

ΕΠΕΚΤΕΙΝΟΝΤΑΣ ΤΟΝ ΚΥΚΛΟ ΖΩΗΣ ΠΟΣΟΤΙΚΩΝ ΕΡΕΥΝΩΝ ΚΑΙ ΔΕΔΟΜΕΝΩΝ

Ο ρόλος των αποθετηρίων δεδομένων

Απόστολος Λιναρδής
Αλέξης Ιωαννίδης

Περίληψη

Η παρούσα εργασία έχει ως στόχο να παρουσιάσει όλες εκείνες τις διαδικασίες που οφείλουν να ακολουθήσουν α) οι καταθέτες για την υποβολή ποσοτικών δεδομένων και ερευνών σε αποθετήρια δεδομένων, β) τα αποθετήρια για τη διαχείριση των δεδομένων που κατατέθηκαν στην υποδομή τους και γ) οι χρήστες που διεξάγουν δευτερογενή ανάλυση για την πρόσβαση και την εξερεύνηση αυτών των δεδομένων. Η ως άνω διαδικασία αποτελεί μεθοδολογική παρέμβαση στη μέχρι πρότινος ερευνητική διαδικασία που περιλάμβανε τέσσερα στάδια: τον σχεδιασμό, τη συλλογή, την ανάλυση των δεδομένων και τη δημοσιοποίηση των αποτελεσμάτων.

Η διαδικασία κατάθεσης των δεδομένων στα αποθετήρια δεν αποτελεί μια απλή διαδικασία. Ο καταθέτης οφείλει να υποβάλει τα δεδομένα στο αποθετήριο ακολουθώντας συγκεκριμένα πρωτόκολλα σε επίπεδο δεδομένων αλλά και σε επίπεδο μεταδεδομένων. Πλην των ποσοτικών ερευνών κατατίθενται στα αποθετήρια και άλλα ποσοτικά αντικείμενα, όπως: κύβοι (cubes), δείκτες και ταξινομήσεις (indices & classifications), μεταδεδομένα στατιστικών ερευνών καθώς και ρέπλικες αναλύσεων. Διεθνώς μάλιστα, παρατηρείται η τάση, στα επιστημονικά περιοδικά με ομότιμη κρίση, να υιοθετούν μια πολιτική δημοσίευσης άρθρων που εμπεριέχουν ανάλυση ποσοτικών δεδομένων υπό την προϋπόθεση ότι θα δημοσιοποιείται ταυτόχρονα η ρέπλικα ανάλυσής τους.

Το αποθετήριο από την άλλη οφείλει να διαθέτει μηχανισμούς για τον έλεγχο δεδομένων και μεταδεδομένων, για την τεκμηρίωση και αρχειοθέτηση αυτών,

καθώς και για τη δημοσιοποίησή τους στους χρήστες που διεξάγουν δευτερογενή ανάλυση. Παράλληλα, η τεκμηριωμένη περιγραφή της παραπάνω διαδικασίας τίθεται ως προαπαιτούμενο χρηματοδότησης εμπειρικών ερευνών από τους χρηματοδοτικούς μηχανισμούς. Χρηματοδότες φορείς έρευνας και επιστημονικά περιοδικά ενισχύουν περαιτέρω τον ρόλο των αποθετηρίων δεδομένων. Από την άλλη και τα αποθετήρια οφείλουν με τη σειρά τους να ενισχύσουν τους μηχανισμούς και τις διαδικασίες τους, ώστε να εξασφαλίζουν στους χρήστες την απρόσκοπτη και μακροπρόθεσμη πρόσβαση στα δεδομένα. Ο κύκλος ζωής των ερευνητικών δεδομένων –μέσω των αποθετηρίων– πλέον επεκτείνεται. Λιγότερο λόγω αλλαγής στην ερευνητική κουλτούρα του διαμοιρασμού των δεδομένων και περισσότερο λόγω των προϋποθέσεων που επιβάλλουν χρηματοδοτικοί μηχανισμοί έρευνας και επιστημονικά περιοδικά.

Λέξεις-κλειδιά: ποσοτικές έρευνες, δείκτες, ρέπλικα, κύβος, διαχείριση δεδομένων.

Abstract

This paper aims to present all the processes that have to be followed a) by depositors for submitting quantitative data and surveys to data repositories, b) by repositories for managing the data deposited in their infrastructure, and c) by users conducting secondary analysis for accessing and exploring these data. The above process is a methodological intervention in the hitherto research process which included four stages: design, data collection, data analysis and reporting of results.

Repository data deposition is not a simple process. The depositor has to submit the data to the repository following specific protocols both at the data level and at the metadata level. In addition to quantitative surveys, other quantitative objects such as: cubes, indices & classifications, statistical survey metadata and quantitative analysis replicates are deposited in the repositories. There is an international trend towards the adoption of a peer-review policy by scientific journals which consists of publishing articles containing quantitative data analysis, only if the data used in the analysis are clearly documented and available, and the empirical findings reproducible.

