

2.2 Creazione di un unico database sulla base del codice identificativo individuale

I tre files Stata corrispondenti alle tre tabelle del db Access sono stati unificati tramite la chiave “ID”, allo scopo di ottenere un singolo database.

La procedura di merge è stata applicata in due 2 fasi, di seguito descritte.

A. Linkage tra il file “Casi” e il file “Persone”

Lo scopo di questo merge consiste nel verificare la consistenza tra identificativi individuali registrati sui due file e, per i casi, di integrare i dati diagnostici con le informazioni anagrafiche, il gruppo di appartenenza (forze armate o corpi civili), il reparto ed il grado. I risultati del linkage sono descritti nella tabella che segue.

Provenienza records	No record
Solo file Casi	0
Solo file Persone	64631
Records linkati	159
Totale	64790

L'esito del primo linkage, dunque, è un db di 64790 record (relativo a 64788¹⁰ diversi soggetti). Questo file include:

- 159 records con esito positivo del linkage (tutti i 159 record del file “Casi” hanno trovato un ID corrispondente nel file “Persone”); questi records contengono variabili provenienti sia dal file “casi”, sia dal file “Persone”;
- 64631 records (relativi a 64631 soggetti) che provengono dal file “Persone” e non hanno trovato alcun ID corrispondente nel file “Casi”; questi record hanno tutti valori mancanti nelle variabili provenienti dal file “Casi”.

B. Linkage tra il file “Casi-Persone” e il file “Missioni”

Gli obiettivi di questo secondo merge consistono nel:

- verificare la consistenza tra identificativi individuali registrati sui due file e
- valutare il numero di missioni per soggetto e
- integrare le informazioni contenute nei file casi-persone con quelle relative alle missioni.

¹⁰ Nel file “Casi” c'erano due record individuali duplicati per i soggetti (ID 21952 e ID 58298) aventi due diverse diagnosi.

Questo merge ha prodotto i seguenti risultati.

Provenienza records	No record
Solo file Casi-Persone	411
Solo file Missioni	0
Records linkati	106993
Totale	107404

In questo linkage, dunque, si sono appaiati 106993 records, mentre ci sono 411 record (relativi ad altrettanti soggetti) che non trovano alcun ID corrispondente nel file “Missioni”.

In attesa di tali chiarimenti, abbiamo per il momento creato un db temporaneo costituito da **64788 soggetti**, i quali hanno svolto un totale di **107402¹¹ missioni** in Bosnia o Kosovo, 411 delle quali prive di qualsiasi informazione su date, luoghi ed altre caratteristiche.

¹¹ I record individuali duplicati per i casi con ID = 21952 e 58298 sono stati trasformati in record individuali singoli (duplicando i campi relativi alla diagnosi del tumore). Poiché i due soggetti hanno effettuato una missione ciascuno, nel db unico finale il totale delle missioni effettuate dai 64788 soggetti è 107402 (e non 107404).

2.3. Quantificazione dei dati mancanti

Le tabelle 1 e 2 riportano il numero di valori mancanti per le variabili e alcune combinazioni di variabili relative ai soggetti con diagnosi di tumore (CASI) e non (Non CASI).

Tabella 1.

Variabile	Decodifica presuntiva	CASI		Non CASI		Totale	
		N°	%	N°	%	N°	%
nato_il	Data di nascita	0	0.0%	46	0.1%	46	0.1%
nato_a	Comune di nascita	3	1.9%	5078	7.9%	5081	7.8%
resid_a	Comune di residenza	8	5.1%	7179	11.1%	7187	11.1%
grado	Grado	0	0.0%	242	0.4%	242	0.4%
rep_app	Reparto appartenenza	2	1.3%	15775	24.4%	15777	24.4%
Numero soggetti		157		64631		64788	

Tabella 2.

Combinazione di variabili	CASI		Non CASI		Totale	
	N°	%	N°	%	N°	%
nato_il & nato_a	0	-	20	0.03%	20	0.03%
nato_a & resid_a	3	1.9%	4959	7.7%	4962	7.66%
nato_il & nato_a & resid_a	0	-	19	0.03%	19	0.03%
rep_app & grado	0	-	9	0.01%	9	0.01%
rep_app & resid_a	1	0.6%	1387	2.1%	1388	2.14%
rep_app & nato_a	0	-	1115	1.7%	1115	1.72%
Numero soggetti	157		64631		64788	

La tabella 3 riporta, per casi e non casi, il numero di valori mancanti per le variabili relative alle 106991 missioni effettuate (escludendo cioè i 411 soggetti apparentemente mai andati nei Balcani).

