

# **INAIL**

ISTITUTO NAZIONALE PER L'ASSICURAZIONE  
CONTRO GLI INFORTUNI SUL LAVORO

## Contarp-Csa



## **COGNITIVE COMPUTING E LINGUAGGIO NATURALE PER L'ANALISI DELL'EVOLUZIONE DEI BANDI ISI 2014-2018**

Applicazione delle risorse cognitive all'analisi dei bandi di finanziamento alle imprese per la realizzazione di interventi in materia di salute e sicurezza sul lavoro in attuazione dell'art. 11, comma 5, del d.lgs. 81/2008, erogati dall'Istituto dal 2014 al 2018

A cura di: Ilaria Barra, Domenico Magnante (Contarp), Giuseppe Morinelli (Csa)

# SOMMARIO

1. Introduzione	6
2. Architettura del dato di riferimento	7
3. approccio e tecnologie utilizzate	8
3.1 Datacap	10
3.2 Watson	10
4. Bandi ISI	12
4.1 Progetti di bonifica amianto	13
4.2 Progetti di riduzione del rischio da vibrazioni	21
5. Conclusioni	26

## Premessa

*L'Inail costituisce uno dei soggetti destinatari degli obiettivi e programmi dell'azione pubblica di miglioramento delle condizioni di salute e sicurezza dei lavoratori, individuati dal Comitato per l'indirizzo e la valutazione delle politiche attive e per il coordinamento nazionale delle attività di vigilanza in materia di salute e sicurezza sul lavoro (art. 5 d.lgs. 81/2008).*

*Le attività promozionali della cultura e delle azioni di prevenzione che ne derivano e che l'Inail mette in atto, si sviluppano anche nel finanziamento, previo trasferimento delle necessarie risorse da parte del Ministero del lavoro, di progetti di investimento in materia di salute e sicurezza sul lavoro da parte delle piccole, medie e micro imprese, i cosiddetti avvisi pubblici ISI. Questo percorso, oramai ultra decennale, è stato sviluppato attraverso sistemi procedurali di accesso a tali finanziamenti che hanno comportato affinamenti continui, sia dal punto di vista amministrativo che dal punto di vista tecnico. Ciò al fine di garantire semplicità di accesso, uniformità di trattamento e particolare attenzione all'individuazione di interventi capaci di incidere sulle condizioni più rischiose e maggiormente diffuse cui sono esposti i lavoratori. Il miglioramento delle condizioni di salute e sicurezza è un presupposto per la concessione del finanziamento, ed è garantito da una accurata ricerca delle tipologie di intervento che Inail propone e riporta negli allegati tecnici degli avvisi pubblici ISI pubblicati annualmente e nella accuratezza della valutazione tecnica di ogni singolo progetto presentato. L'efficacia prevenzionale, pur garantita per ogni singolo progetto, resta comunque un parametro importante da misurare anche nel complesso dell'iniziativa di incentivazione e sostegno.*

*Infatti, la misura di tale efficacia è importante sia per evidenziare la bontà dell'investimento, sia per meglio ulteriormente indirizzare, migliorare e selezionare gli interventi più performanti.*

*Gli infortuni sul lavoro e le malattie professionali, dal punto di vista statistico, sono eventi molto rari, in più le seconde si manifestano con tempi di latenza piuttosto lunghi, da misurare in anni, rispetto alla avvenuta esposizione al rischio.*

*Ciò comporta che i metodi tradizionali di indagine alla base della comprensione della causa scatenante dell'evento non siano del tutto adeguati per intercettare le modalità di accadimento dell'infortunio o dell'esposizione al rischio.*

*Inoltre, le componenti che determinano il rischio si sommano e si intersecano in maniera complessa; per questo motivo risulta particolarmente arduo determinare nel breve l'effettiva riduzione del fenomeno infortunistico o tecnopatico, e questa criticità si presenta sia nell'analisi del complesso dei rischi che in quella connessa agli effetti di un singolo rischio.*

*Quando un evento è definito come "raro" la metodologia di approccio deve sistematicamente essere basata su strumenti di indagine non consueti, nel caso specifico ci si deve basare su una serie di "indizi" che vanno a costituire un cosiddetto "Pattern" probabilistico di riferimento in grado di definire un vettore di variabili che, come una cartina di tornasole, è in grado di identificare con elevata probabilità, un dato che può essere inserito tra quelli idonei a fare Prevenzione. Per la natura stessa dello studio delle cause scatenanti degli eventi infortunistici e tecnopatici ci si è trovati molto spesso a dover fare i conti con una serie di dati che hanno prodotto soltanto "rumore" e quindi non in grado di aggiungere conoscenza a quella preesistente. Tante informazioni producono tanto rumore ma poche informazioni non riescono a consentire un approccio consistente per cui, quando, come in questo caso, lo scenario di riferimento non è uno*

*scenario comodo, occorre settare una serie di parametri in grado di intercettare il maggior numero di informazioni, tenendo ben presente che così facendo il grado di affidabilità decresce. L'ottimo sta nella massimizzazione dei dati e la minimizzazione dell'errore, cose queste che non vanno affatto d'accordo tra di loro. Gli eventi che si sono voluti analizzare sono eventi rari, complessi e intersecati tra di loro, ma non impossibili da analizzare, soprattutto se si guarda agli strumenti informatici che sono oggi in grado di supportare idee e strategie di approccio. Il passo successivo è quello di estrapolare le informazioni per rendere spendibili i dati ricavabili dallo studio, costruire modelli di riferimento e porre l'attenzione soprattutto alle piccole imprese.*

*Nel nostro caso, i dati e le informazioni necessarie a una valutazione di efficacia vanno ricercati all'interno dei documenti che le imprese che hanno richiesto il finanziamento hanno prodotto ed inviato all'Inail e che, almeno fino al 2018, erano informaticamente poco intellegibili e spesso addirittura in forma cartacea. Un lavoro minuzioso di consultazione, ricerca ed estrapolazione di dati destrutturati comporta tempi talmente lunghi che spesso il risultato è anacronistico. Viceversa un'analisi veloce permette la ricerca di dati numericamente insufficienti per fornire un risultato attendibile.*

*Fortunatamente lo sviluppo delle tecnologie ha permesso di affrontare efficacemente tale situazione, permettendo di acquisire una notevole mole di dati in tempi ragionevoli.*

*Le tecnologie di cognitive computing, grazie alla decodifica dei documenti e all'analisi di grandi quantità di dati non strutturati, portano all'apprendimento autonomo (machine learning) e quindi alla possibilità per le macchine di ragionare, comprendere, elaborare e usare il linguaggio naturale dell'uomo. Macchine e strumenti in grado di restituire informazioni utili perché in grado di indagare e auto-apprendere imparando dalle informazioni impartite.*

*Tutto ciò ha permesso quindi di acquisire informazioni relative ai progetti ammessi a finanziamento per diversi anni, andando ad analizzare ad esempio, per alcune tipologie di intervento, le diverse tipologie di materiali e/o macchine da sostituire e/o acquisire e le caratteristiche prestazionali delle stesse in termini di impatto sulla salute e sicurezza dei lavoratori.*

*L'informazione ricavabile che consente di verificare ex-post i risultati degli interventi effettuati non è la riduzione del numero di infortuni e malattie professionali evitate, ma la riduzione effettiva del rischio specifico o i quantitativi di sostanze, attrezzature, modalità operative pericolose rimosse dai luoghi di lavoro.*

*Ovviamente, i dati ricavati dalle richieste presentate dalle imprese per accedere agli incentivi, fanno anche meglio comprendere il mondo lavorativo dei partecipanti e strutturare il bando successivo in maniera più aderente ai rischi sui luoghi e nei processi di lavoro e migliorare le procedure di acquisizione della documentazione da parte dell'Istituto.*

*Lo studio in questione si focalizza sul periodo 2014-2018 e va a verificare quanti materiali contenenti amianto sia stato possibile rimuovere dai luoghi di lavoro e l'entità di riduzione del rischio vibrazioni attraverso la sostituzione di macchine obsolete con altre più moderne e innovative.*

*La scelta di questi due ambiti di interventi è stata fatta in quanto essi costituiscono campioni significativi di intervento; infatti, sono stati richiesti da molte imprese, comprendono una quota significativa dei progetti sottoposti a valutazione e finanziamento, restituiscono informazioni sia su grandezze materiali (m<sup>2</sup> materiali contenenti amianto) sia su grandezze fisiche (vibrazioni*

meccaniche  $m/s^2$ ) e la loro analisi può quindi fungere da base per tipologie di intervento con strutture, parametri e grandezze simili presenti negli Avvisi..

In genere, come si vedrà nel testo, lo studio effettuato conferma la bontà degli assunti di progetto degli avvisi pubblici e dimostra la rimozione di importanti volumi di materiali contenenti amianto e la riduzione significativa delle vibrazioni emesse dai macchinari introdotti ex novo nei cicli produttivi.

Tuttavia, nonostante il ricorso all'intelligenza artificiale, questo lavoro è estremamente impegnativo e oneroso. Pertanto, per consentire una analisi sempre più puntuale e più efficiente degli effetti prevenzionali degli interventi richiesti si è man mano avviato un processo di informatizzazione per l'utenza, introducendo una modalità di inserimento dei dati di richiesta da parte delle aziende e dei tecnici incaricati di progettare gli interventi oggetto di richiesta di finanziamento, di tipo guidato e in grado di restituire dati e informazioni direttamente in forma strutturata.

Ciò renderà più semplice e veloce estrarre i dati di riscontro e operare rendicontazioni sull'efficacia degli avvisi pubblici con maggiore sistematicità e frequenza, in modo da tracciare nel tempo la riduzione dei livelli di rischio.

Nel frattempo è importante evidenziare i dati ed i risultati di questo studio che ha un valore doppio che riguarda i suoi risultati e la modalità con cui sono stati ottenuti, interrogando documenti destrutturati con gli strumenti propri dell'intelligenza artificiale, addestrata dalla mano e dalla mente degli esperti del settore. Questo studio è uno sguardo sul passato che ci dà indicazioni per il futuro.