The repository, on the other hand, must have mechanisms for checking data and metadata, documenting and archiving them and making them available to users conducting secondary analysis. At the same time, a documented description of the above process is a prerequisite for the funding of empirical research by funding mechanisms. Research funders and scientific journals further enhance the role of data repositories. The repositories in turn must strengthen their mechanisms to ensure uninterrupted and long-term access to the data by secondary users. The life cycle of research data is now being extended, less due to a change in the research

culture of data sharing and more due to the conditions imposed by research funding mechanisms and scientific journals.

Keywords: *quantitative surveys, indices, replica, cube, data management.*

Εισαγωγή

Η παρούσα εργασία παρουσιάζει τους μηχανισμούς που διαθέτει η υποδομή SoDaNet για τη διαχείριση συνόλων δεδομένων και μεταδεδομένων (έργων δεδομένων), όπως ποσοτικές έρευνες, κύβους, δείκτες και ταξινομήσεις, μεταδεδομένα στατιστικών ερευνών και ρεπλικές αναλύσεων` καθώς και τον ρόλο των αποθετηρίων και της υποδομής στην επέκταση του κύκλου ζωής αυτών των δεδομένων.

Αρχικά προσδιορίζονται οι πέντε διαφορετικές μορφές συνόλων ποσοτικών δεδομένων και μεταδεδομένων και παρουσιάζεται ο τρόπος με τον οποίο πραγματοποιείται η διαχείρισή τους στην ερευνητική υποδομή του SoDaNet. Οι ποσοτικές έρευνες αφορούν κυρίως μικροδεδομένα που έχουν προκύψει με τη χρήση ερωτηματολογίου ή εντύπων καταγραφής. Οι κύβοι αφορούν αθροιστικά δεδομένα σε πολυδιάστατους πίνακες μεταβλητών, που προκύπτουν από επεξεργασία των μικροδεδομένων ερευνών και χρησιμοποιούνται ως πινακοποιημένα δεδομένα με μεταβλητές ως γραμμές, στήλες και μετρήσεις. Οι δείκτες και οι ταξινομήσεις αφορούν τον ποσοτικό προσδιορισμό κάποιας μεταβλητής-στόχου βάσει πηγαίων μεταβλητών που χρησιμοποιούνται σε έναν αλγόριθμο. Τα στατιστικά μεταδεδομένα αφορούν μεταδεδομένα για δεδομένα που συλλέγονται από φορείς του ελληνικού στατιστικού συστήματος και τεκμηριώνονται βάσει προτύπων που θέτει η Eurostat, ενώ η ρεπλικά ανάλυσης αφορά τον κώδικα και το υποσύνολο δεδομένων που απαιτούνται για την αναπαραγωγή της στατιστικής ανάλυσης η οποία παρουσιάζεται σε κάποια δημοσίευση. Διεθνώς μάλιστα, παρατηρείται η τάση, στα επιστημονικά περιοδικά με ομότιμη κρίση, να υιοθετούν μια πολιτική δημοσίευσης άρθρων που εμπεριέχουν ανάλυση ποσοτικών δεδομένων υπό την προϋπόθεση ότι θα δημοσιοποιείται ταυτόχρονα η ρεπλικά ανάλυσής τους.

Κατόπιν αναφερόμαστε στον ρόλο των αποθετηρίων δεδομένων και ειδικότερα στην Κοινοπραξία των Ευρωπαϊκών Αρχείων Κοινωνικών Δεδομένων, τη CESSDA (χ.η.). Έπειτα προσδιορίζονται οι μηχανισμοί που οφείλει να διαθέτει ένα αποθετήριο για τον έλεγχο δεδομένων και μεταδεδομένων, για την τεκμηρίωση και αρχειοθέτηση αυτών, για τη δημοσιοποίηση αυτών στους χρήστες που διεξάγουν δευτερογενή ανάλυση, καθώς και για τη μακροπρόθεσμη διαχείρισή τους. Η κατάσταση των δεδομένων και μεταδεδομένων στα αποθετήρια δεν αρκεί ώστε τα δεδομένα να δημοσιοποιηθούν μέσω των πληροφοριακών συστημάτων που διαθέτουν τα αποθετήρια. Τα δεδομένα και τα συνοδευτικά τεκμήρια θα πρέπει να πληρούν ποιοτικά κριτήρια και να είναι ανωνυμοποιημένα. Τα δεδομένα θα πρέπει να ελεγχθούν για κωδικούς και