Tabella 3.

Variabile	Decodifica presuntiva	CASI		Non CASI		Totale	
		N°	%	N°	%	N°	%
datainvio	Data inizio missione	0	0%	20	0%	20	0.02%
datarientro	Data rientro missione	0	0%	1347	1%	1347	1%
locimpiego	Località impiego	18	6%	2671	3%	2689	3%
rep_imp	Reparto impiego	71	23%	4578	4%	4649	4%
denom_inc	Denominazione incarico	108	36%	26320	25%	26428	25%
Numero missioni		303		106688		106991	

Nel complesso, sebbene il numero di casi sia esiguo rispetto al totale di soggetti in studio, per alcune variabili riferite ai “soggetti” la registrazione delle informazioni sembrerebbe essere stata più accurata per i “casi” rispetto ai non-casi.

Al contrario, per alcune variabili relative alle missioni la percentuale di valori mancanti è maggiore nei casi rispetto ai soggetti non casi:

2.4. Analisi descrittive

La distribuzione dei soggetti per genere è riportata nella tabella 4. Le donne sono presenti quasi esclusivamente tra il personale dell'esercito (311 su 316 soggetti).

Tabella 4.

Destinazione	CASI		Non CASI		Totale	
	N°	%	N°	%	N°	%
Uomini	156	99%	64316	99.5%	64472	99.5%
Donne	1	1%	315	0.5%	316	0.5%
Totale	157	100%	64631	100%	64788	100%

La distribuzione dei soggetti e delle missioni per forza armata di appartenenza viene illustrata nelle tabelle 5 e 6. Si può osservare una elevata prevalenza di soggetti (e relative missioni) appartenenti all'esercito rispetto alle altre forze armate e l'assenza di civili tra i casi.

Tabella 5.

Forza Armata	CASI		Non CASI		Totale	
	N° soggetti	%	N° soggetti	%	N° soggetti	%
Esercito	111	71%	50642	78%	50753	78%
Carabinieri	21	13%	5744	9%	5765	9%
Aviazione	21	13%	6468	10%	6489	10%
Marina	4	3%	1665	3%	1669	3%
Civili (?)	0	-	112	0.2%	112	0.20%
Totale	157	100%	64631	100%	64788	100%

Tabella 6.

	CASI		Non CASI		Totale	
	N° missioni	%	N° missioni	%	N° missioni	%
Forza Armata						
Esercito	233	76%	86093	80%	86326	80%
Carabinieri	36	12%	8582	8%	8618	8%
Aviazione	30	10%	8621	8%	8651	8%
Marina	9	3%	3635	3%	3644	3%
Civili (?)	0	-	163	0.20%	163	0.20%
Totale	308	100%	107094	100%	107402	100%

Per quanto riguarda la nazione di impiego è stata evidenziata una prevalenza significativamente più elevata di missioni in Kosovo tra i casi rispetto ai non casi (Tabella 7 - Pearson χ^2 (2 g.d.l.) = 66, $p < 0.0001$).

Tabella 7.

	CASI		Non CASI		Totale	
	No missioni	%	No missioni	%	No	%
Bosnia	180	58%	41373	39%	41553	39%
Kosovo	123	40%	65315	61%	65438	61%
Missing	5	2%	406	0.4%	411	0.4%
Totale	308	100%	107094	100%	107402	100%

Per ciò che concerne il numero di missioni per soggetto e per stato di salute (casi/non casi), si rileva che la maggior parte dei soggetti hanno effettuato 1 o 2 missioni (52% nei casi e 58% nei soggetti sani).

