

REPORT

VQR 2020-2024

ARCH-01 – Archeologia

Analisi delle criticità metodologiche, delle difformità valutative e delle anomalie di rappresentazione del dato

1.449 prodotti ARCH-01 <i>perimetro nazionale</i>	0,68 valutazione media <i>dato GSD</i>	57,35% prodotti A+B <i>fasce alte</i>	5,80% prodotti D+E <i>fasce critiche</i>
--	---	--	---

1. Quadro di sintesi

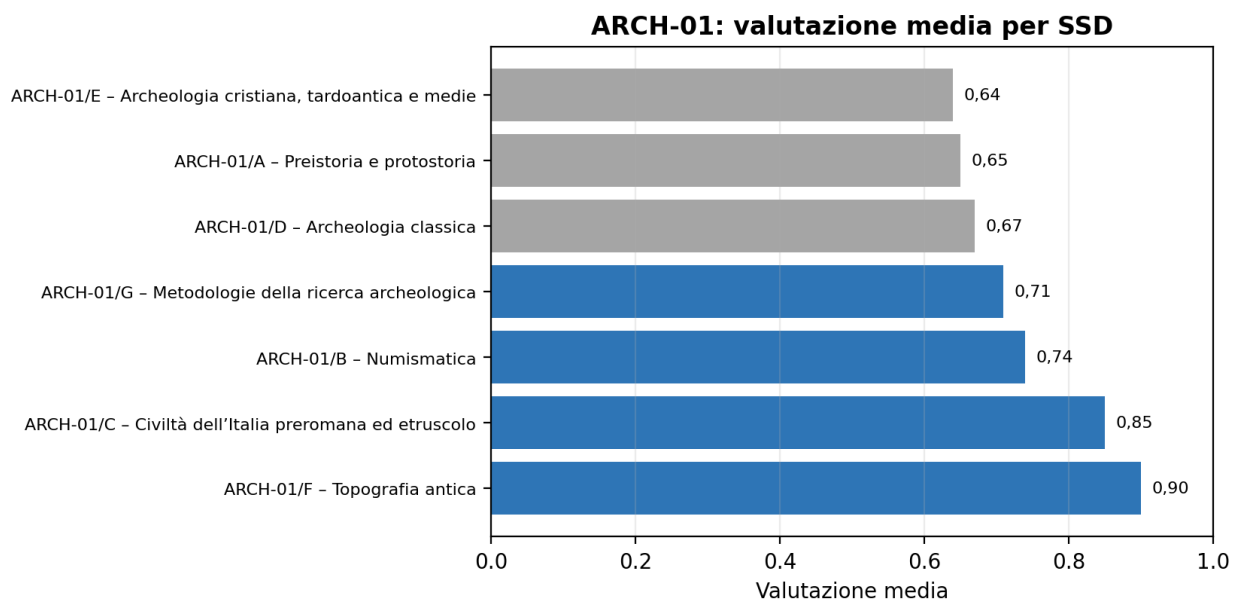
L'analisi dei dati VQR 2020-2024 dell'Area 10, con particolare riferimento al GSD 10/ARCH-01 – Archeologia, fa emergere alcune criticità che non riguardano solo il merito dei singoli prodotti, ma anche la tenuta metodologica del processo valutativo. Il sistema VQR si fonda su criteri comuni — originalità, metodologia e impatto — e su una procedura di peer review che dovrebbe garantire omogeneità di trattamento tra prodotti, settori e istituzioni. Tuttavia, proprio nei dati aggregati si osservano scostamenti tali da richiedere un approfondimento da parte di ANVUR e Ministero.

Il quadro nazionale del GSD Archeologia è articolato: convivono settori molto performanti, settori intermedi e settori più fragili. Il dato complessivo nazionale è pari a circa 1.449 prodotti, con una valutazione media di circa 0,68, una quota A+B del 57,35% e una quota D+E del 5,80%. Nel complesso, quindi, Archeologia non appare come un settore in crisi, ma come un ambito fortemente disomogeneo al suo interno.

