

The Greek Review of Social Research

Vol 147 (2016)

147 B'



The Dream Machine

Απόστολος Λιναρδής, Δήμητρα Κονδύλη

doi: [10.12681/grsr.11650](https://doi.org/10.12681/grsr.11650)

Copyright © 2017, Απόστολος Λιναρδής, Δήμητρα Κονδύλη



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

To cite this article:

Λιναρδής Α., & Κονδύλη Δ. (2017). The Dream Machine. *The Greek Review of Social Research*, 147, 3–39.
<https://doi.org/10.12681/grsr.11650>

Απόστολος Λιναρδής*, Δήμητρα Κονδύλη**

THE DREAM MACHINE¹

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία υιοθετεί ως θεωρητικές συνισταμένες τους κανόνες που τέθηκαν στο έργο *Metanet*, το διεθνές πρότυπο τεκμηρίωσης *Data Documentation Initiative 3*, τις μεθοδολογικές διαδικασίες που ακολουθήθηκαν στη διεθνή έρευνα για τις δεξιότητες ενηλίκων *PIAAC* καθώς και τις νέες προϋποθέσεις του ευρωπαϊκού πλαισίου έρευνας για την επαναχρησιμοποίηση, μακροπρόθεσμη διατήρηση και αρχειοθέτηση των δεδομένων που συλλέχθηκαν, δημιουργήθηκαν ή έτυχαν επεξεργασίας στο πλαίσιο ενός έργου, προκειμένου να περιγράψει ένα ιδεατό σύστημα αναφορικά με τα κοινωνικά δεδομένα. Το ιδεατό σύστημα που επιχειρούμε να περιγράψουμε θα επιτρέπει στους χρήστες τον σχεδιασμό και τη διεξαγωγή μιας έρευνας, την αποτελεσματική αρχειοθέτηση και διαχείριση των δεδομένων της, καθώς και τη διάχυση αποτελεσμάτων, δεδομένων και σχετικής τεκμηρίωσης προς δευτερογενή χρήση.

Λέξεις κλειδιά: ερευνητικές διαδικασίες, μεταδεδομένα, δεδομένα, επαναχρησιμοποίηση

* Κύριος ερευνητής ΕΚΚΕ, μαθηματικός, στατιστικός, e-mail: alinardis@ekke.gr

** Κύρια ερευνήτρια ΕΚΚΕ, κοινωνιολόγος, ανθρωπολόγος, e-mail: dkondyli@ekke.gr

1. Ο τίτλος του άρθρου είναι δάνειο από το άρθρο των Jostein Ryssevik και Simon Musgrave (2001). Αφιερώνεται στον συνάδελφό μας Atle Alvheim του Norwegian Centre for Research Data (NSD) που αυτό το διάστημα εργάζεται με την ομάδα του για το δικό του Dream Machine. Θα θέλαμε επίσης να ευχαριστήσουμε τους ανώνυμους αξιολογητές του άρθρου για τα εποικοδομητικά σχόλιά τους.

Apostolos Linardis, Dimitra Kondyli

THE DREAM MACHINE

ABSTRACT

In this paper, we adopt as theoretical background the main rules set up by Metanet, the “Data Documentation Initiative 3”, along with methodological procedures applied during the implementation of the international survey for adult skills (PIAAC) and the recent European research framework aiming at reuse, long-term preservation and archiving of data collected, created or elaborated within the purposes of a research project aiming at the setting-up of an ideal system related to social data. Hence, we attempt to describe an ideal system which will assist users to design and conduct a study as well as its efficient data archiving and administration together with dissemination outcomes, data and related documentation for secondary use.

Keywords: research procedures, metadata, data, reuse

1. ΕΙΣΑΓΩΓΗ

Η παρούσα εργασία αναδεικνύει ορισμένα από τα ζητήματα που καλείται να αντιμετωπίσει ο ερευνητής που ασχολείται με τα κοινωνικά δεδομένα και, παράλληλα, προσπαθεί να θέσει το πλαίσιο –σε εθνικό και διεθνές επίπεδο– για παραγωγή και αξιοποίηση των παραγόμενων ερευνητικών δεδομένων προς όφελος της ερευνητικής και ακαδημαϊκής κοινότητας με την κεφαλαιοποίηση του τρίπτυχου «υλοποίηση ποσοτικών ερευνών και αποτελεσμάτων ερευνητικών έργων σε περιβάλλον ερευνητικών υποδομών». Ένα από τα βασικά ζητήματα που καλείται να αντιμετωπίσει ο ερευνητής που ασχολείται με κοινωνικά δεδομένα στην Ελλάδα αποτελεί η αποτελεσματική εύρεση των δεδομένων.² Η διαδικασία εντοπισμού των δεδομένων αποτελεί σημαντικό κομμάτι της ερευνητικής διαδικασίας. Για να μπορέσει, στη συνέχεια, ο ερευνητής να επεξεργαστεί δεδομένα δευτερογενώς, πρέπει αυτά να είναι τεκμηριωμένα και ελεύθερα προσβάσιμα/ανοιχτά (open data) (OECD, 2007). Η προσβασιμότητα όμως και η «ανοιχτότητα» (openness) των δεδομένων δεν τα καθιστά αυτομάτως και επαναχρησιμοποιήσιμα. Τα δεδομένα πρέπει να έχουν κατάλληλη μορφή και άδεια χρήσης για την επαναχρησιμοποίησή τους. Επιπρόσθετα, εάν τα δεδομένα διατίθενται σε «συνδεδεμένη μορφή» (linked data format), ανοίγονται νέες δυνατότητες χρήσης, καθώς επιτρέπεται η τεχνική διαλειτουργικότητα των συνόλων δεδομένων. Έτσι, δεδομένα από διαφορετικές πηγές μπορούν εύκολα να «συνδεθούν» μεταξύ τους. Παρακάτω παρατίθενται τα διαφορετικά επίπεδα «ανοιχτότητας» των δεδομένων (Research Data Netherlands, 2016) και η αντιστοιχία με το μοντέλο των 5 αστεριών του Berners-Lee (2009):

- I. Οι χρήστες έχουν πρόσβαση στα μεταδεδομένα αλλά όχι στα ίδια τα δεδομένα
- II. Τα δεδομένα είναι προσπελάσιμα και αναγνώσιμα από οποιονδήποτε.
- III. Τα δεδομένα είναι διαθέσιμα μέσω διαδικτύου σε οποιαδήποτε μορφή και επιτρέπεται η χρήση τους βάσει άδειας ανοιχτότητας (open license), στην οποία καθορίζεται ο βαθμός επαναχρησιμοποίησης των δεδομένων (ένα αστέρι).

2. Η διαδικασία εύρεσης των δεδομένων αποτελεί βασική προτεραιότητα για την ερευνητική κοινότητα και στον ευρωπαϊκό ερευνητικό χώρο, σύμφωνα με πρόσφατη αδημοσίευτη έκθεση του ερευνητικού έργου CESSDA SaW, Horizon 2020.

- IV. Τα δεδομένα είναι διαθέσιμα μέσω μιας δομημένης μορφής (format), και είναι αναγνώσιμα από ηλεκτρονικό υπολογιστή (για παράδειγμα είναι διαθέσιμα μέσω Excel και όχι μέσω σκαναρισμένης εικόνας) (δύο αστέρια).
- V. Οι μορφές δεδομένων που χρησιμοποιούνται δεν υποστηρίζουν συγκεκριμένο προμηθευτή λογισμικού (για παράδειγμα, τα δεδομένα είναι προτιμητέο να υφίστανται σε μορφή csv παρά σε excel) (τρία αστέρια).
- VI. Τα στοιχεία των δεδομένων έχουν ένα URI (Uniform Resource Identifier: URL ή URN) (τέσσερα αστέρια).
- VII. Τα δεδομένα συνδέονται με άλλα δεδομένα (πέντε αστέρια).

Η αυξανόμενη διαβάθμιση ανοιχτότητας διευκολύνει την απόκτηση νέας γνώσης. Παράλληλα, πέραν της ανοιχτότητας των δεδομένων, εγείρονται πλέον νέα ζητήματα ως προς την κοινοχρησία, την αρχειοθέτηση και τη διαχείριση των δεδομένων. Δεν είναι τυχαίο ότι στη θεματική προτεραιότητα «Open Research Data Pilot» του Horizon 2020 είναι απαραίτητη για όλα τα έργα που συμμετέχουν η κατάθεση ενός σχεδίου διαχείρισης δεδομένων (Data Management Plan – DMP), όπου θα αναφέρονται τα πρότυπα μεταδεδομένων που θα χρησιμοποιηθούν, το αποθετήριο κατάθεσης των παραγόμενων δεδομένων, οι διαδικασίες μακρόχρονης διατήρησης των δεδομένων κ.ά. (αναφορά στα DMP's γίνεται παρακάτω). Συνεπώς, το ζήτημα της πρόσβασης στο σώμα των πρωτογενών ή δευτερογενών δεδομένων αποτελεί μια σημαντική πρόκληση για τον κοινωνικό επιστήμονα που ασχολείται με την ποσοτική ανάλυση.

Ένα δεύτερο ζήτημα που συναρτάται με την αποτελεσματική αναζήτηση και περαιτέρω χρήση των δεδομένων είναι η επαρκής τεκμηρίωση ολόκληρου του κύκλου ζωής τους. Στο σημείο αυτό διευκρινίζεται ότι η τεκμηρίωση ποσοτικών ερευνών περιλαμβάνει –πλην της τεκμηρίωσης της ταυτότητας της έρευνας– τουλάχιστον το σύνολο δεδομένων και το ερωτηματολόγιο, τα οποία συγκροτούν σύνθετα ψηφιακά τεκμήρια (αποτελούνται από μεταβλητές και ερωτήσεις, αντιστοίχως) και, ως εκ τούτου, δεν μπορούν να τεκμηριωθούν επαρκώς βάσει γενικών προτύπων τεκμηρίωσης. Δεδομένου ότι η επαρκής τεκμηρίωση συνδέεται με την αναζήτηση και την ανάκτηση, γίνεται σαφές ότι ελλιπής τεκμηρίωση οδηγεί σε προβληματική αναζήτηση (Corti, Eynden, Bishop and Woollard, 2014). Επομένως, η διαδικασία τεκμηρίωσης πρέπει ιδεατά να περιλαμβάνει τον εννοιολογικό προσδιορισμό των δεδομένων, το σχεδιασμό, τη συλλογή

και την κωδικοποίησή τους. Μια σειρά βασικών παραμέτρων πρέπει να ληφθούν υπόψη ως προς τη διαχείριση των δεδομένων σήμερα:

- η συμβολή των διεθνών οργανισμών (Ευρωπαϊκή Επιτροπή, ΟΟΣΑ, ενιαίος Ευρωπαϊκός Ερευνητικός Χώρος, Horizon 2020) στην προώθηση των ελεύθερα προσβάσιμων δεδομένων (αρχή της ανοιχτότητας),
- η δυνατότητα αποθήκευσης των δεδομένων σε αποθετήρια που επιτρέπουν τη μακρόχρονη και απρόσκοπτη διατήρησή τους, ώστε να είναι δυνατή η δευτερογενής χρήση και επαναχρησιμοποίησή τους,
- η ύπαρξη προτύπων μεταδεδομένων όπως το Data Documentation Initiative (DDI), που υποστηρίζουν την τεκμηρίωση ολόκληρου του κύκλου ζωής μιας έρευνας,
- η υποστήριξη της διεξαγωγής ερευνών μέσω CAI (Computer Assisted Interviewing) συστημάτων,
- η υποστήριξη της έρευνας πεδίου από συμπληρωματικά εργαλεία (εργαλεία για την επιλογή του δείγματος, εργαλεία για τη διαχείριση των δεδομένων μετά την έρευνα πεδίου),
- η επανάληψη μεταδεδομένων ερωτήσεων και μεταβλητών σε διαφορετικά λογισμικά με κίνδυνο διπλοεγγραφών και αλλαγών,
- η τεκμηρίωση για την τεκμηρίωση αποτελεί μια μάλλον ανιαρή διαδικασία.

Λαμβάνοντας υπόψη τα παραπάνω, τίθεται το εξής ερώτημα: Γιατί δεν μπορεί να υλοποιηθεί ένα ολοκληρωμένο και ενοποιημένο σύστημα που θα εξυπηρετεί ταυτόχρονα τις ανάγκες των κύριων ερευνητών, των οργανισμών που εμπλέκονται στη συλλογή των δεδομένων καθώς και των δευτερογενών χρηστών, βάσει του οποίου η τεκμηρίωση και, κατ' επέκταση, η δευτερογενής χρήση να προκύπτει ως υποπροϊόν όλων των διαδικασιών που απαιτούνται πριν, κατά τη διάρκεια και μετά το πέρας της έρευνας πεδίου;

Στην παρούσα εργασία θα αναφερθούμε αρχικά στην συνιστάμενη όλων των διαδικασιών που καθιστούν εφικτή και αποδοτική την αναζήτηση, δηλαδή την ποιοτική και αυστηρή τεκμηρίωση. Στη συνέχεια, θα περιγραφούν όλες οι μεθοδολογικές διαδικασίες που απαιτούνται πριν, κατά τη διάρκεια και μετά το πέρας της έρευνας πεδίου. Για κάθε μεθοδολογική διαδικασία, θα αναφερόμαστε στα επιμέρους υποσυστήματα που προκύπτουν.

