

CRUI-Università della Basilicata

Sistema di Supporto alla Valutazione della Produzione Scientifica degli Atenei

Indicazioni per la Risoluzione delle Anomalie

Versione 1.0 del 26/04/18 13:12:00

1. Introduzione

A partire dalla I tornata di valutazione 2018, il Sistema di Supporto adotta una nuova **modalità incrementale per l'acquisizione dei metadati dei prodotti**. Questo consente di ripetere le acquisizioni più frequentemente, e di alleggerire il carico sulle istanze di IRIS. Per semplificare la consultazione dei report delle anomalie, è stata contestualmente introdotta una nuova **funzione di consolidamento dei report**.

Questo documento intende fornire alcune indicazioni utili per affrontare la soluzione delle anomalie utilizzando i report consolidati. In particolare, discute l'impatto delle diverse tipologie di segnalazioni sugli esiti della valutazione, in modo da aiutare gli Atenei a dare priorità alla correzione degli errori più rilevanti.

2. Caratteristiche delle Acquisizioni Incrementali

E' importante sottolineare alcune differenze tra le acquisizioni incrementali e le acquisizioni massive.

Le acquisizioni massive riguardano tutti i prodotti da valutare. Le acquisizioni incrementali, viceversa, riguardano solo un sottoinsieme dei prodotti valutati, per l'appunto quelli che hanno subito delle modifiche tra una acquisizione e un'altra.

Di conseguenza, rispetto alle segnalazioni generate dalle acquisizioni massive, è naturale che le segnalazioni generate da un'acquisizione incrementale siano ridotte. Per essere più chiari:

- consideriamo il caso di un prodotto per cui era stata segnalata in precedenza un'anomalia, e i cui metadati sono stati modificati per risolvere l'anomalia; questo prodotto viene riacquisito nell'acquisizione incrementale successiva, e se l'anomalia non è segnalata vuol dire che è stata risolta;
- viceversa, consideriamo il caso di un prodotto per cui era stata segnalata in precedenza un'anomalia, e i cui metadati NON sono stati modificati per risolvere l'anomalia; questo prodotto non verrà riacquisito; di conseguenza l'anomalia non viene segnalata, ma NON è stata risolta.

Proprio per questa ragione, i prodotti i cui metadati vengono effettivamente acquisiti vengono elencati per esteso nel report.

Per semplificare la lettura e l'interpretazione delle segnalazioni generate dalle acquisizioni incrementali, è stato introdotto lo strumento dei **report consolidati**. Dopo ciascun aggiornamento incrementale, il sistema genera un report riassuntivo consolidando tutte le segnalazioni delle anomalie generate fino a quel punto. Il report consolidato fornisce quindi un quadro unico delle anomalie segnalate, di quelle che sono state risolte, e di quelle che restano da risolvere. Rappresenta quindi lo strumento principale per la correzione dei metadati.

E' opportuno inoltre segnalare una seconda differenza importante tra l'acquisizione massiva e quella incrementale. Le acquisizioni massive effettuano una query per ciascun soggetto valutato. Le acquisizioni incrementali, viceversa, effettuano un'unica query, che richiede ad IRIS tutti i prodotti dell'Ateneo che sono stati aggiunti o eliminati o hanno subito modifiche tra la data iniziale e quella finale.

Di conseguenza, nel corso di un'acquisizione incrementale è possibile che vengano analizzati prodotti i cui metadati sono stati modificati nell'intervallo temporale di riferimento, ma che non riguardano i soggetti valutati. Un tipico esempio sono le tesi di dottorato nel caso di valutazioni relative al personale strutturato (professori e ricercatori), oppure i prodotti di personale cessato che non compare tra i soggetti valutati.