2. Indicatori principali per SSD

SSD	Settore	Prodotti	Media	A+B	D+E	Rev. esterne
ARCH-01/F	Topografia antica	85	0,90	89,41%	0,00%	24,34%
ARCH-01/C	Civiltà dell'Italia preromana ed etruscologia	99	0,85	92,93%	2,02%	40,95%
ARCH-01/B	Numismatica	52	0,74	63,46%	7,69%	37,50%
ARCH-01/G	Metodologie della ricerca archeologica	181	0,71	60,22%	3,31%	7,88%
ARCH-01/D	Archeologia classica	344	0,67	52,91%	6,10%	68,96%
ARCH-01/A	Preistoria e protostoria	509	0,65	49,70%	6,88%	48,94%
ARCH-01/E	Archeologia cristiana, tardoantica e medievale	179	0,64	48,05%	8,94%	32,99%

Nota: i valori evidenziano una forbice interna molto ampia, dal massimo di Topografia antica al minimo di Archeologia cristiana, tardoantica e medievale. La variabilità è esito di processi valutativi differenziati, come mostra, per esempio, il diverso ricorso ai revisori esterni.



3. Anomalie e criticità emerse

3.1 Forte variabilità tra SSD

Il primo elemento critico riguarda la forte variabilità tra settori. Alcuni SSD ottengono risultati molto elevati: Topografia antica ha una media di circa 0,90, quasi il 90% di prodotti A+B e nessun D/E; Civiltà dell'Italia preromana ed etruscologia raggiunge circa 0,85, con oltre il 92% di A+B. All'opposto, Preistoria e protostoria, Archeologia classica e Archeologia cristiana, tardoantica e medievale si collocano in una fascia più bassa, con medie comprese tra circa 0,64 e 0,67.

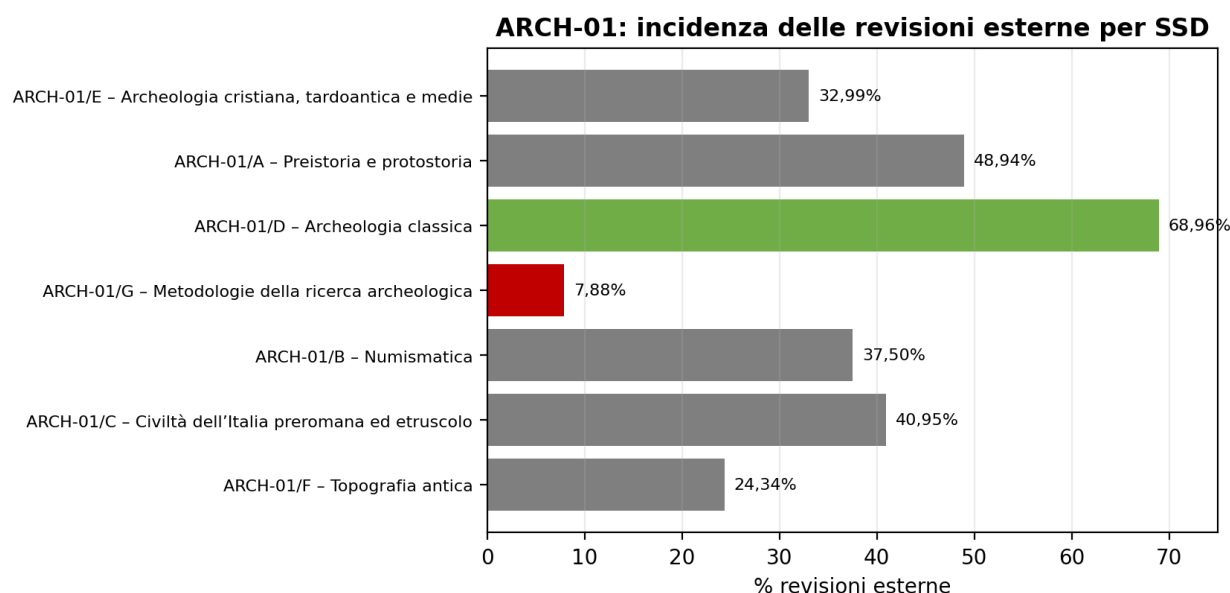
Questa distanza interna è ampia e solleva un problema di comparabilità tra SSD appartenenti allo stesso GSD e un problema di omogeneità valutativa tra settori con affinità scientifiche. Non è possibile stabilire dai dati aggregati se il risultato dipenda da qualità effettivamente inferiore, selezione meno efficace, sedi editoriali percepite come meno competitive o maggiore severità dei valutatori. La forbice osservata suggerisce comunque la necessità di una verifica sulla coerenza applicativa dei criteri.

