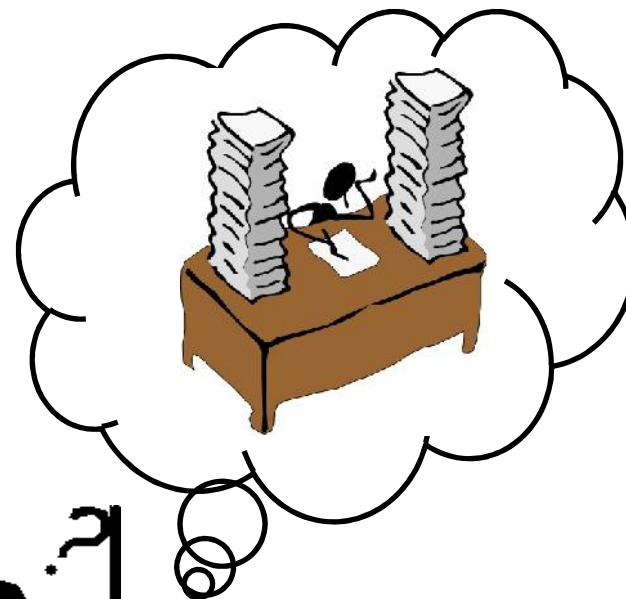
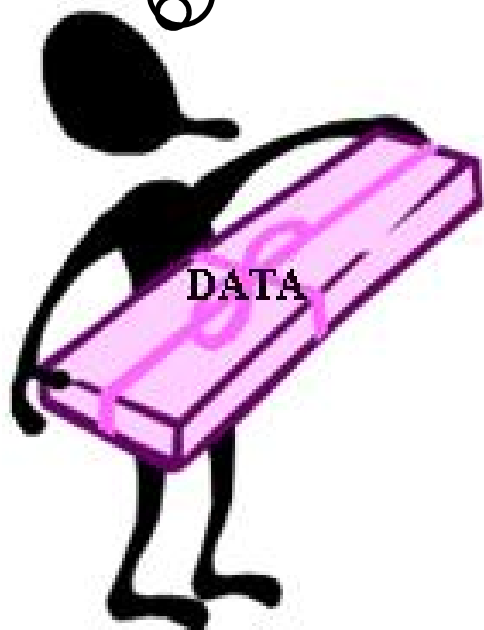
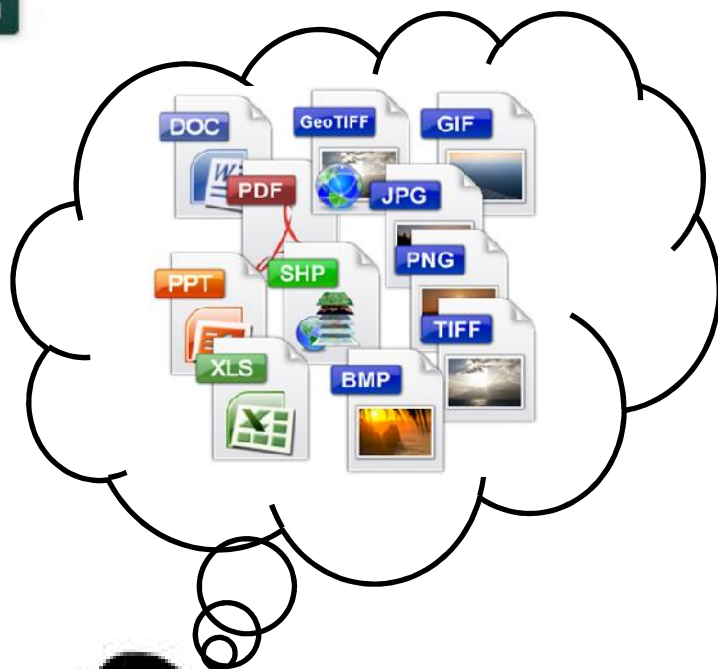




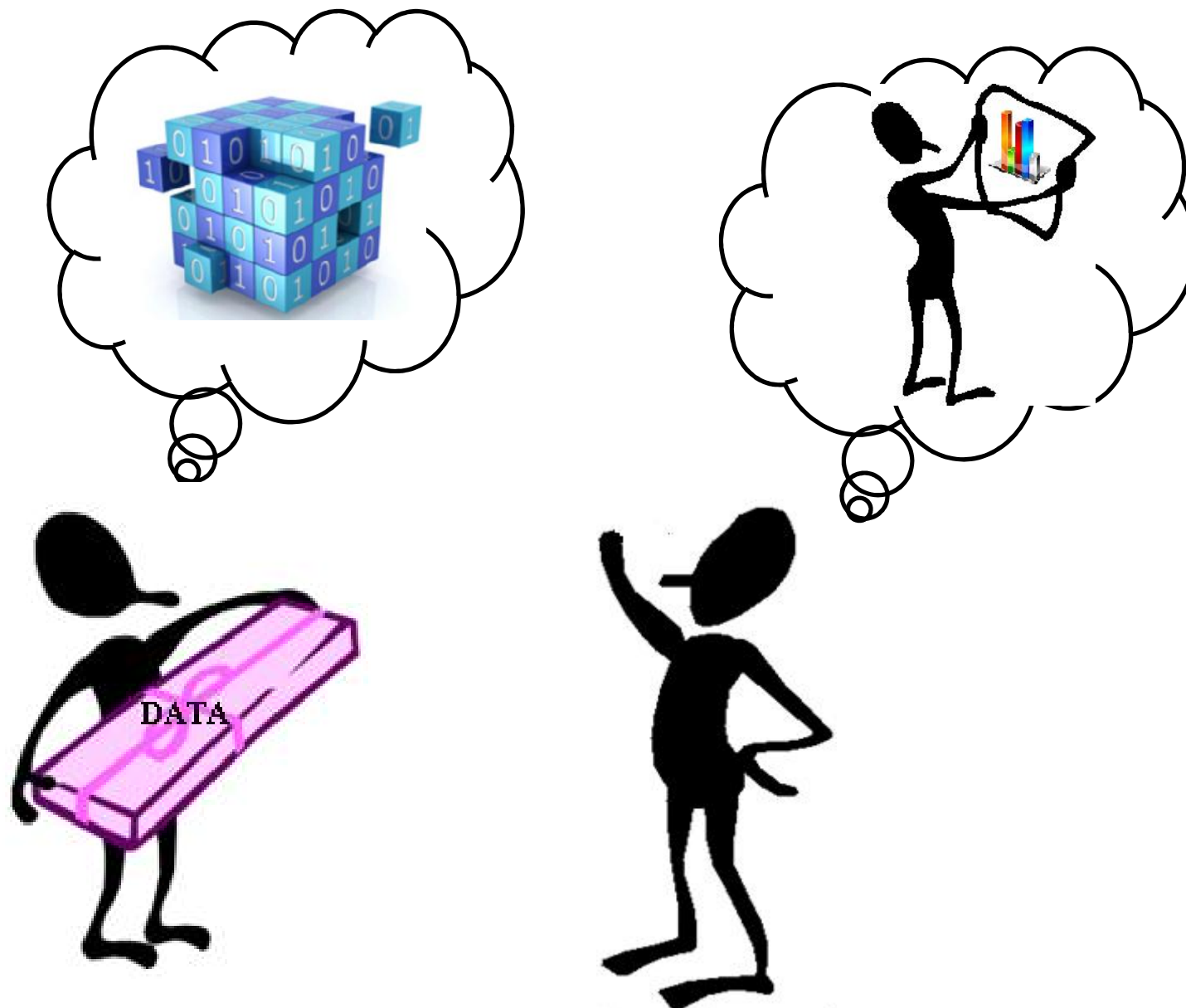
**Διαχείριση δεδομένων και μεταδεδομένων
για την απρόσκοπτη δευτερογενή ανάλυση δεδομένων
ποσοτικών ερευνών**