Il numero di missioni per soggetto (escludendo i 411 soggetti apparentemente mai andati in missione) presenta un certo grado di variabilità in relazione alla forza armata di appartenenza (Tabella 8). Infatti, mentre il personale dell'aeronautica e i civili hanno effettuato prevalentemente 1 o 2 missioni, una quota elevata di graduati dell'esercito e soprattutto della marina ha svolto fino da 3 a 5 missioni. Nessun civile ha partecipato a più di 5 missioni. Il numero di missioni per soggetto in relazione all'area di destinazione (Bosnia o Kosovo) non presenta differenze di rilievo.

Tabella 8.

	Aeronautica		Carabinieri		Civili		Esercito		Marina		Totale	
	N°	%	N°	%	N°	%	N°	%	N°	%	N°	%
1-2	7175	83	5694	68	129	80	53077	62	1474	40	67549	63
3-5	1248	14	2288	27	33	20	31103	36	1769	49	36441	34
6+	202	2	399	5	0	0	2003	2	397	11	3001	3
Totale	8625	100	8381	100	162	100	86183	100	3640	100	106991	100

Al fine di stimare l'età media dei militari andati in missione, è stata calcolata l'età alla data di invio alla prima missione; 458 records non sono analizzabili per assenza di informazioni sulla data di nascita (46 soggetti -47 records-) o sulla data di inizio missione (411 soggetti). La Tabella 9 riporta, per i casi e per i non casi, il valore medio, la deviazione standard (DS), il minimo e massimo dell'età all'inizio della prima missione, separatamente per coloro che hanno effettuato una sola missione o più missioni. Si può osservare che l'età media all'inizio dell'unica o della prima missione risulta più elevata tra i casi rispetto ai non casi (circa 4 anni di differenza).

Tabella 9.

	CASI					Non CASI				
	Età all'inizio in missione (anni)					Età all'inizio in missione (anni)				
	N° missioni	media	DS	Min	Max	N° missioni	media	DS	Min	Max
Solo una missione	75	33.5	9.6	18.7	53.4	39023	27.9	8.4	16.7	76.0
La prima di più missioni	77	29.7	9.6	19.4	51.9	25156	26.2	7.5	15.4	63.3.0

2.4.1. Anno di diagnosi e decesso per i soggetti con diagnosi di tumore

Come descritto precedentemente, i 159 record del file "Casi" si riferiscono a 157 soggetti; a due casi sono state diagnosticate due diverse neoplasie ciascuno. Per questi due soggetti nelle analisi descrittive che seguono è stata considerata la data di diagnosi della prima neoplasia accertata. Per quanto concerne l'anno di diagnosi, escludendo 2 soggetti con dato mancante sulla data di diagnosi, si ottiene la seguente distribuzione di frequenza:

Anno Diagnosi	N° casi	%
1995	1	0.7
1996	2	1.3
1997	1	0.7
1998	8	5.2
1999	9	5.8
2000	25	16.1
2001	17	11.0
2002	28	18.1
2003	20	12.9
2004	36	23.2
2005	8	5.1
Totale	155	100

I 29 decessi (19% dei casi diagnosticati) presentano la seguente distribuzione per anno:

Anno decesso	N° casi	%
1998	1	3.5
1999	1	3.5
2000	7	24.1
2001	2	6.9
2002	2	6.9
2003	6	20.7
2004	8	27.6
2005	2	6.9
Totale	29	100.0

2.5. Problemi legati alla coerenza delle informazioni registrate nel database

Per quanto riguarda i casi con diagnosi di tumore, occorre notare che dal 2000 al 2005 complessivamente **17** soggetti (11% del totale dei casi) hanno stranamente ricevuto una diagnosi di tumore il giorno 1 gennaio:

C'è inoltre **un soggetto**, di cui non si conosce il luogo di nascita, per il quale la data di diagnosi è precedente alla data di inizio dell'unica missione che ha effettuato. In questo caso, occorre chiarire se si tratti di un errore nella registrazione della data di inizio della missione e/o di un errore nella data di diagnosi. Se i dati fossero corretti, questo soggetto non andrebbe conteggiato come caso in un monitoraggio dell'incidenza e dovrebbe essere escluso da uno studio di coorte.

Le date di nascita dei 64788 soggetti inclusi nel database vanno dal **19/8/1924** al **9/5/1987** e ci sono **46** soggetti con data di nascita ignota (tutti soggetti non casi). Occorre anche sottolineare che i valori mancanti per la data di nascita riguardano esclusivamente soggetti che hanno effettuato una sola missione.