**Ing. Ester Rotoli**

Direttore centrale  
prevenzione

**Dr. Stefano Tomasini**

Direttore centrale per  
l'organizzazione digitale

**Dr. Fabrizio Benedetti**

Coordinatore generale della  
Consulenza accertamento  
del rischio e prevenzione

**Dr.ssa Silvia D'Amario**

Coordinatore generale  
della Consulenza  
statistico attuariale

## 1. INTRODUZIONE

Gli Avvisi pubblici a favore delle imprese che investono in sicurezza (nel seguito denominati anche Bandi ISI) costituiscono una interessante opportunità per avviare azioni di prevenzione volte al miglioramento delle condizioni di salute e sicurezza nei luoghi di lavoro.

I Bandi ISI consentono di raccogliere grandi quantità di informazioni sulle imprese che annualmente accedono ai contributi a scopo prevenzionale messi a disposizione dall'Istituto e sui progetti ammessi a finanziamento; tuttavia esiste una gran quantità di dati che non sono immediatamente evincibili da una lettura standardizzata dei campi presenti nella procedura informatica ISI Back Office utilizzata per l'istruttoria dei progetti.

Ogni domanda di finanziamento presentata dalle imprese collocatesi in posizione utile nell'elenco cronologico a seguito del click day, corredata dalla relativa documentazione, è sottoposta a una duplice verifica di congruità: tecnica e amministrativa. Tale duplice verifica viene a sua volta effettuata in due momenti successivi: nella fase di verifica, quando l'impresa richiede il finanziamento, e nella fase di rendicontazione, quando il progetto è stato ultimato e l'impresa deve dimostrarne la realizzazione conforme rispetto a quanto approvato dall'Istituto. In entrambi i momenti viene effettuata un'attenta valutazione della corrispondenza tra quanto richiesto e i requisiti fissati dal bando e dagli allegati tecnici di riferimento. Le verifiche tecniche assicurano il miglioramento delle condizioni di salute e sicurezza legato alla realizzazione del progetto.

Il progetto ha riguardato la verifica della possibilità di effettuare un approccio innovativo allo studio dei progetti ISI, codificando informazioni e dati non codificati e non standardizzati, contenuti nella grande mole di documentazione cartacea prodotta dalle aziende nelle diverse fasi di finanziamento del progetto.

L'estrazione di tali dati, definiti "non strutturati", è stato effettuato con applicativi complessi, basati sull'intelligenza artificiale, ma alimentati e guidati con istruzioni semplici quali quelle riconducibili al linguaggio naturale.

La conoscenza delle informazioni che ne deriva diventa di fondamentale ausilio sia per la messa a punto dei criteri tecnici da implementare nei bandi successivi, sia per l'analisi dell'efficacia prevenzionale dei progetti finanziati.

## 2. ARCHITETTURA DEL DATO DI RIFERIMENTO

L'esperienza maturata nel corso della gestione dei Bandi ISI ha portato a modulare e specializzare sempre di più i parametri di accesso ai finanziamenti, consentendo alle imprese di orientarsi verso interventi sempre più efficaci nel ridurre i rischi infortunistici e tecnopatici.

In particolare, il dimensionamento dei parametri per la partecipazione agli Avvisi e l'orientamento degli stessi in relazione ai rischi cui sono esposti i lavoratori è avvenuto attraverso la simulazione dei possibili scenari e sulla base dei progetti presentati dai partecipanti negli anni precedenti

Sotto il profilo tecnico, il processo di progettazione dei Bandi ISI e in particolare delle singole tipologie di intervento previste dagli allegati, è stato comunque orientato a coniugare il fine prevenzionale dell'Istituto con le esigenze del mondo produttivo.

Le componenti tecniche del gruppo di lavoro deputato alla strutturazione dei Bandi ISI monitorano periodicamente l'andamento dei bandi stessi, analizzando e ridefinendo la distribuzione territoriale dei fondi, le tipologie di imprese beneficiarie sia in termini dimensionali che di settori produttivi e la rischiosità delle lavorazioni ad esso connesse. Le valutazioni riguardano anche lo studio delle tipologie di progetto concretamente realizzate dalle imprese e il miglioramento apportato alle condizioni di rischio aziendali.

La documentazione che le imprese allegano alle domande di finanziamento è spesso voluminosa, articolata e, in passato, anche in buona parte cartacea; da essa è tuttavia possibile estrarre e strutturare una serie di informazioni. Questo lavoro descrive un primo possibile accesso a tutte queste informazioni e dimostra che utilizzando applicativi adeguati è possibile arricchire i database esistenti con ulteriori dati e informazioni preziose.

L'estrazione puntuale di queste informazioni, effettuata manualmente progetto per progetto, richiederebbe ingenti quantità di tempo e di risorse umane. In relazione a ciò, e parallelamente al processo di finanziamento, è stato sviluppato da Inail, attraverso un gruppo di lavoro formato da esperti Contarp, Csa e Dcod, in collaborazione con IBM, un progetto che ha l'obiettivo di acquisire e analizzare grandi quantità di dati che, collegati ai Bandi ISI, non sono immediatamente evincibili dalla lettura dei campi, compreso quelli strutturati e solo teoricamente già disponibili, nelle procedure online normalmente utilizzate.

Non analisi manuale, quindi, ma analisi effettuata con l'ausilio di strumenti informatici che consentano di estrarre conoscenza dai documenti, affiancando, alla lettura del dato strutturato presente nella procedura informatica ISI Back Office, un'analisi del dato di livello superiore che prescinde dagli strumenti cosiddetti tradizionali e strizza l'occhio a tutta una serie di tecnologie d'avanguardia in grado di trovare legami nascosti tra dati apparentemente diversi, di imprese che a prima vista

differiscono tra loro sotto numerosi aspetti (economico, dimensionale, di settore produttivo, di rischio, ecc.). Per fare tutto ciò, il progetto avviato si è avvalso di strumenti software ad hoc in grado di evidenziare legami non evidenti tra i progetti, desunti da dati "non ovvi" presenti nei documenti inviati dalle imprese a corredo delle istanze di finanziamento.

### 3. APPROCCIO E TECNOLOGIE UTILIZZATE

Il metodo migliore per poter monitorare le informazioni nascoste contenute nella documentazione inviata all'Inail dalle imprese è quello di estrarre conoscenza dal loro contenuto testuale. Esistono numerose tecniche e sistemi di estrazione della conoscenza dai testi, che sono in grado di rappresentarne il contenuto in vari modi e che vanno dal modello statistico relativo all'uso dei termini più significativi in esso presenti (reti bayesiane, sistemi di *text mining* e *data mining*), alla descrizione in forma simbolica mediante una serie di proposizioni che ne riassumono il contenuto (*parser* sintattici, sistemi di *knowledge mining*).

Gli attuali metodi statistici e strumenti di *data mining* estraggono pattern ricorrenti da dati e informazioni considerando solo i loro attributi e non la conoscenza di dominio. Tali sistemi non sono, quindi, in grado di assegnare nuovi significati al testo relativo a parole o frasi, perché non sono direttamente ricavabili dal suo contenuto, né riescono a richiamare parte della conoscenza assimilata in precedenza. I recenti strumenti che rientrano nel campo del *knowledge mining* sono invece basati su modelli computazionali che acquisiscono nuova conoscenza a partire sia dai fatti che dalla conoscenza di dominio acquisita precedentemente. Questi strumenti comprendono una base di dati, una base di conoscenza, strumenti per l'analisi di dati e per l'apprendimento induttivo di nuove regole.

Quindi gli strumenti di *knowledge mining* sono particolarmente adatti e promettenti in situazioni in cui la conoscenza è frutto dell'indivisibile insieme di dati strutturati e non strutturati e richiede un modello di interpretazione dello specifico contesto aziendale. Questo tipo di tecnologie consente di dedurre il modello da un limitato insieme di documenti e dati e di applicarlo in modo induttivo all'intero universo di dati, con successivi e continui affinamenti del modello stesso basato su tecniche di *machine learning*.

Il metodo utilizzato per affrontare un progetto di questo tipo, basato su metodiche di estrazione della conoscenza, nel contesto dei Bandi ISI è stato caratterizzato dalle seguenti fasi:

- definizione delle variabili di analisi descrittiva del fenomeno e degli obiettivi di predizione sulla base dell'analisi storica dei dati;

- individuazione delle fonti e mappatura dei contenuti destrutturati che servivano a integrare i dati strutturati;
- *pre-processing* (pulizia dei dati, analisi esplorativa, selezione, trasformazione, formattazione);
- estrazione/acquisizione dei dati (a integrazione dei dati strutturati preesistenti) con tecniche di intelligenza artificiale sia puntuali sul singolo documento sia a livello di concetti sull'intero corpus di documenti;
- esplorazione dei dati con utilizzo di molteplici tecniche di analisi e visualizzazione dei dati;
- interpretazione/valutazione dei risultati.

Lo schema sopra delineato non corrisponde in realtà a una successione lineare, ma richiede un'iterazione a seguito dell'arricchimento, nelle fasi, di correttivi e correlazioni e della pulizia dei dati da parametri che possono confondere il modello (rumore).

Le tecnologie utilizzate per supportare questo metodo sono state:

- riconoscimento intelligente a livello di documento, in grado di riconoscere al suo interno particolari formati, come tabelle, grafici o dati specifici, ed estrarne i dati;
- modelli statistici di riconoscimento dei concetti, entità e relazioni su un set di apprendimento, denominato ground truth, da esportare su tutto il dominio;
- strumenti di esplorazione e filtro delle informazioni in base alle entità ed i dati strutturati modellati;
- strumenti di analisi per la pulizia e correzione dei dati del modello;
- strumenti di rappresentazione descrittiva e grafica dei dati per analisi del fenomeno;
- strumenti per la costruzione di modelli machine learning di interpretazione dati per clusterizzazione ed analisi di correlazione dati.