Ovviamente questi prodotti non riguardano i soggetti di queste valutazioni, e vengono esclusi dall'acquisizione. **I report incrementali – e di conseguenza, i report consolidati – possono quindi contenere segnalazioni relative a prodotti non acquisiti perché tra i coautori non compare nessuno dei soggetti valutati della valutazione.**

3. Tipologie di Segnalazione

Il report consolidato fornisce essenzialmente tre tipologie di segnalazioni:

1. **Informazione:** le righe di questa tipologia non sono finalizzate a segnalare errori, ma solo a fornire un quadro completo di quali prodotti sono stati oggetto di acquisizione o di eliminazione in ciascun aggiornamento incrementale. In altri termini, servono a ricostruire la “storia” delle acquisizioni che hanno riguardato il prodotto. Non rappresentano quindi anomalie da sanare.
2. **Anomalia Risolta:** queste righe, chiaramente indicate con il carattere barrato, rappresentano anomalie che erano state segnalate nel corso di precedenti acquisizioni, ma che non sono state sollevate nelle acquisizioni successive. Questo è il segnale che il problema è stato risolto, oppure era dovuto ad un malfunzionamento temporaneo dei sistemi esterni a cui il Sistema di Supporto accede per l'acquisizione dei metadati (es: temporanea indisponibilità delle API di WOS o di Scopus).
3. Anomalie da risolvere, divise in varie sottocategorie, come **Errore nei Metadati, Inconsistenza nei Metadati, Prodotto non Acquisito** ecc. Queste segnalazioni sono le uniche che rappresentano effettivamente potenziali anomalie da sanare nei metadati.

Di conseguenza, in fase di correzione degli errori, è consigliabile per cominciare utilizzare le funzionalità offerte da Excel per filtrare le righe del report consolidato sulla base della prima colonna, escludendo le righe che ricadono nelle categorie 1. e 2.

Si noti, invece, che ai fini dell'analisi dei problemi collegati ad uno specifico prodotto può invece essere utile visualizzare tutte le righe relative al prodotto, per avere un quadro completo delle anomalie segnalate fino a quel punto, di quelle risolte e di quelle che restano da risolvere.

E' opportuno anche sottolineare subito che **non tutte le segnalazioni emesse dal sistema rappresentano effettivamente errori nei metadati di IRIS, e non tutte le segnalazioni corrispondono esattamente alla tipologia di errore individuato dal sistema.**

Per fornire un esempio del primo caso, quando il sistema segnala differenza nel numero di autori tra IRIS e Scopus, non necessariamente l'errore è nel metadato di IRIS; in alcuni casi, viceversa, l'errore è nei metadati di Scopus. In questi casi l'unica azione che è possibile intraprendere è contattare il servizio clienti di Scopus per segnalare l'errore e richiederne la correzione.

Per fornire un esempio del secondo caso, il sistema utilizza varie tecniche per la ricerca dei potenziali prodotti duplicati, che vanno dal confronto del DOI, a quello dei codici WOS e Scopus, fino al confronto di titolo, autori, tipologia ed anno. Non necessariamente la segnalazione di un possibile duplicato corrisponde in realtà alla presenza di prodotti duplicati nel catalogo: in alcuni casi, la segnalazione potrebbe essere dovuta semplicemente ad una scorretta imputazione del DOI ad uno dei prodotti coinvolti. In questo caso è sufficiente correggere il valore del DOI per eliminare la segnalazione.

In altri termini, **il sistema raccoglie segnali di potenziali errori e li riporta agli utenti**, ma le segnalazioni devono poi essere opportunamente approfondite per appurare la vera causa dell'eventuale problema.

Per tenere traccia delle segnalazioni per le quali si è appurato che non ci sono errori da correggere, suggeriamo di conservare in un file Excel separato dai report una tabella con i codici dei prodotti verificati e un'indicazione delle segnalazioni già verificate. In questo modo, utilizzando le funzioni per la ricerca fornite da Excel, dovrebbe essere relativamente semplice distinguere in un nuovo report le segnalazioni già verificate e che non è necessario approfondire ulteriormente.

4. Comportamento del Sistema in Presenza di Anomalie

E' opportuno sottolineare due principi alla base del funzionamento del Sistema di Supporto in presenza di anomalie nei metadati.