Nel database sono state identificate **121** missioni relative a **119** soggetti la cui età ad inizio missione è minore di 18 anni. E' vero che l'età minima per l'arruolamento è in Italia 16 anni. Tuttavia vorremmo verificare se non si tratti di errori di registrazione nella data di nascita e/o nella data inizio missione.

Per quanto concerne le informazioni desunte attraverso le procedure di linkage tra i tre file disponibili, sono stati identificati **411** record (relativi ad altrettanti soggetti, di cui 5 con diagnosi di tumore) che non trovano alcun ID corrispondente nel file "Missioni". Occorrerà chiarire come mai questi soggetti siano stati inseriti nel database dal momento che, apparentemente, non sono mai stati in missione nei territori della Bosnia-Herzegovina o del Kosovo. Potrebbe darsi, d'altra parte, che si tratti di soggetti effettivamente inviati in missione che non hanno, però, compilato la scheda informativa prevista dal Programma di Monitoraggio.

Questo dato costituisce un primo richiamo alla necessità di ottenere spiegazioni esaustive sulle fonti e sulle procedure utilizzate dalla Sanità Militare per creare e compilare i db che ci sono stati trasmessi.

Non considerando 20 missioni senza data di inizio, le missioni registrate presentano una data di inizio compresa tra il **2-2-1988** ed il **2-7-2005**.

Le date di fine missione vanno dal **22-9-1988** al **29-9-2005**; **1347** record non presentano data di rientro e **82** missioni (tutti non casi) hanno una data di fine anteriore alla data di inizio. Tra queste ultime, e ci sono **6** missioni con date di rientro che presentano errori evidenti di inserimento.

Occorre inoltre osservare che il range osservato di date d'inizio e di fine ha come limiti più remoti date (**3-2-1988** e **22-9-1988**, rispettivamente) non consistenti con l'inizio del periodo d'interesse per il programma di sorveglianza (**1-8-1994**) definito dalla L n. 27/2001. Si noti, in particolare, che nel database sono presenti 7 missioni, riferite a 7 diversi soggetti, con data d'inizio antecedente al 1-8-1994 (tabella successiva). Per queste missioni anche la data di rientro, generalmente, precede il 1 agosto 1994. Inoltre, solo uno di questi soggetti ha effettuato, nella sua attività operativa nei Balcani, solo missioni prima del 1/8/1994.

Nel database è stato identificato un certo numero di records (tutti tra i non casi) con durata della missione pari a giorni 0 (122) e 1 (58).

Sono state, infine, rilevate sovrapposizioni tra le date di inizio e fine missione nelle diverse missioni effettuate dallo stesso soggetto. La quantificazione del numero di records per soggetto con sovrapposizioni tra le date relative a diverse missioni è attualmente in corso.

Sono stati individuate **2765** diverse missioni relative a **1344** soggetti (21 casi e 1323 non casi) che presentano le stesse date di fine

2.6 Problemi legati alla mancata codifica dei campi descrittivi

Per quanto riguarda i campi testo “**grado**”, “**locimpiego**” (località impiego), “**rep_imp**” (reparto impiego), “**rep_app**” (reparto di appartenenza) e “**denom_inc**” (denominazione incarico), l’inserimento delle varie modalità è chiaramente avvenuto senza alcuna standardizzazione e molto spesso non è possibile risalire al significato delle sigle/abbreviazioni utilizzate. Questi problemi rendono al momento molto complesso, se non impossibile, la conduzione di analisi descrittive informative.

I dettagli che seguono permettono di render conto di tale affermazione.

Per quanto concerne la variabile “**grado**”, si tratta di un campo descrittivo che presenta **529** diverse modalità. Inoltre, tale variabile è caratterizzata da elevata eterogeneità nella descrizione di una stessa modalità, ad esempio il grado di “brigadiere” e di “tenente colonnello” possono avere le seguenti forme:

MODALITÀ	OSSERVAZIONI
BRG	15
BRIG	12
BRIG.	79
BRIGADIERE	5
Bri	126
Brig	16
Brig.	88
Brigadiere	10