Δρ. Απόστολος Λιναρδής
Ερευνητής ΕΚΚΕ

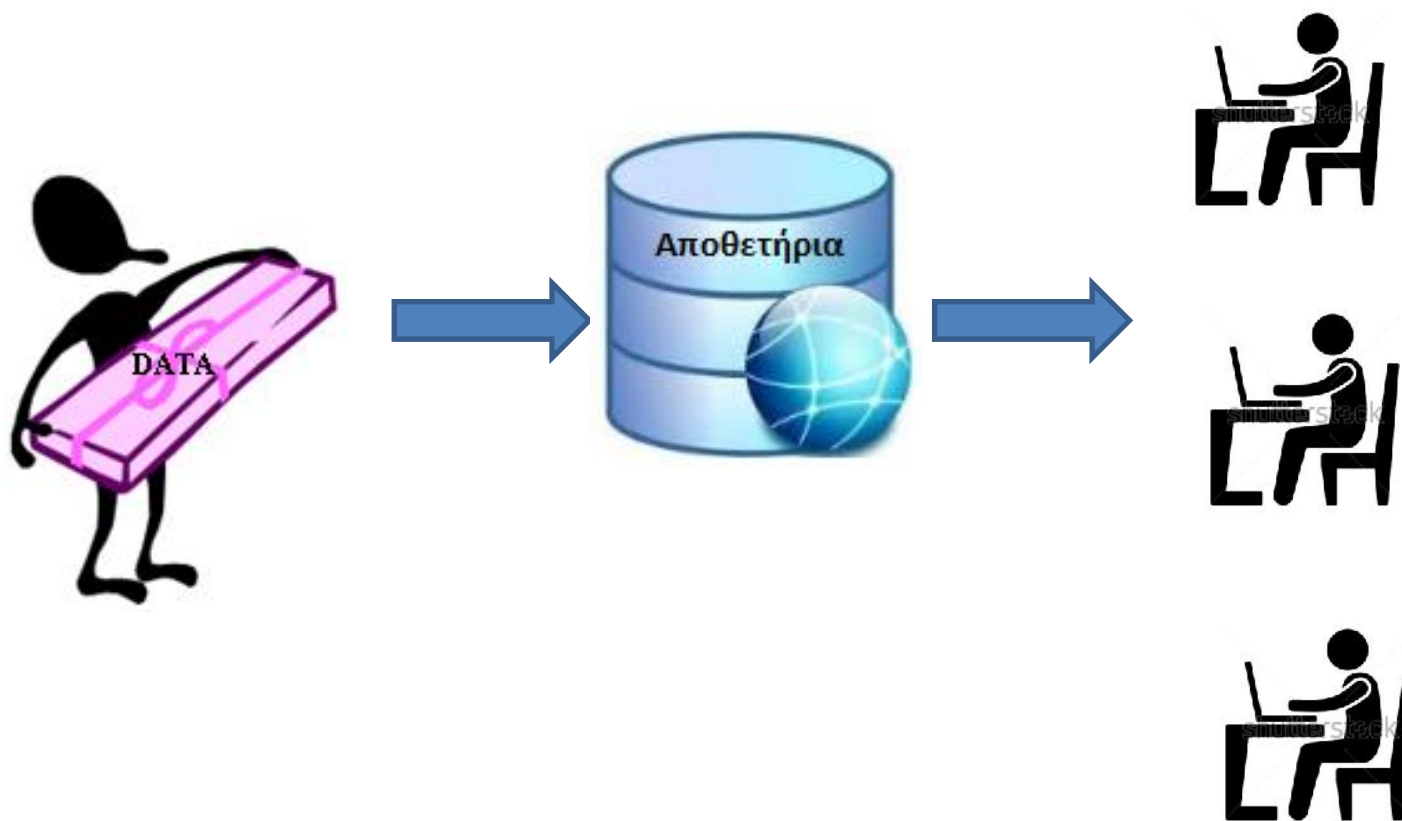
Διαμοιρασμός δεδομένων σε συνεργάτες



Διαμοιρασμός δεδομένων σε συνεργάτες



Διαμοιρασμός δεδομένων μέσω αποθετηρίων δεδομένων





Περιεχόμενα Εισήγησης

1. Ο κύκλος ζωής μίας έρευνας
2. Η δευτερογενής ανάλυση δεδομένων
3. Στρατηγικές μείωσης των σφαλμάτων και των περιορισμών των δεδομένων για την απρόσκοπτη δευτερογενή ανάλυση αυτών
 1. Έλεγχος δομής αρχείου δεδομένων και αντιστοίχισης με ερωτήσεις
 2. Έλεγχος ακεραιότητας αρχείων δεδομένων
 3. Ανωθυμοποίηση δεδομένων
 4. Τεκμηρίωση και παροχή πλαισίου για τα δεδομένα και την έρευνα



Γλωσσάρι

Όρος	Περιγραφή / Ορισμός
Μικροδεδομένα	Είναι τα δεδομένα που περιέχουν πληροφορίες για τα ίδια τα άτομα και όχι συγκεντρωτικά στοιχεία.
Μεταδεδομένα	Δεδομένα που περιγράφουν τα ίδια τα δεδομένα
Αθροιστικά δεδομένα	Πινακοποιημένα δεδομένα που προκύπτουν από επεξεργασία των μικροδεδομένων
Μέθοδος συλλογής δεδομένων	Δια ζώσης (F2F), τηλεφωνικώς, μέσω ταχυδρομείου, μέσω διαδικτύου.
CAI (Computer Assisted Interviewing)	Συστήματα που διευκολύνουν τη διενέργεια της συνέντευξης (CAPI - Personal Interviewing / CATI – Telephone Interviewing / CAWI – Web Interviewing)



ΜΙΚΡΟΔΕΔΟΜΕΝΑ

	id	gender	year_born	age	high_educ_level	marital_status	nationality_initial
1	3	Άνδρας	1978	37	Γυμνάσιο	Έγγαμος/η	Πακιστανική
2	4	Γυναίκα	1976	39	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Ελληνική
3	5	Άνδρας	1924	91	Λύκειο (Γενικά...	Χήρος/α	Ελληνική
4	6	Γυναίκα	1959	56	Λύκειο (Γενικά...	Έγγαμος/η	Ελληνική
5	7	Άνδρας	1965	50	Γυμνάσιο	Έγγαμος/η	Πακιστανική
6	9	Άνδρας	1954	61	ΑΕΙ, Ανώτατες...	Άγαμος/η	Ελληνική
7	11	Γυναίκα	1981	34	ΑΤΕΙ,ΤΕΙ,ΚΑΤ...	Άγαμος/η	Ελληνική
8	12	Άνδρας	1957	58	ΑΕΙ, Ανώτατες...	Διαζευγμέν...	Ελληνική
9	13	Άνδρας	1955	60	Δημοτικό	Έγγαμος/η	Ελληνική
10	14	Άνδρας	1943	72	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Ελληνική
11	15	Γυναίκα	1967	48	ΑΕΙ, Ανώτατες...	Άγαμος/η	Ελληνική
12	16	Άνδρας	1963	52	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Ελληνική
13	17	Άνδρας	1961	54	ΑΤΕΙ,ΤΕΙ,ΚΑΤ...	Έγγαμος/η	Ελληνική
14	18	Γυναίκα	1993	22	Γυμνάσιο	Άγαμος/η	Ρώσικη
15	19	Γυναίκα	1944	71	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Ελληνική
16	20	Γυναίκα	1967	48	ΑΤΕΙ,ΤΕΙ,ΚΑΤ...	Διαζευγμέν...	Ελληνική
17	21	Άνδρας	1957	58	ΑΤΕΙ,ΤΕΙ,ΚΑΤ...	Έγγαμος/η	Ελληνική
18	22	Γυναίκα	1963	52	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Άλλη
19	23	Άνδρας	1976	39	Δεν παρακολ...	Άγαμος/η	Ελληνική
20	24	Άνδρας	1940	75	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Ελληνική
21	25	Άνδρας	1970	45	Δεν παρακολ...	Έγγαμος/η	Πακιστανική
22	26	Άνδρας	1985	30	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Πακιστανική
23	27	Γυναίκα	1960	55	Λύκειο (Γενικά...	Έγγαμος/η	Ελληνική
24	30	Γυναίκα	1964	51	ΑΕΙ, Ανώτατες...	Χήρος/α	Ουκρανική
25	31	Άνδρας	1968	47	Μεταπτυχιακ...	Έγγαμος/η	Ελληνική
26	32	Άνδρας	1959	56	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Ελληνική
27	33	Άνδρας	1953	62	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Αλβανική
28	34	Άνδρας	1958	57	Μεταδευτεροβ...	Έγγαμος/η	Ελληνική
29	35	Άνδρας	1950	65	Διδακτορικό (...)	Έγγαμος/η	Ελληνική
30	36	Άνδρας	1956	59	ΑΕΙ, Ανώτατες...	Έγγαμος/η	Ελληνική