Τέλος, θα επιχειρηθεί η ενοποίηση όλων των εργαλείων σε ένα ενιαίο σύστημα που, δυνητικά, θα αποτελέσει το ιδεατό σύστημα τεκμηρίωσης, αναζήτησης, σχεδιασμού και διεξαγωγής ερευνών και θα βοηθά παράλληλα τους ερευνητές και στην ανάλυση των δεδομένων, αποτελώντας το «dream machine» των κοινωνικών επιστημόνων.

2. Η ΤΕΚΜΗΡΙΩΣΗ ΠΟΣΟΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Η τεκμηρίωση ποσοτικών δεδομένων ή ερευνητικών συνιστωσών που χρησιμοποιούνται σε ποσοτικές έρευνες επιτυγχάνεται με τη χρήση των μεταδεδομένων. Τα μεταδεδομένα παρέχουν τον ουσιαστικό σύνδεσμο μεταξύ του πρωτογενούς υλικού και της δευτερογενούς χρήσης του. Τα δεδομένα αποτελούν το τελικό προϊόν της εμπειρικής κοινωνικής έρευνας, αλλά για να είναι αξιοποιήσιμα από τρίτους θα πρέπει να συνοδεύονται από ένα πρόσθετο σώμα δεδομένων που ονομάζουμε «μεταδεδομένα». Οι Grossmann (1999) και Sundgren (1996) ορίζουν τα μεταδεδομένα ως δεδομένα που περιγράφουν τα ίδια τα δεδομένα (data about data). Αρχικά, τα μεταδεδομένα αφορούσαν ποσοτικά δεδομένα που προέκυπταν για τον προσδιορισμό αθροιστικών – πινακοποιημένων δεδομένων και για τις ανάγκες επικοινωνίας και εναρμονισμένης ανταλλαγής στατιστικών δεδομένων των στατιστικών υπηρεσιών των χωρών μελών με την Eurostat (2000) και τον ΟΟΣΑ (OECD, 1999· Moorman, 2000). Τη δεκαετία του 1960 δημιουργούνται επίσης τα Αρχεία Δεδομένων που μετασχηματίζονται σταδιακά σε ερευνητικές υποδομές και τεκμηριώνουν και διαχέουν μικροδεδομένα με μεταδεδομένα στην ακαδημαϊκή κοινότητα (Κονδύλη και Λιναρδής, 2004).

Στο παρελθόν, τα μεταδεδομένα των μικροδεδομένων δεν ήταν προσβάσιμα στον χρήστη (πέραν της ελλιπούς τεκμηρίωσης των μεταβλητών στο αρχείο δεδομένων). Η «ανασυγκρότηση» των δεδομένων μπορούσε να γίνει μόνο από τον παραγωγό του πρωτογενούς υλικού, για όσο διάστημα μπορούσε να ανακαλέσει τη διαδικασία στη μνήμη του ή μέσα από κάποιες σημειώσεις και σχετικά αρχεία, που όμως βασίζονταν σε αναπαραστάσεις ελεύθερου κειμένου, ήταν αδόμητα και μη επεξεργάσιμα από πληροφοριακά συστήματα (Kleiner, Renschler, Wernli, Farago and Joye, 2013· Farago, 2014). Ως προς τα στατιστικά δεδομένα, τα μεταδεδομένα αφορούσαν κάποιες απλές υποσημειώσεις που αναφέρονταν σε πίνακες.

Η ελλιπής τεκμηρίωση γίνεται ιδιαίτερα προβληματική όταν συστηματοποιούνται οι ανταλλαγές δεδομένων στους κόλπους των διεθνών οργανισμών μετά τη δεκαετία του 1970. Οι Eurostat και ΟΟΣΑ, για λόγους συγκρισιμότητας των εθνικών δεδομένων από πολλαπλές πηγές, έθεσαν τα πλαίσια χρήσης κοινών μεταδεδομένων στα δεδομένα που τους παρείχαν οι εθνικές στατιστικές υπηρεσίες των χωρών μελών. Παρά τα κοινά μεταδεδομένα, η εννοιολογική διαφοροποίηση των ορισμών δυσχέραινε την ανάλυση/συγκρισιμότητα των αποτελεσμάτων. Τότε, αρχίζουν οι προσπάθειες για την από κοινού δόμηση των μεταδεδομένων και χρήση κοινής ορολογίας. Ο Sundgren (Sundgren, 1999) προτείνει τη χρήση ενιαίων φορμών-πινάκων (templates) για τη δομημένη εισαγωγή των στατιστικών μεταδεδομένων. Η τεχνική των φορμών αποτέλεσε μια σαφή βελτίωση από την απλή παρουσίαση μεταδεδομένων, δεδομένου ότι εμπεριέχει μια στοιχειώδη δόμηση υπό κοινό σχεδιασμό (Froeschl, Yamada and Kudrna, 2002). Εντούτοις, η πρόταση παρουσιάζει περιορισμούς (Papageorgiou, Vardaki and Pentaris, 2000). Ο πιο σημαντικός συνίσταται στο ότι οι φόρμες χρησιμοποιούνται για την απλή εισαγωγή των μεταδεδομένων και όχι για την αυτοματοποιημένη χρήση τους σε πληροφοριακά συστήματα, αφού δεν υπάρχει η πληροφορία του πώς να χειριστούν τα συστήματα αυτή τη μεταπληροφορία. Επιπρόσθετα, τα συστήματα δεν μπορούν να αντιλαμβάνονται το νόημα της αποθηκευμένης πληροφορίας, οπότε τη χρησιμοποιούν ως ελεύθερο κείμενο. Για την εναρμονισμένη επικοινωνία ανθρώπων και συστημάτων, απαιτείται η χρήση προτύπων μεταδεδομένων.

3. ΤΑ ΠΡΟΤΥΠΑ ΤΕΚΜΗΡΙΩΣΗΣ ΠΟΣΟΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Στην παρούσα ενότητα περιγράφεται συνοπτικά το διεθνές πρότυπο μεταδεδομένων DDI, που χρησιμοποιείται ευρέως για την τεκμηρίωση ολόκληρου του κύκλου ζωής μιας έρευνας, καθώς και βασικοί κανόνες που πρέπει να τηρούνται σε σχέση με τα μεταδεδομένα, όπως προέκυψαν από το ερευνητικό έργο Metanet.

Το διεθνές πρότυπο μεταδεδομένων που χρησιμοποιείται για την τεκμηρίωση αθροιστικών δεδομένων από τις στατιστικές υπηρεσίες είναι το Statistical Data and Metadata eXchange standard (SDMX, 2009), ωστόσο και το DDI3 τεκμηριώνει αθροιστικά δεδομένα (κύβους). Όπως επισημαίνεται από τους Gregory and Heus (2007) πρόκειται για συμπληρωματικά και όχι για ανταγωνιστικά πρότυπα. Το DDI

έχει ως στόχο την επίλυση των προβλημάτων με την τεκμηρίωση της έρευνας και για όλη τη διάρκεια ζωής των μικροδεδομένων και προέρχεται από τον κόσμο των Αρχείων Δεδομένων και των ερευνητών κοινωνικών επιστημών, ενώ το SDMX ασχολείται με τη δημιουργία αποδόσεων γύρω από την ανταλλαγή των συγκεντρωτικών στοιχείων και δημιουργήθηκε για τις ανάγκες των στατιστικών υπηρεσιών.

3.1. Το πρότυπο DDI2/3

Το OSIRIS, που δημιουργήθηκε το 1970, αποτέλεσε το πρώτο πρότυπο τεκμηρίωσης μικροδεδομένων εμπειρικών ερευνών και αντικαταστάθηκε το 2000 από το DDI.³ Ο πρωταρχικός στόχος του DDI ήταν να αντικαταστήσει το ευρέως χρησιμοποιούμενο βιβλίο κωδικοποίησης (codebook) OSIRIS με ένα εκσυγχρονισμένο πρότυπο, ώστε να υποστηρίζει ευρύτερη τεκμηρίωση των ερευνών. Το OSIRIS παρείχε τεκμηρίωση των μεταβλητών μιας έρευνας, ενώ το DDI κατόρθωσε να τεκμηριώσει τον σχεδιασμό καθώς και τη διαδικασία συλλογής της έρευνας. Το DDI έχει επεκταθεί πολλαπλά από την πρώτη έκδοση και μετά. Πιο συγκεκριμένα, οι εκδόσεις 1 και 2 του προτύπου DDI συγκροτούνται από τους ακόλουθους 5 βασικούς τομείς (Λιναρδής, 2006):

- Document Description (περιγραφή εγγράφου): περιέχει βιβλιογραφική πληροφορία του ίδιου του XML εγγράφου.
- Study Description (περιγραφή έρευνας): περιέχει αναλυτική περιγραφή της έρευνας και του σχεδιασμού της.
- Data Files Description (περιγραφή αρχείων δεδομένων): η περιγραφή αρχείων δεδομένων αναφέρεται στα αρχεία δεδομένων από τα οποία απαρτίζεται η έρευνα.
- Variable Description (περιγραφή μεταβλητών): περιλαμβάνει περιεκτική περιγραφή των μεταβλητών των αρχείων δεδομένων.
- Other Study Related Materials (άλλα τεκμήρια σχετικά με την έρευνα): αποθήκη ηλεκτρονικών υλικών που περιλαμβάνει πρόσθετο υλικό για την τεκμηρίωση της έρευνας και των μεταβλητών.

Η πρόσφατη έκδοση DDI 3 καλύπτει πλέον ολόκληρο τον κύκλο ζωής μιας έρευνας. Το DDI 3 τεκμηριώνει μια έρευνα από τη σύλλη-

3. <http://www.ddialliance.org>

ψη της ιδέας της έρευνας μέχρι την ανάλυση και την επαναστόχευση, επιτρέπει την επαναχρησιμοποίηση ερευνητικών συνιστωσών, όπως εννοιών, ερωτήσεων και μεταβλητών, την τεκμηρίωση της συγκριτικής έρευνας και τη διαχείριση εκδόσεων των ίδιων ερευνητικών συνιστωσών, τεκμηριώνει το όργανο μέτρησης και τη λογική αλληλουχία των ερωτήσεων⁴ καθώς και σύνθετα ιεραρχικά δεδομένα (Green, 2008). Το DDI3 έχει δημιουργηθεί με στόχο να συνεργάζεται καλύτερα με υπολογιστές (machine actionable), παρά να παρέχει πληροφορία εύκολα αναγνώσιμη από τον άνθρωπο (human readable) (Thomas, 2010). Προσπαθεί, δηλαδή, να τυποποιήσει τιμές πεδίων, ώστε η τυποποίηση αυτή να μπορεί να χρησιμοποιηθεί από έναν υπολογιστή αντί κειμένων τα οποία είναι μεν περισσότερο κατανοητά στον άνθρωπο, αλλά δεν μπορούν να χρησιμοποιηθούν εύκολα από τον υπολογιστή.

Το DDI3, αν και αποτελεί μετεξέλιξη του DDI2 και διαθέτει περισσότερες δυνατότητες τεκμηρίωσης, εντούτοις θεωρείται πιο δυσνόητο στη σύλληψη και πιο δύσκολα εφαρμόσιμο. Αρκετοί από τους κανόνες που παρουσιάζονται στο έργο Metanet υιοθετήθηκαν από το DDI3.

4. ΤΟ ΕΡΓΟ METANET

Ένα σημαντικό ευρωπαϊκό έργο για μοντέλα και πρότυπα μεταδεδομένων στατιστικών δεδομένων είναι το έργο Metanet (A network of excellence for harmonizing and synthesising the development of statistical metadata). Το Metanet⁵ ήταν ένα δίκτυο αριστείας για την εναρμόνιση και αξιοποίηση των στατιστικών μεταδεδομένων. Παρακάτω, παρατίθενται ορισμένοι κανόνες που πρέπει να τηρεί ένας σχεδιαστής μοντέλων και συστημάτων μεταδεδομένων, όπως αυτοί τέθηκαν από τον Sundgren (2003) στο πλαίσιο του εν λόγω έργου:

- *Αντιμετωπίστε την εργασία σε σχέση με τα μεταδεδομένα ως ένα ολοκληρωμένο τμήμα των επιχειρησιακών διαδικασιών του οργανισμού σας:* θα πρέπει να αποφευχθεί η διαχείριση των μεταδεδομένων ως μια ξεχωριστή δραστηριότητα του οργανισμού, καθώς καταναλώνονται πόροι ενώ θεωρείται ταυτόχρονα από τους περισσότερους εμπλεκόμενους διαδικασία που απαιτεί όγκο δουλειάς

4. Απόδοση του αγγλικού όρου «questionnaire workflow»

5. <http://www.ces.ed.ac.uk/research/Epros%20projects/metanet/>

χωρίς ανάλογο ερευνητικό ενδιαφέρον. Τα μεταδεδομένα πρέπει να παράγονται ως φυσικά υποπροϊόντα άλλων διαδικασιών που θα εκτελεσθούν οπωσδήποτε.

- *«Συλλαμβάνετε» τα μεταδεδομένα στις φυσικές πηγές τους, και κατά προτίμηση ως υποπροϊόντα άλλων διαδικασιών:* για παράδειγμα, κατά τη διάρκεια του σχεδιασμού συστημάτων παραγωγής στατιστικών, πολλές αποφάσεις λαμβάνονται σχετικά με το περιεχόμενο και τις διαδικασίες του συστήματος που κατασκευάζεται. Εάν αυτές οι αποφάσεις καταγράφονται κατάλληλα, μπορούν να παρέχουν την κρίσιμη μάζα της βασικής τεκμηρίωσης του συστήματος.