MODALITÀ	OSSERVAZIONI
T COL	2
T.COL	2
T.COL.	5
T.Col.	99
TC	128
TCL	1101
TCOL	3
TEN COL	7
TEN. COL	2
TEN. COL.	14
TEN.COL.	406
TENENTE COLONNELLO	4
Ten Col.	2
Ten. Col.	5
Ten.Col	3
Ten.Col.	87
Tenente Colonnello	109

La variabile “**locimpiego**” è un campo descrittivo che presenta **256** diverse modalità, con elevata eterogeneità nella descrizione della stessa modalità; ad esempio le località “Sarajevo” e “Pristina” possono assumere le seguenti forme:

MODALITÀ	OSSERVAZIONI
SARAEVO	1
SARAEVO	5
SARAJEVO	34085
SARAJEVO (BOSNIA)	14
SARAJEVO	2
SARAYEVO	4
SARJEVO	1
SARYEVO	1
PRISTINA	2912
PRISTINA (KOSOVO)	9
PRISTINA (KOSSOVO)	1
PRISTINA KFOR	1
PRISTINA(KOSOVO)	1
PRISTJNA	1
Pristina	781

Sembra, poi, che durante la fase di registrazione del db “Missioni” ci sia stata una certa confusione nella compilazione dei campi “**locimpiego**” e “**naz_imp**” (nazione di destinazione?): infatti, nel campo “locimpiego” 56 records riportano la sola locuzione “Bosnia” e in ben 1467 records compare solo “Kosovo”.

Il campo “**rep_imp**” è un campo descrittivo che presenta **2014** diverse modalità, con elevata eterogeneità nella descrizione della stessa modalità; ad esempio il “*reparto comando e supporto tattico Taurinense*” può essere presente come:

MODALITÀ	OSSERVAZIONI
R.C.S.T. "TAURINENSE"	1
RCST "TAURINENSE"	6
RCST TAURINENSE	726
RCST"TAURINENSE"	1
RECO SUTAT "TAURINENSE"	1
RCST. "TAURINENSE"	1
RE.CO.SU.TA TAURINENSE	40
RE.CO.SU.TAT. "TAURINENSE"	4
RCST "TAURINENSE"	1
RECO SUTAT TAURINENSE	1
REP. COMANDO E SUPPORTI T	3
REP.CDO E SPT.TAT."TAURIN	1
ReCoSuTat Taurinense	1
SFOR RCST "TAURINENSE"	300
SFOR RCST"TAURINENSE"	1
SFOR RE CO TATT TAURINENS	1

Sembra, inoltre, che anche il campo “**rep_imp**” venga spesso confuso con la variabile “**naz_imp**” (nazione di destinazione?) oppure con la variabile “**locimpiego**”. Infatti si osservano le seguenti modalità (chiaramente erranee) della variabile “**rep_imp**”:

MODALITÀ	OSSERVAZIONI
KOSOVO	14
PRISTINA	57
PRISTINA(KOSOVO)	28
Kosovo	599
BOSNIA	2632

Per quanto concerne la variabile “**rep_app**”, si tratta di un campo descrittivo che presenta **4355** diverse modalità, con elevata eterogeneità nella descrizione della stessa modalità che comporta problematiche simili a quelle emerse per la variabile “**rep_imp**”.

Occorre inoltre osservare che i campi **grado**, **residenza**, e **reparto di appartenenza**, facendo parte del file a record singoli “Persone con ID”, sono stati evidentemente compilati al momento della creazione di tale file e quindi, sebbene siano suscettibili di modifiche nel tempo (i.e. un militare avanza di grado, può cambiare reparto e residenza), tali variabili andrebbero connotate con la data alla quale si riferisce la rilevazione e aggiornate nel corso degli anni. Queste considerazioni valgono in particolare per i soggetti che hanno effettuato varie missioni.

Per quanto concerne, infine, la variabile “**denom_inc**” (denominazione incarico), si tratta di un campo descrittivo che presenta **10736** diverse modalità (praticamente come se uno stesso incarico non fosse stato svolto in più di 10 diverse missioni), con elevata eterogeneità nella descrizione della stessa modalità.

Si riportano, ad esempio, le 26 diverse modalità di inserimento per l’incarico “aiutante di sanità” individuate nel file.