Per supportare e velocizzare i cicli di iterazione si è tenuto conto di un disegno integrato di una piattaforma di analisi, in modo da facilitarne l'utilizzo autonomo da parte dell'analista dati esperto di dominio.

Vengono di seguito specificate le tecnologie e gli strumenti utilizzati nel progetto, corredate da una breve descrizione del loro funzionamento.

### 3.1 DATACAP

*Datacap* è una soluzione completa per l'acquisizione di dati e documenti che permette di scansionare, classificare, riconoscere, validare, verificare ed esportare i dati e i documenti.

*Datacap* combina diversi motori di riconoscimento: *Optical Character Recognition* (OCR), *Intelligent Character Recognition* (ICR) e *Optical Mark Recognition* (OMR), permettendo di catturare dati a partire sia da documenti strutturati (variabili a piacere), che da documenti non strutturati. Inoltre, l'utilizzo di OCR, ICR e OMR permette di trattare testi sia dattiloscritti che scritti a mano, contenenti diversi tipi di codici a barre e/o caselle di tipo check box.

### 3.2 WATSON

Watson è una linea di prodotti AI per il business a marchio IBM, che includono un insieme diversificato di funzionalità quali *text mining*, *reasoning*, *decision support system* (DSS), *visual recognition* e *virtual assistant*.

Watson Studio (WS) è una piattaforma che unisce ambiente di sviluppo e ambiente di produzione e raccoglie allo stesso tempo tutti gli strumenti necessari per un set di progetti complessi di *data science* in cloud, per cui è richiesta l'integrazione di diverse fonti dati.

All'interno di questo servizio è possibile scegliere tra molteplici tipi di asset per l'analisi e la visualizzazione delle informazioni, per la pulizia e la manipolazione, per lo storage e per la creazione di modelli di *machine learning* (ML). Tra le risorse disponibili su WS sono presenti: Data Refinery Tool, Jupyter Notebook, ML e *deep learning model* (DLM), *classification tool* e *dashboard* interattive per la presentazione dei dati.

Tali strumenti consentono di acquisire nuova conoscenza a partire sia dai dati in essere (strutturati e non strutturati) che dalla conoscenza di dominio acquisita precedentemente.

Questo tipo di tecnologie consente di dedurre il modello da un limitato insieme di documenti e dati (un campione) e di applicarlo in modo induttivo all'intero universo di dati, con successivi e continui affinamenti del modello stesso basato su tecniche di *machine learning*.

Nella figura sottostante sono rappresentate le tecnologie e il flusso di informazioni utilizzato nel presente progetto.

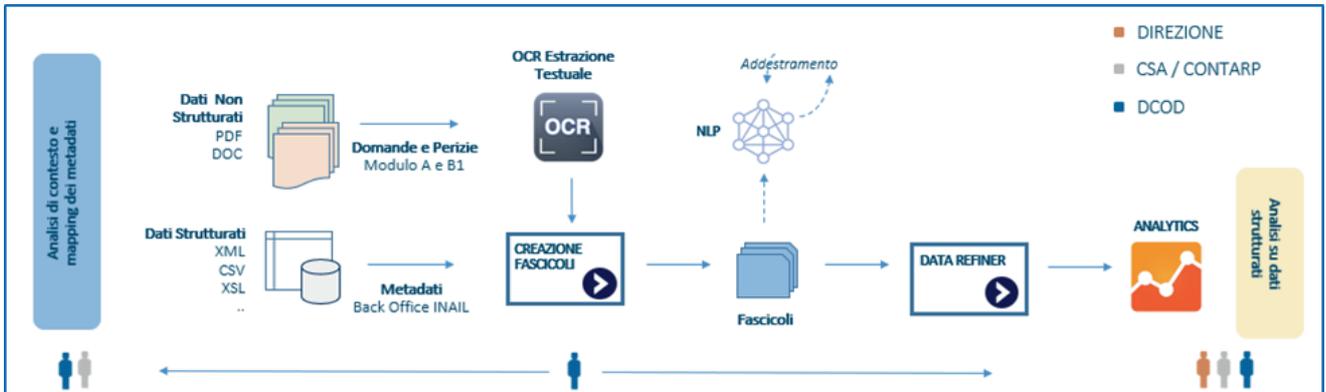


Figura 1 - Tecnologie e flusso operativo

Lo schema applicato all'analisi dei singoli progetti è riassunto nella figura seguente.

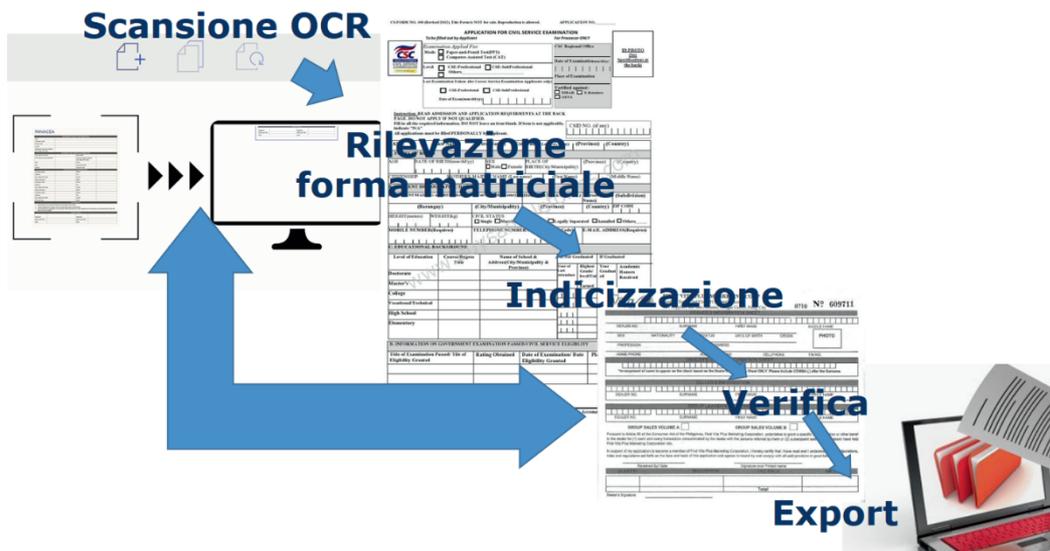




Figura 2 - Schema di analisi della singola pratica

Il processo inizia con la scansione della singola pratica cartacea.

Attraverso gli algoritmi collaudati si prova a intercettare la frase esatta portatrice dell'informazione non strutturata nascosta; se non la si trova si passa a un altro approccio con l'algoritmo successivo. Se anche in questo caso non si riesce a intercettare l'informazione, si divide il documento in una serie di matrici nelle quali la singola suddivisione viene analizzata alla ricerca di informazioni utili alla decodifica. Una volta intercettata l'informazione, si passa alla cosiddetta verifica in cui, confrontando il dato appena acquisito con le ontologie contenute nella libreria di controllo, si cerca di autenticare il dato nuovo che è stato appena acquisito (ad esempio: il dato è stato preso due volte).

Il meccanismo viene reiterato su tutte le pratiche cartacee sottoposte a analisi.

#### 4. BANDI ISI

I modelli sopra menzionati sono stati applicati in fase sperimentale ai progetti presentati per il Bando ISI 2014; in particolare è stato chiesto al motore dell'applicativo di intercettare tutte le variabili legate agli interventi/progetti portati a termine per la rimozione e sostituzione del materiale contenente amianto nei luoghi di lavoro. Successivamente, conclusasi positivamente la fase sperimentale, tali modelli sono stati applicati su tutti i progetti di rimozione amianto e ai progetti volti alla riduzione del rischio da vibrazioni per i bandi dal 2014 al 2017.

Di seguito si illustrano alcune informazioni ottenute con tale tecnologia, rilevanti ai fini della presente disamina nei due ambiti di rischio:

- rischio legato alla bonifica da materiali contenenti amianto;
- rischio legato a livelli di vibrazioni emessi da macchine, dannosi per l'uomo.

La scelta ha riguardato questi due ambiti sia perché facevano parte dei progetti maggiormente richiesti negli anni in esame, sia perché costituiscono un duplice approccio; in considerazione dei parametri tecnici previsti, in un caso grandezze fisiche da indagare, nell'altro grandezze qualitative. In tal modo possiamo essere in grado di porre le basi per costruire il percorso di analisi a tutte le altre tipologie di intervento presenti negli Avvisi pubblici.

#### 4.1 PROGETTI DI BONIFICA AMIANTO

L'analisi che ha riguardato i progetti riferiti all'amianto ha voluto intercettare, nell'ambito di ciascun progetto presentato, il numero di metri quadri di materiali contenenti amianto rimossi; in una fase immediatamente successiva si è voluta intercettare l'informazione circa la tipologia di sostituzione/rimozione e il tipo di struttura sostituita.

Nel Bando ISI 2014 la seconda tipologia di intervento prevenzionale più richiesta dalle imprese è risultata quella relativa alla bonifica da amianto (4.000 progetti su quasi 23mila totali); pertanto, a partire dal bando 2015 è stato previsto un asse dedicato a questa specifica tipologia di intervento.

Da allora, il 27% del budget complessivo messo a disposizione dall'Istituto nel quadriennio che abbraccia i Bandi ISI 2015-2018 è stato destinato a questa finalità di intervento; il totale dei progetti specifici presentati ha coperto mediamente il 19% di tutte le richieste.

Nel grafico che segue è rappresentato il dettaglio dei singoli bandi del quadriennio.