3.2 Ricorso disomogeneo ai revisori esterni

Un secondo profilo critico riguarda il ricorso disomogeneo ai revisori esterni. Archeologia presenta nel complesso una quota di revisioni esterne pari a circa 42,65%, inferiore alla media dell'Area 10. Tuttavia, il dato varia fortemente tra SSD: Archeologia classica arriva a circa 68,96%, mentre Metodologie della ricerca archeologica registra solo 32 revisioni esterne su 406, pari al 7,88%.

Caso procedurale rilevante: ARCH-01/G

Metodologie non presenta una performance negativa: media circa 0,71, oltre 60% A+B e quota D+E contenuta. La criticità non è quindi legata all'esito, ma al processo: un settore valutato quasi interamente all'interno del GEV risulta meno confrontabile con settori sottoposti a un ricorso molto più ampio alla revisione esterna. Tale anomalia, pur se meno rilevante si riscontra anche per il settore di Topografia antica (24, 34%).



3.3 Caso CNR e rappresentazione del dato

Un terzo elemento riguarda il caso CNR, particolarmente rilevante per Preistoria e protostoria, ma con effetti sulla lettura di tutta Archeologia. Il CNR incide per circa il 24,3% sul totale dell'area archeologica, con 353 prodotti su 1.449. Nelle tabelle VQR, quasi tutto l'apporto archeologico del CNR risulta concentrato e riversato in ARCH-01/A.

L'articolazione scientifica del CNR in ambito archeologico comprende tuttavia anche archeologia classica, topografia, metodologie, archeologia tardoantica e medievale. Secondo la ricostruzione svolta, tale ripartizione sarebbe stata considerata nel processo valutativo, ma nel riversamento dei dati negli output tabellari i valori risultano aggregati al settore di Preistoria e protostoria. Questo scarto tra classificazione tabellare e profilo scientifico effettivo produce una rappresentazione potenzialmente distortiva: ARCH-01/A appare sovradimensionato, mentre altri SSD risultano sottorappresentati nel contributo CNR.

353	24,3%	ARCH-01/A
prodotti CNR su ARCH-01	peso CNR nel GSD	SSD assorbente
<i>sui 1.449 totali</i>	<i>incidenza nazionale</i>	<i>output tabellare</i>

3.4 Effetti sugli indicatori e sulle classifiche

Gli indicatori sintetici — R, valutazione media, quota A+B — sono utili, ma non bastano a garantire una lettura robusta degli esiti. Nei settori con pochi prodotti, poche valutazioni basse possono modificare sensibilmente medie e posizioni. Nei settori più ampi, invece, effetti di classificazione e conferimento possono condizionare la rappresentazione nazionale del peso disciplinare.

Le classifiche producono effetti reputazionali e strategici rilevanti. Per questo devono essere lette insieme alla numerosità dei prodotti, alla quota di revisioni esterne, alla distribuzione disciplinare dei conferimenti e alla presenza di eventuali anomalie di rappresentazione del dato.

4. Conclusione

Sintesi conclusiva

Archeologia mostra almeno tre anomalie sistemiche: forte disomogeneità tra SSD, uso molto variabile dei revisori esterni e attribuzione CNR concentrata quasi interamente in Preistoria. Questi elementi non dimostrano automaticamente errori puntuali, ma indicano che il sistema, così applicato, non garantisce sempre una lettura pienamente equilibrata della qualità scientifica né un utilizzo pienamente efficace del processo valutativo per la crescita della ricerca.

La conclusione più prudente è che la VQR restituisce un quadro utile, ma non autosufficiente. Gli indicatori di Archeologia vanno interpretati insieme alla numerosità dei prodotti, alla quota di revisioni esterne, alla distribuzione disciplinare dei conferimenti e al peso degli enti di ricerca. Senza questi correttivi, il rischio è che la VQR misuri non soltanto la qualità della ricerca, ma anche effetti di classificazione, calibrazione interna e diversa gestione valutativa tra settori.

Appendice – Punti in discussione

Si propone di seguito un elenco di questioni emerse dall'analisi dei dati VQR relativi ad Archeologia. I punti riprendono, in forma più estesa, le criticità sintetizzate nel testo principale e possono costituire una base per eventuali osservazioni o proposte da sottoporre ad ANVUR e al Ministero.