τιμές εκτός του αποδεκτού εύρους, να υποστούν ελέγχους συνοχής, λογικούς ελέγχους, ελέγχους ως προς το ταίριασμα των εγγραφών και ελέγχους σε σχέση με την ανωνυμοποίηση των δεδομένων. Αν και ο καταθέτης θα πρέπει να έχει μεριμνήσει εκ των προτέρων για την κατάθεση δεδομένων στο αποθετήριο βάσει των ως άνω ποιοτικών κριτηρίων, εν τούτοις ο τελικός έλεγχος πρέπει να γίνεται από τα αποθετήρια. Επιπροσθέτως, τα δεδομένα θα πρέπει να τεκμηριώνονται βάσει προτύπων μεταδεδωμένων, όπως το DDI (Data Documentation Initiative) σε επίπεδο έργου δεδομένων, σε επίπεδο πόρου, καθώς και σε επίπεδο μεταβλητών (Λιναρδής, Αλεξανδρής & Κληρονόμος, 2022). Κατόπιν αναφερόμαστε στις υπόλοιπες «τεχνικές» διαδικασίες που εκτελούν τα αποθετήρια μέχρι και τη διάχυση των δεδομένων –όπως αποθήκευση, αντίγραφα ασφαλείας (backups) και ασφάλεια των δεδομένων, άδειες χρήσης και κατηγορίες πρόσβασης– και την απόδοση παραπομπών και μόνιμων ταυτοποιητών. Κατηγορίες πρόσβασης και άδειες χρήσης δεδομένων ορίζονται κατά την κατάθεση δεδομένων από τους καταθέτες, αλλά τα αποθετήρια δεδομένων δύνανται να έχουν πολιτικές που εν μέρει επιβάλλονται στους καταθέτες. Επίσης, τα αποθετήρια αποδίδουν μόνιμους ταυτοποιητές είτε σε επίπεδο έργου, είτε ακόμα και σε επίπεδο πόρου και δημιουργούν προτεινόμενες παραπομπές σε έργα ή/και σε πόρους.

Μετά τη δημοσιοποίηση των δεδομένων και των μεταδεδωμένων, οι χρήστες που διεξάγουν δευτερογενή ανάλυση, ανακαλύπτουν τα δεδομένα μέσω διαδικασιών που ομοιάζουν με τη βιβλιογραφική επισκόπηση και υπό προϋποθέσεις τα ανακτούν. Τα πληροφοριακά συστήματα των αποθετηρίων δίνουν τη δυνατότητα στους χρήστες που διενεργούν δευτερογενή έρευνα ή ανάλυση να προηγηθούν και να εξερευνήσουν τα μεταδεδωμένα των έργων/πόρων/μεταβλητών, να αναλύσουν online τα δεδομένα και υπό προϋποθέσεις να ανακτήσουν τοπικά τα δεδομένα ή τα αποτελέσματα της online ανάλυσης. Στην ενότητα που πραγματεύεται θέματα δευτερογενούς χρήσης παρατίθενται επίσης κάποιες βασικές πηγές απ' όπου μπορούν να ανακτηθούν κοινωνικο-οικονομικά δεδομένα. Η δευτερογενής χρήση και ανάλυση αποτελεί μια εύκολη επιλογή για τους ερευνητές, διότι δεν προϋποθέτει τη φάση συλλογής των δεδομένων. Από την άλλη πλευρά, οι χρήστες που διενεργούν δευτερογενή ανάλυση υποστηρίζουν ότι αφιερώνουν πολύ χρόνο στην κατανόηση των δεδομένων πριν από τη δευτερογενή ανάλυση. Πολλές φορές δε, η ανάλυση αυτή δεν είναι δυνατή, είτε γιατί ο τύπος των διαθέσιμων δεδομένων δεν το επιτρέπει, είτε διότι τα χωροχρονικά κριτήρια που επιθυμεί ο ερευνητής δεν ικανοποιούνται πλήρως, είτε διότι οι κατηγορίες των μεταβλητών δεν είναι προσανατολισμένες στα δευτερογενή ερευνητικά ερωτήματα του ερευνητή, ή τέλος διότι δεν έχουν καταγραφεί καλώς τα μεταδεδωμένα ώστε να γίνονται τα δεδομένα πλήρως κατανοητά σε τρίτους. Καθίσταται προφανές ότι τα αποθετήρια οφείλουν να δημοσιοποιούν δεδομένα που εξυπηρετούν την απρόσκοπτη δευτερογενή ανάλυση, τηρώντας παράλληλα όλους τους κώδικες δεοντολογίας για τη μη δημοσιοποίηση προσωπικών δεδομένων.

Τέλος αναφερόμαστε στον κύκλο ζωής των δεδομένων, στα Σχέδια Διαχείρισης Δεδομένων (ΣΔΔ - Data Management Plans) και στις πολιτικές των περιοδικών με ομότιμη κρίση. Το ΣΔΔ περιγράφει τον κύκλο ζωής και διαχείρισης των δεδομένων που θα συλλεγούν, θα δημιουργηθούν ή θα υποβληθούν σε επεξεργασία σε ένα έργο με απώτερο στόχο την επαναχρησιμοποίηση αυτών από τρίτους, μετά την ολοκλήρωση του έργου/έρευνας. Αποτελεί επίσημο έγγραφο και πλέον προαπαιτούμενο από πολλούς φορείς χρηματοδότησης της έρευνας σε ευρωπαϊκό και διεθνές επίπεδο για την υλοποίηση ερευνητικών έργων όλων των επιστημονικών κλάδων. Πιο συγκεκριμένα, ο ανάδοχος ενός έργου που περιλαμβάνει εμπειρική έρευνα ή χρήση δεδομένων καλείται να υποβάλει στην αναθέτουσα αρχή το σχέδιο διαχείρισης δεδομένων, στο οποίο περιγράφονται εξαρχής όλα τα στάδια της ερευνητικής διαδικασίας όπως θα υλοποιηθεί έτσι, ώστε με τη λήξη του έργου, τα δεδομένα να είναι προσβάσιμα από τρίτους μέσω αποθετηρίων δεδομένων.