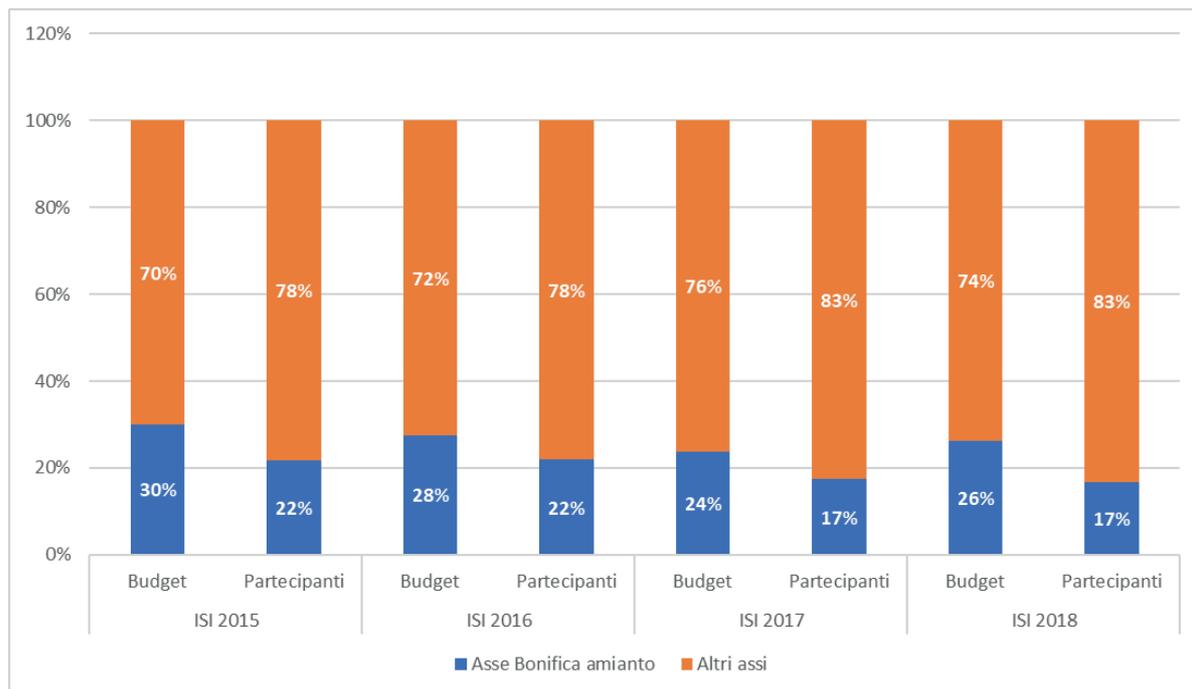


Figura 3 - Incidenza dei progetti di bonifica da materiali contenenti amianto

È opportuno precisare che i progetti di bonifica dall'amianto afferiscono a una delle molteplici tipologie di progetti ammessi a finanziamento e che gli stessi presentano importi mediamente più elevati rispetto alle altre tipologie.

Nelle quattro edizioni dei Bandi analizzate, l'Inail ha quindi erogato quasi 162 milioni per la realizzazione di progetti atti a contrastare il rischio amianto attraverso la sua rimozione dai siti produttivi (bonifica).

I documenti su cui si è concretamente attuata l'indagine con gli strumenti esposti nel capitolo precedente sono il Modulo A di Domanda (presentato dall'impresa) e il Modulo B di Perizia Giurata (elaborata dal perito/professionista che ha elaborato il progetto); i dati estratti da questi testi "non strutturati" sono stati poi integrati con quelli "strutturati" provenienti dagli archivi Inail e forniti come dati contenuti nella procedura ISI Back Office che viene utilizzata dal personale tecnico e amministrativo dell'Inail per le attività istruttorie.

Nella figura che segue è riportato un esempio dei testi contenuti nei documenti analizzati, con l'evidenza delle parti di testo su cui ci si è concentrati.

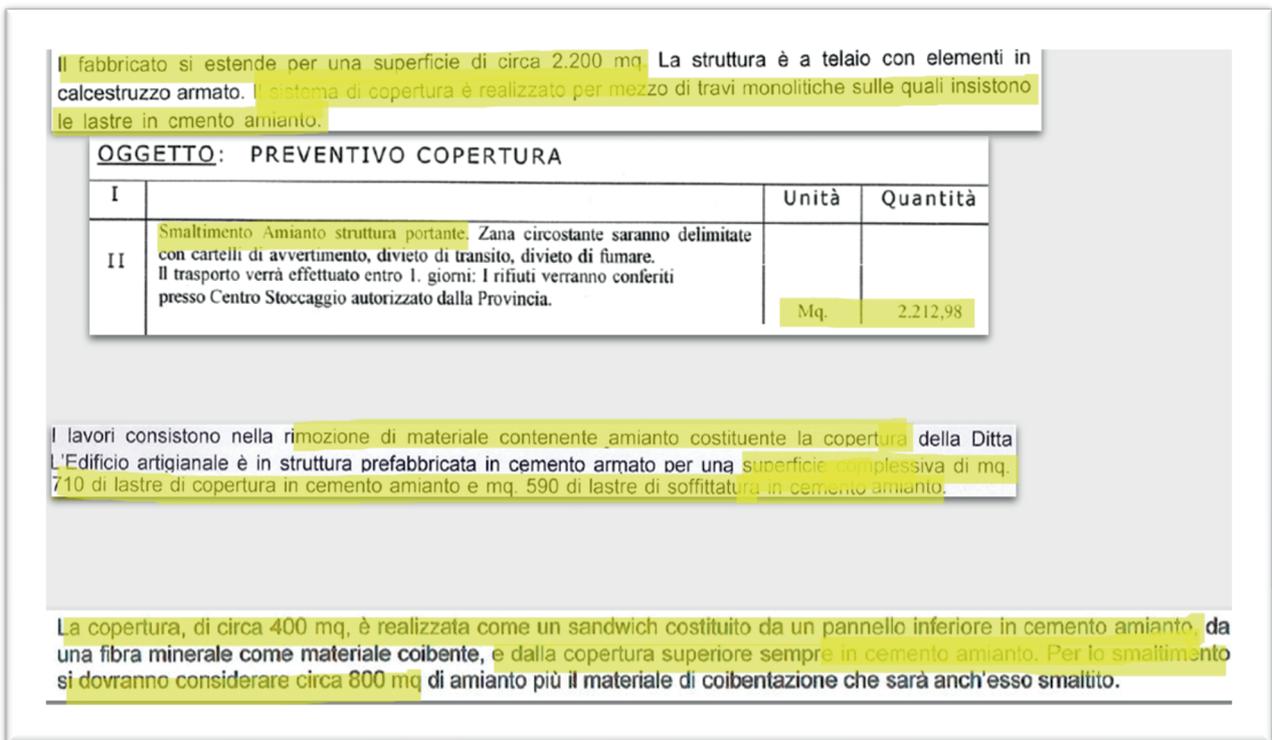


Figura 4 - Dati non strutturati tratti da perizie relative a progetti di bonifica da amianto

I dati estratti da ciascun progetto sono stati quindi raccolti in un file (dati) CSV. Il file ha costituito il primo prodotto dell'analisi e la base principale di informazioni cui attingere per ulteriori indagini, analisi e rappresentazioni (*dashboard*).

Nel caso dell'amianto, i principali dati estratti dalla perizia di ciascun progetto sono principalmente costituiti da:

- struttura sostituita (controsoppito, copertura, sottocopertura, etc.);
- metri quadrati di amianto asportato per ciascuna struttura;
- tipologia di intervento (rimozione, sostituzione).

Per ottenere le strutture e la tipologia di informazioni si è proceduto utilizzando dei dizionari che contenessero le diverse espressioni di questi elementi e cercando tutte le occorrenze/conessioni; queste informazioni comparivano vicino a elementi del testo che indicavano la presenza di amianto (anche per l'amianto è stato generato un dizionario ad hoc). Lo stesso ragionamento è stato utilizzato per i metri quadrati di amianto oggetto dell'intervento, caratterizzati da un dato di tipo numerico (individuato con una *character rule*) e da un elemento di tipo unità di misura di superficie (anch'essa individuata nel testo grazie a dizionari), entrambi associati a un'espressione di tipo "amianto".

Negli anni le possibili espressioni dei singoli elementi sono cambiate, così come la struttura del documento che le conteneva; tale strutturazione documentale è andata mano a mano verso una forma più standardizzata in termini di rappresentazione dei dati di interesse. Nel 2014 i documenti (perizia, ecc.) avevano generalmente strutture diverse, anche per la stessa tipologia di intervento; la forma del documento infatti era fortemente dipendente dalle decisioni personali del singolo perito, portando quindi a una grande variabilità nell'esposizione delle informazioni. Ciò ha richiesto una particolare attenzione nel momento della ricerca delle informazioni nei testi, ad esempio il dato dei metri quadrati di amianto sostituiti; questi talvolta venivano riportati in forme semplici, talvolta aggregate, talvolta venivano duplicati in ragione delle superfici di materiale da rimuovere e di quelle riferite al materiale sostitutivo. In questi casi si è reso necessario addestrare il sistema con diverse regole di decodifica, come associazioni e/o distinzioni di superfici, tolleranze del loro valore, somme o scarto degli elementi consecutivi separati dal simbolo di addizione, ecc.

L'utilizzo di strumenti che ricorrono all'intelligenza artificiale ha permesso da un lato l'estrazione dei dati non strutturati precedentemente descritti, dall'altro ha reso possibile porre le basi per agevolare in futuro la fase di verifica tecnico amministrativa delle domande di finanziamento. È stato possibile infatti estrarre altre informazioni presenti e in varia forma nei documenti a corredo dei progetti, come ad esempio la congruità economica del progetto, le caratteristiche dell'intervento (es. rimozione, incapsulamento), ecc.

L'analisi cognitiva è stata condotta sulle perizie giurate delle domande che hanno superato la verifica tecnico amministrativa, vale a dire sulla documentazione relativa a 4.046 progetti ammessi potenzialmente a finanziamento; purtroppo, nel 7% dei casi (286 progetti) non è stato possibile recuperare in modo "automatizzato" l'informazione relativa ai metri quadrati interessati dal progetto di bonifica a causa della bassa qualità della scansione dei documenti analizzati. C'è peraltro da osservare che per oltre la metà di queste pratiche scartate dalla decodifica informatica neanche un operatore umano sarebbe riuscito a desumere la quantità di metri quadrati di amianto rimossi a causa della scarsissima qualità della fotocopia. In questi casi i funzionari dell'Istituto incaricati dell'istruttoria hanno chiesto chiarimenti all'impresa, ma la relativa documentazione inviata dalle imprese esulava dal progetto in esame.

