla società.

Proposte

Si suggerisce di attribuire un maggior peso della valutazione della Terza missione degli Atenei per assicurare un adeguato impegno nell'impatto sociale. Nell'ambito di questo aumento di risorse finanziarie richiesto, si propone, inoltre, che la quota destinata alla Terza missione non eroda quella della ricerca, ma che sia specificata e non metta in competizione ulteriore gli Atenei anche su questo fronte (FFO premiale o altri strumenti come quelli legati alla Programmazione Triennale).

Per contrastare le possibili distorsioni nell'indicatore IRAS4 - basato sulla numerosità del personale dell'Ateneo quando invece il numero di casi studio attesi è stato proporzionale al numero di Dipartimenti - si potrebbe utilizzare per entrambi un criterio comune.

Si suggerisce di fornire i risultati di R anche in modo complessivo, rispetto alla media nazionale, indipendentemente dalla tipologia di Struttura (Università statali, non statali, Scuole e Enti di ricerca).

8. VQR Terza missione e PNRR

Proposte

Per l'esercizio della prossima VQR, occorre considerare che gli Atenei italiani sono attualmente impegnati nei progetti della Missione 4.2 del Piano Nazionale di Ripresa e Resilienza ("dall'Università all'impresa"). Si tratta di progetti che si configurano come strumenti di collaborazione con il territorio su temi di ricerca applicata (es. Ecosistemi dell'Innovazione, Partenariati, ecc...).

Rispetto a questa fattispecie, si potrebbe ipotizzare una scelta di esclusione dai casi di studio connessi al PNRR, vista la natura complessa e la presenza di forme di collaborazione ampia tra molti Atenei.

Tuttavia, a nostro avviso ciò potrebbe rivelarsi estremamente penalizzante per il sistema universitario, oltre che per il sistema-Paese. Riteniamo piuttosto che accettare casi legati al PNRR, comunque da conferire nell'ambito dei campi di azione della tassonomia ordinaria, potrebbe rafforzare l'incentivo degli Atenei a usare correttamente le risorse acquisite, in collaborazione con il territorio e a maturare ulteriormente l'approccio alla Terza missione.

Qualora si decidesse di includere nella VQR anche progetti di Terza missione connessi al PNRR, occorrerebbe prendere in considerazione le seguenti criticità:

- chiarire come considerare la compartecipazione di più Atenei agli output/outcome finali (sarà possibile presentare lo stesso caso studio da parte di più Atenei oppure il singolo Ateneo può

focalizzarsi su parti/ambiti di un progetto in cui il suo contributo risulta ben definibile e prevalente? Come impatterà la valutazione a livello di FFO qualora vi fosse una inscindibile compartecipazione all'impatto finale da parte di più Atenei?). Una soluzione potrebbe essere quella di consentire la sottomissione del caso studio PNRR alternativamente da parte del Coordinatore (come caso sull'intero progetto) o dei responsabili di singoli spoke (come caso studio di spoke) oppure dei singoli Atenei (come caso di studio per parti/ambiti ben specifici, in cui sia ben delineabile il contributo del singolo Ateneo) oppure ancora elaborando una procedura nella quale si individui il contributo di ogni Ateneo. La scelta potrebbe essere lasciata ai partecipanti al progetto, con dichiarazione da rendere all'atto dell'invio del caso;

- valutare i casi studio PNRR in modo diverso da quelli "ordinari", considerando che gli stessi beneficiano di ingenti risorse finanziarie per il raggiungimento degli impatti attesi;
- per la valutazione dell'impatto si potrebbero utilizzare gli stessi indicatori di risultato/impatto su cui il partenariato si è impegnato verso il MUR/Commissione Europea.