Οι διάφορες μορφές συνόλων ποσοτικών δεδομένων και μεταδεδομένων

Στην υποδομή SoDaNet, όλες οι μορφές συνόλων ποσοτικών δεδομένων και μεταδεδομένων που θα παρουσιαστούν παρακάτω τεκμηριώνονται σε τουλάχιστον δύο επίπεδα: σε επίπεδο έργου δεδομένων και σε επίπεδο πόρου. Στο έργο δεδομένων πραγματοποιείται η γενική περιγραφή του αντικειμένου και για κάθε πόρο (αρχείο) τίθενται γενικά, ειδικά και τεχνικά χαρακτηριστικά αυτού. Εφόσον ο πόρος ενέχει μεταβλητές, τότε υφίσταται τεκμηρίωση και σε τρίτο επίπεδο, αυτό των μεταβλητών. Η τεκμηρίωση στο SoDaNet είναι συμβατή με το πρότυπο μεταδεδομένων DDI 2.5 και το λογισμικό που χρησιμοποιείται για τη διαχείριση των έργων δεδομένων είναι το Dataverse, που υλοποιήθηκε από το πανεπιστήμιο του Χάρβαρντ. Περισσότερα σε σχέση με την υλοποίηση του νέου αποθετηρίου του SoDaNet και την επέκταση των λειτουργιών του Dataverse βρείτε στο Λιναρδής κ.ά. (2022).

Οι ποσοτικές έρευνες

Οι ποσοτικές έρευνες προσδιορίζονται από την ταυτότητα και τη γενικότερη περιγραφή της έρευνας. Σε επίπεδο έργου δεδομένων εισάγεται πληροφορία για πεδία όπως: τίτλος, περίληψη, λέξεις-κλειδιά (Εικόνα 1), χρόνος και τόπος συλλογής των δεδομένων, εφαρμοζόμενη δειγματοληψία, μέθοδος συλλογής δεδομένων, δειγματοληπτικό σφάλμα, στάθμιση κ.ά.

Για κάθε πόρο/αρχείο προσδιορίζεται ο τίτλος του, η περιγραφή του, η γλώσσα συγγραφής του, ο τύπος του, η περίοδος εμπάργκο κ.ά. Για κάθε ποσοτική έρευνα ιδεατά πρέπει να υφίστανται τουλάχιστον τρεις πόροι (Εικόνα 2): το αρχείο δεδο-

Το Διαδίκτυο στην Ελλάδα 2019

Share it: [f](#) [t](#) [in](#) [sk](#)



Δεμερτζής Νίκος; Τσέκρης Χαράλαμπος, 2021, "Το Διαδίκτυο στην Ελλάδα 2019", <https://doi.org/10.17903/FK2/DZCJTC>, Κατάλογος Δεδομένων SoDaNet, 1.1, UNF:6-vvYBaUMeTS30mncvz7mlkIQ== [fileUNF]

Cite Data Project | [Learn about Data Citation Standards.](#) | [Contact Owner](#) | [Share](#)

Data Project Metrics 📊 | 41 Downloads 📄

Abstract 📄

Το Εθνικό Κέντρο Κοινωνικών Ερευνών (ΕΚΚΕ) παρακολουθεί συνεχώς τις κοινωνικές και κοινωνιοτεχνολογικές εξελίξεις με στόχο να δώσει ώθηση στη χάραξη πολιτικών βάσει τεκμηρίων και να ανταποκριθεί στην ανάγκη για έγκαιρη και αξιόπιστη πληροφόρηση σχετικά με τη χρήση του διαδικτύου στην Ελλάδα. Ως εκ τούτου, έχει αναλάβει από το 2015 την πρωτοβουλία να διεξάγει μεγάλη κλίμακας έρευνα με κύριο πεδίο μελέτης την καταγραφή της εισόδου και των πολλαπλών επιδράσεων των διαδικτυακών υποδομών στην ελληνική κοινωνία, φιλοδοξώντας σε μια επιστημονικά ελέγξιμη και φιλοσοφικά ανασταχασμένη θεώρηση των φαινομένων του ψηφιακού κόσμου μας. Συγκεκριμένα, η εμπειρική δειγματοληπτική έρευνα World Internet Project Greece υλοποιείται από το ΕΚΚΕ και εντάσσεται στη σύμπραξη World Internet Project, μια διεθνή σύμπραξη πανεπιστημίων και ερευνητικών κέντρων που αποτελείται από περισσότερες από 30 χώρες διαφορετικών ηπείρων. Πρόκειται για ένα διεθνές ερευνητικό πρόγραμμα που ξεκίνησε το 1999 και διευθύνεται από το Annenberg School Center for the Digital Future του Πανεπιστημίου της Νότιας Καλιφόρνιας (ΗΠΑ), υπό τον Καθηγητή Jeff Cole. Στην παρούσα έκθεση παρουσιάζονται αναλυτικά τα περιεχόμενα του τελευταίου γύρου του World Internet Project Greece (2019), που αφορούν στην ανάδειξη κρίσιμων ζητημάτων του «ελληνικού διαδικτύου», όπως το ψηφιακό χάσμα, οι χρήσεις και η πληροφοριακή/ψυχολογική αξία του διαδικτύου, το κοινωνικό κεφάλαιο και ο βαθμός αξιοπιστίας, ελευθερίας της έκφρασης, πολιτικής επάρκειας, θυματοποίησης και παραβίασης της ιδιωτικότητας. (2020-06-02)