- *Μη «συλλαμβάνετε» τα ίδια μεταδεδομένα δύο φορές:* συνίσταται τα μεταδεδομένα να αποθηκεύονται μόλις «συλληφθούν» και να οργανώνονται έτσι ώστε να είναι προσβάσιμα και επαναχρησιμοποιήσιμα.

- *Αποφύγετε την ασυντόνιστη σύλληψη παρόμοιων μεταδεδομένων- δημιουργήστε αλυσίδες σχετικών μεταδεδομένων:* παρόμοια αλλά μη ταυτόσημα μεταδεδομένα μπορεί να απαιτηθούν για διαφορετικούς λόγους. Η σύλληψη και η διατήρηση παρόμοιων μεταδεδομένων χωρίς επαρκή συντονισμό μπορεί να προκαλέσει αλόγιστες προσπάθειες τεκμηρίωσης και, στη χειρότερη περίπτωση, να οδηγήσουν σε μεταδεδομένα που εν τέλει δεν έχουν σχέση το ένα με το άλλο. Το πρόβλημα αυτό επιλύεται με τη δημιουργία αλυσίδων μεταδεδομένων.

- *Όταν δημιουργούνται οι ανάγκες για νέα μεταδεδομένα, προσπαθήστε να βρείτε απαντήσεις στις νέες ανάγκες χρησιμοποιώντας ήδη υπάρχοντα μεταδεδομένα και εμπλουτίζοντάς τα με πρόσθετα και μη πλεονάζοντα μεταδεδομένα.*

- *Μετατρέψτε τα δεδομένα και τα συνοδεύοντα μεταδεδομένα σε συγχρονισμένες παράλληλες διαδικασίες, πλήρως αυτοματοποιημένες, όπου αυτό είναι δυνατό:* οι διαδικασίες διαχείρισης δεδομένων εντός των στατιστικών υπηρεσιών έχουν τυποποιηθεί και αυτοματοποιηθεί εδώ και πολύ καιρό. Αντίθετα, η διαχείριση των απαραίτητων συνοδευτικών μεταδεδομένων αντιμετωπίζεται συνήθως χειρωνακτικά και οι χειρωνακτικοί μετασχηματισμοί μεταδεδομένων δεν έχουν συντονιστεί επαρκώς με τους αυτοματοποιημένους μετασχηματισμούς δεδομένων. Για παράδειγμα, για την ίδια ερώτηση μπορεί να έχουν δοθεί διαφορετικά λεκτικά είτε από τον σχεδιαστή του ερωτηματολογίου, είτε από τον προγραμματιστή του λογισμικού συλλογής και διαχείρισης δεδομένων.

- Τα μεταδεδομένα ενημερώνονται, αλλάζουν και δημιουργούνται καινούργιες εκδόσεις τους. Σε ένα σύστημα, όμως, πρέπει να διατηρούνται ταυτόχρονα και οι προηγούμενες εκδόσεις αυτών, διότι πολλές έρευνες μπορεί να έχουν χρησιμοποιήσει προηγούμενες εκδόσεις: για παράδειγμα, οι παλαιές εκδόσεις μιας κατάταξης καθώς επίσης και οι πίνακες αντιστοίχισής τους πρέπει να είναι διαθέσιμοι στο σύστημα, ιδιαίτερα όταν επαναχρησιμοποιούνται παλαιά δεδομένα καθώς και όταν διαμορφώνονται χρονοσειρές δεδομένων που προήλθαν από διαφορετικές εκδόσεις της ταξινόμησης.

5. ΑΡΧΕΣ ΣΥΓΚΡΟΤΗΣΗΣ ΤΟΥ DREAM MACHINE ΤΩΝ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

Έχοντας ως αφετηρία τους παραπάνω κανόνες του Metanet και το πρότυπο μεταδεδομένων DDI3 (Thomas, ό.π.), θα επιχειρήσουμε μια πρώτη προσέγγιση του *ιδεατού συστήματος (Dream Machine, στο εξής DM)*, που θα εξυπηρετεί τις διαδικασίες που απαιτούνται σε όλες τις φάσεις μιας έρευνας, και βάσει των μεθοδολογικών διαδικασιών που ακολουθήθηκαν στην έρευνα *Programme for the International Assessment of Adult Competencies (PIAAC) του ΟΟΣΑ (OECD, 2014)*. Το DM θα υποστηρίζει την κατάθεση του σχεδίου διαχείρισης των δεδομένων και, επιπλέον, τον καθορισμό του προβλήματος, τη διατύπωση υποθέσεων, την επιλογή της ερευνητικής μεθόδου, την έρευνα πεδίου, την κωδικοποίηση των δεδομένων και την ανάλυσή τους. Επίσης, θα υποστηρίζει την αρχειοθέτηση και τη διαχείριση των δεδομένων, ώστε να καθιστά διαθέσιμα για δευτερογενή χρήση όλα τα δεδομένα της ερευνητικής διαδικασίας. Συνισταμένη αποτελεί η διαδικασία της αυστηρής τεκμηρίωσης που είναι υποτιμημένη στην Ελλάδα και συχνά ανιαρή όταν στοχεύει μόνο στην ίδια την τεκμηρίωση. Στόχος εδώ είναι η τεκμηρίωση να προκύπτει ως υποπροϊόν άλλων διαδικασιών.

Στην πράξη, το DM δύναται να επικοινωνεί με εξωτερικά εργαλεία όπως στατιστικά πακέτα, εργαλεία διενέργειας ερευνών ή άλλα χειριστικά εργαλεία. Παρόλα αυτά, για τις ανάγκες της παρούσας εργασίας, τα επιμέρους εργαλεία θα θεωρούνται ως υποσυστήματα του ιδεατού συστήματος. Επίσης, η διεξαγωγή της έρευνας πεδίου θα περιγραφεί βάσει συστημάτων CAI και θεωρώντας έρευνες όσες εμπλέκουν ερευνητές (συνεντευκτές) και επόπτες. Έρευνες πεδίου στις οποίες δεν συμμετέχουν ερευνητές και επόπτες, καθώς και

έρευνες που διεξάγονται μέσω PAPI (Paper and Pencil Interviewing) εκλαμβάνονται ως υποπεριπτώσεις των μεθοδολογικών διαδικασιών που θα περιγραφούν παρακάτω. Τέλος, στο σύστημα θα εμπλέκονται πολλοί χρήστες με διαφορετικούς ρόλους (κύριος ερευνητής, ερευνητές, επόπτες, συνεντευκτές, δευτερογενής χρήστης, πληροφορικός, μεταφραστής κ.α.). Σε κάθε υποσύστημα που θα προκύπτει, θα αναφέρονται οι δυνητικοί χρήστες. Θα εξαιρεθεί ο ρόλος του τεκμηριωτή, καθώς η τεκμηρίωση θα απορρέει ως υποπροϊόν των διαδικασιών. Επίσης, ο ερευνητής ή η ερευνητική ομάδα που σχεδιάζει/ουν την έρευνα θα αναφέρονται ως κύριοι ερευνητές.

5.1. Το σχέδιο διαχείρισης δεδομένων

Η κατάθεση του σχεδίου διαχείρισης των δεδομένων απαιτείται πλέον στο «Horizon 2020» για όλα τα έργα στη θεματική προτεραιότητα: «Open Research Data Pilot», ενώ είναι προαιρετική για άλλα έργα (European Commission, 2016). Το Σχέδιο Διαχείρισης Δεδομένων (Data Management Planning, στο εξής DMP) περιγράφει τον κύκλο ζωής διαχείρισης των δεδομένων για όλα τα σύνολα δεδομένων που θα συλλεχθούν, δημιουργηθούν ή επεξεργαστούν σε ένα έργο. Σε όλα τα έργα του «Open Research Data Pilot» ή σε όσα έχει αποφασιστεί ότι θα κατατεθεί DMP, υπάρχει η υποχρέωση κατάθεσης του DMP εντός των πρώτων έξι μηνών. Προς το παρόν, η Ευρωπαϊκή Επιτροπή δεν παρέχει εργαλείο συγγραφής ενός DMP. Οι υποψήφιοι ανάδοχοι θα πρέπει είτε να δημιουργήσουν ένα ad hoc DMP βάσει ενός προτύπου που παρέχεται από την Ευρωπαϊκή Επιτροπή, είτε να χρησιμοποιήσουν το εργαλείο DMP on line⁶ που παρέχεται από το Digital Curation Centre.

Βάσει του προτύπου της Ευρωπαϊκής Επιτροπής, κάθε σχέδιο διαχείρισης δεδομένων πρέπει να καλύπτει τα κάτωθι (στο ίδιο):

- Το όνομα και τα πεδία ταυτοποίησης των συνόλων δεδομένων.
- Την περιγραφή των συνόλων δεδομένων και συναφών δεδομένων καταγράφοντας τις δυνατότητες επαναχρησιμοποίησης, ολοκλήρωσης και αξιοποίησής τους σε επιστημονικές δημοσιεύσεις.

6. Το DMP on line είναι διαθέσιμο στον ακόλουθο σύνδεσμο: <https://dmponline.dcc.ac.uk/>

- Αναφορά σε κατάλληλα πρότυπα τεκμηρίωσης. Όταν δεν ακολουθούνται πρότυπα, να αναφέρεται ένα περιγράμμα των μεταδεδομένων που θα χρησιμοποιηθούν.
- Κοινοχρησία δεδομένων: περιγραφή του τρόπου με τον οποίο τα δεδομένα θα διαμοιραστούν, διαδικασίες πρόσβασης, περιόδους εμπάργκο, όρους χρήσης (ελεύθερα προσβάσιμα ή περιορισμένης χρήσης), μηχανισμούς διάχυσης, λογισμικά και εργαλεία για την επαναχρησιμοποίηση των δεδομένων, ταυτοποίηση του αποθετηρίου που θα αποθηκευτούν τα δεδομένα. Εάν τα δεδομένα δεν μπορούν να χρησιμοποιηθούν, να γίνεται ρητή αναφορά των λόγων (ηθικοί, προσωπικά δεδομένα, πνευματική ιδιοκτησία, εμπορικοί λόγοι κ.ά.).
- Αρχαιοθήκη και διατήρηση (εμπεριέχεται αποθήκευση και back up): περιγραφή των διαδικασιών για τη μακρόχρονη διατήρηση των δεδομένων, καθώς και εκτίμηση δαπάνης και τρόποι κάλυψής της.

DM1: Σύστημα διαχείρισης «σχεδίων διαχείρισης δεδομένων» (DMP's). / *Χρήστες:* Κύριος Ερευνητής

5.2. Πριν την έρευνα

Ο κύριος ερευνητής, σε συνεργασία με τον γενικό διαχειριστή του συστήματος, καταγράφει πληροφορίες που αφορούν στους φορείς, στα άτομα και στον ρόλο τους στην έρευνα. Σε κάθε έρευνα εμπλέκονται πολλοί χρήστες (κύριος ερευνητής, γενικός διαχειριστής, πληροφορικός, μεταφραστής, στατιστικός, επόπτες, συνεντευκτές κ.ά.). Κάθε κατηγορία χρήστη έχει πρόσβαση σε διαφορετικά υποσυστήματα και λειτουργίες του DM. Για αυτό, η καταγραφή των φορέων δράσης (φορείς ή άτομα) που εμπλέκονται στην έρευνα, πρέπει να συνοδεύεται από μια κατάλληλη εφαρμογή διαχείρισης των χρηστών που θα προσδιορίζει ρόλους και θα οριοθετεί την πρόσβαση στις εφαρμογές όχι μόνον ανά έρευνα αλλά και ανά συλλογή δεδομένων (βλ. παρακάτω).

Αρχικά, ο κύριος ερευνητής καταγράφει την ταυτότητα της έρευνας, όπως τίτλο και σκοπό, καθώς και τον πληθυσμό που απευθύνεται. Στη συνθετότερη περίπτωση, μια έρευνα μπορεί να αφορά πολλές «συλλογές δεδομένων». Ενδεικτικά, μια έρευνα σε σχολεία

δύναται να αφορά στους διευθυντές, στους καθηγητές και στους μαθητές, χρησιμοποιώντας διαφορετικό ερωτηματολόγιο ανά πληθυσμό. Σε αυτήν την περίπτωση, ενδέχεται να διαφοροποιούνται τόσο τα άτομα όσο και οι φορείς που εμπλέκονται σε κάθε συλλογή δεδομένων. Άλλη τέτοια περίπτωση είναι οι μεγάλες συγκριτικές διακρατικές έρευνες, π.χ. European Social Survey (ESS), PIAAC, Ευρωβαρόμετρο, ISSP. Η έρευνα είναι επίσης μία, αλλά οι «συλλογές δεδομένων» πολλές (μία ή περισσότερες ανά χώρα). Ως εκ τούτου, η έρευνα πρέπει να διαφοροποιείται από τη συλλογή δεδομένων. Στην έρευνα καταγράφονται εν τέλει τα γενικά στοιχεία, ενώ στη «συλλογή δεδομένων» τα επιχειρησιακά δεδομένα και μεταδεδομένα που αφορούν στην έρευνα πεδίου.

Τέλος, ο κύριος ερευνητής ενημερώνει το σύστημα με στοιχεία της προετοιμασίας συλλογής των δεδομένων, όπως η μεθοδολογία και δειγματοληψία, η μέθοδος ή οι μέθοδοι συλλογής δεδομένων, οι ενέργειες ελαχιστοποίησης της μη απόκρισης κ.ά.