ΜΕΤΑΔΕΔΟΜΕΝΑ ΜΕΤΑΒΛΗΤΩΝ

	Name	Label	Values	Width	Decimals	Measure	Missing	Type
1	id	Κωδικός ερωτώμενου	None					
2	gender	Φύλο	{1, Άνδρ...					
3	year_born	Έτος γεννήσεως	None					
4	age	Ηλικία	None					
5	high_educ_i...	Υψηλότερο επίπεδο εκπαίδευσης που έχετε ...	ισης}...					
6	marital_status	Οικογενειακή Κατάσταση	{1, Άγαμ...					
7	nationality_i...	Υπηκοότητα	{1, Ελλη...					
8	nationality_...	Άλλη υπηκοότητα	None					
9	nationality_...	Υπηκοότητα [κωδικοποιημένη]	{1, Ελλη...					
10	employment...	Κατάσταση απασχόλησης	{1, Εργα...					
11	work_status	Κατάσταση εργασίας	{1, Αυτο...					
12	occupation_...	Επάγγελμα στη σημερινή εργασία	None					
13	ISCO08_C	Επάγγελμα κύριας δραστηριότητας στη σημε...	{01, Αξι...					
14	industry_cur...	Κύρια δραστηριότητα επιχείρησης στη σημερι...	None	194	0	Nominal	None	String
15	NACE2_C	Κύρια δραστηριότητα επιχείρησης στη σημερι...	{01, Φυτι...	8	0	Nominal	998	String

Value Labels

Value Labels

Value:

Label:

Spelling...

Add

Change

Remove

1 = "Δεν παρακολούθησα ποτέ καμία βαθμίδα"

2 = "Προσχολική εκπαίδευση"

3 = "Μερικές τάξεις δημοτικού"

4 = "Δημοτικό"

5 = "Γυμνάσιο"

6 = "Λύκειο (Γενικό, Ενιαίο, Τεχνικό Επαγγελματικό)"

OK Cancel Help



ΑΘΡΟΙΣΤΙΚΑ ΔΕΔΟΜΕΝΑ

Πίνακας 1
Μόνιμος Πληθυσμός κατά φύλο και Περιφέρεια

Περιγραφή	Σύνολα			Ποσοστά επί συνόλου Περιφέρειας	
	Σύνολο	Άρρενες	Θήλεις	Άρρενες	Θήλεις
ΣΥΝΟΛΟ ΧΩΡΑΣ	10.816.286	5.303.223	5.513.063	49,0	51,0
ΠΕΡΙΦΕΡΕΙΑ ΑΝΑΤΟΛΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ ΚΑΙ ΘΡΑΚΗΣ	608.182	299.643	308.539	49,3	50,7
ΠΕΡΙΦΕΡΕΙΑ ΚΕΝΤΡΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ	1.882.108	912.693	969.415	48,5	51,5
ΠΕΡΙΦΕΡΕΙΑ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ	283.689	141.779	141.910	50,0	50,0
ΠΕΡΙΦΕΡΕΙΑ ΗΠΕΙΡΟΥ	336.856	165.775	171.081	49,2	50,8
ΠΕΡΙΦΕΡΕΙΑ ΘΕΣΣΑΛΙΑΣ	732.762	362.194	370.568	49,4	50,6
ΠΕΡΙΦΕΡΕΙΑ ΣΤΕΡΕΑΣ ΕΛΛΑΔΑΣ	547.390	277.475	269.915	50,7	49,3
ΠΕΡΙΦΕΡΕΙΑ ΙΟΝΙΩΝ ΝΗΣΩΝ	207.855	102.400	105.455	49,3	50,7
ΠΕΡΙΦΕΡΕΙΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ	679.796	339.310	340.486	49,9	50,1
ΠΕΡΙΦΕΡΕΙΑ ΠΕΛΟΠΟΝΝΗΣΟΥ	577.903	291.777	286.126	50,5	49,5
ΠΕΡΙΦΕΡΕΙΑ ΑΤΤΙΚΗΣ	3.828.434	1.845.663	1.982.771	48,2	51,8
ΠΕΡΙΦΕΡΕΙΑ ΒΟΡΕΙΟΥ ΑΙΓΑΙΟΥ	199.231	99.984	99.247	50,2	49,8
ΠΕΡΙΦΕΡΕΙΑ ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ	309.015	155.865	153.150	50,4	49,6
ΠΕΡΙΦΕΡΕΙΑ ΚΡΗΤΗΣ	623.065	308.665	314.400	49,5	50,5