Keyword 📄

ΧΡΗΣΗ ΔΙΑΔΙΚΤΥΟΥ, ΠΡΟΣΒΑΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΤΕΧΝΟΛΟΓΙΑ ΚΑΙ ΚΑΙΝΟΤΟΜΙΑ, ΕΓΓΡΑΜΜΑΤΟΣΥΝΗ ΣΤΗΝ ΤΕΧΝΟΛΟΓΙΑ ΤΩΝ ΥΠΟΛΟΓΙΣΤΩΝ, ΚΟΙΝΩΝΙΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ, ΔΙΑΔΙΚΤΥΟ, ΠΡΟΣΒΑΣΗ ΣΤΟ ΔΙΑΔΙΚΤΥΟ, ΗΛΕΚΤΡΟΝΙΚΕΣ ΑΓΟΡΕΣ, ΔΙΑΔΙΚΤΥΑΚΟΣ ΕΚΦΟΒΙΣΜΟΣ, ΗΛΕΚΤΡΟΝΙΚΟ ΤΑΧΥΔΡΟΜΕΙΟ




Εικόνα 1: Έργο δεδομένων ποσοτικής έρευνας στο αποθετήριο του ΕΚΚΕ στο SoDaNet

Metadata
Resources
Terms

🔍 Search

Filter by 📏 Sort

1 to 3 of 3 Resources

	<p>Ερωτηματολόγιο/codebook της έρευνας "Διαδίκτυο στην Ελλάδα 2019"</p> <p>Description: Είναι το ερωτηματολόγιο / codebook της έρευνας με τίτλο "Διαδίκτυο στην Ελλάδα 2019" σε μορφή doc.</p> <p>Resource Category: Tool: Questionnaire: Structured</p> <p>Filename: ques_codebook_WIP3_GR.doc</p> <p>Other Information: MS Word - 344.0 KB - Jan 10, 2021 - 15 Downloads</p>	Access 📄
	<p>Ερωτηματολόγιο/codebook της έρευνας "Διαδίκτυο στην Ελλάδα 2019"</p> <p>Description: Είναι το ερωτηματολόγιο / codebook της έρευνας με τίτλο "Διαδίκτυο στην Ελλάδα 2019" σε μορφή pdf.</p> <p>Resource Category: Tool: Questionnaire: Structured</p> <p>Filename: ques_codebook_WIP3_GR.pdf</p> <p>Other Information: Adobe PDF - 626.6 KB - Jan 10, 2021 - 5 Downloads</p>	Access 📄
	<p>Αρχείο δεδομένων της έρευνας με τίτλο: "Το Διαδίκτυο στην Ελλάδα 2019".</p> <p>Resource Category: Data: Microdata</p> <p>Filename: World_Internet_Project_2019-GR.tab</p> <p>Data Ingestion Details: 144 Number of Variables, 1208 Number of Cases - UNF:6-vvYBaUMeTS30mncvz7mlkIQ==</p> <p>Other Information: Tabular Data - 338.4 KB - Jan 10, 2021 - 21 Downloads</p>	🔍 Explore upon login Access 📄

Εικόνα 2: Οι πόροι ενός έργου δεδομένων ποσοτικής έρευνας

Search all fields: <input type="text" value="Enter keyword"/>							
	Label	Name	Type	Geographical	Role	Additivity	Measure
<input type="radio"/>	Μεταβλητή στάθμησης	weight	NUMERIC	-	Weight		Scale
<input type="radio"/>	Κωδικός ατόμου	id	NUMERIC	-			Scale
<input type="radio"/>	Αποκεντρωμένη Διοίκηση	dec_admin	NUMERIC	true			Nominal
<input type="radio"/>	Περιφέρεια	region	NUMERIC	Περιφέρειες Ελλάδας			Nominal
<input type="radio"/>	Πόσα άτομα ζουν σε αυτό το σπίτι και είναι 15 ετών και άνω;	q1	NUMERIC	-			Scale
<input type="radio"/>	Από αυτά τα άτομα (15 και άνω), πόσα είναι στο σπίτι αυτή τη στιγμή;	q2	NUMERIC	-			Scale
<input type="radio"/>	Είστε μαθητής ή φοιτητής;	q3	NUMERIC	-			Nominal

[View Categories / Values](#)

Εικόνα 3: Η τεκμηρίωση ενός πόρου/αρχείου μεταβλητών

μένων με τα μικροδοδεμένα της έρευνας, το ερωτηματολόγιο και το βιβλίο κωδικοποίησης. Προφανώς η έρευνα μπορεί να εμπλουτίζεται και με πρόσθετους πόρους, όπως είναι η αναφορά της έρευνας (report), οι οδηγίες προς τους συνεντευκτές, οι φόρμες επαφής, η φόρμα συναίνεσης, τα δελτία τύπου, οι επιστολές προς τους ερευνώμενους κ.ά.