9. Valorizzazione e visibilità esterna del quadro risultante delle attività di Terza missione del sistema universitario italiano

Proposte

Si suggerisce l'opportunità di una maggiore valorizzazione e visibilità esterna del quadro delle attività di Terza missione del sistema universitario italiano che risulterà dal prossimo esercizio della VQR, mediante eventuale elaborazione a posteriori da parte di ANVUR, con il possibile supporto della CRUI, di un Rapporto - espressamente orientato, tarato e finalizzato all'opinione pubblica - con la presentazione (supportata da semplici e chiare elaborazioni statistiche) della varietà ed interdisciplinarietà dei casi di studio di Terza missione degli Atenei italiani, eventualmente dettagliando talune "buone pratiche" rappresentative per tipologie di azioni di Terza missione perseguite e per classi di riferimento degli Atenei (mega, grandi, medi, piccoli, non statali).

Dopo la pubblicazione del Rapporto, si suggerisce di creare, attraverso CRUI, degli appuntamenti (ad esempio, ogni due mesi) e un forum tematico volto allo scambio di buone prassi tra delegati alla Terza missione per l'impostazione del prossimo esercizio VQR. Il panorama italiano presenta una serie di associazioni con competenze specifiche nei diversi ambiti della terza missione, le quali potrebbero fornire un contributo attivo.

Parte IV – Reforming Research Assessment

Contesto

Il 20 Luglio 2022, La Commissione Europea (CE) ha reso disponibile in forma definitiva un documento (AGREEMENT ON REFORMING RESEARCH ASSESSMENT) sulle modalità di valutazione della ricerca che ha il potenziale di incidere profondamente sui criteri attualmente usati nelle università Italiane per la valutazione delle strutture di ricerca e per il reclutamento e le carriere di ricercatori e docenti (<https://research-and-innovation.ec.europa.eu/system/files/2022-07/rra-agreement-2022.pdf>).

Il contenuto del documento, la cui bozza circolava già da diversi mesi, è stato recepito positivamente sia dalla League of European Research Universities (LERU) (<https://www.leru.org/publications/a-pathway-towards-multidimensional-academic-careers-a-leru-framework-for-the-assessment-of-researchers>) che dalla European University Association (EUA) (<https://eua.eu/news/922:reforming-research-assessment-the-agreement-is-now-final.html>), con l'espresso invito, diretto agli atenei europei, a **firmare l'agreement e ad impegnarsi a mettere in atto i nuovi criteri in esso contenuti**. L'Agreement ha ottenuto la piena approvazione della Commissaria Europea a Innovazione, Ricerca, Cultura, Istruzione e Giovani Mariya Gabriel, del Presidente di EUA Michael Murphy e del Presidente di Scienza Europa **Marc Schiltz** [1].

Il GdL N. 4 ha preso atto del contenuto dell'Agreement, condividendone in linea di massima le proposte relative alla valutazione delle strutture di ricerca, dei progetti di ricerca e delle carriere dei ricercatori, ma al tempo stesso realizzando la difficoltà ad applicarle immediatamente al sistema Italia, attualmente basato su criteri di valutazione centralizzata che sono, per vari aspetti, molto diversi da quelli proposti dalla CE. Il GdL non ritiene che al momento le Università italiane possano allinearsi alle nuove linee guida proposte senza un'adeguata revisione dei criteri di valutazione, dalla quale peraltro dipende in maniera significativa il finanziamento ministeriale agli atenei. Il GdL si è quindi orientato a recepire le differenze fra il nuovo sistema di valutazione suggerito dalla CE e quello Italiano, e a proporre una serie di modifiche da discutere con gli organi deputati alla valutazione della ricerca.

Con questo documento, partendo dai criteri impiegati per l'ultimo esercizio di valutazione VQR2015-2019, il GdL esprime quindi una serie di riflessioni e suggerimenti su:

- la valutazione delle strutture, e,
- reclutamento e valutazione delle carriere di ricercatori e docenti.