SURVEY METHODS

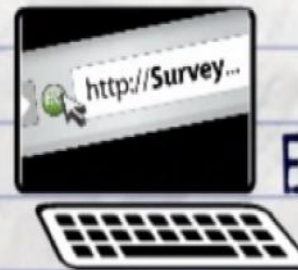
TELEPHONE



PERSONAL



MAIL



ELECTRONIC

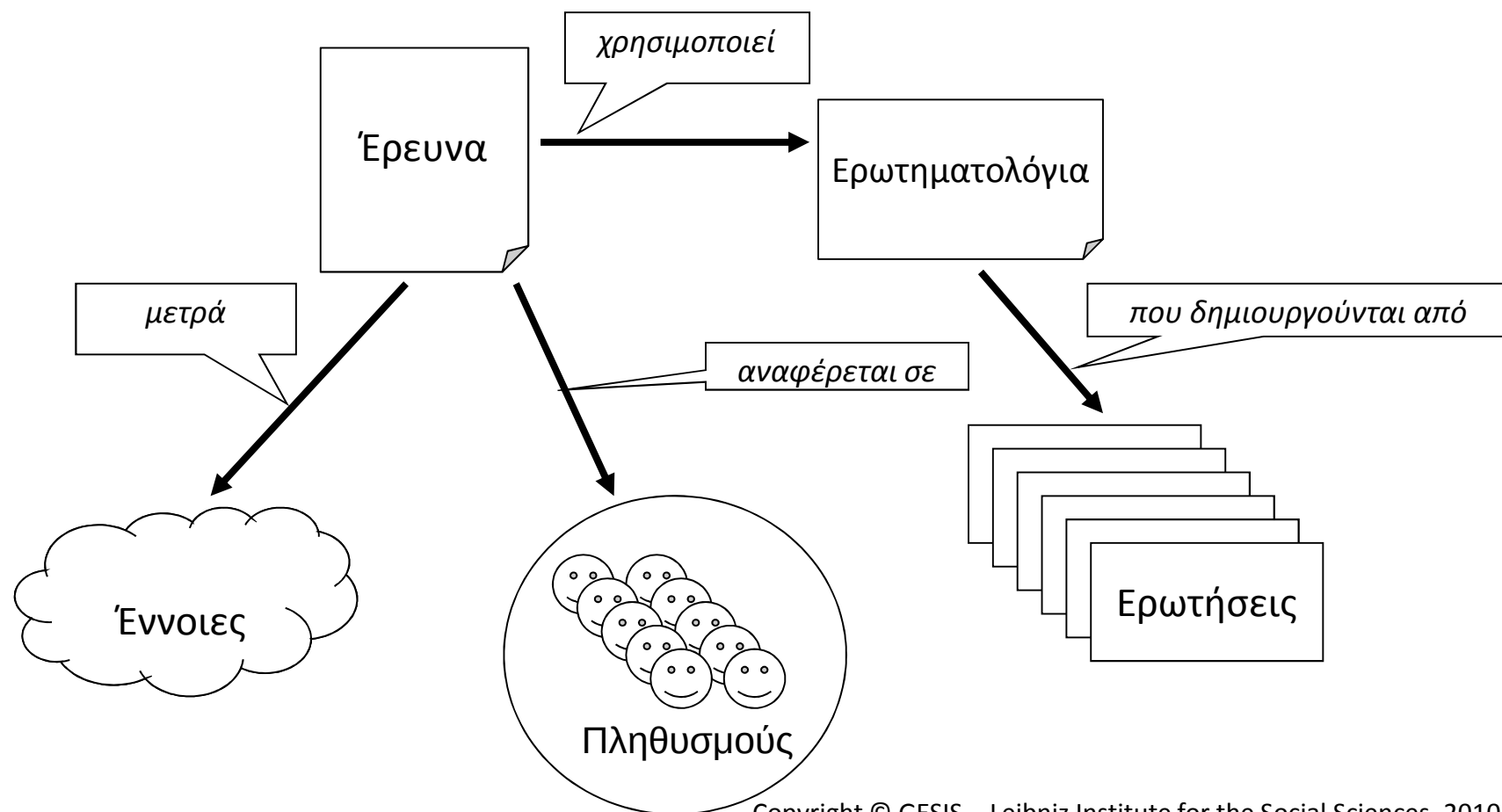
CREATED USING
BoWToon



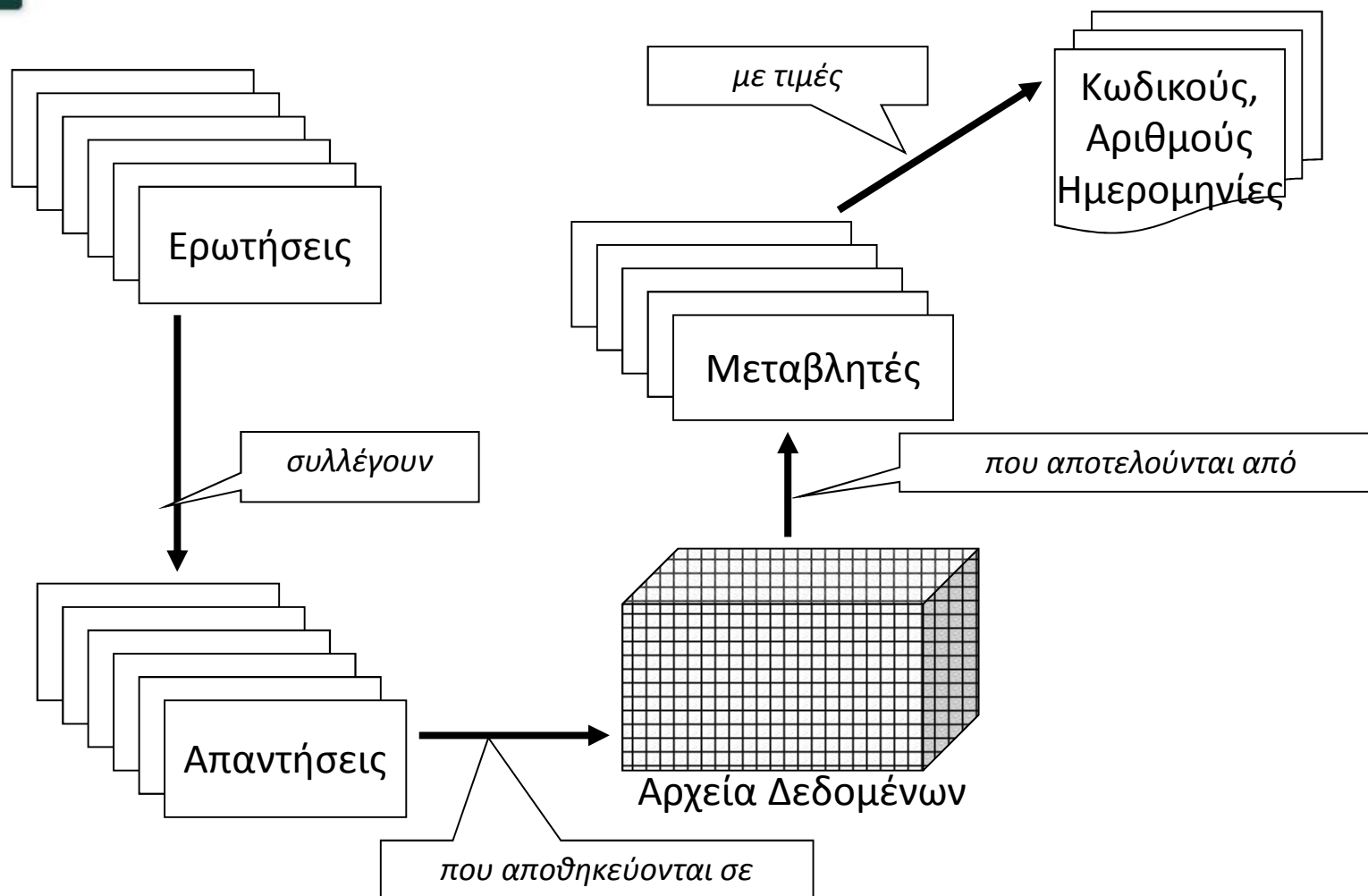
Μέρος 1:

Ο κύκλος ζωής μίας έρευνας

Η ΠΟΣΟΤΙΚΗ ΕΡΕΥΝΑ

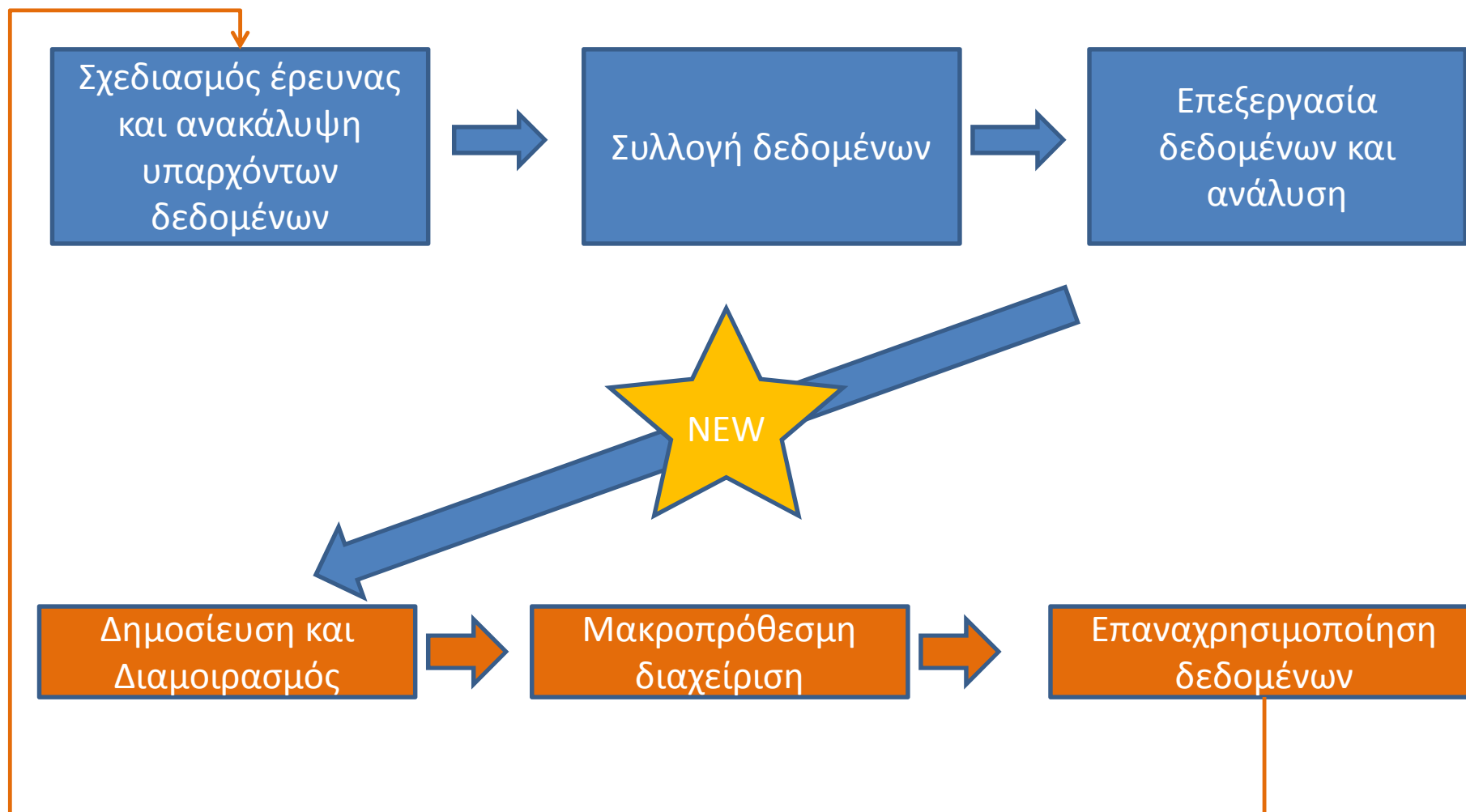


Η ΠΟΣΟΤΙΚΗ ΕΡΕΥΝΑ





Ο ΚΥΚΛΟΣ ΖΩΗΣ ΜΙΑΣ ΕΡΕΥΝΑΣ



Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: A guide to good practice*. Sage.



ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΑΝΑΚΑΛΥΨΗ
Σχεδιασμός έρευνας
Σχεδιασμός στρατηγικής διαχείρισης δεδομένων
Σχεδιασμός στρατηγικής συγκαταθέσεων για συλλογή δεδομένων
Σχεδιασμός συλλογής δεδομένων, επεξεργασία πρωτοκόλλων και προτύπων
Εύρεση και ανακάλυψη υπαρχουσών πηγών δεδομένων
ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ
Συλλογή δεδομένων μέσω καταγραφής, παρατήρησης, μέτρησης, πειραματισμού ή προσωμοίωσης
Δημιουργία και "σύλληψη" μεταδεδομένων
Απόκτηση υπαρχόντων δεδομένων από τρίτες πηγές
ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΝΑΛΥΣΗ
Εισαγωγή δεδομένων, ψηφιοποίηση, μεταγραφή και μετάφραση
Έλεγχος, επικύρωση, καθαρισμός και ανωνυμοποίηση δεδομένων
Δημιουργηθείσες μεταβλητές και δεδομένα
Ανάλυση δεδομένων
Ερμηνεία δεδομένων
Παραγωγή ερευνητικών αποτελεσμάτων
Συγγραφή δημοσιεύσεων
Παράθεση πηγών δεδομένων
Διαχείριση και αποθήκευση δεδομένων

ΔΗΜΟΣΙΕΥΣΗ ΚΑΙ ΔΙΑΜΟΙΡΑΣΜΟΣ
Θέσπιση πνευματικών δικαιωμάτων (copyright) των δεδομένων
Δημιουργία μεταδεδομένων ανακάλυψης και τεκμηρίωσης για τον χρήστη
Διανομή δεδομένων
Έλεγχος πρόσβασης στα δεδομένα
Προώθηση των δεδομένων
ΜΑΚΡΟΠΡΟΘΕΣΜΗ ΔΙΑΧΕΙΡΙΣΗ
Μετάβαση δεδομένων στο καλύτερο δυνατό format
Μετάβαση δεδομένων στο καταλληλότερο μέσο αποθήκευσης
Back up και αποθήκευση δεδομένων
Συλλογή και παραγωγή μεταδεδομένων και τεκμηρίωσης
Διατήρηση και διαχείριση δεδομένων
ΕΠΑΝΑΧΡΗΣΙΜΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ
Διενέργεια δευτερογενούς ανάλυσης
Διενέργεια έρευνας που ξεκινά από εκεί που σταμάτησε η προηγούμενη
Διενέργεια ανασκόπησης ερευνών
Λεπτομερής εξέταση των αποτελεσμάτων
Χρήση των δεδομένων για εκπαίδευση και εκμάθηση



Μέρος 2:

Η δευτερογενής ανάλυση των μικροδεδομένων



Η δευτερογενής ανάλυση των δεδομένων

1) Νέα ερευνητικά ερωτήματα που απαντώνται από υφιστάμενα δεδομένα

Μόλις καθοριστεί και διευκρινιστεί σαφώς το ερευνητικό πρόβλημα, η ερευνητική προσπάθεια προσανατολίζεται λογικά στη συλλογή των δεδομένων. Ο αναμενόμενος πειρασμός για τους αρχάριους ερευνητές είναι να υποστηρίξουν κάποιο είδος πρωτογενούς έρευνας μεταξύ των κατάλληλων ομάδων ερωτωμένων. Αυτό θα πρέπει να είναι η τελευταία επιλογή και όχι η πρώτη. Ο ερευνητής θα πρέπει να έχει κατά νου ότι και κάποιος άλλος μπορεί να έχει ερευνήσει αυτό το πρόβλημα ή παρόμοιο στο παρελθόν. Ο ερευνητής σε αυτό το σημείο καλείται να αναζητήσει υπάρχοντα δεδομένα (μικροδεδομένα ή αθροιστικά δεδομένα), τα οποία θα χρησιμοποιήσει για να απαντήσει στα ερευνητικά του ερωτήματα, τα οποία δύναται να διαφέρουν από αυτά που έθεσε ο αρχικός ερευνητής που ήταν υπεύθυνος για τη διεξαγωγή της αρχικής έρευνας. Παρόλα αυτά οι μεταβλητές της έρευνας δύναται εν τέλει να υποστηρίζουν τα νέα ερευνητικά ερωτήματα, τα οποία δεν αφορούσαν τα ερευνητικά ενδιαφέροντα του αρχικού ερευνητή.

2) Νέες μέθοδοι που εφαρμόζονται σε υφιστάμενα δεδομένα και για τα ίδια ή παρόμοια ερευνητικά ερωτήματα



Πλεονεκτήματα / περιορισμοί δευτερογενούς ανάλυσης δεδομένων

Πλεονεκτήματα

- Δεν απαιτείται διεξαγωγή έρευνας με προφανές αποτέλεσμα την οικονομία σε «κόστος» και χρόνο
- προτάσεις για βελτιωμένες στατιστικές μεθόδους που θα εφαρμοστούν στα δεδομένα
- παραγωγή αποτελεσμάτων που μπορούν να χρησιμοποιηθούν ως συγκριτικά με αυτά της πρωτογενούς ανάλυσης

Περιορισμοί

- τύπος δεδομένων προς διάθεση (μικροδεδομένα ή αθροιστικά δεδομένα;)
- προβλήματα εφαρμογής (κατηγορίες, χρόνος διεξαγωγής και αναφοράς ερευνών)
- Τα δεδομένα περιέχουν δομικά λάθη καθώς και λάθη εισαγωγής δεδομένων
- Δεν έχει καταγραφεί πρόσθετη πληροφορία, ώστε να γίνονται τα δεδομένα κατανοητά [Ποιος ήταν ο σκοπός της έρευνας; Πότε διεξηχθεί; Ποιός ήταν ο πληθυσμός: Είχαν συμπεριληφθεί οι ερωτώμενοι που έπρεπε στην έρευνα; Πώς λήφθηκαν οι πληροφορίες (τηλεφωνικώς, διαδικτυακά, προσωπική συνέντευξη ή με ταχυδρομείο); Ποιό ήταν το ποσοστό άρνησης; Υπάρχουν οδηγίες επεξεργασίας των δεδομένων;]



Μέρος 3:

**Στρατηγικές μείωσης των περιορισμών των δεδομένων για την
απρόσκοπτη δευτερογενή ανάλυση των μικροδεδομένων**



ΕΛΕΓΧΟΣ ΔΟΜΗΣ ΑΡΧΕΙΟΥ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΝΤΙΣΤΟΙΧΗΣΗΣ ΜΕ ΤΟ ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

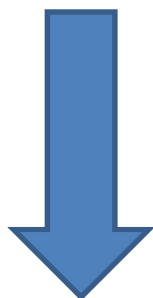
Τύπος ερώτησης	Πλήθος μεταβλητών
Κλειστή ερώτηση που επιτρέπει μόνο μία απάντηση	Μία
Κλειστή ερώτηση που επιτρέπει πολλαπλές απαντήσεις	Τόσες διχοτομικές μεταβλητές όσες και οι δυνατές απαντήσεις
Ημιανοιχτή ερώτηση	Δύο μεταβλητές: μία αριθμητική και μία κειμενική
Ανοιχτή ερώτηση	Τουλάχιστον δύο μεταβλητές: Κειμενική μεταβλητή και αριθμητική
Ερώτηση τύπου πλέγματος με υποερωτήματα	Τόσες μεταβλητές όσες και τα υποερωτήματα



ΑΝΤΙΣΤΟΙΧΙΣΗ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΩΝ

1. Φύλο

- ο Άνδρας
- ο Γυναίκα



The dialog box shows the configuration for the 'gender' variable. The 'Value' field is empty, and the 'Label' field is empty. The list of values and labels shows '1 = "Άνδρας"' and '2 = "Γυναίκα"'. The 'Add', 'Change', and 'Remove' buttons are visible. The 'OK', 'Cancel', and 'Help' buttons are at the bottom.



Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
gender	Numeric	1	0	Φύλο	{1, Άνδρα...	None	8	≡ Right	Nominal



ΑΝΤΙΣΤΟΙΧΙΣΗ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΩΝ

5. Ποιο είναι η υπηκοότητά σας;

- ☐ Ελληνική
- ☐ Αλβανική
- ☐ Βουλγάρικη
- ☐ Ρουμάνικη
- ☐ Πακιστανική
- ☐ Ινδική
- ☐ Μπαγκλαντέζικη
- ☐ Ρώσικη
- ☐ Μολδαβική
- ☐ Γεωργιανή
- ☐ Ουκρανική
- ☐ Άλλο:



The Value Labels dialog box in SPSS shows the following mappings:

Value	Label
1	"Ελληνική"
2	"Αλβανική"
3	"Βουλγάρικη"
4	"Ρουμάνικη"
5	"Πακιστανική"
6	"Ινδική"
7	"Μπαγκλαντέζικη"



nationality_initial	Numeric	2	0	Υπηκοότητα	{1, Ελληνι...	None	10	Right	Nominal
nationality_other	String	48	0	Άλλη υπηκοότητα	None	None	18	Left	Nominal
nationality_coded	Numeric	2	0	Υπηκοότητα [κωδικοποιημένη]	{1, Ελληνι...	None	28	Right	Nominal

ΑΝΤΙΣΤΟΙΧΙΣΗ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΩΝ

8. Να περιγράψετε, με τη μεγαλύτερη δυνατή λεπτομέρεια, το είδος της κύριας δραστηριότητας (επάγγελμα) που έχετε στη σημερινή εργασία σας (π.χ. υπάλληλος λογιστηρίου, οικιακή βοηθός κ.λπ.)

D

ISCO_08 (2 digits)

- 01 - Commissioned armed forces officers
- 02 - Non-commissioned armed forces officers
- 03 - Armed forces occupations, other ranks
- 11 - Chief executives, senior officials and legislators
- 12 - Administrative and commercial managers
- 13 - Production and specialised services managers
- 14 - Hospitality, retail and other services managers
- 21 - Science and engineering professionals

occup_current	String	150	0	Επάγγελμα (τρέχουσα εργασία)	None
ISCO_08_1_digit	String	8	0	ISCO08 (Πρώτο ψηφίο)	{0, Armed Forces}...
ISCO_08_2_digits	String	8	0	ISCO08 (Δεύτερο ψηφίο)	{01, Commissioned armed forces officers}...

C

ISCO_08 (1 digit)

- 0 - Armed forces occupations
- 1 - Managers
- 2 - Professionals
- 3 - Technicians and associate professions
- 4 - Clerical support workers
- 5 - Service and sales workers
- 6 - Skilled agricultural, forestry and fishery workers
- 7 - Craft and related trades workers



ΑΝΤΙΣΤΟΙΧΙΣΗ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΩΝ

19. Συνέβη μήπως οποιοδήποτε από τα ακόλουθα κατά τη διάρκεια της συνέντευξης (σημειώστε ότι ισχύει);

- ο Ο/Η ερωτώμενος/η είχε συνομιλία με κάποιον άλλο στο σπίτι εκτός από τον/την συνεντευκτή/τρια
- ο Ο/Η ερωτώμενος/η απάντησε στο τηλέφωνο, σε ηλεκτρονικό μήνυμα ή σε ηλεκτρονικό ταχυδρομείο (e-mail)
- ο Ο/Η ερωτώμενος/η πρόσεχε παιδιά
- ο Ο/Η ερωτώμενος/η έκανε οικιακές εργασίες, όπως μαγείρεμα ή πλύσιμο
- ο Στο άμεσο περιβάλλον του/της ερωτώμενου/ης λειτουργούσαν τηλεόραση, ραδιόφωνο, συστήματα μουσικής ή ηλεκτρονικά παιχνίδια
- ο Ο/Η ερωτώμενος/η διέκοπτε τη συνένταυξη εξαιτίας άλλης δραστηριότητας, εργασίας ή άλλου συμβάντος

1 = "Έχει επιλεγθεί"
2 = "Δεν Έχει επιλεγθεί"

Name	Type	Width	De...	Label	Values
speak_other	Numeric	8	0	Ο/ Η ερωτώμενος/η είχε συνομιλία με ...	{1, Έχει ε...
answered_other	Numeric	8	0	Ο/ Η ερωτώμενος/η απάντησε στο τη...	{1, Έχει ε...
take_care_children	Numeric	8	0	Ο/Η ερωτώμενος/η πρόσεχε παιδιά	{1, Έχει ε...
do_homeworks	Numeric	8	0	Ο/Η ερωτώμενος/η έκανε οικιακές ερ...	{1, Έχει ε...
turn_on_electronic_...	Numeric	8	0	Στο άμεσο περιβάλλον του/της ερωτώ...	{1, Έχει ε...
interrupt_interview	Numeric	8	0	Ο/Η ερωτώμενος/η διέκοπτε τη συνέντ...	{1, Έχει ε...

ΑΝΤΙΣΤΟΙΧΙΣΗ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΜΕΤΑΒΛΗΤΩΝ

*** Ε. ΕΜΠΙΣΤΟΣΥΝΗ ΣΤΟΥΣ ΘΕΣΜΟΥΣ ***

14. Παρακαλώ πείτε μου τώρα, πόσο εμπιστεύεστε εσείς προσωπικά κάθε ένα από τα παρακάτω που θα σας διαβάσω. Το 0 σημαίνει ότι δεν έχετε καμία εμπιστοσύνη και το 10 σημαίνει ότι έχετε απόλυτη εμπιστοσύνη στους θεσμούς στην Ελλάδα.

	0 Καμία Εμπιστοσύνη	1	2	3	4	5	6	7	8	9	10 Απόλυτη Εμπιστοσύνη	Δ.Γ	Δ.Α
Βουλή													
Πολιτικά Κόμματα													
Δικαιοσύνη (Δικαστήρια)													
Σώματα Ασφαλείας													
Ένοπλες Δυνάμεις													
Μη Κυβερνητικές Οργανώσεις													
Εκκλησία													

Missing Values

☒ Discrete missing values

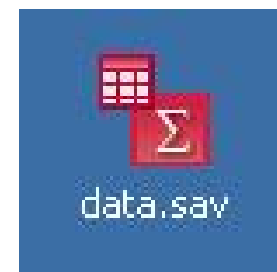
11 12

0 = "0 καμία εμπιστοσύνη"
 1 = "1"
 2 = "2"
 3 = "3"
 4 = "4"
 5 = "5"
 6 = "6"
 7 = "7"
 8 = "8"
 9 = "9"
 10 = "10 απόλυτη εμπιστοσύνη"
 11 = "Δ.Γ"
 12 = "Δ.Α"

Name	Type	Width	De...	Label	Values	Missing
trst_parlament	Numeric	2	0	[Βουλή] Εμπιστοσύνη στους θεσμούς	{0, 0 καμί...	11, 12
trst_parties	Numeric	2	0	[Πολιτικά κόμματα] Εμπιστοσύνη στο...	{0, 0 καμί...	11, 12
trst_justice	Numeric	2	0	[Δικαιοσύνη (Δικαστήρια)] Εμπιστοσ...	{0, 0 καμί...	11, 12
trst_security_forces	Numeric	2	0	[Σώματα Ασφαλείας] Εμπιστοσύνη σ...	{0, 0 καμί...	11, 12
trst_armed_forces	Numeric	2	0	[Ένοπλες Δυνάμεις] Εμπιστοσύνη σι...	{0, 0 καμί...	11, 12
trst_ngo	Numeric	2	0	[Μη Κυβερνητικές Οργανώσεις] Ερπι...	{0, 0 καμί...	11, 12
trst_church	Numeric	2	0	[Εκκλησία] Εμπιστοσύνη στους θεσμ...	{0, 0 καμί...	11, 12



ΕΛΕΓΧΟΣ ΑΚΕΡΑΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ



Η μεταγραφή των πληροφοριών από ένα έντυπο ερωτηματολόγιο σε ψηφιακή και δομημένη μορφή μπορεί να εισάγει διάφορους τύπους λαθών, συμπεριλαμβανομένων των σφαλμάτων δακτυλογράφησης, κωδικούς που δεν έχουν νόημα και εγγραφές που δεν ταιριάζουν. Για το λόγο αυτό, συνίσταται μια στρατηγική συλλογής δεδομένων που καταγράφει δεδομένα απευθείας κατά τη διάρκεια της διαδικασίας της συνέντευξης. Έλεγχοι συνέπειας μπορούν στη συνέχεια να ενταχθούν στη διαδικασία συλλογής των δεδομένων μέσω της χρήσης του λογισμικού / CAPI, CATI για να διορθωθούν τα προβλήματα κατά τη διάρκεια μιας συνέντευξης.

Ωστόσο, ακόμη και αν τα δεδομένα έχουν μεταγραφεί (είτε από φόρμες επαφής ή πίνακες), αρκετά μέτρα μπορούν να ληφθούν εκ των προτέρων για να μειωθεί η συχνότητα εμφάνισης τέτοιων σφαλμάτων.