Metadata	Variable Description	Variable Groups	Questions	Versions
<ul style="list-style-type: none"> • Q01) Πόσα άτομα ζουν σε αυτό το σπίτι και είναι 15 ετών και άνω;... • Q02) Από αυτά τα άτομα (15 και άνω), πόσα είναι στο σπίτι αυτή τη στιγμή;... • Q03) Είστε μαθητής ή φοιτητής;... • Q04) Εργάζεστε;... • Q05) Ποιο είναι ακριβώς το επάγγελμά σας;... • Q06) Είστε μερικής ή πλήρους απασχόλησης;... • Q07) Είστε;... • Q07_other) Άλλο ... προσδιορίστε... • Q08) Τον καιρό αυτό χρησιμοποιείτε το διαδίκτυο;... • Q09) Ποιος είναι ο κυριότερος λόγος για τον οποίο ΔΕΝ χρησιμοποιείτε το διαδίκτυο;... • Q09_other) Άλλο: π;... • Q10) Έχετε χρησιμοποιήσει το διαδίκτυο ποτέ στο παρελθόν;... • Q11_01) Φόβος/σύγχυση με την τεχνολογία... • Q11_02) Δεν είχα καλό υπολογιστή... • Q11_03) Δεν ξέρω να το χρησιμοποιώ... • Q11_04) Δεν με ενδιέφερε... <p><i>Prerequisite:</i> Ποιοι θα λέγατε ότι ήταν οι κυριότεροι λόγοι που σταματήσατε να χρησιμοποιείτε το διαδίκτυο;</p> <p><i>Literal:</i> Δεν με ενδιέφερε</p> <p><i>Instructions:</i> ΑΠΑΝΤΑΝΕ ΟΙ ΜΗ ΧΡΗΣΤΕΣ ΔΙΑΔΙΚΤΥΟΥ ΠΟΥ ΣΤΗΝ Q10 ΑΠΑΝΤΗΣΑΝ ΝΑΙ ΔΩΣΤΕ ΕΩΣ ΤΕΣΣΕΡΙΣ ΑΠΑΝΤΗΣΕΙΣ</p> <p><i>Variable:</i> Δεν με ενδιέφερε (Κυριότεροι λόγοι που σταματήσατε να χρησιμοποιείτε το διαδίκτυο) (q11_4)</p>				

Εικόνα 4: Η τεκμηρίωση των ερωτήσεων από τις οποίες προήλθαν οι μεταβλητές

Όταν κάποιος πόρος περιέχει μεταβλητές, όπως το αρχείο μικροδεδομένων, τότε στο αποθετήριο καταγράφονται μεταδεδομένα σε επίπεδο μεταβλητής, όπως όνομα, ετικέτα, τύπος, μέτρηση, τιμές/κατηγορίες μεταβλητής κ.ά. (Εικόνα 3).

Αν η έρευνα έχει προκύψει μέσω ερωτηματολογίου, τότε για το αρχείο μικροδεδομένων και ειδικότερα για τις μεταβλητές αυτού προσδιορίζονται και οι ερωτήσεις (Εικόνα 4).

Για κάθε ερώτηση εισάγεται μια αρίθμηση, η κύρια ερώτηση (literal question), τα κείμενα που ενδεχομένως υπάρχουν πριν ή/και μετά από την κύρια ερώτηση καθώς και οι οδηγίες προς τον συνεντευκτή ή την αυτοσυμπλήρωση. Επιπλέον καθορίζεται η μεταβλητή που προήλθε από την ερώτηση.

Το σχετικό έργο δεδομένων που ονομάζεται «Το διαδίκτυο στην Ελλάδα 2019» είναι διαθέσιμο στον σύνδεσμο: <https://doi.org/10.17903/FK2/DZCJTC>

Οι κύβοι

Ένας κύβος είναι ένας πολυδιάστατος πίνακας που περιέχει αθροιστικά δεδομένα και δομείται από ένα πλήθος μεταβλητών που εκφράζουν διαστάσεις (dimensions) και τουλάχιστον μία μεταβλητή που εκφράζει μέτρηση (measure) (Nesstar 2008, 2011). Οι μεταβλητές που εκφράζουν διαστάσεις (π.χ. φύλο με δύο επίπεδα άντρας-γυναίκα, ηλικιακές ομάδες) διαμορφώνουν τη δομή του κύβου, ενώ οι μεταβλητές που εκφράζουν μέτρηση αφορούν τα περιεχόμενα και τις αριθμητικές τιμές που υπάρχουν στα κελιά του κύβου. Ως μεταβλητές διάστασης μπορούν να χρησιμοποιηθούν μία ή περισσότερες γεωγραφικές μεταβλητές (Εικόνα 5).