A premessa delle raccomandazioni qui sotto riportate, si precisa che il GdL ritiene che attività di ricerca e di trasferimento tecnologico (TT) siano in continuità bidirezionale l'una con l'altra, in accordo con le linee di indirizzo del PNR (Programma Nazionale della Ricerca²⁰²¹⁻²⁰²⁷) e del PNRR. In tal senso, alcuni dei nuovi indicatori suggeriti per la valutazione della ricerca sono comuni anche alla valutazione delle attività di TT.

Aspetti specifici di osservazione

1. Processo di valutazione della produzione scientifica delle strutture di ricerca (criteri di valutazione con particolare riferimento alle pubblicazioni scientifiche, indicatori, ecc.)

Punti di forza

Il processo di valutazione condotto dai GEV ha assicurato l'anonimato e la mancanza di conflitto di interesse.

Punti di debolezza

Il GdL ritiene innanzitutto che, per le **pubblicazioni scientifiche**, a causa della complessità e della difficoltà legate all'individuazione di referee adeguati per tutte le pubblicazioni da esaminare, **il meccanismo della peer review non possa rappresentare la modalità più adeguata per valutare su larga scala la produzione scientifica delle istituzioni**, soprattutto nelle aree bibliometriche.

La peer review deve essere evitata per vari motivi: non è scalabile, ovvero è difficile trovare revisori per un grande numero di prodotti; i componenti di un GEV non possono mai coprire tutte le aree di ricerca che possono essere presenti nei prodotti inviati; si rischia inoltre di invalidare il lavoro di peer review già svolto dalle riviste (internazionali e indicizzate) che hanno accettato la pubblicazione dei lavori analizzati. Invece, è opportuno avvalersi delle metriche e dei criteri comunemente accettati dalla comunità scientifica internazionale e disponibili sotto forma di database come, ad esempio, Scopus. Il GdL ritiene che la valutazione delle pubblicazioni scientifiche dovrebbe essere completamente - o perlomeno il più possibile - automatica, e per il minimo necessario manuale tramite peer review, da riservarsi solo a casi eccezionali, associati ad elementi di controversia chiaramente documentabili.

Proposte

Sulla base di quanto riportato sopra, dal punto di vista pratico, si suggerisce di adottare per l'ambito bibliometrico una classificazione automatica per lavori ad alto indice citazionale e ottimo impact factor, limitando l'uso della peer review solo ai casi in cui gli indicatori bibliometrici forniscano risultati incerti (si veda anche pag. 6 del presente documento).

Il GdL ritiene inoltre che per tutti i lavori scientifici vada privilegiata e opportunamente valutata la qualità, l'originalità e l'innovatività.

Tuttavia, proprio perché **le metriche non possono rappresentare l'unico parametro per giudicare l'output dell'attività di ricerca¹**, il Gruppo propone che ai parametri relativi alla sede editoriale e alle citazioni delle pubblicazioni scientifiche, vengano affiancati **altri parametri di qualità della ricerca che siano facilmente misurabili**, da normalizzare alle dimensioni della struttura. Questi possono includere almeno 3-5 parametri fra quelli qui sotto elencati:

- l'entità dei finanziamenti ottenuti dalla struttura attraverso bandi competitivi emessi da enti di ricerca pubblici o privati nazionali e internazionali;
- l'eventuale attività di servizio svolta dalla struttura per enti e aziende esterni attraverso consulenze e ricerca commissionata (caratteristiche dell'attività di servizio: da definirsi);
- l'istituzione di laboratori congiunti con aziende o altre istituzioni incluso il Terzo Settore, a testimonianza della capacità di innovazione;
- l'attuazione delle Raccomandazioni della CE dell'11/3/2005 riguardante la Carta Europea dei Ricercatori attraverso l'adozione del Codice di Condotta per l'assunzione dei Ricercatori. Indicatore misurabile di tale attività è, ad esempio il conseguimento dello Human Resources Strategy for Researchers (HRS4R) logo, attribuito dalla CE, che identifica le istituzioni di ricerca e le organizzazioni che attuano politiche di qualità nel reclutamento e nell'ambiente della ricerca. Il conseguimento di tale certificazione europea rende le Istituzioni più attrattive per i ricercatori sia Italiani che stranieri in cerca di una destinazione e testimonia un impegno multidimensionale, a lungo termine e a tutti i livelli delle Istituzioni;
- le politiche di open science, quali ad esempio la costituzione di applicazioni web open source per la condivisione, conservazione, citazione, esplorazione e analisi dei dati della ricerca, o gli

investimenti in contratti trasformativi con le case editrici, normalizzati alla dimensione dell'Ateneo;