MODALITÀ	OSSERVAZIONI
2 (AIUTANTE DI SANITA')	8
2 - AITANTE DI SANITA'	4
2 - AIUTANTE DI SANITA	1
2 - AIUTANTE DI SANITA'	7
2 - AIUTANTE DI SANITA'	6
2 - Aiutante di sanita'	1
2 - Aiutante di sanità	5
2 - Aiutante di sanità'	62
2 AIUTANTE DI SANITA'	19
2 ASA	2
2- AIUTANTE DI SANITA'	1
2-AIUTANTE DI SANITA'	4
2-Aiutante di Sanita'	2
2-Aiutante di Sanità	1
2-Aiutante di sanità	6
AIUTANTE DI SANITA	6
AIUTANTE DI SANITA'	383
AIUTANTE DI SANITA' - 2	1
AIUTANTE DI SANITA'	1
AIUTANTE DI SANITÀ	2
Aiut.sanità	1
Aiutante di Sanità	6
Aiutante di sanità	15
AIUTANTE D SANITA	2
AIUTANTE DI ASANITA'	1
AIUTANTE SANITA'	12

Nella sezioni successive verranno descritte alcune procedure di standardizzazione delle modalità inserite nei campi “Denominazione Incarico” e “Grado”.

3. Verifica di coerenza interna e procedura di codifica dei dati relativi alla variabile “Denominazione di Incarico”

L’analisi del database contenente i records relativi alle “Missioni” ha evidenziato la necessità di procedere ad una standardizzazione preliminare e ad una successiva codifica del campo riportante l’informazione sull’incarico (“Denominazione Incarico”) svolto dal personale militare impegnato in missione, in quanto l’eccessivo dettaglio e la mancanza di criteri nella classificazione degli incarichi nel database originario non consentono di utilizzare tale variabile nelle analisi statistiche.

L’entità del problema si apprezza chiaramente dalla tabella seguente, che riporta la descrizione dei 50 incarichi più frequenti su 10736 descritti.

Denominazione Incarico	Frequenza	% su totale
1 missing	26428	24.7
2 INCARICO OPERATIVO	22444	21.0
3 30/A	1640	1.5
4 FUCILIERE	1592	1.5
5 ADDETTO	882	0.8
6 18/A	857	0.8
7 30/A - Fuciliere	852	0.8
8 54/G	782	0.7
9 CONDUTTORE	764	0.7
10 ADDETTO AI COMANDI	572	0.5
11 ELETTROGENISTA	531	0.5
12 GUASTATORE	529	0.5
13 MECCANICO	491	0.5
14 CONDUTTORE AUTOMEZZI	428	0.4
15 ARMAIOLO	415	0.4
16 18/A - Conduttore automezzi vari	393	0.4
17 AIUTANTE DI SANITA'	383	0.4
18 107 - Esploratore	378	0.4
19 OPERATIVO	358	0.3
20 CAU	346	0.3
21 CUCINIERE	337	0.3
22 18/A - CONDUTTORE AUTOMEZZI	320	0.3
23 PILOTA	320	0.3
24 54/G - Pilota VTC	319	0.3
25 OPERATORE DI POLIZIA MILITARE	301	0.3
26 ESPLORATORE BLINDO LEGGERE	271	0.3
27 INFERMIERE	270	0.3
28 PILOTA VTC	270	0.3
29 MUSICANTE	263	0.2
30 107 ESPLORATORE	256	0.2
31 MOT.: COND.AUTOMEZZI	239	0.2
32 111/A	236	0.2
33 CARRISTA D'EQUIPAGGIO	236	0.2
34 Operatore Radiotelescrivente	220	0.2
35 CARRISTA DI EQUIPAGGIO	211	0.2
36 IDRAULICO	203	0.2
37 216/B	202	0.2
38 51	200	0.2
39 Operatore per le TLC	198	0.2
40 COM/TE SQUADRA	196	0.2
41 PIONIERE	195	0.2
42 CONDUTTORE AUTOMEZZI VARI	191	0.2
43 MORTAISTA	189	0.2
44 C.TE DI PLOTONE	179	0.2
45 INFERMIERE PROFESSIONALE	178	0.2
46 PILOTA MEZZI CINGOLATI	178	0.2
47 OPERATORE MACCHINE STRADALI	172	0.2
48 DISPOSIZ. DEL C/DO	162	0.2
49 RADIOFONISTA	162	0.2
50 PLOTONE SICUREZZA	153	0.1

Si è proceduto alla codifica del campo “Denominazione Incarico” utilizzando un programma SAS appositamente elaborato, basato sulla selezione di sotto-stringhe della denominazione di incarico. Allo stato attuale si è pervenuti ad una classificazione degli incarichi che prevede **29 categorie**.