Ένας κύβος είναι ένα αρχείο με μεταβλητές (πόρος) που περιλαμβάνει αθροιστικά δεδομένα και για το οποίο ορίζεται η αρχική του διάταξη (layout), όπου προσδιορίζονται οι μεταβλητές γραμμές, στήλες και μετρήσεις. Ο κύβος της Εικόνας 5 δημιουργήθηκε από διαδικασία «διασταύρωσης» των μεταβλητών ενός αρχείου δεδομένων με μεταβλητές όπως παρουσιάζονται στην Εικόνα 6 και με αρχική διάταξη μεταβλητών την ακόλουθη: γραμμές: GEO1, GEO2, GEO3, στήλες: PARTY, και μέτρηση: VOTES.

Στην ερευνητική υποδομή του SoDaNet και σε επίπεδο έργου δεδομένων ως κύβος νοείται μια συλλογή από κύβους. Ειδικότερα για το παραπάνω παράδειγμα, η συλλογή των κύβων –το έργο δεδομένων– αφορά εκλογικά δεδομένα του 1958 για όλη την Ελλάδα, ενώ οι πόροι αφορούν τα εκλογικά δεδομένα για την Κρήτη, την Αττική κ.λπ. Το σχετικό έργο δεδομένων για τα δεδομένα του 1958 είναι διαθέσιμο στον σύνδεσμο: <https://doi.org/10.17903/FK2/RZPQM3>.

Σε επίπεδο έργου δεδομένων, η συλλογή των κύβων τεκμηριώνεται με πεδία όμοια με αυτά της ποσοτικής έρευνας, όπως τίτλος, περίληψη, λέξεις-κλειδιά κ.ά., ενώ σε επίπεδο πόρου χρησιμοποιούνται πεδία όπως τίτλος, σημειώσεις, γλώσσα,

Name	Label
GEO_1	ΓΕΩΓΡΑΦΙΚΟ ΔΙΑΜΕΡΙΣΜΑ
GEO_2	ΝΟΜΟΣ
GEO_3	ΕΠΑΡΧΙΑ
GEO_4	ΔΗΜΟΙ - ΧΩΡΙΑ
GEO_5	ΕΚΛΟΓΙΚΟ ΤΜΗΜΑ
PARTY	ΚΟΜΜΑ
VOTES	ΠΛΗΘΟΣ ΨΗΦΩΝ

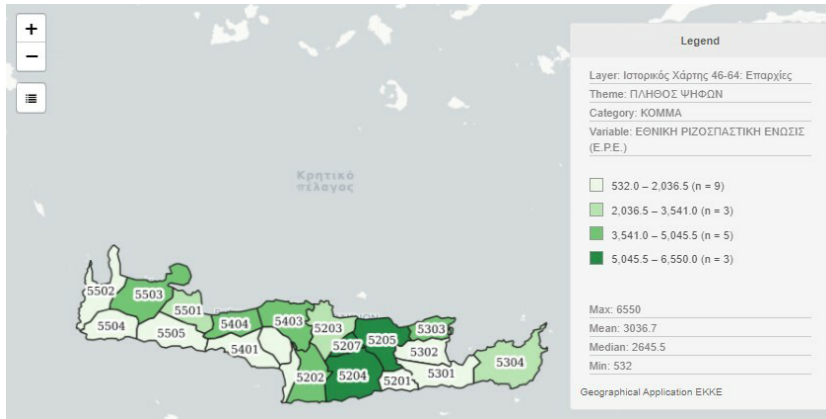
Table		VOTES	GEO_5	GEO_4
Sum	VOTES	PARTY		