- l'analisi dello stato di avanzamento delle azioni disposte nei Gender Equality Plan;
- il Coordinamento di Consorzi di progetti UE (H2020 e/o Horizon Europe);
- il numero di Progetti ERC (sia vinti da docenti già membri dell'ateneo che attratti dall'esterno) e di progetti FIS(Fondo Italiano per la Scienza);
- il numero di progetti Marie Curie Individual Fellowships attratti dalla sede come host;
- il numero di Dottori di Ricerca (in termini di PhD conclusi) con accordi internazionali Joint/Double, ovvero con titolo riconosciuto da almeno un'altra Università internazionale (criteri di identificazione del panorama internazionale di riferimento da definire in una fase successiva della stesura del documento);
- il numero di eventi ad elevato impatto organizzati nell'ambito di azioni Public Engagement (criteri di identificazione della tipologia di evento, del gruppo di interesse a cui è rivolto e delle metriche associate all'impatto saranno da definire in una fase successiva della stesura del documento e ispirate al contesto internazionale);
- l'impatto della ricerca svolta dalla struttura sulle policies e pratiche adottate da decisori, società scientifiche, istituzioni nazionali e internazionali e governi nei campi applicativi di interesse. Relativamente a questo parametro, il Gruppo ritiene che sarà necessario condurre approfondimenti sugli indicatori da utilizzare.

Il GdL propone inoltre che i parametri selezionati per la valutazione relativi ad un certo orizzonte temporale vengano proposti *ex ante* e con ampio anticipo agli Atenei, per dare tempo alle università di verificare la loro posizione e prevedere adeguati sistemi di rilevazione.

2. Reclutamento e valutazione delle carriere di ricercatori e docenti

Il GdL ha anche approfondito il tema del reclutamento e della valutazione delle carriere di ricercatori e docenti, allo scopo di individuare le modalità più corrette per analizzare la produzione scientifica individuale, analizzando vantaggi e svantaggi del sistema attuale.

Punti di forza

Il GdL ritiene all'unanimità che l'introduzione del sistema dell'Abilitazione Scientifica Nazionale (ASN) abbia contribuito a migliorare la qualità dei reclutamenti da parte degli atenei, adeguandola maggiormente agli standard degli altri Paesi Europei. Sulla suddivisione delle discipline in SSD e sulle modalità del reclutamento a livello locale (e in generale sui concorsi universitari) il GdL si riserva di condurre un'analisi dettagliata e formulare proposte in tempi successivi.

Punti di debolezza

Il GdL ritiene che le norme attuali privilegino maggiormente la quantità delle pubblicazioni scientifiche rispetto alla qualità, e che non siano ancora state messe in atto norme efficaci per limitare l' uso scorretto degli indicatori bibliometrici.

Proposte

Allo scopo di incentivare la promozione della qualità delle pubblicazioni scientifiche e del reale apporto (impatto) del docente/ricercatore all'avanzamento delle conoscenze e alla progressione della sua disciplina, il GdL propone:

- la scelta, da parte del candidato al concorso o alla progressione di carriera, di un **numero limitato di pubblicazioni scientifiche**, ove il candidato identifichi egli/ella stesso/a il proprio contributo personale più originale e più importante, giustificandone la motivazione;
- una più appropriata **valorizzazione dell'interdisciplinarietà** delle pubblicazioni scientifiche, a dimostrazione della capacità di interazione del candidato/a con esperti di altre discipline e di un approccio integrato e globale alle tematiche di ricerca trattate. Il GdL propone che, se necessario, il reale valore aggiunto dell'interdisciplinarietà della produzione scientifica del candidato/a venga valutato da una sottocommissione ad hoc, come peraltro succede in altri paesi europei.