Il primo è un principio di **inclusività della valutazione**: il sistema tende a non escludere nessun prodotto dalla valutazione. Viceversa, tenta comunque di effettuare la valutazione anche in presenza di anomalie. Di conseguenza – tranne che in casi ovvi, come quelli in cui il prodotto non viene acquisito per nulla – la presenza di segnalazioni non impedisce la valutazione del prodotto.

Il secondo è un principio di **correlazione dei dati per la correzione delle anomalie**: in presenza di metadati mancanti o inconsistenti, il sistema utilizza le restanti informazioni disponibili in merito al prodotto per cercare di sanare l'anomalia in fase di valutazione. Questo vale in particolare per i problemi collegati al numero di autori:

- Nel caso in cui il metadato sia mancante nella scheda IRIS del prodotto, il sistema cerca in ogni caso di calcolarlo analizzando la stringa autori.
- Nel caso in cui ci siano valori inconsistenti del metadato tra IRIS e i fornitori di dati bibliometrici (WOS e Scopus), il sistema corregge il numero di autori utilizzando il valore più alto tra quello di IRIS e quello riportato da Scopus, se disponibile, o in alternativa da WOS.

5. Impatto delle Anomalie e Priorità di Risoluzione

Alla luce delle considerazioni riportate nella Sezione 4, dovrebbe essere chiaro il fatto che non tutte le anomalie hanno la stessa incidenza sulla valutazione dei prodotti. Alcune, in realtà, non hanno alcuna incidenza, dal momento che il sistema corregge il metadato prima di utilizzarlo in fase di valutazione.

L'obiettivo di questa sezione è discutere l'impatto delle diverse tipologie di errori sui risultati della valutazione. Se è ovvio, che in linea di principio sarebbe opportuno correggere tutti gli errori presenti nei metadati, è altrettanto vero che gli Atenei possono trovarsi nella condizione di dover attribuire delle priorità. La discussione che segue dovrebbe fornire un ausilio in questa direzione.

Per cominciare, enunciamo alcune regole di buon senso:

- E' sempre opportuno **dare priorità alla correzione dei metadati dei prodotti degli ultimi anni** (ad esempio, nel 2018, ai prodotti dal 2012 al 2018). Questi prodotti, infatti, sono coinvolti tanto nella valutazione con parametri FFABR/VQR-Like, tanto in quella con parametri ASN. Di conseguenza, le inconsistenze che li riguardano rischiano di pregiudicare entrambe le valutazioni. I prodotti più vecchi, invece, sono coinvolti solo nella valutazione ASN.
- Fino a che la valutazione FFABR si baserà esclusivamente sui dati bibliometrici di Scopus, è **opportuno dare priorità alla correzione degli errori che riguardano i dati bibliometrici di Scopus** rispetto a quelli di WOS. I primi, infatti, hanno impatto sia sulla valutazione con parametri FFABR/VQR-Like, sia su quella con parametri ASN, i secondi solo sulla valutazione ASN.

Tenendo in mente queste indicazioni, di seguito presentiamo le principali tipologie di segnalazioni emesse dal sistema, discutendone l'impatto sui risultati della valutazione. L'elenco è riportato in ordine decrescente di impatto, per cui le prime segnalazioni sono quelle a cui è opportuno dare la priorità più alta.

1. **Prodotti duplicati**: questa è probabilmente la tipologia di errori che ha l'impatto più alto della valutazione, dal momento che altera i punteggi e i valori degli indicatori. Dovrebbe essere la prima ad essere eliminata.
2. **Errori nei codici Scopus e WOS**: dal momento che tutti i modelli di valutazione si basano pesantemente sull'utilizzo del numero di citazioni ricevute dai prodotti, associare ad un prodotto un codice Scopus o WOS scorretto può avere un impatto notevole sui risultati della valutazione. Queste segnalazioni

dovrebbero quindi rappresentare la seconda priorità da risolvere in fase di correzione dei metadati. Si noti che ci sono due grandi tipologie di errori che ricadono in questa categoria:

- a. **Errori nell'acquisizione dei dati bibliometrici per il prodotto**, tipicamente dovuti a codici Scopus o WOS indicati in forma scorretta. La conseguenza di questi errori è tipicamente di rendere il prodotto non valutabile. Si noti che, in altri casi, questi errori sono dovuti semplicemente a malfunzionamenti temporanei nelle API degli editori al momento dell'acquisizione. In questo caso, l'anomalia si risolve spontaneamente all'acquisizione successiva.
 - b. **Inconsistenze nei codici Scopus/WOS**, tipicamente dovuti all'associazione al prodotto di codici Scopus o WOS di prodotti diversi. La conseguenza di questi errori è tipicamente di alterare la valutazione del prodotto, in meglio o in peggio, associandogli un numero di citazioni scorretto.
3. **Metadati mancanti che possono pregiudicare la valutazione del prodotto**: il sistema riporta varie segnalazioni di metadati mancanti. Non tutte queste segnalazioni hanno la stessa incidenza sulla valutazione. Tra le segnalazioni che possono pregiudicare la valutazione del prodotto, segnaliamo la mancanza di codici **ISBN/ISSN**. I criteri di valutazione ANVUR tendono a distinguere i prodotti valutabili da quelli non valutabili sulla base della presenza o meno di questi codici. E' opportuno quindi destinare a questi prodotti la dovuta attenzione.
 4. **Prodotti non acquisiti o eliminati**: i prodotti non acquisiti dal sistema ovviamente non vengono valutati. Nella maggior parte dei casi, questo dipende dal fatto che il prodotto non è in uno stato accettato dal sistema (ad esempio: non è inviabile a loginmiur). In altri casi, il soggetto valutato non è tra gli autori riconosciuti del prodotto. E' opportuno valutare queste situazioni, per evitare che il sistema ometta prodotti che sarebbero invece da valutare.
 5. **Impossibile riconoscere la posizione dell'autore**: i criteri di valutazione relativi al FFABR richiedono, in alcune aree, di attribuire un coefficiente moltiplicativo pari al 10% del punteggio del prodotto agli autori che compaiono in prima ed in ultima posizione. Per attribuire questo punteggio è necessario riconoscere la posizione di ciascun soggetto valutato all'interno della stringa autori. Sottolineiamo che questa procedura non ha nulla a che fare con l'attribuzione del prodotto al soggetto valutato. Si tratta piuttosto di un'elaborazione aggiuntiva, finalizzata semplicemente a riconoscere se il soggetto compare come primo o ultimo autore del prodotto. Nei casi in cui la stringa autori non rispetta il formato standard di IRIS – del tipo Cognome, Nome; Cognome, Nome; ... – il sistema può avere difficoltà a riconoscere la posizione del coautore, e quindi ad attribuire correttamente il coefficiente relativo. Si tratta quindi di anomalie da sanare. D'altro canto, queste incidono solo per il 10% del punteggio del prodotto, e di conseguenza possono essere considerate a più bassa priorità rispetto a quelle precedenti.
 6. **Numero di autori mancante o inconsistente**: queste segnalazioni, come discusso nella Sezione 4, in molti casi non hanno alcun effetto sulla valutazione del prodotto, perché il sistema provvede autonomamente a correggere il metadato. Nel caso in cui la valutazione attribuita sulla base del metadato corretto dal sistema sia accurata, è possibile evitare di correggere l'errore. Quando, invece, questo non è vero, è opportuno procedere alla correzione del metadato.
 7. **Metadati mancanti o inconsistenti che non pregiudicano la valutazione del prodotto**: ogni volta che il sistema trova un metadato mancante – ad esempio, la casa editrice di un volume – oppure palesemente scorretto – ad esempio, un numero di pagine che non è un valore numerico – segnala per completezza l'anomalia. D'altro canto, in entrambi questi casi l'errore non ha alcun impatto sulla valutazione del prodotto. Di conseguenza, anomalie di questo tipo possono essere considerate a bassa priorità.