DM2: Τράπεζα Ερευνών / *Χρήστες:* Κύριος Ερευνητής

DM3: Τράπεζα «Συλλογών δεδομένων» / *Χρήστες:* Κύριος Ερευνητής

DM4: Τράπεζα Φορέων Δράσης (Φορείς ή άτομα) και καταγραφή ρόλων (Εφαρμογή διαχείρισης χρηστών) / *Χρήστες:* Γενικός Διαχειριστής

Κατόπιν, ο κύριος ερευνητής θα θελήσει να αποτυπώσει τις έννοιες που προτίθεται να μετρήσει. Οι έννοιες μπορεί να είναι μονοδιάστατες ή πολυδιάστατες. Η εννοιολόγηση αφορά στην επεξεργασία και εξειδίκευση αφηρημένων εννοιών και παράγει ένα συμβατικό νόημα που εξυπηρετεί τη συγκεκριμένη έρευνα. Ο πρώτος ορισμός στην έννοια είναι ονομαστικός. Στη συνέχεια, τίθεται ένας λειτουργικός ορισμός για τη μέτρηση της έννοιας (ανάγκες λειτουργικοποίησης). Για παράδειγμα, η ανεργία αφορά «την κατάσταση ενός ατόμου που, ενώ είναι ικανό, πρόθυμο και διαθέσιμο να εργαστεί, δεν μπορεί να βρει εργασία». Ο λειτουργικός ορισμός της όμως περιλαμβάνει τουλάχιστον τον προσδιορισμό του χρόνου ανεργίας.

DM5: Τράπεζα εννοιών: Τεκμηρίωση εννοιών-ορισμών / *Χρήστες:* Κύριος Ερευνητής

Η επιλογή της ποσοτικής ερευνητικής μεθόδου καταλήγει στο σχεδιασμό ενός δομημένου ερωτηματολογίου που αποτελείται από ερωτήσεις διαφόρων τύπων (για παράδειγμα η πρόσφατη έκδοση του λογισμικού Limesurvey διαχειρίζεται εικοσιεννιά τύπους ερωτήσεων. Συνηθέστερα στις ποσοτικές έρευνες οι ερωτήσεις ή/και οι υποερωτήσεις είναι κλειστές και συνοδεύονται από απαντητικές κατηγορίες και αναλογούντες κωδικούς. Οι απαντητικές κατηγορίες δύναται να αφορούν περισσότερες από μία ερωτήσεις που μπορούν να επαναχρησιμοποιούνται. Για παράδειγμα, πενταβάθμιες κλίμακες συμφωνίας μπορεί να χρησιμοποιούνται σε πολλές ερωτήσεις ως απαντητικές κατηγορίες. Η ορθή τεκμηρίωση αφορά την άπαξ καταγραφή της προαναφερόμενης κλίμακας και της επαναχρησιμοποίησής της, όπου απαιτείται. Στη συνέχεια, οι παραγόμενες κατηγοριοποιήσεις συνδέονται με το κείμενο της ερώτησης ή της υποερώτησης ώστε να προκύψει η συνολική ερώτηση. Οι ερωτήσεις μπορούν επίσης να συνδεθούν με τους λειτουργικούς ορισμούς που τέθηκαν σε προηγούμενη φάση.

DM6: Τράπεζα κατηγοριοποιήσεων (κωδικών – κατηγοριών)/
Χρήστες: Κύριος Ερευνητής

DM7: Τράπεζα ερωτήσεων: τεκμηρίωση ερωτήσεων -
υποερωτήσεων / σύνδεση ερωτήσεων - υποερωτήσεων με
έννοιες και ορισμούς. / *Χρήστες:* Κύριος Ερευνητής

Στη συνέχεια, οι ερωτήσεις πρέπει να ενσωματωθούν σε ένα εργαλείο μέτρησης, το ερωτηματολόγιο, όπου πλέον αρχίζουν και αποκτούν επιπλέον ιδιότητες και χαρακτηριστικά (διάταξη, κανόνες εγκυρότητας, skip logic, καταγραφή χρόνου ανά ερώτηση κ.ά.). Το σύστημα θα πρέπει να υποστηρίζει: α) το σχεδιασμό και τη δημιουργία του ερωτηματολογίου βάσει διαφορετικών τύπων ερωτήσεων (απλής απάντησης, πολλαπλής απάντησης, ανοιχτής απάντησης, ερωτήσεις τύπου πίνακα κ.ά.), β) τη διεξαγωγή της έρευνας μέσω του ερωτηματολογίου και γ) την καταγραφή των απαντήσεων του ερωτώμενου. Πρόκειται, συνεπώς, για ένα σύστημα CAI, το οποίο μπορεί είτε να δημιουργηθεί εξαρχής, είτε να τροποποιηθεί χρησιμοποιώντας εργαλεία ανοιχτού κώδικα (π.χ. Limesurvey) είτε να επιτρέψει την ενσωμάτωση εμπορικών προϊόντων CAI όπως: Blaise, CASES, Nipo, CS Pro κ.ά. Επειδή στα CAI συστήματα τα τρία προαναφερθέντα παρέχονται με ενιαίο τρόπο, στην παρούσα εργασία

θα τα περιγράψουμε σαν να τα δημιουργούμε εξ αρχής. Το ερωτηματολόγιο δεν αποτελεί ακολουθία ερωτήσεων, αλλά έναν αλγόριθμο ερωτήσεων ως προς το σχεδιασμό και έναν «κώδικα» γραμμένο σε γλώσσα προγραμματισμού ως προς τη διεξαγωγή της έρευνας. Ο αλγόριθμος και ο κώδικας δημιουργούν την ακολουθία των ερωτήσεων που, ανάλογα με τις απαντήσεις των ερωτώμενων σε προηγούμενες ερωτήσεις αλλά και τη χρήση του ηλεκτρονικού γραφικού περιβάλλοντος, μπορεί να διαφοροποιείται. Δημιουργεί, δηλαδή, διαφορετικό «μονοπάτι» ερωτήσεων ανάλογα με τον ερωτώμενο και τις ενέργειες του συνεντευκτή. Επίσης, ενέχει ελέγχους εγκυρότητας των δεδομένων, οδηγίες προς συνεντευκτή και προς ερωτώμενο, δυνατότητα δημιουργίας μεταβλητών (derived variables), ακόμα και παραμετροποίηση του λεκτικού των ερωτήσεων που προκύπτουν βάσει των απαντήσεων του ερωτώμενου. Επιπρόσθετα, είναι δυνατή η καταγραφή διαχειριστικών δεδομένων της συνέντευξης όπως: ημερομηνία και χρόνος έναρξης-λήξης συνέντευξης, χρόνοι που απαιτήθηκαν για να απαντηθεί μια ερώτηση ή υποερώτηση, κωδικοί διάθεσης (αποτέλεσμα συνέντευξης – disposition codes), ενέργειες του χρήστη (ερωτώμενου ή συνεντευκτή) στο σύστημα CAI κ.α. –δεδομένα που μπορούν να χρησιμοποιηθούν από τους επόπτες για τον ποιοτικό έλεγχο των συνεντεύξεων– αλλά και δεδομένα τα οποία προκύπτουν έπειτα από υπολογισμό βάσει απαντήσεων του ερωτώμενου.

Τα πλέον σύγχρονα συστήματα CAI υποστηρίζουν από κοινού όλους τους τρόπους συνέντευξης, τηλεφωνικές συνεντεύξεις (CATI), δια ζώσης (CAPI) καθώς και αυτοσυμπληρούμενων ερωτηματολογίων μέσω web (CAWI) (multi mode data collection). Η μέθοδος διεξαγωγής της συνέντευξης επηρεάζει εν γένει και τη διαμόρφωση του λεκτικού των ερωτήσεων. Για τη διεξαγωγή των ερευνών και τη συλλογή των απαντήσεων θα πρέπει να εμφανίζεται ένα αντίγραφο του ερωτηματολογίου, είτε στους συνεντευκτές (CAPI-CATI) είτε στους ερωτώμενους (CAWI). Στην περίπτωση που υπάρχει δυνατότητα σύνδεσης με το διαδίκτυο, αυτό μπορεί να γίνεται μέσω ενός μοναδικού συνδέσμου που παράγει το CAI για το τελικώς διαμορφούμενο ερωτηματολόγιο κάθε έρευνας. Σε διαφορετική περίπτωση, το ηλεκτρονικό ερωτηματολόγιο εγκαθίσταται στις φορητές συσκευές των συνεντευκτών και η αποστολή των δεδομένων στην κεντρική βάση δεδομένων πραγματοποιείται μόλις ο συνεντευκτής συνδεθεί με το διαδίκτυο.

Συχνά, ένα ερωτηματολόγιο μεταφράζεται σε διαφορετικές γλώσσες εντός μιας χώρας, είτε γιατί ο πληθυσμός της έρευνας μπορεί να

είναι ανομοιογενής σε σχέση με την ομιλούμενη γλώσσα, είτε γιατί οι επίσημες γλώσσες είναι περισσότερες της μίας. Το ιδεατό σύστημα θα διευκόλυνε τους χρήστες ως προς τη μετάφραση, παρουσιάζοντας αρχικά τις ερωτήσεις στην κυρίαρχη γλώσσα και προσφέροντας τη δυνατότητα μετάφρασης στις άλλες γλώσσες. Επίσης, στις διεθνείς συγκριτικές έρευνες το αρχικό ερωτηματολόγιο εκπονείται στην αγγλική γλώσσα (πρότυπο ερωτηματολόγιο – source questionnaire), το οποίο στη συνέχεια μεταφράζεται στις γλώσσες των συμμετεχουσών χωρών (ερωτηματολόγιο στόχος – target questionnaire). Στη φάση της μετάφρασης πραγματοποιούνται οι τυχόν εθνικές προσαρμογές (π.χ. εκπαιδευτικό επίπεδο, μητρική γλώσσα, γλώσσες επικοινωνίας μέσα στο νοικοκυριό, χώρα απόκτησης τίτλου σπουδών, χώρα γέννησης, πολιτικά κόμματα κ.α.). Επιπρόσθετα, κάθε χώρα μπορεί να συμπεριλάβει επιπλέον ερωτήσεις «εθνικού ενδιαφέροντος».

DM8: Τράπεζα ερωτηματολογίων CAI: δημιουργία ερωτηματολογίων και σχετικού αλγόριθμου / *Χρήστες:* Κύριος Ερευνητής & Πληροφορικός (δημιουργός ηλεκτρονικού ερωτηματολογίου)

DM9: Σύστημα μετάφρασης ερωτήσεων, απαντητικών κατηγοριών και ερωτηματολογίων / *Χρήστες:* Μεταφραστής

Σε αυτό το σημείο αντιστοιχίζονται οι ερωτήσεις με τις μεταβλητές. Η σχέση ερώτησης και μεταβλητής δεν είναι πάντα σχέση ένα προς ένα, διότι σε μία ερώτηση μπορεί να αντιστοιχίζονται περισσότερες μεταβλητές. Εδώ προστίθενται ονόματα και περιγραφές των μεταβλητών (ετικέτες / κατηγορίες και κωδικοί – με δυνατότητα επαναχρησιμοποίησης κωδικών-κατηγοριών) καθώς και η αντιστοιχισή τους με τις ερωτήσεις. Όπως προαναφέρθηκε, στις ερωτήσεις του ερωτηματολογίου –εκτός των απαντήσεων– καταγράφονται και άλλα συμβάντα και λειτουργίες. Ως εκ τούτου, το βιβλίο κωδικοποίησης μπορεί να περιέχει και άλλες μεταβλητές, στις οποίες αποθηκεύονται και άλλα δεδομένα όπως: διαχειριστικά δεδομένα της συνέντευξης (ημερομηνία και χρόνος έναρξης-λήξης συνέντευξης, κωδικοί διάθεσης (disposition codes), ενέργειες του χρήστη στο σύστημα CAI κ.α.), μεταβλητές που προκύπτουν μετά από υπολογισμό (derived variables) και από συνδυασμό απαντήσεων του ερωτώμενου (στην τελευταία περίπτωση προστίθεται ο κώδικας υπολογισμού των νέων μεταβλητών).

Ο πληροφορικός παραλαμβάνει το ερωτηματολόγιο, το βιβλίο κωδικοποίησης και την αντιστοίχιση μεταβλητών και ερωτήσεων και δημιουργεί το ηλεκτρονικό ερωτηματολόγιο που θα επιτρέπει την πλοήγηση στις ερωτήσεις και την καταγραφή των απαντήσεων.

- DM10:** Τράπεζα Μεταβλητών: Τεκμηρίωση μεταβλητών και κατηγοριοποιήσεων (κωδικών & κατηγοριών). Αντιστοίχιση μεταβλητών με ερωτήσεις. Δημιουργία νέων μεταβλητών και καταγραφή του σχετικού αλγόριθμου / Χρήστες: Κύριος Ερευνητής & Στατιστικός
- DM 11:** Δημιουργία συστήματος CAI διενέργειας ερευνών / Χρήστες: Συνεντευκτής Ερωτώμενος.