Di seguito si riportano graficamente i dati che, in tal modo, è stato possibile estrarre.

In figura 5 sono riportati: l'importo finanziato, il totale dei metri quadrati rimossi, il numero di addetti che hanno beneficiato dell'intervento e la distribuzione dei metri quadrati di amianto rimossi a livello regionale nei 4 bandi oggetto di indagine. Le tonalità di blu delle Regioni sono state rappresentate in colore più scuro in funzione dei quantitativi crescenti di amianto rimosso. Le superfici interessate, ovviamente, risentono in buona parte delle somme messe a disposizione dai Bandi ISI per le varie Regioni, che sono direttamente correlate al numero di aziende in esse presenti.

# 204 Mln    6,59 Mln    63 mila

importo finanziabile

mq amianto rimossi

numero addetti



Figura 5 - Distribuzione regionale dei metri quadrati di amianto rimossi

La figura 6 riporta l'andamento delle richieste negli ultimi quattro anni nelle quattro regioni italiane che maggiormente hanno usufruito di tale intervento.

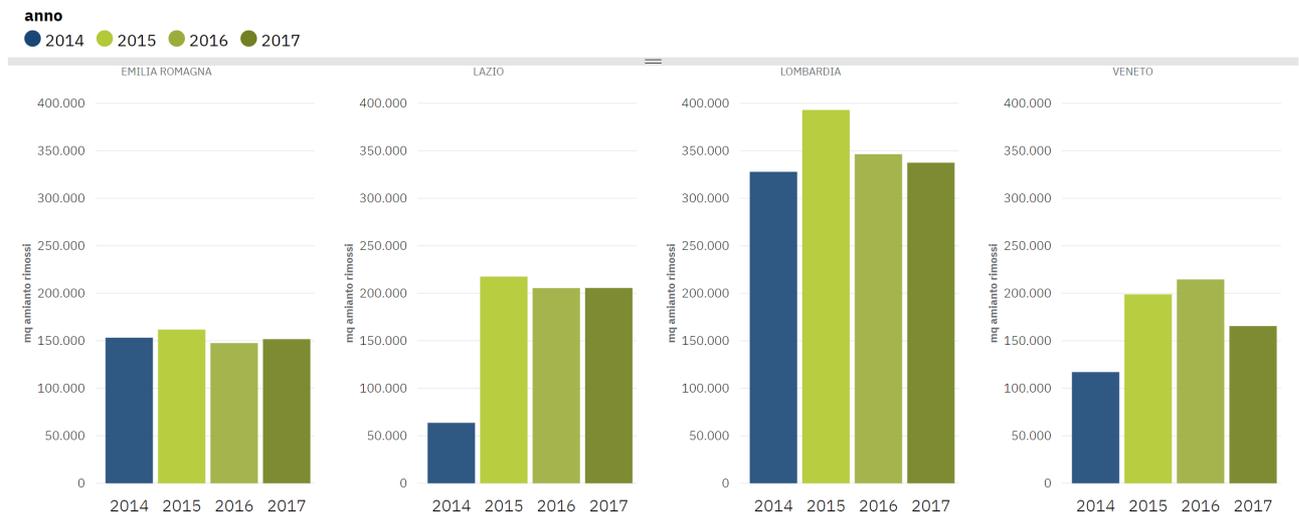


Figura 6 - Andamento dei metri quadrati rimossi nelle prime 4 Regioni dal 2014 al 2017

In figura 7 è rappresentato invece l'andamento nel tempo, per tutte le Regioni, degli importi medi richiesti in fase di domanda, comprensivi delle operazioni di bonifica e di rifacimento.

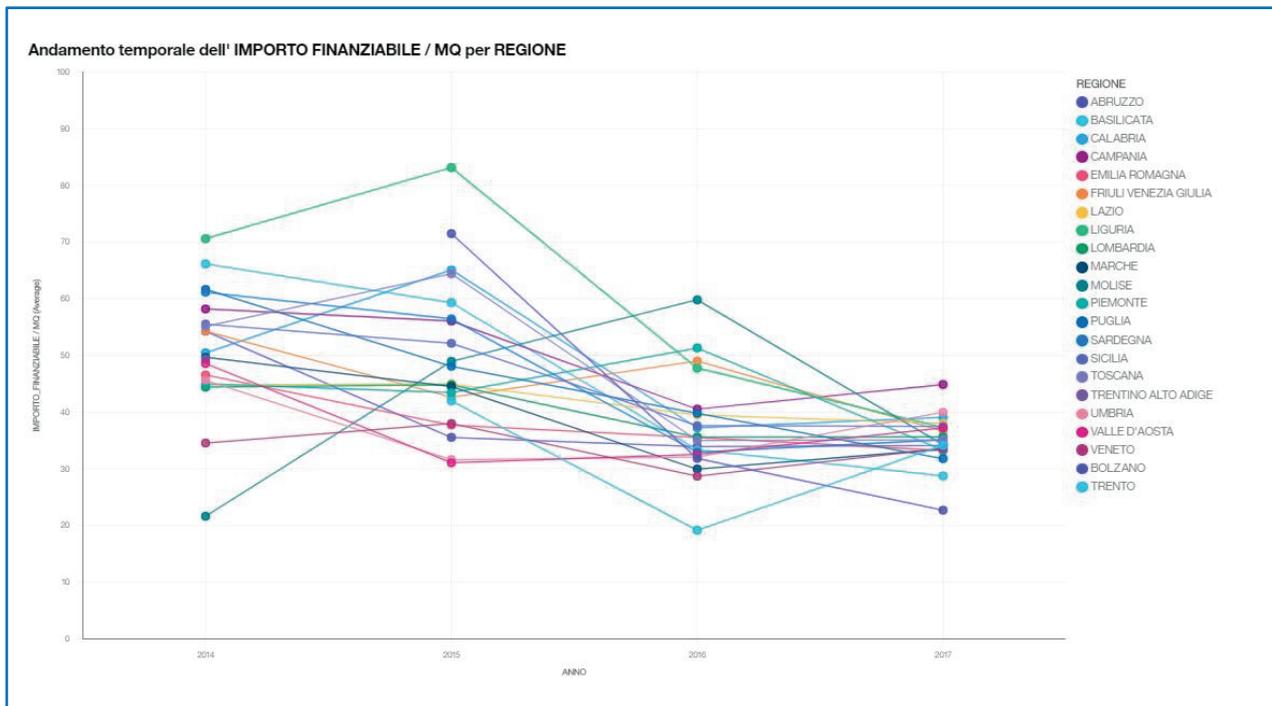


Figura 7 - Andamento temporale importo al mq per regione 2014-2017

Quasi tutte le Regioni si attestano nell'intorno di uno stesso costo, con andamento mediamente decrescente negli anni. Nel Bando ISI 2016 è stato introdotto un limite alla spesa finanziabile per la nuova copertura. La linea turchese in alto, corrispondente alla Liguria, si discosta di molto dalle altre Regioni, in particolare per i valori del 2015. Per tale motivo, in figura 8 è stata selezionata la sola Liguria, con l'evidenza di tutti i progetti di questa Regione suddivise negli anni, in modo da avere evidenza di quelli per i quali si è riscontrato un valore particolarmente alto rispetto alla media del costo al metro quadrato.

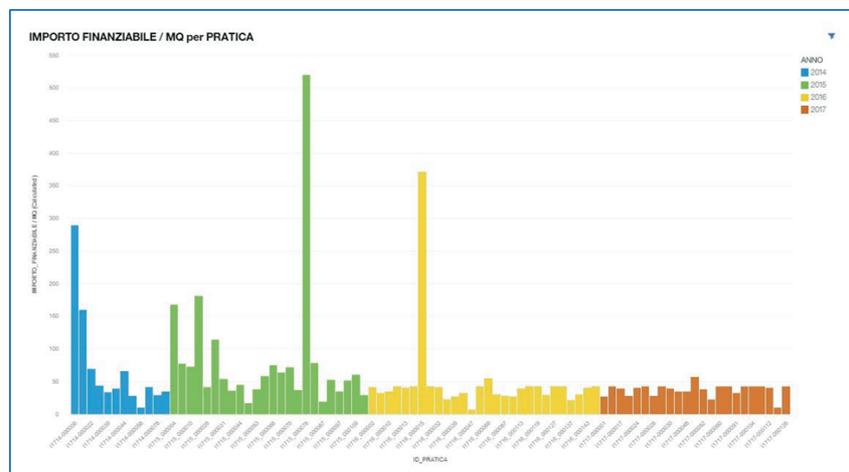


Figura 8 - Andamento temporale del costo medio di bonifica per mq per progetto, in Liguria, dal 2014 al 2017

Questa particolare rappresentazione permette immediatamente di risalire al numero della pratica e di andare ad analizzare il contenuto del documento corrispondente, in modo da poter rintracciare eventuali anomalie e cause che ne hanno determinato la particolarità (estensioni particolarmente ridotte, difficoltà tecniche di realizzazione, accesso difficoltoso alla struttura, necessità di apprestamenti di cantiere onerosi, ecc.).

Un'analisi simile (*scatter*) è ottenibile per le stesse motivazioni e scopi in figura 9.