Difformità di calibrazione tra settori e GEV

I dati mostrano scostamenti molto marcati tra SSD appartenenti alla stessa area e allo stesso GSD, non sempre spiegabili con la sola qualità dei prodotti conferiti. In alcuni casi, anche il miglior risultato istituzionale resta appena in linea con la media d'area, suggerendo una possibile applicazione più severa dei criteri. Ciò pone un problema di omogeneità valutativa e di comparabilità nazionale.

Eccessiva internalizzazione e peso della soggettività valutativa

La bassa quota di revisioni esterne in alcuni settori, unita al forte peso del giudizio interno al GEV, può amplificare l'effetto della soggettività del valutatore o del gruppo di valutazione. Alcuni casi, con meno dell'8% di revisioni esterne, segnalano una limitata pluralizzazione del giudizio e una possibile autoreferenzialità del processo, soprattutto in settori scientificamente articolati come quelli archeologici.

Attribuzioni disciplinari incoerenti o distorsive

Il caso CNR in Archeologia evidenzia una concentrazione quasi totale dei prodotti valutati presso l'ente di ricerca nel SSD di Preistoria e protostoria, pur in presenza di competenze scientifiche più ampie. Questo può alterare la rappresentazione reale dei settori, gonfiando artificialmente alcuni SSD e sottorappresentandone altri. La VQR rischia così di fotografare il meccanismo di conferimento più che la struttura scientifica effettiva.

Fragilità degli indicatori e delle classifiche

Indicatori come R, valutazione media e quota A+B sono utili, ma non sufficienti a garantire una lettura robusta degli esiti. Nei settori con pochi prodotti, poche valutazioni basse possono modificare sensibilmente medie e posizioni. Le classifiche producono effetti reputazionali e strategici rilevanti, ma non sempre appaiono sostenute da una base statistica e procedurale pienamente solida, soprattutto quando settori piccoli e settori ampi sono confrontati con logiche valutative non del tutto omogenee.

Insufficiente trasparenza e verificabilità del processo

Le tabelle finali non consentono di ricostruire il peso effettivo delle revisioni, le eventuali divergenze tra valutatori, le modalità di riallineamento interno o le ragioni delle attribuzioni SSD. Senza dati più disaggregati e controlli pubblicamente verificabili, il sistema non garantisce pienamente qualità, terzietà e comparabilità tra settori, istituzioni e prodotti.

Composizione, coordinamento e funzionamento dei GEV

Una criticità rilevante riguarda la possibile difformità di valutazione tra SSD, che potrebbe dipendere da una valutazione eccessivamente sviluppata a livello disciplinare e non sufficientemente armonizzata a livello d'area. Questo aspetto pone il problema della composizione dei sub-GEV, del coordinamento tra valutatori e del ruolo dei profili scientifici dei componenti e dei coordinatori nel garantire criteri realmente condivisi.

Controllo sull'attribuzione dei prodotti agli SSD effettivi

Alla luce delle anomalie emerse nel caso di Archeologia preistorica e protostorica, appare necessario sviluppare una filiera più chiara e trasparente del prodotto: attribuzione disciplinare, conferimento, valutazione e restituzione del dato. Una maggiore verificabilità dell'assegnazione dei prodotti agli SSD effettivi ridurrebbe il rischio di distorsioni nella rappresentazione dei settori.

Formazione dei GEV e chiarezza dei processi valutativi

Pur in presenza di attività di formazione e informazione rivolte ai soggetti coinvolti nella VQR, appare opportuno rafforzare tali strumenti. È necessario garantire maggiore coerenza tra criteri dichiarati, priorità metodologiche e applicazione concreta della valutazione. Questo riguarda sia i valutatori sia il personale docente e ricercatore chiamato a conferire i prodotti.

Necessità di verifiche ex post e correttivi di sistema

Le criticità individuate non implicano necessariamente errori puntuali nelle singole valutazioni, ma segnalano l'esigenza di controlli ex post più strutturati. Sarebbe opportuno prevedere verifiche sulle attribuzioni disciplinari, sulla distribuzione delle revisioni esterne, sulla coerenza tra SSD e sulla calibrazione dei giudizi, così da rafforzare la credibilità complessiva del sistema VQR.