5.3. Επαναχρησιμοποίηση ερευνητικών συνιστωσών: το πακέτο μεταδεδομένων

Σε ιδεατές συνθήκες, ο ερευνητής μπορεί να ανατρέχει σε προηγούμενες έρευνες και στις ερευνητικές συνιστώσες τους ώστε να επιλύει ζητήματα που αφορούν στο σχεδιασμό της έρευνας και του ερωτηματολογίου και να αποφεύγει διπλοεγγραφές μεταδεδομένων στο σύστημα. Για παράδειγμα, η λεκτική απόδοση εννοιών υπό τη μορφή ερωτήσεων είναι γνωστό πρόβλημα που αντιμετωπίζουν οι κοινωνικοί επιστήμονες κατά τη φάση σχεδιασμού ενός ερωτηματολογίου. Επίσης, ο υπολογισμός της μέτρησης σύνθετων εννοιών που απορρέουν από άλλες διαστάσεις – μεταβλητές δυσχεραίνει το έργο του κοινωνικού επιστήμονα κατά τη φάση της ανάλυσης των δεδομένων και, άρα, κατά τη φάση του σχεδιασμού πρέπει να προβληματιστεί για την προσθήκη των κατάλληλων ερωτήσεων. Η ύπαρξη ενός συστήματος όπου είναι καταγεγραμμένες μεμονωμένες ή συσχετισμένες ερευνητικές συνιστώσες όπως έννοιες, τρόποι μέτρησης εννοιών (κλίμακες, δείκτες, ταξινομήσεις), ερωτήσεις και μεταβλητές, θα διευκόλυνε τον ερευνητή ως προς το σχεδιασμό του ερωτηματολογίου, αλλά και ως προς τη στατιστική ανάλυση. Οι ερευνητές μπορεί να ενδιαφέρονται είτε για μεμονωμένες ερευνητικές συνιστώσες (π.χ. μια ταξινόμηση) είτε για συσχετισμένες ερευνητικές συνιστώσες (π.χ. να εντοπίσουν βάσει μιας έννοιας, τους τρόπους μέτρησής της, τη λεκτική απόδοσή της, αλλά και αλγόριθμο μέτρησης στην περίπτωση σύνθετης έννοιας που προκύπτει από άλλες).

Ένα σύνολο από συσχετισμένες ή μη ερευνητικές συνιστώσες αφορά ένα πακέτο μεταδεδομένων, το οποίο εξυπηρετεί διττό στόχο. Ο πρώτος αφορά στην επαναχρησιμοποίηση των ερευνητικών συνιστωσών από τους κύριους ερευνητές κατά το σχεδιασμό της έρευνας, ενώ ο δεύτερος αφορά στη διάχυση ενός πακέτου μεταδεδομένων, ως αυτόνομο έργο με στόχο τη δευτερογενή χρήση. Παρακάτω, παρατίθενται δύο παραδείγματα πακέτων μεταδεδομένων.

Το πρώτο πακέτο μεταδεδομένων που αφορά στην αποτύπωση και στον υπολογισμό των διαστάσεων του κοινωνικού κεφαλαίου περιλαμβάνει: «Έννοιες (υποέννοιες) – Ερωτήσεις – Μεταβλητές – Κατηγοριοποιήσεις (ταξινομήσεις) – Αλγόριθμους υπολογισμού μεταβλητών». Αρχικά, τίθενται οι παρακάτω υποέννοιες που προσδιορίζουν την έννοια του κοινωνικού κεφαλαίου (Pippa and James, 2007· Kaasa, 2009):

Κοινωνικό κεφάλαιο

- Γενικευμένη εμπιστοσύνη
- Εμπιστοσύνη στους θεσμούς
- Κοινωνική ανοχή
- Τυπικά Δίκτυα
- Άτυπα Δίκτυα
- Πολιτική συμμετοχή
- Νόρμες κοινωνικής συμπεριφοράς

Σε δεύτερη φάση, θα μπορούσαν να προσδιοριστούν οι ερωτήσεις που χρησιμοποιούνται για να μετρηθούν οι υφιστάμενες έννοιες. Για παράδειγμα, ερωτήσεις μέτρησης της έννοιας της γενικευμένης εμπιστοσύνης είναι οι ακόλουθες:

Ερώτηση	Κωδικοί – Κατηγορίες
<p>Γενικά μιλώντας, θα λέγατε ότι μπορούμε να έχουμε εμπιστοσύνη στους περισσότερους ανθρώπους, ή ότι πρέπει να είμαστε πάντα επιφυλακτικοί στις σχέσεις μας με τους ανθρώπους;</p>	<p>0 → «πρέπει να είμαστε πάντα επιφυλακτικοί στις σχέσεις μας με τους ανθρώπους»</p> <p>1</p> <p>2</p> <p>3</p> <p>4</p> <p>5</p> <p>6</p> <p>7</p> <p>8</p> <p>9</p> <p>10 → «μπορούμε να έχουμε εμπιστοσύνη στους περισσότερους ανθρώπους».</p> <p>99 → Δ.Γ /Δ.Α</p>
<p>Πιστεύετε ότι οι περισσότεροι άνθρωποι θα προσπαθούσαν να σας εκμεταλλευτούν, αν είχαν την ευκαιρία, ή ότι θα προσπαθούσαν να είναι δίκαιοι απέναντί σας;</p>	<p>0 → «οι περισσότεροι άνθρωποι θα προσπαθούσαν να με εκμεταλλευτούν, αν είχαν την ευκαιρία»</p> <p>1</p> <p>2</p> <p>3</p> <p>4</p> <p>5</p> <p>6</p> <p>7</p> <p>8</p> <p>9</p> <p>10 → «οι περισσότεροι άνθρωποι θα προσπαθούσαν να είναι δίκαιοι απέναντί μου».</p> <p>99 → Δ.Γ /Δ.Α</p>
<p>Θα λέγατε ότι τις περισσότερες φορές οι άνθρωποι προσπαθούν να βοηθήνε τον άλλο, ή ότι νοιάζονται κυρίως για τον εαυτό τους;</p>	<p>0 → «Οι άνθρωποι νοιάζονται κυρίως για τον εαυτό τους»</p> <p>1</p> <p>2</p> <p>3</p> <p>4</p> <p>5</p> <p>6</p> <p>7</p> <p>8</p> <p>9</p> <p>10 → «οι άνθρωποι προσπαθούν να βοηθήνε τον άλλον».</p> <p>99 → Δ.Γ /Δ.Α</p>

Κατόπιν προσδιορίζονται οι μεταβλητές οι οποίες προκύπτουν από τις παραπάνω ερωτήσεις, ως εξής:

Όνομα	Ετικέτα	Κωδικοί-Κατηγορίες	Μέτρηση
Gen_trst	Γενικευμένη Εμπιστοσύνη	Όπως στις ερωτήσεις	Κλίμακας
Trst_people	Εμπιστοσύνη στους περισσότερους ανθρώπους ή επιφυλακτικότητα	Όπως στις ερωτήσεις	Διάταξης
Justice_people	Εκμετάλλευση από τους άλλους ανθρώπους ή δικαιοσύνη	Όπως στις ερωτήσεις	Διάταξης
Help_people	Οι άνθρωποι βοηθάνε τον άλλο ή νοιάζονται κυρίως για τον εαυτό τους	Όπως στις ερωτήσεις	Διάταξης

Τέλος, προσδιορίζεται ο αλγόριθμος που θα χρησιμοποιηθεί για τη μέτρηση των υφιστάμενων εννοιών, για παράδειγμα η μεταβλητή της γενικευμένης εμπιστοσύνης μπορεί να υπολογιστεί ως ένας δείκτης που προκύπτει από το άθροισμα των τιμών των μεταβλητών: Trst_people, Justice_people και Help_people.

Ένα άλλο παράδειγμα πακέτου μεταδεδομένων αφορά την Ευρωπαϊκή κοινωνικο-οικονομική ταξινόμηση (European Socio-Economic Classification – ESeC), η οποία σχεδιάστηκε ώστε να είναι εφικτή η μέτρηση της κοινωνικο-οικονομικής τάξης ενός ατόμου (βλ. Harrison and Rose, 2006· Rose and Harrison, 2007). Οι ταξικές κατηγορίες κατά ESEC είναι οι ακόλουθες δέκα (Εμμανουήλ, 2016, σελ. 50):

	Κατηγορία βάσει της ESeC
1	Μεγάλοι εργοδότες, ανώτεροι μάνατζερ, ανώτεροι επαγγελματίες
2	Κατώτεροι μάνατζερ/ επαγγελματίες, ανώτεροι επόπτες/τεχνικοί
3	Ενδιάμεσα επαγγέλματα (ανώτεροι υπάλληλοι ρουτίνας)
4	Μικροί εργοδότες και αυτοαπασχολούμενοι (εκτός γεωργίας)
5	Μικροί εργοδότες και αυτοαπασχολούμενοι (γεωργία κλπ)
6	Κατώτεροι τεχνίτες και επόπτες
7	Κατώτεροι υπάλληλοι ρουτίνας σε εμπόριο και υπηρεσίες
8	Κατώτερα τεχνικά επαγγέλματα (ειδικευμένοι χειρώνακτες)
9	Εργάτες ρουτίνας (ημι-ειδικευμένοι και ανειδίκευτοι)
10	Δεν δούλεψαν ποτέ ή μακροχρόνια άνεργοι

Η κοινωνικο-οικονομική τάξη ενός ατόμου μπορεί να μετρηθεί με δύο τρόπους βάσει ESEC:

- Ο πρώτος και απλούστερος τρόπος μέτρησης προϋποθέτει την ύπαρξη μιας μεταβλητής επαγγέλματος που έχει κωδικοποιηθεί βάσει της ταξινόμησης ISCO88 σε τριψήφιο επίπεδο.
- Ο δεύτερος και συνθετότερος τρόπος προϋποθέτει την ύπαρξη μεταβλητής επαγγέλματος βάσει της ταξινόμησης ISCO88 σε τριψήφιο επίπεδο καθώς και μεταβλητή που αφορά τη θέση απασχόλησης (employment status) του ερωτώμενου. Αρχικά, προσδιορίζεται η θέση απασχόλησης του ερωτώμενου βάσει των παρακάτω μεταβλητών:
 - τη σχέση απασχόληση (employment relation) (μισθωτός / αυτοαπασχολούμενος, οικογενειακός βοηθός)
 - το μέγεθος της επιχείρησης (0, 1-9, 10+) καθώς και
 - το αν ο εργαζόμενος έχει ή όχι εποπτικό ρόλο

Στη συνέχεια, η θέση απασχόλησης κωδικοποιείται σε πέντε κατηγορίες, όπως παρακάτω (στο ίδιο, σελ. 52):

1. Αυτοαπασχολούμενοι (εργοδότες) με μεγάλες μονάδες (10+ υπαλλήλους)
2. Αυτοαπασχολούμενοι (εργοδότες) με μικρές μονάδες (1-9 υπαλλήλους)
3. Αυτοαπασχολούμενοι (0 και χωρίς στοιχεία μεγέθους)
4. Μισθωτοί (και οικογενειακοί βοηθοί) επόπτες
5. Μισθωτοί (και οικογενειακοί βοηθοί) μη - επόπτες.

Κατόπιν, η περίπτωση εντάσσεται σε μία από τις 10 κατηγορίες της ESEC ανάλογα με τη θέση απασχόλησης και το επάγγελμα βάσει ενός αλγόριθμου⁷.

Για την παραπάνω ταξινόμηση, μπορεί να δημιουργηθεί ένα πακέτο μεταδεδομένων που να εμπεριέχει τις μεταβλητές που χρησιμοποιούνται για τον υπολογισμό της ESEC και τις κατηγορίες τους και, επιπλέον, την ίδια την ταξινόμηση υπό μορφή μεταβλητής και συνοδευόμενη από τον αλγόριθμο (ή SPSS syntax) που επιτρέπει τον υπολογισμό αυτής από τις προαναφερθείσες μεταβλητές. Το πακέτο

7. Οι αλγόριθμοι υπολογισμού των κατηγοριών ESEC, βάσει του πρώτου και δεύτερου τρόπου μέτρησης παρατίθενται σε SPSS syntax στον παρακάτω σύνδεσμο: <https://www.iser.essex.ac.uk/archives/esec/matrices-and-syntax>

μεταδεδομένων θα μπορούσε να γίνει ακόμα πιο σύνθετο, αν συμπεριλάμβανε ερωτήσεις και έννοιες.

Η Eurostat μέσω ενός εξυπηρετητή δεδομένων, του Ramon,⁸ παρέχει πρόσβαση σε πολλά ταξινομητικά σχήματα όπως ταξινομήσεις, γλωσσάρια, θησαυρούς και ελεγχόμενα λεξιλόγια, καθώς και σε έννοιες και ορισμούς. Επίσης, στο πλαίσιο του Ευρωπαϊκού ερευνητικού έργου «CESSDA Preparatory Phase Project⁹» της CESSDA,¹⁰ επιχειρήθηκε η δημιουργία ενός λογισμικού διαχείρισης εννοιών, ταξινομήσεων και αλγορίθμων (3CDB – Concept Classification and Conversion Database) καθώς και εννοιών – ερωτήσεων (Question Data Bank – QDB) (βλ. Bourmpos and Linardis, 2009· Kallas and Linardis, 2010). Τέτοια ταξινομητικά σχήματα, γλωσσάρια και θησαυροί θα μπορούσαν να ενσωματωθούν στην υποδομή με στόχο την επαναχρησιμοποίησή τους.

DM12: Τράπεζα Πακέτων Μεταδεδομένων (συσχετισμένων ή μη ερευνητικών συνιστωσών) *Χρήστες:* Κύριος Ερευνητής

Ακολουθεί η ενσωμάτωση του δειγματοληπτικού πλαισίου εντός της βάσης δεδομένων του συστήματος, αν υπάρχει. Το δειγματοληπτικό πλαίσιο αποτελεί έναν κατάλογο ολόκληρου ή σχεδόν ολόκληρου του πληθυσμού της έρευνας και, ανάλογα με το αν υπάρχει ή όχι, διαφοροποιείται και η μέθοδος δειγματοληψίας. Σε αυτό αποθηκεύονται, επίσης, τα στοιχεία «επικοινωνίας» των δυνητικά ερωτώμενων (διεύθυνση, όταν η έρευνα διεξάγεται μέσω CAPI, τηλέφωνο για CATI συστήματα, e-mail για CAWI συστήματα). Ένα μη τρώο ατόμων, ο τηλεφωνικός ή ο εκλογικός κατάλογος αποτελούν εν δυνάμει δειγματοληπτικό πλαίσιο από το οποίο επιλέγεται ένα αντιπροσωπευτικό δείγμα βάσει διαφόρων τεχνικών δειγματοληψίας. Ανάλογα με την τεχνική δειγματοληψίας, μπορεί να προκύψει αυτοματοποιημένα το δείγμα της έρευνας από το σύστημα. Η κάθε περίπτωση που θα συμπεριληφθεί στο δείγμα ταυτοποιείται μοναδικά μέσω ενός κωδικού.