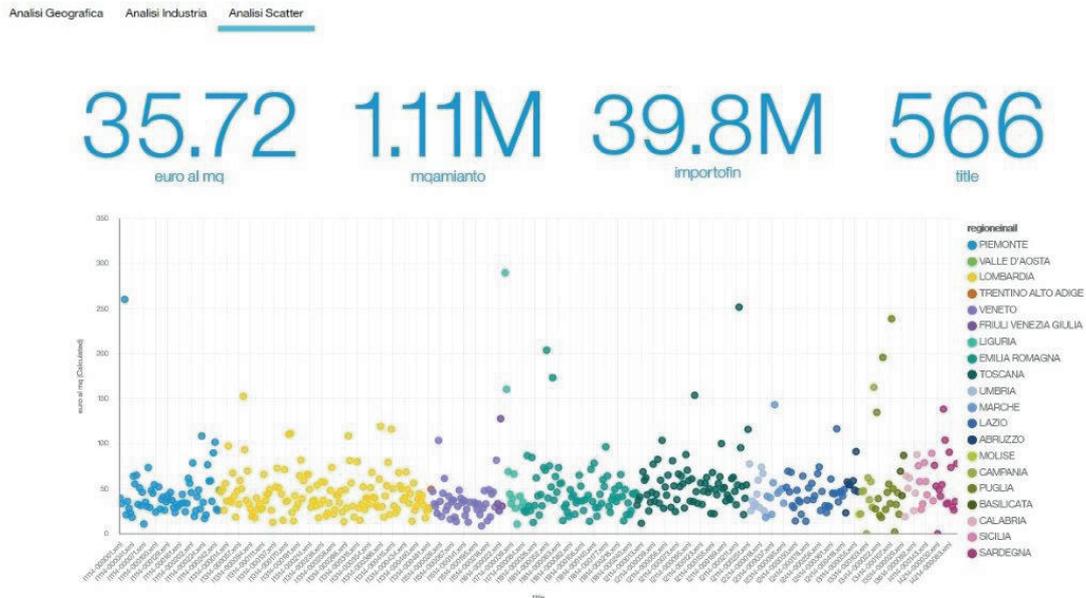


Figura 9 - Costo medio di bonifica per mq per progetto a livello regionale nel 2014

In figura 10 è riportato invece l'andamento dei metri quadrati rimossi in funzione del settore produttivo per il Bando ISI 2014. La figura fornisce un'indicazione immediata dei settori che richiedono e beneficiano maggiormente di tale tipologia di intervento; in questo caso si evidenzia che, per l'anno in questione, l'interesse è stato manifestato soprattutto dalle attività manifatturiere e del commercio. Tale tipologia di informazione può risultare particolarmente rilevante ai fini di una corretta progettazione dei bandi futuri, in modo tale da poter orientare diversamente i finanziamenti, andando ad incidere sui settori particolarmente rischiosi che ancora non hanno fatto ricorso a tale finanziamento.

48,2 Mln

IMPORTO\_FINANZIABILE

1,56 Mln

TOT\_MQ\_RIMOSI

15,9 mila

NR\_ADDETTI

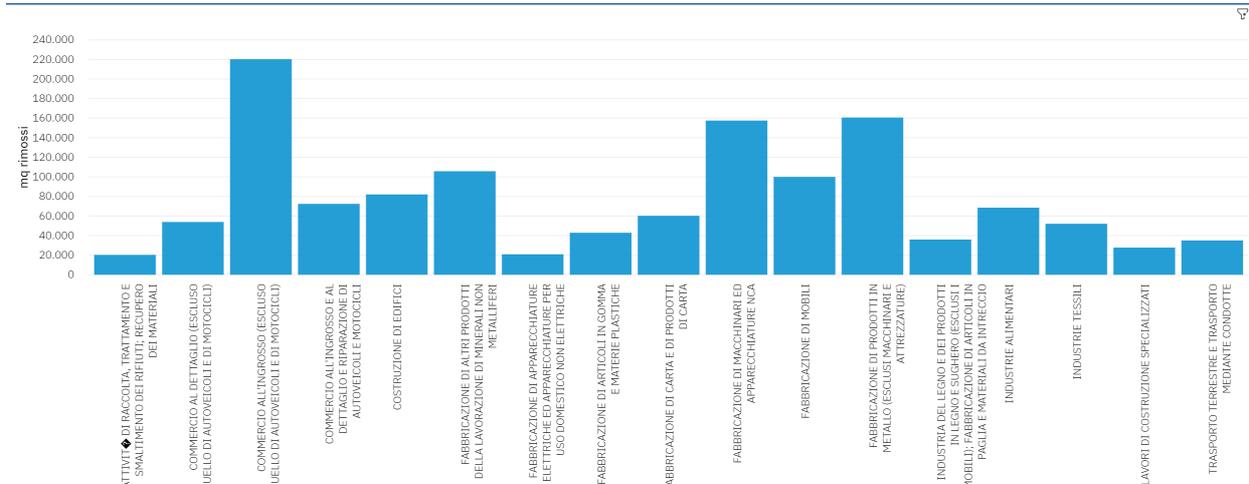


Figura 10 - Andamento per settore produttivo nel 2014

Infine, è stato possibile localizzare gli ambiti territoriali e le perizie giurate a corredo dei progetti. In figura 11 sono riportati i riferimenti e il numero di pratiche (colori più scuri corrispondono a un maggior numero di pratiche) suddivisi per Regione. Con questo tipo di dato si ha un'immediata evidenza delle Regioni, il tasso di richiesta e quello dei vari progetti finanziabili. In tal modo è possibile ad esempio individuare le Regioni nelle quali sarebbe necessaria un'attività di divulgazione e informazione mirata al fine di migliorare il livello di conoscenza della tipologia di intervento, del rischio che si vuole aggredire e dello strumento di finanziamento.

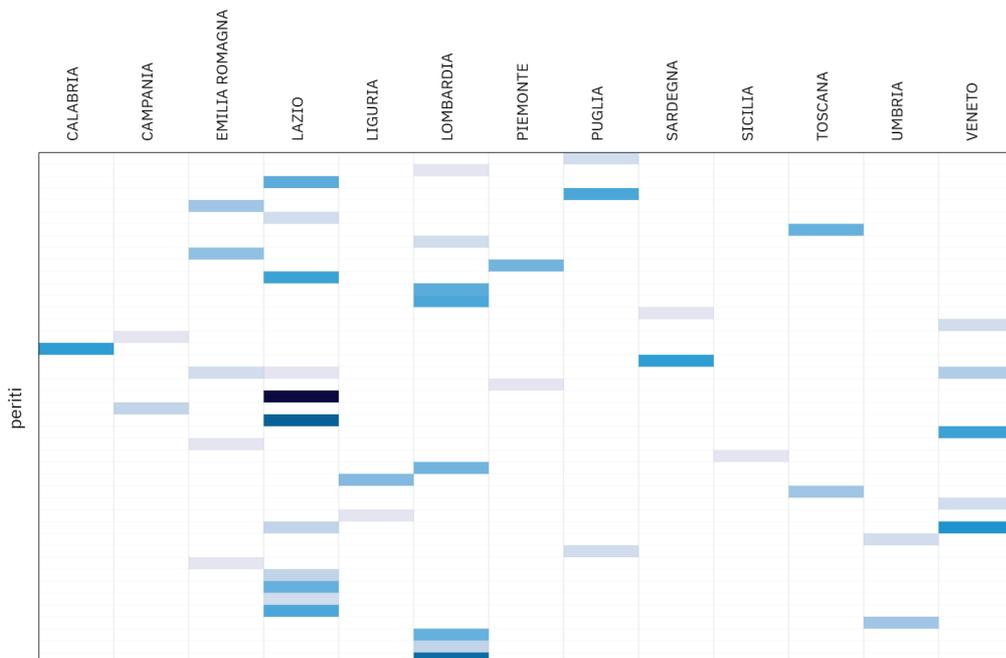
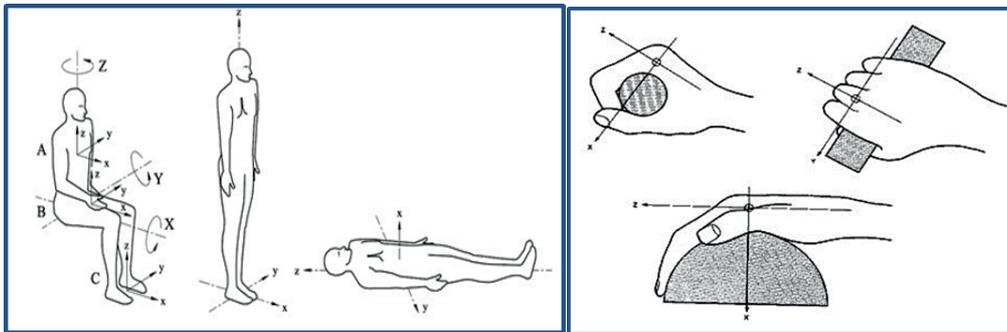


Figura 11 - Distribuzione dei progetti di bonifica da amianto per Regione e per periti

## 4.2 PROGETTI DI RIDUZIONE DEL RISCHIO DA VIBRAZIONI

Il rischio derivante dall'esposizione a vibrazioni meccaniche è associabile a macchine che i lavoratori utilizzano. La tipologia di rischio è distinta in rischio da vibrazioni al corpo intero e rischio da vibrazioni al sistema mano-braccio. Nel primo caso viene investito tutto il corpo del lavoratore e ciò avviene quando lo stesso opera/conduce la macchina, seduto o in piedi su di essa. Nel secondo caso il rischio si concentra sugli arti superiori e ciò è determinato dall'utilizzo di macchine gestite manualmente. Nei Bandi ISI, la tipologia di richiesta per rischio al sistema mano braccio è residuale rispetto a quella al corpo intero. La grandezza fisica che è alla base della valutazione e definizione del rischio è l'accelerazione "a" dell'oggetto che vibra, misurata in metri al secondo al quadrato ( $m/s^2$ ).

L'accelerazione viene valutata nelle sue componenti assiali, singolarmente o combinate, rispetto al sistema di riferimento. Di seguito se ne riporta una loro rappresentazione grafica.



Le accelerazioni prese in considerazione hanno diverse apici a seconda che siano rivolte al corpo intero o mano braccio ( $a_{w,x,y,z}; a_{hv}$ ). In generale il riferimento per la valutazione dell'esposizione è quello della così definita A(8), ovvero il valore delle vibrazioni cui è esposto il lavoratore riferito ad un periodo normalizzato sulle 8 ore.