La tabella che segue presenta i risultati preliminari della procedura di standardizzazione.

Tabella 10. Risultati preliminari della procedura di standardizzazione dell'incarico

	Classe Incarico	Frequenza	% su totale	Tipologie rientranti nella classe
1	Incarico Operativo	22649	21.2	
2	Addetto	8417	7.9	ai comandi, ai mezzi di lancio, alla vigilanza, etc..
3	Operatore	3986	3.7	una parte va attribuita alla classe "incarico operativo"
4	Comandante	3974	3.7	plotone, alpieri, rgt, squadra etc..
5	Conduttore Automezzi	3361	3.1	
6	Capo (vari incarichi)	2104	2.0	capo cellula, blindo, servizio, sezione, ufficio
7	Fuciliere	1941	1.8	
8	Meccanico	1932	1.8	
9	Esploratore	1553	1.5	blindo leggere, pesanti
10	Pilota	1506	1.4	VTC, mezzi cingolati
11	Sotto-Ufficiale	1163	1.1	
12	Ufficiale	975	0.9	
13	Carrista	768	0.7	
14	Aiutante	764	0.7	di sanità, etc.
15	Tecnico (vari)	743	0.7	
16	Infermiere	664	0.6	
17	Elettrogenista	594	0.6	
18	Armaiolo	588	0.5	
19	Guastatore	576	0.5	
20	Cau	478	0.4	
21	Specialista	436	0.4	
22	Cuciniere	375	0.4	
23	Musicante	270	0.3	
24	Mortaista	208	0.2	
25	Paracadutista	159	0.1	
26	Osservatore	123	0.1	
27	Elettricista	109	0.1	
28	Altro	20147	18.8	
29	Non specificato	26428	24.7	
	Totale	106991	100.0	

Un primo problema è rappresentato dall'elevato numero di casi per il quali l'incarico nella missione non viene specificato (**missing**), che risulta essere pari a 26428 (circa il 25% del totale delle missioni).

Per questi casi è ipotizzabile mettere a punto un criterio di imputazione del dato mancante, utilizzando le informazioni presenti nel database contenente i records relativi alle *persone*, o quelle desumibili dai documenti informativi sulle missioni disponibili sui siti militari e del Ministero della Difesa.

Sono candidati ad analoghe procedure di recupero informazioni sull'incarico anche i 22444 records (21% delle missioni) con informazione limitata ad "**incarico operativo**" (la cui valenza informativa appare praticamente nulla).

Inoltre, alcune delle categorie individuate tramite la procedura di codifica risultano troppo ampie ed eterogenee (esempio: addetto, operatore, etc.); in questi casi è possibile effettuare un'analisi della distribuzione degli incarichi più frequenti presenti al loro interno.

Permane, infine, un cospicuo numero di records (19% del totale) con denominazioni di incarico meno frequenti, che al momento sfuggono alla procedura di codifica. A tal proposito, sono attualmente in corso di implementazione affinamenti della procedura.

4. Verifica di coerenza interna e procedura di codifica dei dati relativi alla variabile “Grado”

La variabile “grado”, si tratta, come descritto precedentemente, di un campo che presenta 529 diverse modalità di tipo descrittivo, in formato testo.

Si è inizialmente proceduto ad una decifrazione delle diverse modalità registrate, alla verifica di validità delle codifiche originali e, successivamente, ad una ricodifica della variabile “grado”. Per la realizzazione di tali procedure sono stati innanzitutto consultati i seguenti siti internet al fine di identificare correttamente la tipologia e la gerarchia dei gradi nelle diverse armi:

www.hyperreview.com/Militari/Gradi%20Militari.htm

www.aeronautica.difesa.it/SMA/1Rep/1U/Gradi/

www.esercito.difesa.it/root/uniformi/gradi_index.asp

www.esercito.difesa.it/root/vario/sigle.asp

In particolare, la consultazione del sito [http://it.wikipedia.org/wiki/Grado_\(militare\)](http://it.wikipedia.org/wiki/Grado_(militare)) ha consentito di definire la corrispondenza tra la gerarchia dei gradi nelle Forze Armate italiane (compresa la correlazione con i gradi NATO). Tale strumento ha permesso di evidenziare alcune discrepanze tra la modalità del grado e quella della forza armata registrate nel database.