8. http://ec.europa.eu/eurostat/ramon/index.cfm?TargetUri=DSP_PUB_WELC

9. <http://ppp.CESSDA.net/>

10. <http://www.CESSDA.net>

Στην περίπτωση που η δειγματοληψία επιλογής αφορά δειγματοληψία με προκαθορισμένα ποσοστά (quota sampling), είναι βοηθητικό να σωρευτούν υπό μορφή πινάκων τα ποσοστά των μεταβλητών του γενικού πληθυσμού που αφορούν τα quotas. Ακόμα και αν δεν χρησιμοποιηθούν quotas, θα μπορούσε σε αυτό το στάδιο να σωρευτούν κύβοι (πινακοποιήσεις) του γενικού πληθυσμού για τη μετέπειτα στάθμιση και τον έλεγχο του δείγματος.

Επιπλέον καταγράφεται και ενσωματώνεται στο σύστημα το υποστηρικτικό υλικό για τη διεξαγωγή της έρευνας, όπως φόρμες επαφής, κάρτες ερωτηματολογίου, οδηγίες προς ερευνητές και επόπτες, συνοδευτικές επιστολές, φυλλάδια κ.ά. Το υποστηρικτικό υλικό συνδέεται με την εκάστοτε συλλογή δεδομένων και μπορεί να ενταχθεί σε μία «τράπεζα τεκμηρίων».

DM13: Σύστημα διαχείρισης δειγματοληπτικού πλαισίου και επιλογή δείγματος βάσει μεθόδου δειγματοληψίας).

Χρήστες: Στατιστικός

DM14: Τράπεζα τεκμηρίων. *Χρήστες:* Κύριος Ερευνητής

DM15: Τράπεζα Κύβων. *Χρήστες:* Κύριος Ερευνητής – Στατιστικός

5.4. Έρευνα πεδίου – Συλλογή Δεδομένων

Στην έρευνα πεδίου, ο συντονιστής επόπτης αναθέτει περιπτώσεις στους επόπτες. Αυτοί με την σειρά τους αναθέτουν περιπτώσεις στους συνεντευκτές για τη διεξαγωγή των συνεντεύξεων. Σε έρευνες με χρήση έντυπων υλικών, οι συνεντευκτές χρησιμοποιούν έντυπη φόρμα επαφών για να καταγράψουν τις διαδικασίες για την οριστικοποίηση της συνέντευξης. Μια περίπτωση οριστικοποιείται επιτυχώς εάν επιλεγεί ερωτώμενος ή οριστικοποιείται ανεπιτυχώς εάν δεν επιλεγεί για οποιονδήποτε λόγο.

Ηλεκτρονικά, η διαδικασία αυτή επιτυγχάνεται μέσω ενός συστήματος διαχείρισης περιπτώσεων (Case Management System - CMS). Το CMS καταγράφει σε μια βάση δεδομένων το σύνολο των ενεργειών (ραντεβού, επισκέψεις, τηλεφωνικές επαφές, αποστολή email κ.ά.) μέχρι να οριστικοποιηθεί μια περίπτωση και εξασφαλίζει διαφορετικές λειτουργίες για διαφορετικούς ρόλους χρηστών (συντονιστής επόπτης, επόπτες, συνεντευκτές). Το CMS θα πρέπει να υποστηρίζει

τουλάχιστον τη διαχείριση των δύο τελευταίων μονάδων που επιλέγονται σε μια πολυσταδιακή στρωματοποιημένη δειγματοληψία. Για παράδειγμα, στην έρευνα PIAAC_GR¹¹ εφαρμόστηκε μια πολυσταδιακή στρωματοποιημένη δειγματοληψία με πρωτογενή μονάδα έρευνας την επιφάνεια (ένα ή περισσότερα οικοδομικά τετράγωνα ή ένας μικρός οικισμός), δευτερογενή μονάδα το νοικοκυριό και τελική μονάδα έρευνας το άτομο ηλικίας 16-65 ετών. Στην προτεινόμενη μέθοδο δειγματοληψίας, το CMS θα πρέπει να διαχειρίζεται ταυτόχρονα τα νοικοκυριά και τα άτομα που ανήκουν σε ένα νοικοκυριό.

Το CMS ενημερώνεται για τους χρήστες και τους ρόλους τους μέσω του συστήματος διαχείρισης χρηστών. Επίσης, ενημερώνεται για τις περιπτώσεις του δείγματος που επελέγησαν, οι οποίες ταυτοποιούνται μοναδικά σε όλα τα επιμέρους συστήματα.

Ακολουθούν οι λειτουργίες ενός CMS ανά χρήστη (δεν θα αναφερθούμε εδώ στην τελική και στην προ της τελικής μονάδα έρευνας, αλλά σε νοικοκυριό και άτομο) για την καλύτερη κατανόηση. Έτσι, μέσω του CMS (Client CMS), ο συνεντευκτής:

- Πληροφορείται για τα στοιχεία των νοικοκυριών που του έχουν ανατεθεί από τον επόπτη, καθώς και για τις περιπτώσεις (νοικοκυριά ή άτομα) που έχει ολοκληρώσει ή εκκρεμούν. Κάθε νοικοκυριό έχει μοναδικό κωδικό ταυτοποίησης σε όλα τα επιμέρους συστήματα.
- Καταγράφει τις προσπάθειες επαφής για ένα νοικοκυριό (ημερομηνία, ώρα και σχόλια). Στην περίπτωση ανεπιτυχών επαφών, δίνει κάποιο προσωρινό κωδικοποιημένο αποτέλεσμα της επαφής (π.χ. Κανείς στο νοικοκυριό / Επιστροφή στο νοικοκυριό με ραντεβού/ Επιστροφή στο νοικοκυριό χωρίς ραντεβού/ Άρνηση - επιδέχεται επανεπίσκεψη κ.ά.).
- Καταγράφει τις λεπτομέρειες του ραντεβού όταν οριστικοποιηθεί (ημερομηνία, ώρα, σχόλια, τόπος).
- Οριστικοποιεί ένα νοικοκυριό για μη διεξαγωγή της συνέντευξης δίνοντας έναν τελικό κωδικό (Screener Disposition Code) που επεξηγεί τους λόγους (Άρνηση από μέλος νοικοκυριού / Γλωσσικό πρόβλημα/ Μέγιστος αριθμός προσπαθειών επαφής / Κενή κατοικία/ Προβλήματα ακοής κ.ά.).

11. Πρόγραμμα για τη Διεθνή αποτίμηση των Ικανοτήτων των Ενηλίκων (PIAAC) www.piaac.gr

- Όταν ο εκπρόσωπος του νοικοκυριού αποφάινεται θετικά, τότε ο συνεντευκτής ενεργοποιεί τη διαδικασία screening – καταγράφει δηλαδή τα άτομα του νοικοκυριού (όνομα, επώνυμο, φύλο, ημερομηνία γέννησης κ.ά) και επιλέγει αυτοματοποιημένα μέσω Kish grid¹² τον ερωτώμενο. Το κάθε άτομο δέχεται μοναδικό κωδικό ταυτοποίησης, που παραπέμπει στον μοναδικό κωδικό ταυτοποίησης του νοικοκυριού, ώστε να είναι εύκολος ο εντοπισμός του ατόμου εντός του νοικοκυριού.
- Όταν επιλεγεί το άτομο και είναι άμεσα διαθέσιμο, μπορεί να γίνει έναρξη της διαδικασίας συνέντευξης (ερωτηματολόγιο CAI) μέσω του CMS. Το άτομο οριστικοποιείται, εφόσον επιβεβαιώνονται τα στοιχεία που συλλέχθηκαν κατά τη διαδικασία screening και δίνεται τελικός κωδικός (Case Initialization Disposition Code).
- Αν το άτομο δεν είναι άμεσα διαθέσιμο, θα πρέπει να δίνεται ένας προσωρινός κωδικός (απουσία ερωτώμενου/ επιστροφή στον ερωτώμενο με ραντεβού / επιστροφή στον ερωτώμενο χωρίς ραντεβού/ άρνηση - επιδέχεται επανεπίσκεψη κ.ά.).
- Αν δεν πραγματοποιηθεί τελικά η συνέντευξη με το άτομο που επιλέχθηκε, δίνεται ένας τελικός κωδικός σε επίπεδο ατόμου επεξηγώντας τους λόγους.
- Το σύστημα CATI διενεργεί τις παραπάνω διαδικασίες καθώς και κλήση τηλεφώνου με αυτοματοποιημένο τρόπο, χωρίς ο συνεντευκτής να πληκτρολογεί τον τηλεφωνικό αριθμό.

Μέσω του CMS (Admin CMS) ο επόπτης:

- Πληροφορείται για τις περιπτώσεις που του έχουν ανατεθεί από τον συντονιστή επόπτη.
- Αναθέτει περιπτώσεις στους συνεντευκτές που εποπτεύει.
- Ελέγχει (τηλεφωνικά ή επιτόπου) την ποιότητα εργασίας των συνεντευκτών που εποπτεύει με βάση τις πληροφορίες (reports) που εξάγονται από το σύστημα με κριτήρια (π.χ. ανά ημέρα / ανά συνεντευκτή). Καταγράφει στο CMS το αποτέλεσμα του ποιοτικού ελέγχου (Quality Control – Flag).

12. Μέθοδος επιλογής μελών νοικοκυριού με προκαθορισθέντες αριθμούς που δημιουργήσε ο στατιστικός Leslie Kish και χρησιμοποιείται ευρέως στην εμπειρική έρευνα.

Ο συντονιστής επόπτης παρακολουθεί και έχει πρόσβαση στις περιπτώσεις που έχουν ανατεθεί σε όλους τους συνεντευκτές. Στην περίπτωση CAWI, δεν εμπλέκονται επόπτες και συνεντευκτές και η αποστολή των e-mails μπορεί να γίνεται από ένα άτομο (π.χ. το συντονιστή επόπτη).

Έπειτα οι συνεντευκτές ή οι ερωτώμενοι χρησιμοποιούν το DM11, δηλαδή το σύστημα CAI.

DM16: Σύστημα διαχείρισης περιπτώσεων CMS (Admin/ Client) /
Χρήστες: Συντονιστής επόπτης, επόπτες, συνεντευκτές

Ορισμένες ερωτήσεις ενός ερωτηματολογίου ενδέχεται να είναι ανοιχτές. Για παράδειγμα, σε γενικού ενδιαφέροντος έρευνες οι ερωτήσεις του επαγγέλματος, του κλάδου οικονομικής δραστηριότητας της επιχείρησης στην οποία εργάζεται/εργαζόταν ο ερωτώμενος, δεν προβλέπουν απαντητικές κατηγορίες. Για τις προαναφερόμενες ερωτήσεις, υφίστανται διεθνείς ταξινομήσεις (π.χ. International Standard Classification of Occupations – ISCO, International Standard Industrial Classification of All Economic Activities – ISIC). Η εργασία της εκ των υστέρων κωδικοποίησης των ανοιχτών ερωτήσεων πραγματοποιείται από εξειδικευμένους επαγγελματίες που αντιστοιχούν τις ανοιχτές ερωτήσεις σε ανάλογο κωδικό της συγκεκριμένης ταξινόμησης. Το σύστημα θα πρέπει να προβλέπει: α) την προβολή επιλεγμένων μεταβλητών που διευκολύνουν την εκ των υστέρων κωδικοποίηση των επαγγελματιών (οι μεταβλητές που θα εμφανίζονται θα είναι «μόνο για ανάγνωση» και δεν επιτρέπεται να αλλάζουν) και β) την καταχώρηση των κωδικών σε νέες μεταβλητές βάσει των υφιστάμενων ταξινομήσεων.

Η διαδικασία αυτή, λόγω του ποιοτικού ελέγχου των δεδομένων, εντάσσεται ως παράλληλη διαδικασία με τη διαξαγωγή της έρευνας πεδίου, ενώ μπορεί να γίνεται και μετά την ολοκλήρωσή της.

DM17: Σύστημα κωδικοποίησης ανοιχτών ερωτήσεων /
Χρήστες: Κωδικοποιητής

Οι επόπτες και ο κύριος ερευνητής (μέσω των reports που παράγονται από το CMS) έχουν πρόσβαση στη διαχειριστική πληροφορία

των συνεντεύξεων και όχι στα ίδια τα δεδομένα που παράγονται κατά τη διεξαγωγή της έρευνας πεδίου. Γι' αυτό απαιτείται ένα εργαλείο, το οποίο να ενοποιεί τα δεδομένα από τις διάφορες βάσεις δεδομένων στα επί μέρους συστήματα (δειγματοληπτικό πλαίσιο, CMS, CAI, σύστημα κωδικοποίησης ανοιχτών ερωτήσεων). Η διαδικασία συνένωσης των επιμέρους βάσεων δεδομένων σε μία επιτυγχάνεται εύκολα, λόγω των μοναδικών κωδικών ταυτοποίησης που υφίστανται για την τελική και την προ της τελικής, μονάδα έρευνας.