Come per la bonifica da materiali contenenti amianto, i documenti non strutturati su cui si è concretamente attuata l'indagine con gli strumenti esposti nel capitolo precedente sono il Modulo A di Domanda (presentato dall'impresa) e il Modulo B di Perizia Giurata (elaborata dal perito/professionista che ha elaborato il progetto) relativi ai progetti ammessi a finanziamento.

I dati sono stati raccolti in file CSV utilizzati per generare le *dashboard*.

Nel caso delle vibrazioni, le informazioni sono costituite dall'insieme dei seguenti dati per ciascuna perizia:

- il riferimento alla macchina (sia esso da sostituire o da acquistare) costituito dalla sua tipologia, dal suo produttore e dal modello;

- il valore della vibrazione per ciascuna macchina, espressa come valore di accelerazione  $(a_{w,max})^1 / (a_{hv})^2$ ;
- la modalità di alienazione scelta dall'impresa per la macchina sostituita (rottamazione o vendita/permuta).

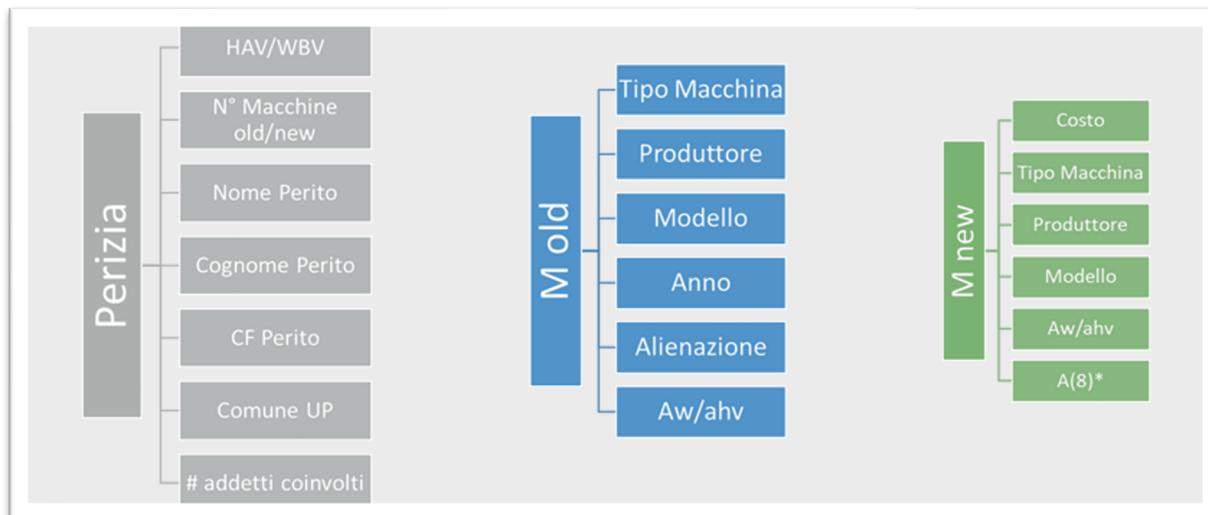


Figura 12 - Distribuzione dei progetti riguardanti il rischio vibrazioni per perizia

Le tipologie di macchine, i relativi produttori e le modalità di alienazione sono state ricavate utilizzando dei dizionari, mentre il modello (solitamente espresso con una parola di pochi caratteri alfanumerici) è stato ricercato nel testo tramite una specifica *character rule* quando compariva associato a una parola quale "tipologia di macchina" o "produttore". Per quanto riguarda il valore di  $a_{wmax} / a_{hv}$ , come per i metri quadrati di amianto, è stato individuato tramite una *character rule* di tipo numerico quando risultava accompagnato da un'espressione di unità di misura di accelerazione associata a un macchinario (almeno uno tra tipologia di macchinario, produttore e modello). Particolare attenzione è stata richiesta per risolvere il problema dell'identificazione dei macchinari da sostituire e quelli da acquistare.

La soluzione trovata è stata di evidenziare nel testo tutti i macchinari (completi di tutte le informazioni) e andare a sottrarre a questo insieme il sottoinsieme di macchine presenti nella "Tabella riassuntiva delle voci di spesa previste"; in questo modo, quelli che rimanevano, potevano essere assegnati alla categoria dei macchinari da sostituire.

<sup>1</sup> L'entità della vibrazione trasmessa al corpo intero corrisponde in genere al valore più elevato dell'accelerazione ponderata in frequenza determinato su uno dei tre assi ortogonali ( $1,4a_{wx}$ ,  $1,4a_{wy}$  o  $a_{wz}$ ) per un lavoratore in posizione seduta o eretta; in alternativa, quando non è chiaramente individuare un asse predominante, l'emissione vibratoria si può riferire ad un valore ottenuto come somma vettoriale delle tre componenti assiali.

<sup>2</sup> Il rischio di vibrazione al sistema mano-braccio si basa sul valore totale dell'accelerazione ponderata in frequenza  $a_{hv}$  che equivale alla radice quadrata della somma dei quadrati dell'accelerazione ponderata in frequenza determinati sui tre assi ortogonali x, y e z.

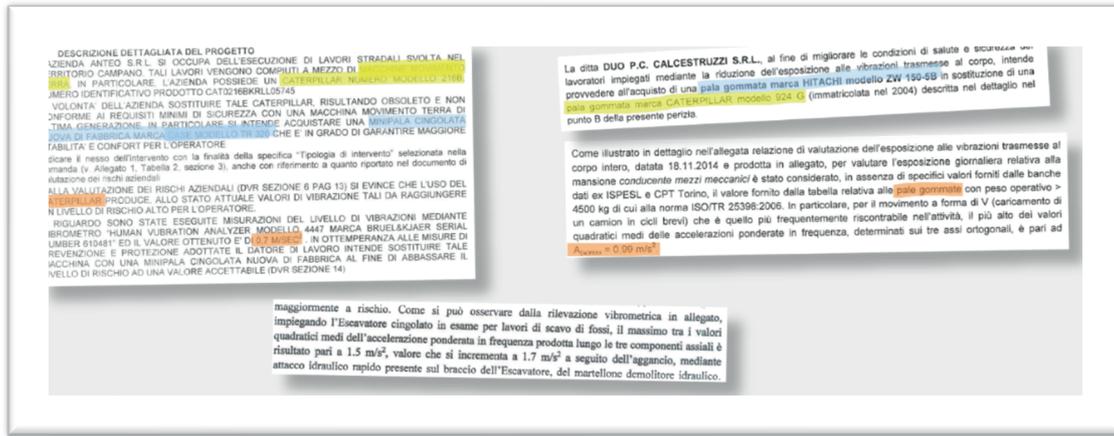


Figura 13 - Analisi dei dati non strutturati relativi a progetti per il rischio vibrazioni

Come visto per l'altro asse, anche per i documenti relativi alle pratiche di rischio vibrazioni, seppur più lentamente, è stata operata una standardizzazione nel corso degli anni, passando da un'esposizione dei dati in forma libera a una forma più tabellare, schematica e guidata. Ciò, se da un lato ha aiutato durante la fase di annotazione a chiarire l'esposizione delle informazioni, dall'altro ha presentato alcune complicazioni per l'estrazione delle informazioni.

Si riportano di seguito i dati, strutturati e non, relativi ai progetti volti alla riduzione dei rischi da vibrazioni, finanziati dall'Istituto con i Bandi ISI dal 2014 al 2017. In figura 14 sono riportati gli importi finanziabili richiesti per tale tipologia di intervento, gli addetti coinvolti e la distribuzione regionale. Il numero di lavoratori che utilizzano la macchina interessata non è un dato noto, in quanto dipende dall'attività aziendale e dalle sue dimensioni. Il dato riportato nei grafici riguarda tutti i dipendenti dell'impresa.

**195 Mln** **38,3 mila**

importo finanziabile

numero addetti

Distribuzione del numero di pratiche per regione

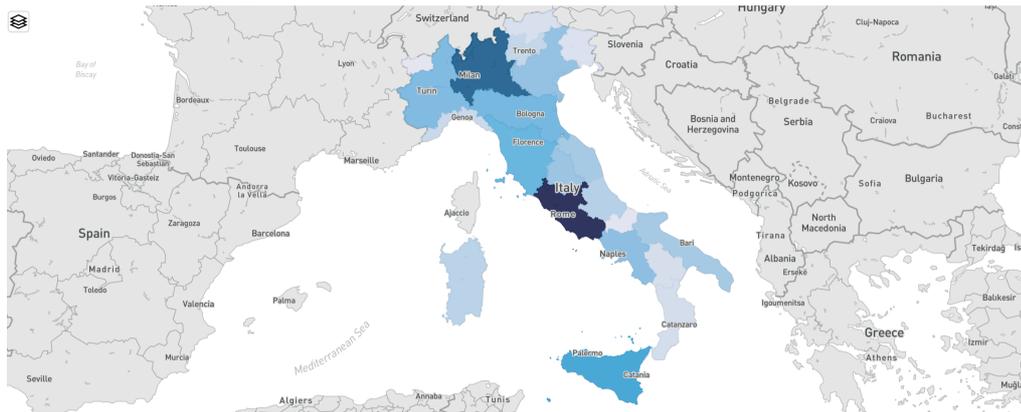


Figura 14 - Distribuzione regionale dei progetti di riduzione del rischio vibrazioni

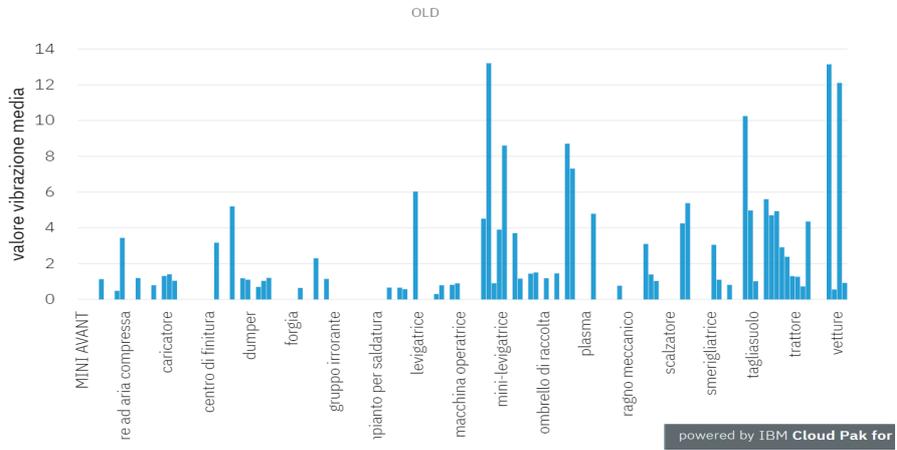
Nel grafico seguente (figura 15) si è risaliti invece alla riduzione del rischio vibrazioni a seguito dell'intervento, confrontando le caratteristiche tecniche, rispettivamente, delle macchine acquistate e sostituite.