Παρά τις καλύτερες προσπάθειες, τα λάθη είναι βέβαιο ότι θα συμβούν ανεξάρτητα από τον τρόπο συλλογής των στοιχείων. Παρακάτω παρατίθεται μια λίστα ελέγχων που πρέπει να διενεργηθούν σε σχέση με το αρχείο δεδομένων .



ΛΙΣΤΑ ΜΕ ΕΛΕΓΧΟΥΣ ΑΚΕΡΑΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ

ΕΛΕΓΧΟΣ ΑΚΕΡΑΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ

Έλεγχοι κωδικών και τιμών εκτός αποδεκτού εύρους. Παράδειγμα: ενώ η μεταβλητή gender δέχεται τιμές 1 → Άντρας και 2 → Γυναίκα, η κατανομή συχνοτήτων μέσω της διαδικασίας Frequencies ενδέχεται να δείξει ότι υφίστανται τιμές εκτός των δύο αυτών κωδικών που πρέπει εκ των υστέρων να διορθωθούν. Επίσης η κατανομή συχνοτήτων θα αναδείξει και λάθη «εκτός του αποδεκτού εύρους»: για παράδειγμα 99 παιδιά (που πιθανά να προέκυψε από λάθος πληκτρολόγησης της τιμής 9).

Έλεγχοι συνοχής: οι έλεγχοι συνοχής απαιτούν ουσιαστική γνώση της μελέτης. Συνήθως, αυτοί περιλαμβάνουν συγκρίσεις μεταξύ των μεταβλητών. Οι έλεγχοι αυτοί μπορούν να αποκαλύψουν ασυμφωνίες μεταξύ των απαντήσεων σε ερωτήσεις φίλτρα και ερωτήσεων που εξαρτώνται από την απάντηση στο φίλτρο. Για παράδειγμα, ένας ερωτώμενος δηλώνει ότι δεν πήγε στην εκκλησία την προηγούμενη εβδομάδα αλλά υφίσταται απάντηση στην ερώτηση σχετικά με τη συχνότητα εκκλησιασμού την προηγούμενη εβδομάδα ενώ στη θέση της απάντησης θα έπρεπε να υπάρχει ελλείπουσα τιμή με λεκτικό «Μη εφαρμόσιμη».



ΛΙΣΤΑ ΜΕ ΕΛΕΓΧΟΥΣ ΑΚΕΡΑΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ

ΕΛΕΓΧΟΣ ΑΚΕΡΑΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ

- **Λογικοί έλεγχοι:** προκύπτουν από τη λογική διασταύρωση κάποιων μεταβλητών π.χ.: ανώτατο επίπεδο εκπαίδευσης και ηλικία

- **Έλεγχοι συνοχής:** οι έλεγχοι συνοχής απαιτούν ουσιαστική γνώση της μελέτης. Συνήθως, αυτοί περιλαμβάνουν συγκρίσεις μεταξύ των μεταβλητών. Οι έλεγχοι αυτοί μπορούν να αποκαλύψουν ασυμφωνίες μεταξύ των απαντήσεων σε ερωτήσεις φίλτρα και ερωτήσεων που εξαρτώνται από την απάντηση στο φίλτρο. Για παράδειγμα, ένας ερωτώμενος δηλώνει ότι δεν πήγε στην εκκλησία την προηγούμενη εβδομάδα αλλά υφίσταται απάντηση στην ερώτηση σχετικά με τη συχνότητα εκκλησιασμού την προηγούμενη εβδομάδα ενώ στη θέση της απάντησης θα έπρεπε να υπάρχει ελλείπουσα τιμή με λεκτικό «Μη εφαρμόσιμη».

- **Ταίριασμα εγγραφών και έλεγχος πλήθους.** Παράδειγμα ιεραρχικών αρχείων: νοικοκυριού ατόμων με κωδικούς ταυτοποίησης ατόμων και νοικοκυριών.



ΑΝΩΝΥΜΟΠΟΙΗΣΗ

- Πριν τα δεδομένα μίας έρευνας μοιραστούν με άλλους ερευνητές ή προβούν σε διαδικασία αρχειοθέτησης, μπορεί να χρειαστεί να ανωνυμοποιηθούν έτσι ώστε να μην μπορούν να εντοπιστούν τα άτομα, οι οργανισμοί ή οι επιχειρήσεις που έλαβαν μέρος στην έρευνα.
- Από την άλλη πρέπει να διατηρηθούν τα δεδομένα με όσο το δυνατόν περισσότερη ουσιαστική πληροφορία.
- Η ανωνυμοποίηση απαιτείται για ηθικούς λόγους, ώστε να προστατευτούν οι ταυτότητες των ατόμων στην έρευνα και για νομικούς λόγους ώστε να μην αποκαλυφθούν προσωπικά δεδομένα.
- Τα δεδομένα προσωπικού χαρακτήρα δεν πρέπει ποτέ να αποκαλύπτονται από τις υπόλοιπες πληροφορίες και μεταβλητές της έρευνας, εκτός αν ένας συμμετέχων έχει δώσει τη ρητή συγκατάθεση του (ιδεατά γραπτώς).
- Οι διαδικασίες για την ανωνυμοποίηση των δεδομένων πρέπει πάντοτε να εξετάζονται παράλληλα με την απόκτηση συγκατάθεσης για τον διαμοιρασμό δεδομένων ή/και την επιβολή περιορισμών στην πρόσβαση.
- Ανωνυμοποίηση των δεδομένων της έρευνας μπορεί να είναι χρονοβόρα και επομένως δαπανηρή, αλλά ο έγκαιρος προγραμματισμός βοηθά στη μείωση αυτών των δαπανών.



Αναγνωριστικά (Identifiers)

ΑΝΩΝΥΜΟΠΟΙΗΣΗ

Η ταυτότητα ενός ατόμου μπορεί να αποκαλυφθεί από:

- άμεσα αναγνωριστικά (direct identifiers) όπως: ονόματα, διευθύνσεις, αριθμούς τηλεφώνου, εικόνες, αριθμούς ταυτότητας, που επιτρέπουν την άμεση αναγνώριση του ερωτώμενου και τα οποία δεν είναι απαραίτητα για στατιστικούς ή ερευνητικούς σκοπούς και ως εκ τούτου πρέπει να αφαιρεθούν από το σύνολο δεδομένων που πρόκειται να δημοσιοποιηθεί. Τα άμεσα αναγνωριστικά στοιχεία συλλέγονται συχνά ως μέρος της διαδικασίας διαχείρισης της έρευνας.
- έμμεσα αναγνωριστικά (indirect identifiers): τα οποία είναι χαρακτηριστικά που όταν συνδυαστούν με άλλες διαθέσιμες πληροφορίες μπορούν να καταλήξουν στην ταυτοποίηση ενός ερωτώμενου. Για παράδειγμα, ο συνδυασμός των μεταβλητών: περιοχή κατοικίας, ηλικία, φύλο και επάγγελμα θα μπορούσε να καταλήξει σε ένα άτομο από τη συγκεκριμένη περιοχή. Τα έμμεσα αναγνωριστικά είναι απαραίτητα για ερευνητικούς σκοπούς.



Τεχνικές Ανωνυμοποίησης (1/2)

ΑΝΩΝΥΜΟΠΟΙΗΣΗ

Οι τεχνικές ανωνυμοποίησης αφορούν στην εξάλειψη μεταβλητών, στην συνάθροιση τιμών και κατηγοριών των μεταβλητών ή στην μείωση της ακρίβειας και της λεπτομέρειας της κειμενικής πληροφορίας μίας μεταβλητής. Πιο συγκεκριμένα οι τεχνικές αφορούν στα κάτωθι:

- Αφαίρεση άμεσων αναγνωριστικών (direct identifiers). Παράδειγμα: απομάκρυνση των ονομάτων των ερωτώμενων ή αντικατάσταση με κάποιον κωδικό. Αφαίρεση διευθύνσεων, φορέων, τηλεφώνων κ.α.
- Ομαδοποίηση κατηγοριών ή μείωση της ακρίβειας μιας μεταβλητής, όπως η ηλικία και ο τόπος διαμονής. Κατά γενικό κανόνα, αναφέρεται το χαμηλότερο επίπεδο γεωαναφοράς που δεν αποκαλύπτει στοιχεία, στάσεις και αντιλήψεις κάποιου ερωτώμενου. Το ακριβές επίπεδο εξαρτάται από το είδος των δεδομένων που συλλέγονται, αλλά λεπτομερείς γεω-αναφορές, όπως ταχυδρομικοί κώδικες ή ονόματα μικρών πόλεων ή χωριών είναι συνήθως προβληματικές. Στην περίπτωση που εντοπιστεί πρόβλημα η λύση αφορά στην επανακωδικοποίηση των μεταβλητών σε ευρύτερες κατηγορίες. Παράδειγμα: Καταγράψτε το έτος γέννησης και όχι την ημέρα, το μήνα και το έτος, ή τα πρώτα 2-3 ψηφία του TK και όχι τον πλήρη κώδικα, ή καταγράψτε την ταξινόμηση των επαγγελματιών ISCO σε μονοψήφιο ή διψήφιο επίπεδο. Αποφύγετε τον τριψήφιο ή τετραψήφιο κωδικό.