Sulla base dei contenuti del documento della Commissione Europea sopra citato, il Gruppo ritiene inoltre opportuno suggerire che, nella valutazione della produzione scientifica, vengano inclusi **tutti gli output della ricerca** condotta dal candidato, quali:

- tutti i prodotti della ricerca (es, la messa in opera di prototipi o software, l'inserimento nelle banche dati di riferimento di nuove informazioni importanti quali sequenze di DNA o strutture virali, campagne di esperimenti o osservazioni, anche attraverso infrastrutture assegnate in modo competitivo - come acceleratori, telescopi, centri di supercalcolo), l'adozione di linee guida e di policies basate sui risultati della ricerca, la messa a punto di nuove terapie, etc);
- l'attività di mentoring e di formazione di ricercatori più giovani;
- le attività di servizio che contribuiscono ad incrementare la qualità della produzione scientifica, quali la partecipazione ai presidi di qualità degli atenei, le attività di tutoraggio di tesisti e dottorandi, il coordinamento dottorale, etc, soprattutto se svolte in un contesto di confronto internazionale.
- i brevetti e le partecipazioni a spin-off universitarie (in quota maggioritaria o da definirsi in maniera opportuna)
- l'impatto sul territorio, per il quale andranno definiti nuovi indicatori appropriati.
- la divulgazione dei risultati e le attività di public engagement.

A tal proposito, il Gruppo propone che venga stilato un form (**portfolio**) che possa contenere tutte le informazioni relative al contributo forniti dal ricercatore/docente con le sue attività, facilmente compilabile dai membri di tutte le aree disciplinari e facilmente comprensibile a tutti, che possa essere adottato a livello locale e nazionale per facilitare e uniformare la valutazione delle carriere scientifiche.

3. Raccomandazioni generali

Il GdL ritiene che qualsiasi parametro di valutazione individuale dovrebbe essere normalizzato all'età accademica del candidato, per non favorire i senior anche su posizioni/iniziative che sono invece pensate per i giovani.

Infine, il Gruppo sottolinea la necessità di una tempistica precisa e prevedibile della valutazione scientifica nazionale grazie alla quale le strutture possano organizzarsi in modo appropriato per la produzione e la raccolta dei dati.

[1] Mariya Gabriel, Commissioner for Innovation, Research, Culture, Education and Youth said: *"I believe in a research culture that recognises a diversity of contributions to science and society; that celebrates high quality and impactful research; and that values sharing, collaboration, integrity and engagement with society, transmitting knowledge from generation to generation."*

Michael Murphy, President of the European University Association, EUA, said: *"It is now time to go beyond existing declarations and define clearly what we want for the future of research assessment. Universities, researchers and all stakeholders need to choose how they want to be assessed and need to choose now."*

Marc Schiltz, President of Science Europe, said *"'Publish or perish' and metrics have led us into a blind alley. Let's start recognizing the full breadth of value created by researchers."*

Conclusioni

Ogni attività di indirizzo, programmazione e monitoraggio è efficace se basata sulla comprensione delle regole di funzionamento alla base del meccanismo di valutazione. A tal fine, si auspica che venga posto in essere un processo di decodificazione dei risultati della VQR3, affinché gli Atenei possano comprendere i punti di forza e di debolezza dei prodotti e dei casi studio presentati e attuare gli opportuni provvedimenti correttivi.

In relazione al processo di Formazione dei GEV ed alla valutazione dei prodotti scientifici presentati dalle Istituzioni, il gruppo di lavoro rileva e apprezza l'elevato grado di consapevolezza e responsabilità maturato dalle Istituzioni nei confronti del processo di valutazione, anche in considerazione delle positive ricadute che questo può e deve avere nell'indirizzare in modo virtuoso le politiche di investimento e programmazione.