Macchinari OBSOLETI



1,283

valore di vibrazione media



Macchinari NUOVI



0,558

valore di vibrazione media

Valore di vibrazione media per tipologia di macchinario

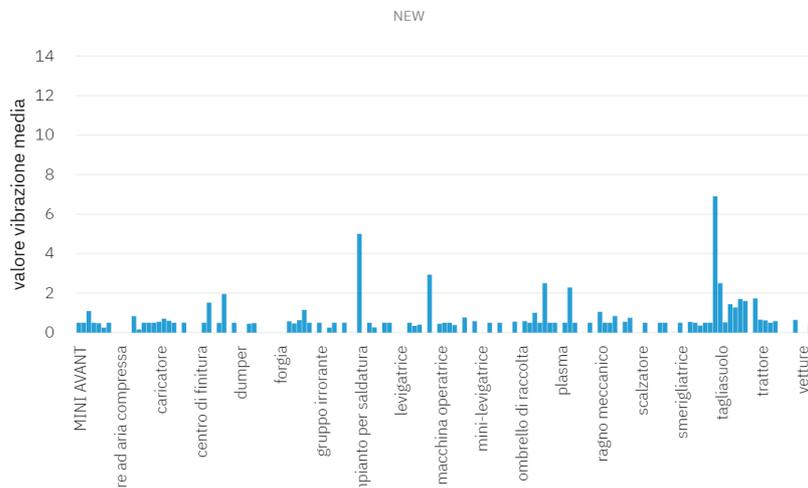


Figura 15 - Confronto del valore di accelerazione delle macchine sostituite e delle macchine nuove

Dal grafico emerge una riduzione consistente del rischio associato alla sola emissione della macchina: il valore medio delle vibrazioni meccaniche (accelerazione) trasmesse dalle macchine al corpo umano (principalmente corpo intero  $a_{w \max}$  rispetto a mano braccio  $a_{hv}$ ) diminuisce di oltre il 50%.

Infine si è effettuata un'estrazione delle informazioni relative alle principali tipologie di macchine richieste (figura 16). Tale informazione risulta utile anche ai fini della progettazione dei bandi futuri. Da essa infatti emerge una richiesta rilevante di escavatori, seguita dalle pale caricatrici e altre macchine di dimensioni spesso notevoli (massa operativa) caratterizzate da importi ingenti.

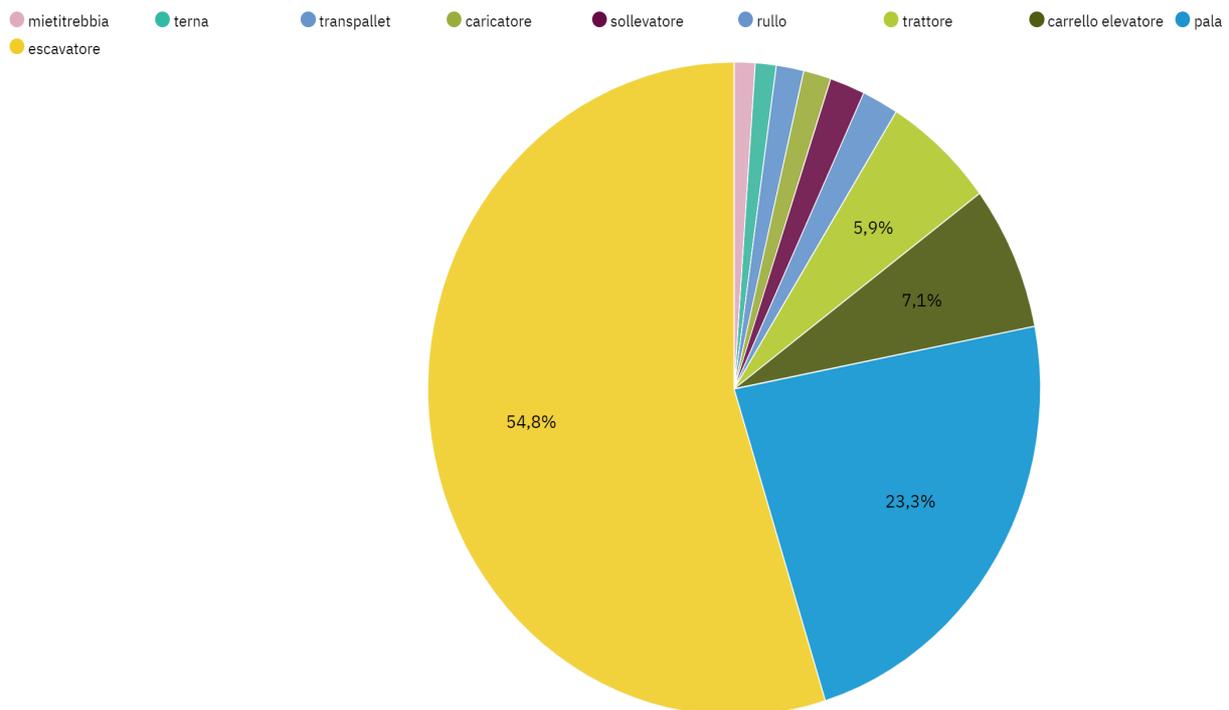


Figura 16 - Principali tipologie di macchine richieste per la riduzione del rischio vibrazioni

Questo dato, unito all'esame di ulteriori caratteristiche delle macchine che però esula dal presente progetto, ha permesso di formulare considerazioni che hanno portato nei bandi successivi a quello del 2020 a introdurre una modifica alla dimensione/massa operativa massima ammissibile per tali macchine; ciò al fine di aumentare, a parità di risorse economiche messe a disposizione, il numero di imprese che possono accedere ai finanziamenti e quindi di migliorare le condizioni di salute e sicurezza per un numero maggiore di lavoratori.

## 5. CONCLUSIONI

Le analisi condotte attraverso gli strumenti cognitivi hanno permesso di superare le difficoltà legate all'analisi di una notevole mole di documenti, che avrebbe altrimenti richiesto un impiego ingente di risorse umane. I vari *stress test* effettuati sullo strumento hanno permesso di stimarne l'affidabilità del risultato come pari o superiore all'80%.

Le informazioni desunte hanno quindi fornito i primi riscontri utili per ottenere una misura di massima dell'efficacia degli incentivi per il miglioramento della salute e sicurezza dei lavoratori nei due ambiti indagati. I dati relativi ai metri quadrati di amianto rimossi e alla riduzione delle vibrazioni connesse alla sostituzione delle macchine, sono un esempio di misura quantitativa dell'efficacia delle risorse messe a disposizione nei Bandi ISI.

Un aspetto altrettanto importante messo in evidenza dall'analisi condotta, che avvalorata da un diverso punto di vista i miglioramenti delle condizioni di salute e sicurezza attesi, è dato dall'utilità delle informazioni ricavate nell'individuare i punti di possibile miglioramento delle tipologie di intervento finanziabili. Da tali informazioni ha tratto giovamento in particolare la progettazione dei bandi successivi sia in termini di efficacia prevenzionale che di semplificazione e chiarezza per le imprese. Si tratta di miglioramenti che hanno portato alla riduzione del numero di progetti non ammessi.

L'utilizzo delle informazioni ha permesso inoltre di effettuare valutazioni in merito alla qualità e al costo degli interventi ammessi a finanziamento. Come già anticipato, l'analisi delle macchine richieste per mitigare il rischio da vibrazioni meccaniche ha permesso di apportare alcune modifiche ai Bandi ISI, per aumentare la distribuzione delle risorse e quindi il numero di lavoratori interessati dal miglioramento delle condizioni di lavoro. Parimenti, per le operazioni di bonifica da materiali contenenti amianto, tale strumento ha sostenuto le attività tecniche tese ad individuare, nell'Avviso, un prezzo medio congruo per metro quadrato, limitando così il verificarsi di situazioni anomale e possibili sperequazioni e agevolando al contempo le fasi di istruttoria tecnica.

Le risultanze del presente progetto permetteranno infine di avviare una standardizzazione degli strumenti cognitivi utilizzati, sia all'interno degli applicativi istituzionali esistenti sia all'interno di applicativi specifici creati ad hoc, che potrà risultare indispensabile per l'estrazione in tempo reale di tutte le informazioni tecniche e gestionali necessarie alle attività dell'Istituto. Nel corso degli ultimi anni l'Istituto ha portato avanti un processo di informatizzazione, delle varie fasi del Bando e realizzato un percorso guidato per l'inserimento delle richieste di finanziamento. In questo contesto tale tecnologia sarebbe utilissima se utilizzata direttamente durante la fase di inserimento delle informazioni e dati relativi ai progetti poiché consentirebbe di rilevare, monitorare e risolvere nel breve le criticità, limitando i tempi persi su problematiche immediatamente risolvibili con una comunicazione più rapida ed efficace con l'utenza.