Τεχνικές Ανωνυμοποίησης (2/2)

- Γενίκευση του νοήματος κειμενικής πληροφορίας στην περίπτωση που ο ερωτώμενος αναφέρεται σε λεπτομέρειες που μπορεί να αποκαλύψουν την ταυτότητά του. Παράδειγμα: Αναφορά σε λεπτομερούς τομείς της ιατρικής ειδικότητας θα μπορούσαν να αποκαλύψουν την ταυτότητα του γιατρού.
- Περιορισμός του εύρους μίας συνεχούς μεταβλητής, ώστε να μην αποκαλύπτονται οι ακραίες τιμές – είτε πολύ μεγάλες είτε πολύ μικρές. Οι ασυνήθιστα μεγάλες ή μικρές τιμές μπορεί να επανακωδικοποιηθούν σε νέες κατηγορίες. Παράδειγμα: Το ετήσιο εισόδημα θα μπορούσε να είναι κωδικοποιημένο για τα υψηλόμισθα άτομα, ως εξής: «100.000 ευρώ ή άνω».
- Ανωνυμοποίηση σχεσιακών δεδομένων, όπου υφίστανται σχεσιακά ή συνδεδεμένα σύνολα δεδομένων.
- Ανωνυμοποίηση δεδομένων γεωγραφικής αναφοράς συντεταγμένων – σημείου μέσω της αντικατάστασης αυτών με χαρακτηριστικά που δεν αποκαλύπτουν την ταυτότητα (π.χ. Αντικατάσταση με πολύγωνα ή με μεταβλητές που χαρακτηρίζουν την επιλεχθείσα περιοχή, όπως δείκτη φτώχειας, πυκνότητα πληθυσμού κ.α. Κατά προτίμηση όμως κρατήστε τις γεω-αναφορές ως είναι και επιβάλετε περιορισμούς πρόσβασης στα δεδομένα.



Τεκμηρίωση σε επίπεδο έρευνας

Η καλού επιπέδου τεκμηρίωση των δεδομένων σε «Επίπεδο Έρευνας» θα πρέπει να περιλαμβάνει την παρακάτω πληροφορία:

- Τίτλο έρευνας, περίγραμμα σχεδιασμού έρευνας και συλλογής δεδομένων: ιστορικό της έρευνας, στόχοι, αντικείμενα και ερευνητικές υποθέσεις, φορείς και ερευνητές δράσεις και χρηματοδότες
- Μεθόδους συλλογής δεδομένων: πρωτόκολλα συλλογής δεδομένων, σχεδιασμό δειγματοληψίας, ροές εργασιών, όργανα μέτρησης που χρησιμοποιήθηκαν, λογισμικό και υλικό που χρησιμοποιήθηκε, γεωγραφική και χρονική κάλυψη
- Δομή των αρχείων δεδομένων και πρόσθετη τεκμηρίωση με το σύνολο των περιπτώσεων και των μεταβλητών που περιλαμβάνουν, σχέσεις μεταξύ αρχείων με την καταγραφή των πρωτευόντων και δευτερευόντων κλειδιών
- Δευτερογενείς πηγές δεδομένων που χρησιμοποιήθηκαν και προέλευση αυτών
- Διαδικασίες επικύρωσης, ελέγχου, καθαρισμού δεδομένων στο πλαίσιο των ποιοτικών ελέγχων που διεξήχθησαν
- Καταγραφή των αλλαγών που διενεργήθηκαν στα δεδομένα από τη στιγμή που δημοσιοποιήθηκαν, ταυτοποίηση και αναγνώριση των διαφορετικών εκδόσεων των συνόλων δεδομένων



Τεκμηρίωση σε επίπεδο έρευνας

ΤΕΚΜΗΡΙΩΣΗ

- για χρονοσειρές ή διαχρονικές έρευνες, καταγράφονται οι αλλαγές που διενεργήθηκαν στη μεθοδολογία, στο περιεχόμενο των μεταβλητών, στο κείμενο της ερώτησης, στις ετικέτες των μεταβλητών, στις μετρήσεις και στη δειγματοληψία και στην περίπτωση ερευνών panel καταγράφεται ο τρόπος με τον οποίο έγινε η διαχείριση των ατόμων στον χρόνο και μεταξύ των κυμάτων
- καταγράφεται πληροφορία σε σχέση με την πρόσβαση και την εμπιστευτικότητα των δεδομένων και συνθήκες χρήσης
- Δημοσιεύσεις, παρουσιάσεις και άλλα ερευνητικά αποτελέσματα



Τεκμηρίωση σε επίπεδο δεδομένων

Ονόματα, ετικέτες και περιγραφές των μεταβλητών: Οι ετικέτες των μεταβλητών θα πρέπει να είναι σύντομες και όπου αυτό είναι εφικτό θα πρέπει να καταγράφεται στην ετικέτα η μονάδα μέτρησης και η αναφορά στον αριθμό της ερώτησης του ερωτηματολογίου.

Παράδειγμα: μεταβλητή με όνομα: 'q11hexw' και ετικέτα: «Q11: hours spent taking physical exercise in a physical week' όπου η ετικέτα εμπεριέχει και τη μονάδα μέτρησης αλλά και αναφορά στον κωδικό της ερώτησης από την οποία προήλθε.

Κωδικοί και ετικέτες κατηγοριών: για παράδειγμα για την μεταβλητή q1sex = 'sex of respondent' οι κωδικοί και οι κατηγορίες δύναται να είναι οι: 1 → female, 2 → male, 8 → Don't know, 9 → not answered



Τεκμηρίωση σε επίπεδο δεδομένων

Κωδικοποίηση και ταξινομητικά σχήματα (τα οποία πρέπει να επεξηγούνται πλήρως): ISCO 2008 για την ταξινόμηση των επαγγελμάτων, ISO 3166 κωδικοί χωρών με 2 γράμματα. Παρατηρείστε ότι κάποιες από τις ταξινομήσεις αλλάζουν μέσα στον χρόνο.

Κωδικοί για ελλείπουσες τιμές (missing values): κενά, ελλείπουσες τιμές συστήματος και τα μηδενικά θα πρέπει να αποφεύγονται.

Μεταβλητές που παρήχθησαν μέσω ενός σύνθετου αλγόριθμου θα πρέπει να συμπεριλαμβάνουν τον ίδιο τον αλγόριθμο, ως πρόσθετη τεκμηρίωση.

Μεταβλητές στάθμισης και οδηγίες για τη χρήση τους

Καθορισμός πληθυσμού που αποκλίνει από τον γενικό πληθυσμό της έρευνας και σε επίπεδο μεταβλητής π.χ. σε ερωτήσεις που ρωτούνται μόνο οι εργαζόμενοι ή οι άνεργοι.



Μεταδεδομένα για την ορθή παραπομπή σε δεδομένα (DataCite 2011)

ΤΕΚΜΗΡΙΩΣΗ

DataCite mandatory metadata elements are	Υποχρεωτικά πεδία μεταδεδομένων
Identifier	Ταυτοποιητής
Creator	Δημιουργός
Title	Τίτλος
Publisher	Εκδότης
Publication year	Έτος έκδοσης
DataCite Optional elements	Προαιρετικά πεδία μεταδεδομένων
Subject	Θέμα
Contributor	Συνεισφέρων
date	ημερομηνία
language	γλώσσα
resource type	Τύπος πηγής
alternate identifier	Εναλλακτικός ταυτοποιητής
related identifier	Σχετικός ταυτοποιητής
size	Μέγεθος
format	Format
version	Έκδοση
rights	Δικαιώματα
description	Περιγραφή
DataCite administrative metadata elements are:	Διαχειριστικά μεταδεδομένα (DataCite):
Date of last metadata update	Ημερομηνία τελευταίας ενημέρωσης μεταδεδομένων
metadata version number	Αριθμός έκδοσης μεταδεδομένων