Ο διαχειριστής δεδομένων είναι αρμόδιος για τη λειτουργία του εργαλείου. Παράγει κατανομές ανά μεταβλητή και διασταυρώσεις μεταβλητών και δύναται να περιλαμβάνει προκαθορισμένους ελέγχους πληρότητας και ορθότητας των μεταβλητών. Παράγει reports και εξάγει τα δεδομένα σε μορφές κατάλληλες (π.χ. csv, SPSS, SAS) για τον περαιτέρω έλεγχο των δεδομένων. Τα reports και τα δεδομένα αποστέλλονται στους κύριους ερευνητές και στους επόπτες.

DM18: Σύστημα διαχείρισης δεδομένων / Χρήστες:
Διαχειριστής δεδομένων

5.5. Μετά την έρευνα

Τα δεδομένα που προέκυψαν από την έρευνα πεδίου αποθηκεύονται «ως τελικά» στο σύστημα διαχείρισης δεδομένων και παράγεται αντίγραφο αυτών. Ο διαχειριστής, μέσω του συστήματος διαχείρισης δεδομένων, παρεμβαίνει στο αντίγραφο για τη διόρθωση σφαλμάτων που προκύπτουν από τον ποιοτικό έλεγχο καθώς και άλλα λογικά σφάλματα. Σε συνεργασία με τον στατιστικό καταλογίζει δεδομένα (imputation), όταν μεταβλητές εμφανίζουν υψηλό ποσοστό ελλειψών τιμών. Ακολουθεί η ανωνυμοποίηση των δεδομένων μέσω της διαγραφής μεταβλητών που επιτρέπουν την ταυτοποίηση των υποκειμένων της έρευνας όπως: όνομα, διεύθυνση, τηλέφωνο, e-mail, καθώς επίσης και μέσω της μετατροπής διαχειριστικών μεταβλητών σε γενικότερες μεταβλητές, όταν εντοπίζονται περιπτώσεις στις οποίες μια μονάδα μπορεί να ταυτοποιηθεί (π.χ. στην περίπτωση που υφίσταται γεωγραφική μεταβλητή σε επίπεδο νομού, δημιουργείται μεταβλητή στο γενικότερο επίπεδο της γεωγραφικής περιφέρειας).

Επιπρόσθετα, το σύστημα μπορεί να παράγει αυτοματοποιημένα τη μεταβλητή στάθμισης των δεδομένων βάσει επιλεγμένων μεταβλητών και των αντίστοιχων κύβων (πινάκων) του γενικού πληθυ-

σμού, που έχουν σωρευτεί στην τράπεζα κύβων του συστήματος. Το νέο σύνολο δεδομένων που προκύπτει είναι αυτό που μπορεί να διατεθεί στους δευτερογενείς χρήστες για περαιτέρω επεξεργασία και ανάλυση (PUF: Public Use Files).

Στη συνέχεια, ενημερώνεται η «συλλογή δεδομένων» με πρόσθετες πληροφορίες, όπως το διάστημα διεξαγωγής της έρευνας, τις αποκλίσεις από τον σχεδιασμό, τον υπολογισμό του ρυθμού απόκρισης κ.ά. Τέλος, το βιβλίο κωδικοποίησης μπορεί να μεταφραστεί σε διάφορες γλώσσες για περαιτέρω διάχυση.

DM19: Σύστημα διενέργειας σταθμίσεων / Χρήστες:

Στατιστικός

DM20: Σύστημα μετάφρασης βιβλίου κωδικοποίησης /

Χρήστες: Μεταφραστής

5.6. Προετοιμασία για διάχυση της έρευνας

Η έρευνα ολοκληρώνεται με τη συγγραφή του report (εισάγεται στο σύστημα τεκμηρίων), καθώς και με τη δημιουργία πινάκων (κύβων) ιδιαίτερου ενδιαφέροντος. Οι κύβοι που θα προκύψουν σε αυτό το στάδιο τυγχάνουν διαφορετικής διαχείρισης από τους κύβους που χρησιμοποιήθηκαν κατά τις διαδικασίες δειγματοληψίας και στάθμισης. Αυτό συμβαίνει διότι οι κύβοι αυτοί προκύπτουν άμεσα από τα ίδια τα δεδομένα της έρευνας, διασταυρώνοντας μεταβλητές και θεωρώντας μια συνάρτηση υπολογισμού μεταξύ τους, σε αντίθεση με τους κύβους που χρησιμοποιούνται στις διαδικασίες δειγματοληψίας και στάθμισης που έχουν προκύψει από άλλες έρευνες και απογραφές, τα πρωτογενή δεδομένα των οποίων δεν εμπεριέχονται κατ' ανάγκη στο σύστημα.

Αν στα δεδομένα εμπεριέχονται γεωγραφικές μεταβλητές, ιδεατά το σύστημα θα μπορούσε να υποστηρίξει τη δημιουργία χαρτών. Θα υπάρχει δυνατότητα προσυνδεδεμένες μεταβλητές με χάρτες να τεθούν ως πακέτο μεταδεδομένων και να επαναχρησιμοποιούνται σε όλες τις έρευνες και για την αυτοματοποιημένη δημιουργία χαρτών. Ουσιαστικά, η λογική των γεωγραφικών μεταβλητών προσυνδεδεμένων με χάρτη αφορά στη δημιουργία ενός κύβου στον οποίο συμμετέχει η γεωγραφική μεταβλητή και στην απεικόνιση των δεδομένων του σε χάρτη.

Οι ενέργειες που παρατίθενται στη συνέχεια προετοιμάζουν τη δευτερογενή χρήση δεδομένων:

- «Καθάρισμα» τεκμηρίωσης. Το σύστημα σε πλήρη ανάπτυξη θα υποστηρίζει διαφορετικά ακροατήρια σε ορισμένα περιγραφικά πεδία.
- Δημιουργία άδειας ανοιχτής πρόσβασης (open license) και ένταξη στο σύστημα τεκμηρίων.
- Δημιουργία επιπέδων πρόσβασης σε δεδομένα, μεταδεδομένα και τεκμήρια βάσει της άδειας ανοιχτής πρόσβασης.
- Σύνδεση έρευνας με σειρά ερευνών (linked data).
- Ορισμός πακέτων μεταδεδομένων προς επαναχρησιμοποίηση.
- Δημιουργία μοναδικού URL για τις βασικές οντότητες: έρευνες, κύβοι, πακέτα μεταδεδομένων, σειρές.
- Πρόβλεψη εξαγωγής δεδομένων σε μορφές που δεν υποστηρίζουν συγκεκριμένους προμηθευτές λογισμικού (για παράδειγμα, τα δεδομένα είναι προτιμητέο να εξάγονται σε μορφή csv παρά σε μορφή excel ή SPSS).
- Ένταξη της έρευνας, κύβων, πακέτων μεταδεδομένων και σειρών σε θεματικές κατηγορίες.

Στη συνέχεια, η έρευνα δημοσιεύεται στο σύστημα και η τεκμηρίωση ή τα δεδομένα δεν τροποποιούνται, παρά δέχονται νέες εκδόσεις.

- DM21:** Σύστημα δημιουργίας κύβων από μικροδεδομένα / Χρήστες: Στατιστικός – Κύριος Ερευνητής
- DM22:** Σύστημα δημιουργίας χαρτών/ Χρήστες: Στατιστικός
- DM23:** Σύστημα δημιουργίας επιπέδων πρόσβασης σε δεδομένα, μεταδεδομένα και τεκμήρια / Χρήστες: Διαχειριστής
- DM24:** Τράπεζα σειρών ερευνών / Χρήστες: Διαχειριστής & Κύριος Ερευνητής
- DM25:** Σύστημα διαχείρισης θησαυρών όρων και ένταξης των βασικών οντοτήτων σε θησαυρούς / Χρήστες: Διαχειριστής & Κύριος Ερευνητής

5.7. Αναζήτηση και ανακάλυψη ερευνών και δεδομένων για δευτερογενή χρήση

Οι δευτερογενείς χρήστες αναζητούν έρευνες με ερωτήσεις όπως: «Εντόπισε εκείνες τις έρευνες ή τις μεταβλητές που αναφέρονται σε συγκεκριμένο τόπο, χρόνο και θεματική», οδηγώντας στην εμφάνιση πολλών ερευνών και μεταβλητών. Στο σημείο αυτό, ο χρήστης θα διερευνήσει πλέον τα δεδομένα μέσω της περαιτέρω τεκμηρίωσης: «Από ποια ερώτηση προήλθε η συγκεκριμένη μεταβλητή; Ποια είναι η έννοια που καλείται να μετρήσει; Πώς κατασκευάστηκε μια μεταβλητή που έχει προέλθει από επεξεργασία άλλων μεταβλητών; Ποιες είναι οι κατηγορίες της μεταβλητής; Είναι δυνατή μια on line στατιστική ανάλυση, ώστε ο χρήστης να ανακαλύψει περαιτέρω τα δεδομένα»;

Ταυτόχρονα, ο δευτερογενής χρήστης μπορεί να ενδιαφέρεται και για πινακοποιημένα δεδομένα που παρέχουν ευρύτερη εικόνα ενός πληθυσμού. Τα στατιστικά δεδομένα αποτελούν την πρώτη ύλη για την παραγωγή δευτερογενούς πληροφόρησης, κατάλληλης για την ερμηνεία των φαινομένων και την άσκηση πολιτικής (Ρόντος, 2011). Ερευνητές, αναλυτές, η επιστημονική κοινότητα εν γένει μπορεί να ενδιαφέρονται και για τα πακέτα μεταδεδομένων τα οποία, υπό ευρεία έννοια, συγκροτούν δυνητικά αυτόνομα δημοσιευμένα έργα ισότιμης αξίας με επιστημονική δημοσίευση.

Οι εφαρμογές που υποστηρίζουν αναζήτηση και ανακάλυψη ενέχουν τις ακόλουθες βασικές λειτουργικές προδιαγραφές:

- Αναζήτηση βάσει συγκεκριμένων πεδίων όπως: τίτλος, πληθυσμός αναφοράς, τόπος, χρόνος, θεματική
- Σύνθετη αναζήτηση
- Πλοήγηση εντός των μεταδεδομένων
- Δημιουργία αιτήματος πρόσβασης στα δεδομένα όταν δεν είναι ελεύθερης πρόσβασης
- Ανάκτηση δεδομένων
- Ανακάλυψη δεδομένων χρησιμοποιώντας on line εργαλεία στατιστικής ανάλυσης και χάρτες.

DM27: Αναζήτηση και ανακάλυψη ερευνών, κύβων, σειρών και πακέτων μεταδεδομένων / Χρήστες: Δευτερογενείς χρήστες

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η συνεχής ανάπτυξη της τεχνολογίας και του διαδικτύου επέδρασε στη μεθοδολογία των κοινωνικών επιστημών και διεύρυνε τους ορίζοντες χρήσης και διαχείρισης των κοινωνικών δεδομένων. Η επιστήμη εξελίσσεται, ενώ, παράλληλα, ο παγκοσμιοποιημένος κόσμος δημιουργεί την ανάγκη συνεχούς ροής πληροφοριών θέτοντας ταυτόχρονα χρηματοδοτικούς και δημοσιονομικούς περιορισμούς σε εθνικό επίπεδο. Οι διεθνείς οργανισμοί συνεργάζονται ανταλλάσσοντας δεδομένα και προάγοντας την παραγωγή και επαναχρησιμοποίησή τους για ερευνητικούς και επιχειρησιακούς σκοπούς. Οι επιστήμονες παγκοσμίως συγκροτούν κοινότητες ανά τομέα ερευνητικού ενδιαφέροντος. Μεταξύ αυτών, η κοινότητα των κοινωνικών επιστημόνων που ασχολείται με τα δεδομένα προσπαθεί με συστηματικό τρόπο να παρέχει γνώση, εργαλεία και υπηρεσίες στην ακαδημαϊκή κοινότητα και κάθε ενδιαφερόμενο χρήστη.

Ο ευρωπαϊκός χώρος έρευνας μέσω του Horizon 2020 προωθεί για πρώτη φορά την άμεση δημοσιοποίηση των δεδομένων και ερευνητικών αποτελεσμάτων ως προαπαιτούμενο. Η διάχυση, φιλοξενία και συντήρηση των δεδομένων σε κατάλληλες υποδομές-αποθετήρια μεγιστοποιεί το όφελος των ερευνητικών υποδομών των κοινωνικών επιστημών, όπως η Σύμπραξη των Ευρωπαϊκών Αρχείων Κοινωνικών Δεδομένων (CESSDA). Η Ελλάδα συμμετέχει ως μέλος μέσω της εθνικής ερευνητικής υποδομής So.Da.Net¹³.

Ως ερευνητές, επιχειρούμε την προαγωγή της γνώσης και της μεθοδολογίας της κοινωνικής έρευνας σε ένα διαρκές ταξίδι προς νέες ανακαλύψεις και ανασυνθέσεις (δεδομένων), διατηρώντας ταυτόχρονα τη ματιά και στο παρελθόν (διατήρηση της μνήμης μέσω των δεδομένων) και, κυρίως, κτίζοντας τις ερευνητικές πρακτικές με συνεργασίες με τον επιστημονικό κόσμο. Το dream machine είναι το προϊόν εμπειρίας, συνεργασιών, χρήσης δεδομένων και επιθυμίας αλλαγής της κουλτούρας κοινοχρησίας και ανταλλαγής δεδομένων διότι, εν τέλει, ό,τι διαχέεται ελέγχεται μέσα από τη χρήση, την κριτική και την υπαγωγή των δεδομένων στην επιστημονική κοινότητα προς όφελος της κοινωνίας (Κονδύλη, 2008).