Ne deriva l'assoluta necessità di un sistema che sia al contempo trasparente e replicabile e soprattutto basato su indicatori definiti a priori, precocemente, e pertanto monitorabili nel tempo dalle istituzioni.

La composizione dei GEV assume in questa ottica un ruolo fondamentale. I membri dei GEV e i potenziali revisori esterni devono godere di una riconosciuta autorevolezza presso la comunità scientifica della quale sono chiamati a gestire il processo di valutazione; oltre ad un adeguato curriculum scientifico-disciplinare, essi dovrebbero possedere una comprovata esperienza in attività di valutazione della ricerca. A tal fine si auspica, ove possibile e praticabile, anche il ricorso a revisori esterni stranieri.

Nel processo di valutazione, infine, in presenza di eccellenti valori degli indicatori bibliometrici dei prodotti presentati dalle istituzioni, si auspica che nei futuri esercizi di valutazione si trovino i meccanismi adeguati affinché il ruolo della peer review non porti a radicali capovolgimenti della qualità attesa, se non a fronte di adeguate e condivise motivazioni.

In relazione al finanziamento dei Dipartimenti di Eccellenza, il gruppo di lavoro raccomanda una riconsiderazione dell'indicatore standardizzato di performance dipartimentale (ISPD) che ne affronti i problemi di stabilità, riproducibilità, polarizzazione dei valori, in quanto questi problemi inducono nei ricercatori sfiducia nel processo di valutazione e hanno l'indesiderata conseguenza di deresponsabilizzare le strutture valutate. Inoltre, si suggerisce una possibile revisione delle modalità di distribuzione dei finanziamenti nelle diverse tornate.

La Terza missione è meritevole di un maggior riconoscimento istituzionale e finanziario perché attiva processi di condivisione di know-how e di co-produzione di innovazione e conoscenza, che possono favorire lo sviluppo sostenibile dei territori e contrastare le disuguaglianze, ispirandosi a obiettivi di giustizia sociale e ambientale e di apertura internazionale dei sistemi locali. La necessità di individuare degli indicatori di impatto misurabili in maniera quantitativa nella valutazione della Terza missione non sempre ha favorito la selezione dei migliori casi studio. Pertanto, è necessario un approfondimento marcato volto all'individuazione di indicatori specifici e facilmente applicabili, che tengano conto dei diversi orizzonti temporali della pubblicazione scientifica e del suo impatto sul territorio, da rendere noti con lauto anticipo e che tengano conto anche della variegata situazione degli Atenei. Oltre al confronto sulle tematiche contenute nel documento, il Gruppo di lavoro ritiene indispensabile riflettere anche sulla denominazione di "Terza missione", al fine di evitare il rischio che venga percepita come un insieme di attività residuali rispetto alla Didattica e alla Ricerca. Un successivo confronto potrebbe, quindi, in linea con quanto già avviene negli altri Paesi europei, avere per oggetto anche la scelta di un nome diverso per la "Terza missione", che ne valorizzi la continuità bidirezionale con la ricerca fondamentale e la natura di attività di scambio, collaborazione

e co-produzione della conoscenza con le comunità locali (in una logica di Knowledge Exchange), in cui si manifestano pienamente le responsabilità sociali e l'impegno pubblico degli Atenei in direzione di uno sviluppo sostenibile dei territori.

Infine, il Gruppo di lavoro non ritiene che al momento le Università italiane possano allinearsi alle nuove linee guida per la valutazione proposte dalla Commissione Europea nel Luglio 2022 senza un'adeguata revisione dei criteri di valutazione.

Il Gruppo di lavoro elogia l'iniziativa della Commissione Ricerca CRUI di attivare una riflessione sull'esperienza passata della VQR3 e manifesta, quindi, la più completa disponibilità per un successivo confronto con l'ANVUR.