La situazione di partenza si presentava come qui di seguito riportata:

	N° modalità	N° records
Modalità originali della variabile grado	529	107111
"Missing"	1	291
Totale	530	107402

Al termine della procedura di analisi, 458 (87%) delle 529 modalità originali sono state accorpate in 90 nuove modalità consentendo di classificare correttamente 83802 (78%) dei 107111 records. La seguente tabella fornisce il quadro completo della riclassificazione delle modalità originali della variabile “grado”.

Tabella 11. Riclassificazione delle modalità originali in nuove modalità (da 458 a 90 modalità)

N° modalità originali	N° record codificati	Descrizione univoca adottata per la nuova modalità
9	782	1° aviere
3	21	1° aviere scelto
14	11700	1° caporal maggiore
10	1397	1° maresciallo
6	8	1° maresciallo luogotenente
1	1	2° maresciallo
5	1692	aiutante
3	53	allievo maresciallo
1	1	allievo sergente
1	14	allievo ufficiale
1	3	ammiraglio di squadra
5	1114	appuntato
19	632	appuntato scelto
2	2	aspirante guardiamarina
6	191	aviere capo
1	12	aviere capo scelto
4	199	aviere scelto
8	351	brigadiere
7	18	brigadiere capo
6	75	brigadiere generale
7	3905	capitano
2	53	capitano di corvetta
2	33	capitano di fregata
2	4	capitano di vascello
1	2	capitano medico
4	97	capo prima classe
4	575	capo seconda classe
1	2	capo seconda classe fucilieri di marina/anfibi
3	158	capo terza classe
13	9102	caporal maggiore
1	13	caporal maggiore capo scelto
15	7704	caporal maggiore scelto
2	8	caporal maggiore scelto volontario in servizio permanente
1	37	caporal maggiore volontario in ferma breve
1	43	caporal maggiore volontario in servizio permanente
6	11738	caporale
3	91	caporale scelto
1	8	caporale volontario in ferma breve
1	4	caporale volontario in servizio permanente
9	2085	carabiniere
1	2	carabiniere in servizio permanente
22	383	carabiniere scelto
1	9	civile
6	341	colonnello
3	230	comune di prima calsse
4	24	comune di seconda calsse
2	49	comune di terza calsse
1	6	contrammiraglio
1	3	dipendente civile
1	4	generale

2	7	generale brigata aerea
1	2	generale di brigata
1	5	generale maggiore
1	4	guardiamarina
1	1	impiegato
1	1	impiegato civile
7	115	luogotenente
2	16	maggior generale
9	913	maggiore
10	3270	maresciallo
1	2	maresciallo aiutante
13	217	maresciallo aiutante sostituto ufficiale di pubblica sicurezza
17	2389	maresciallo capo
13	4684	maresciallo ordinario
13	563	maresciallo prima classe
13	2471	maresciallo seconda classe
11	416	maresciallo terza classe
6	136	operaio
5	18	operaio civile
4	122	primo maresciallo
1	1	primo maresciallo luogotenente
2	11	secondo capo
8	3998	sergente
1	2	sergente fucilieri di marina/anfibi
6	896	sergente maggiore
4	156	soldato
4	704	sottocapo
2	119	sottocapo di seconda classe
3	485	sottocapo di terza classe
11	1062	sottotenente
4	79	sottotenente di vascello
1	1	sottoufficiale
8	3063	tenente
17	1979	tenente colonnello
4	199	tenente di vascello
2	6	tenente generale (o generale di corpo d'armata)
1	1	tenente medico
1	1	tenente ruolo normale
1	1	tenente tecnico logistico
13	707	vice brigadiere
Modalità originali = 458		Totale records = 83802
		Totale nuove modalità = 90