Συνοψίζοντας, στον παρακάτω πίνακα παρατίθενται τα 26 επιμέρους συστήματα που συγκροτούν το «dream machine» με τους εν δυ-

13. <http://www.sodanet.gr>

νάμει χρήστες, επιδιώκοντας η τεκμηρίωση και η δευτερογενής χρήση να αποτελούν υποπροϊόν των διαδικασιών που απαιτούνται πριν, κατά τη διάρκεια και μετά την ολοκλήρωση της έρευνας πεδίου. Επιπλέον, στη στήλη «Ερευνητική πρόταση» εντάσσουμε εκείνα τα συστήματα που θεωρούμε ότι, δυνητικά, αυτοματοποιούν την υποβολή μας ερευνητικής πρότασης προς έναν χρηματοδοτικό φορέα.

A/A	Περιγραφή	Χρήστες	Ερευνητική πρόταση	Πρώτη χρήση
DM1	Σύστημα «σχεδίων διαχείρισης δεδομένων» (DMP's)	Κύριος Ερευνητής	X	Πριν την έρευνα
DM2	Τράπεζα Ερευνών	Κύριος Ερευνητής	X	Πριν την έρευνα
DM3	Τράπεζα «Συλλογών δεδομένων»	Κύριος Ερευνητής	X	Πριν την έρευνα
DM4	Τράπεζα Φορέων Δράσης (Φορείς ή άτομα) και καταγραφή ρόλων (Εφαρμογή διαχείρισης χρηστών)	Διαχειριστής	X	Πριν την έρευνα
DM5	Τράπεζα εννοιών: Τεκμηρίωση εννοιών – ορισμών	Κύριος Ερευνητής	X	Πριν την έρευνα
DM6	Τράπεζα κατηγοριοποιήσεων (κωδικών – κατηγοριών)	Κύριος Ερευνητής	X	Πριν την έρευνα
DM7	Τράπεζα ερωτήσεων: τεκμηρίωση ερωτήσεων / σύνδεση ερωτήσεων με έννοιες και ορισμούς.	Κύριος Ερευνητής	X	Πριν την έρευνα
DM8	Τράπεζα ερωτηματολογίων CAI: δημιουργία ερωτηματολογίων και σχετικού αλγόριθμου	Κύριος Ερευνητής & Πληροφορικός		Πριν την έρευνα
DM9	Σύστημα μετάφρασης ερωτήσεων, απαντητικών κατηγοριών και ερωτηματολογίων	Μεταφραστής		Πριν την έρευνα
DM10	Τράπεζα Μεταβλητών: Τεκμηρίωση μεταβλητών και κατηγοριοποιήσεων (κωδικών & κατηγοριών). Αντιστοίχιση μεταβλητών με ερωτήσεις. Δημιουργία νέων μεταβλητών και καταγραφή του σχετικού αλγόριθμου.	Κύριος Ερευνητής & Στατιστικός	X	Πριν την έρευνα

A/A	Περιγραφή	Χρήστες	Ερευνητική πρόταση	Πρώτη χρήση
DM11	Σύστημα διενέργειας ερευνών CAI	Πληροφορικός		Πριν την έρευνα
DM12	Τράπεζα Πακέτων Μεταδεδομένων (συσχετισμένων ή μη ερευνητικών συνιστωσών)	Κύριος Ερευνητής		Οποτεδήποτε
DM13	Σύστημα διαχείρισης δειγματοληπτικού πλαισίου και επιλογή δείγματος βάσει τεχνικής δειγματοληψίας	Στατιστικός		Πριν την έρευνα
DM14	Τράπεζα τεκμηρίων	Κύριος Ερευνητής		Πριν την έρευνα
DM15	Τράπεζα Κύβων	Κύριος Ερευνητής & Στατιστικός		Πριν την έρευνα
DM16	Σύστημα διαχείρισης περιπτώσεων CMS (Admin/Client)	Γενικός επόπτης, επόπτες, συνεντευκτές		Κατά τη διάρκεια της έρευνας πεδίου
DM17	Σύστημα κωδικοποίησης ανοιχτών ερωτήσεων	Κωδικοποιητής		Κατά τη διάρκεια της έρευνας πεδίου
DM18	Σύστημα διαχείρισης δεδομένων	Διαχειριστής δεδομένων		Κατά τη διάρκεια της έρευνας πεδίου
DM19	Σύστημα διενέργειας σταθμίσεων	Στατιστικός		Μετά την έρευνα
DM20	Σύστημα μετάφρασης βιβλίου κωδικοποίησης	Μεταφραστής		Μετά την έρευνα
DM21	Σύστημα δημιουργίας κύβων από μικροδεδομένα	Στατιστικός / Κύριος Ερευνητής		Μετά την έρευνα
DM22	Σύστημα δημιουργίας χαρτών	Στατιστικός		Μετά την έρευνα
DM23	Σύστημα δημιουργίας επιπέδων πρόσβασης σε δεδομένα, μεταδεδομένα και τεκμήρια	Διαχειριστής		Μετά την έρευνα

A/A	Περιγραφή	Χρήστες	Ερευνητική πρόταση	Πρώτη χρήση
DM24	Τράπεζα σειρών ερευνών	Διαχειριστής & Κύριος Ερευνητής		Μετά την έρευνα
DM25	Σύστημα διαχείρισης θησαυρών και ένταξης των βασικών οντοτήτων σε θησαυρούς	Διαχειριστής & Κύριος Ερευνητής		Μετά την έρευνα
DM26	Αναζήτηση και ανακάλυψη ερευνών, κύβων, σειρών και πακέτων μεταδομένων	Δευτερογενείς χρήστες		Μετά τη δημοσιοποίηση της έρευνας

BIBΛΙΟΓΡΑΦΙΑ

Ελληνόγλωσση

- Εμμανουήλ, Δ. (2016). Κοινωνικοοικονομικές τάξεις και στρώματα status στην Αθήνα. Μια διερεύνηση της νεοβεμπεριανής προσέγγισης με τα διαθέσιμα σύνολα δεδομένων. Στο Δ. Εμμανουήλ (επιμ.), *Κοινωνικές τάξεις και κατανάλωση* (σελ. 39-86). Αθήνα: Αλεξάνδρεια.
- Κονδύλη, Δ. (2008). Ζητήματα προστασίας και διάχυσης των κοινωνικών δεδομένων. Στο Γ. Κάλλας, Δ. Κονδύλη και Γ. Καραγιάννης (επιμ.), *Μεθοδολογικά ζητήματα και ερευνητικές υποδομές των κοινωνικών επιστημών* (σελ. 109-125). Αθήνα: Ποταμός.
- Κονδύλη, Δ. και Λιναρδής Α. (2004). Ερευνητικές υποδομές: το ψηφιακό περιβάλλον της συγκριτικής έρευνας. *Επιθεώρηση Κοινωνικών Ερευνών*, 115, σελ. 119-138. doi: <http://dx.doi.org/10.12681/grsr.9347>
- Λιναρδής, Α. (2006). *Η αναγκαιότητα ανστηρής και ποιοτικής τεκμηρίωσης στις κοινωνικές έρευνες. Το πρότυπο τεκμηρίωσης Data Documentation Initiative*. Κείμενα εργασίας 2006/14 ΕΚΚΕ. Ανακτήθηκε από <http://www.ekke.gr/publications/wp/wp14.pdf>
- Ρόντος, Κ. (2011). Σύγχρονες μέθοδοι συλλογής πληθυσμιακών δεδομένων: η περίπτωση των πληθυσμιακών μητρώων. Στο Γ. Τσιώλης, Ν. Σερντεδάκις και Γ. Κάλλας (επιμ.), *Ερευνητικές υποδομές και δεδομένα στην εμπειρική κοινωνική έρευνα* (σελ. 267-292). Αθήνα: Νήσος.

Ξενόγλωσση

- Research Data Netherlands (2016). Essential 4 Data. Retrieved from <http://datasupport.researchdata.nl/en/start-de-cursus/i-definitions/open-data/>
- Berners-Lee, T. (2009). Linked data: design issues. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>
- Bourmpou, M. and Linardis, A. (2009). Functional and technical specifications of 3CDB

- (D9.2a). CESSDA Preparatory Phase Project. Retrieved from http://ppp. cessda.net/doc/D9.2a_3CDB_functional_specification.pdf
- Cortie, L., Eynden, V.D., Bishop, L. and Woollard, M. (2014). *Managing and sharing research data. A guide to good practice*. London: Sage.
- European Commission (2016). Guidelines on Data Management in Horizon 2020. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Eurostat (2000). *The metadata problem in a European context*. Working paper No.1.1.1. Retrieved from <http://www.unece.org/fileadmin/DAM/stats/documents/2000/11/metis/crp.1.e.pdf>
- Farago, P. (2014) Understanding how research infrastructures shape the social sciences: impact, challenges and outlook. In A. Duşa, D. Nelle, G. Stock and G.G. Wagner (eds), *Facing the future: European research infrastructures for the humanities and social sciences* (pp. 21-34). Berlin: SCIVERO.
- Froeschl, K.A., Yamada, T. and Kudrna, R. (2002). Industrial statistics revisited: From footnotes to meta-information management. *Austrian Journal of Statistics*, 31(1), pp. 9-34. doi: <http://dx.doi.org/10.17713/ajs.v31i1.466>
- Green, A. (2008, February 5-6). Data Documentation Initiative. Paper presented at the DataShare project meeting, University of Edinburgh, Scotland. Retrieved from http://www.disc-uk.org/docs/DDI_Green.pdf
- Gregory, A. and Heus, P. (2007). DDI and SDMX: Complementary, not competing, standards. Retrieved from http://www.opendatafoundation.org/papers/DDI_and_SDMX.pdf
- Grossmann, W. (1999). Metadata. In S. Kotz (ed), *Encyclopedia of Statistical Sciences*, vol. 3 (pp. 811-815). New York, NY: Wiley.
- Harrison, E. and Rose, D. (2006). The European socio-economic classification (ESeC) user guide. Retrieved from <http://www.iser.essex.ac.uk/archives/esec/user-guide>
- Kaasa, A. (2009). Effects of different dimensions of social capital on innovative activity: Evidence from Europe at the regional level. *Technovation*, 29 (3), pp. 218-233. doi: 10.1016/j.technovation.2008.01.003.
- Kallas, J. and Linardis, A. (2010). A question documentation model on the needs of comparative research. *Archival Science*, 10 (1), pp. 65-83. doi: 10.1007/s10502-010-9115-0.
- Kleiner, B., Renschler, I., Wernli, B., Farago, P., and Joye, D. (2013). Introduction. In B. Kleiner, I. Renschler, B. Wernli, P. Farago and D. Joye (eds), *Understanding research infrastructures in the social sciences*. Zürich: Seismo.
- OECD (1999). Metadata for international comparisons and adherence to international standards. UN/ECE Work Session on Statistical Metadata, Working Paper No.3. Retrieved from <http://www.unece.org/fileadmin/DAM/stats/documents/1999/09/metis/3.e.pdf>
- Moorman, D. (2000). Social sciences for a digital world. Building infrastructure for the future. In OECD (ed.), *Social sciences for a digital world. Building infrastructure and databases for the future*. OECD.
- OECD (2007). OECD principles and guidelines for access to research data from public funding. Retrieved from <http://www.oecd.org/science/sci-tech/38500813.pdf>
- OECD (2014). PIAAC Technical standards and Guidelines. Retrieved from [http://www.oecd.org/skills/piaac/PIAACNPM\(2014_06\)PIAAC_Technical_Standards_and_Guidelines.pdf](http://www.oecd.org/skills/piaac/PIAACNPM(2014_06)PIAAC_Technical_Standards_and_Guidelines.pdf)
- Papageorgiou, H., Vardaki, M. and Pentaris, F. (2000). Data and metadata transformations. *Research in Official Statistics*, 3(2), pp. 27-43.
- Pippa, N. and James, D. (2007). A continent divide? Social capital in the US and Europe. In R. Jowell, C. Roberts, R. Fitzgerald and G. Eva (eds), *Measuring attitudes cross-nationally – Lessons from the European Social Survey* (pp. 239-264). London: Sage.

-
- Rose, D. and Harrison, E. (2007). The European socio-economic classification: A new social class schema for comparative European research. *European Societies*, 9 (3), pp. 459-490. doi: 10.1080/14616690701336518.
- Ryssevik, J. and Musgrave, S. (2001). The social science Dream Machine. Resource discovery, analysis and delivery on the Web. *Social Science Computer Review*, 19 (2), pp. 163-174. doi: 10.1177/089443930101900203.
- SDMX (2009). User guide. Retrieved from http://sdmx.org/?page_id=38
- Sundgren, B. (1996). Making statistical data more available. *International Statistical Review*, 64, pp. 23-38.
- Sundgren, B. (1999). *Information systems architecture for national and international statistical offices. Guidelines and recommendations*. Paper presented at the Conference of European Statisticians, UNECE, Statistical Standards and Studies, paper No. 51, Geneva. Retrieved from https://sites.google.com/site/bosundgren/my-life/Isa_e51.pdf?attredirects=0
- Sundgren, B. (2003). Developing and implementing statistical metadata systems. MetaNet WG3 Deliverable D6. Retrieved from <http://sites.google.com/site/bosundgren/my-life/WP3-no7-Metanet.doc>
- Thomas, W. (2010). Course on DDI 3: Putting DDI to work for you. Retrieved from <http://slideplayer.com/slide/5961884/>