Composizione dei Gruppi di lavoro

Gruppo di lavoro 1 “Peer Review e GEV”

Antonino Squillace, Università degli Studi di Napoli Federico II (Coordinatore)
Alessandro Ambrosi, Università degli Studi Vita-Salute S. Raffaele di Milano
Debora Berti, Università degli Studi di Firenze
Paola Butelli, Politecnico di Milano
Alberto Credi, Università degli Studi di Bologna
Paola Di Donato, Università degli Studi di Napoli Parthenope
Pietro Gargiulo, Università degli Studi di Teramo
Claudia Martini, Università degli Studi di Pisa
Paolo Mariani, Università Politecnica delle Marche
Nicola Moraci, Università degli Studi Mediterranea di Reggio Calabria
Rossella Bonito Oliva, Università degli Studi di Napoli L'Orientale
Marco Oliverio, Università degli Studi di Roma La Sapienza
Paolo Oliveri, Università degli Studi di Genova
Gianmario Raimondi, Università della Valle d'Aosta
Sara Rainieri, Università degli Studi di Parma
Angeloantonio Russo, Università LUM Bari
Mario Savino, Università degli Studi della Tuscia
Laura Scomparin, Università degli Studi di Torino
Davide Scotti, Università degli Studi di Bergamo
Gianluca Setti, Politecnico di Torino
Daniela Vuri, Università degli Studi di Roma Tor Vergata

Gruppo di lavoro 2 “ISPD e Dipartimenti di Eccellenza”

Elisa Roma, Università degli Studi di Pavia (Coordinatrice)
Lucia Altucci, Università degli Studi della Campania "L. Vanvitelli"
Alessandro Baraldi, Università degli Studi di Trieste
Mauro Barni, Università degli Studi di Siena
Luciano Colombo, Università degli Studi di Cagliari
Federico Forneris, Università degli Studi di Pavia
Nadia Gatti, Università degli Studi di Teramo
Cesare Indiveri, Università della Calabria
Antonina Maria Marino, Politecnico di Torino
Nicola Moraci, Università degli Studi Mediterranea di Reggio Calabria
Marco Oliverio, Università degli Studi di Roma La Sapienza
Anna Paganoni, Politecnico di Milano
Alessandro Petrolini, Università degli Studi di Genova
Giuseppe Pulina, Università degli Studi di Sassari
Cristina Prandi, Università degli Studi di Torino
Alessandro Trovarelli, Università degli Studi di Udine

Gruppo di lavoro 3 “Terza missione”

Elisa Giacosa, Università degli Studi di Torino (Coordinatrice)
Anna Antola, Politecnico di Milano
Fabrizio Benente, Università degli Studi di Genova
Giuseppe Ciccarone, Sapienza Università di Roma
Andrea De Bortoli, Università degli Studi di Torino e APENet
Caterina Falbo, Università degli Studi di Trieste
Lelio Iapadre, Università degli Studi dell’Aquila
Massimo Lauria, Università degli Studi Mediterranea di Reggio Calabria
Patrizia Lombardi, Politecnico di Torino e Rete delle Università per lo Sviluppo sostenibile (RUS)
Francesco Mauriello, Università degli Studi Mediterranea di Reggio Calabria
Fabrizio Pilo, Università degli Studi di Cagliari
Francesca Spigarelli, Università degli Studi di Macerata
Fabio Tatano, Università degli Studi di Urbino “Carlo Bo”

Gruppo di lavoro 4 “Reforming Research Assessment”

Maria Pia Abbracchio, Università degli Studi di Milano (Coordinatrice)
Alessandro Ambrosi, Università Vita e Salute San Raffaele
Francesca Demichelis, Università degli Studi di Trento
Laura Gaggero, Università degli Studi di Genova
Johann Gamper, Libera Università di Bolzano
Alberto Guadagnini, Politecnico di Milano
Manuela Minozzi, Roma Link Campus
Alessandro Vercelli, Università degli Studi di Torino
Fabio Zwirner, Università degli Studi di Padova