			PARTY	ΑΝΕΞΑΡΤΗΤΟΙ	ΕΘΝΙΚΗ ΡΙΖΟΣΠΑΣΤΙΚΗ ΕΝΩΣΙΣ (Ε.Ρ.Ε.)	ΕΝΩΙΑ ΔΗΜΟΚΡΑΤΙΚΗ ΑΡΙΣΤΕΡΑ (Ε.Δ.Α)	ΕΝΩΣΙΣ ΛΑΙΚΩΝ ΚΟΜΜΑΤΩΝ (Ε.Λ.Κ)	ΚΟΜΜΑ ΦΙΛΕΛΕΥΘΕΡΩΝ	ΠΡΟΟΔΕΥΤΙΚΗ ΑΓΡΟΤΙΚΗ ΔΗΜΟΚΡΑΤΙΚΗ ΕΝΩΣΙΣ (Π.Α.Δ.Ε)	Totals
GEO_1	GEO_2	GEO_3								
ΚΡΗΤΗ	ΝΟΜΟΣ ΗΡΑΚΛΕΙΟΥ	ΒΙΑΝΝΟΥ	0.00	647.00	717.00	645.00	1.839.00	74.00	3,922.00	
		ΚΑΙΝΟΥΡΓΙΟΥ	0.00	3,629.00	1,187.00	5.00	3,517.00	2,258.00	10,596.00	
		ΜΑΛΕΒΥΣΙΟΥ	3.00	2,632.00	846.00	4.00	6,008.00	2,051.00	11,544.00	
		ΜΟΝΟΦΑΤΣΙΟΥ	0.00	5,511.00	1,326.00	9.00	6,374.00	1,204.00	14,424.00	
		ΠΕΔΙΑΔΟΣ	3.00	6,234.00	2,574.00	109.00	12,263.00	1,578.00	22,751.00	
		ΠΥΡΓΙΟΤΙΣΣΗΣ	0.00	940.00	382.00	0.00	1,349.00	1,939.00	4,610.00	
		ΤΕΜΕΝΟΥΣ	4.00	6,550.00	9,296.00	82.00	7,969.00	2,147.00	26,068.00	
	ΝΟΜΟΣ ΛΑΣΙΘΙΟΥ	ΙΕΡΑΠΕΤΡΑΣ		1,893.00	2,146.00	54.00	6,105.00	159.00	10,357.00	
		ΛΑΣΙΘΙΟΥ			1,721.00	309.00	6.00	2,397.00	107.00	4,540.00
		ΜΙΡΑΜΠΕΛΟΥ		4,999.00	1,378.00	190.00	6,562.00	45.00	13,174.00	
	ΝΟΜΟΣ ΡΕΘΥΜΝΗΣ	ΣΗΤΕΙΑΣ		2,659.00	1,327.00	66.00	9,286.00	343.00	13,681.00	
		ΑΓΙΟΥ ΒΑΣΙΛΕΙΟΥ			1,932.00	884.00		4,321.00		7,137.00
	ΝΟΜΟΣ ΡΕΘΥΜΝΗΣ	ΑΜΑΡΙΟΥ			1,972.00	207.00		3,416.00		5,595.00
		ΜΥΛΟΠΟΤΑΜΟΥ			3,046.00	1,481.00		7,287.00		12,714.00
		ΡΕΘΥΜΝΗΣ			3,931.00	2,366.00		7,967.00		14,264.00
	ΝΟΜΟΣ ΧΑΝΙΩΝ	ΑΠΟΚΟΡΩΝΟΥ			3,460.00	2,483.00		4,971.00	405.00	11,317.00
		ΚΙΣΣΑΜΟΥ			1,764.00	3,375.00		8,452.00	1,303.00	14,894.00
		ΚΥΔΟΝΙΑΣ			4,702.00	9,536.00		15,096.00	1,917.00	31,251.00
		ΣΕΙΝΙΟΥ			532.00	725.00		2,586.00	1,304.00	5,147.00
			ΣΦΑΚΙΩΝ		1,090.00	168.00		478.00	39.00	1,775.00
Totals			10.00	60,734.00	42,713.00	1,170.00	118,263.00	16,871.00	239,761.00	

Εικόνα 5: Εκλογικά δεδομένα 1958 για την Κρήτη, τους Νομούς και τις Επαρχίες της: Κύβος με μεταβλητές-γραμμή τα Γεωγραφικά Διαμερίσματα, τους Νομούς και τις Επαρχίες, μεταβλητή τη στήλη το κόμμα και μεταβλητή μέτρησης τους ψηφισαντες. Αποθετήριο EKKE / SoDaNet

Metadata	Variable Description	Variable Groups	Questions	Versions		
Search all fields: <input type="text" value="Enter keyword"/>						
Label	Name	Type	Geographical	Role	Additivity	Measure
<input type="radio"/> ΓΕΩΓΡΑΦΙΚΟ ΔΙΑΜΕΡΙΣΜΑ	GEO_1	NUMERIC	Ιστορικός Χάρτης 46-64: Γεωγραφικά Διαμερίσματα			Nominal
<input type="radio"/> ΝΟΜΟΣ	GEO_2	NUMERIC	Ιστορικός Χάρτης 46-64: Νομοί			Nominal
<input type="radio"/> ΕΠΑΡΧΙΑ	GEO_3	NUMERIC	Ιστορικός Χάρτης 46-64: Επαρχίες			Nominal
<input type="radio"/> ΔΗΜΟΙ - ΧΩΡΙΑ	GEO_4	CHARACTER	true			Scale
<input type="radio"/> ΕΚΛΟΓΙΚΟ ΤΜΗΜΑ	GEO_5	CHARACTER	true			Scale
<input type="radio"/> ΚΟΜΜΑ	PARTY	NUMERIC	-			Nominal
<input type="radio"/> ΠΛΗΘΟΣ ΨΗΦΩΝ	VOTES	NUMERIC	-			Scale

Εικόνα 6: Η δομή των μεταβλητών του κύβου που παρουσιάζεται στην Εικόνα 5



Εικόνα 7: Χάρτης με την κατανομή των ψήφων της ΕΡΕ στις επαρχίες της Κρήτης (1958)

τύπος, η τεκμηρίωση των μεταβλητών, όπως αναφέρθηκε πριν, αλλά και η αρχική διάταξη του κύβου (Cube Layout).

Επισημαίνεται ότι ως πόρος κύβου μπορεί να θεωρηθεί και ένα αρχείο της μορφής της Εικόνας 5, π.χ. σε μορφότυπο excel, αλλά, αν ο πόρος δεν έχει δομηθεί, όπως στην Εικόνα 6, δεν είναι δυνατή η online τροποποίηση της δομής του κύβου ή η απεικόνιση των δεδομένων του σε χάρτες.

Οι δείκτες και οι ταξινομήσεις

Η τεκμηρίωση δεικτών και ταξινομήσεων αφορά κυρίως τεκμηρίωση μεταδεδομένων. Προσδιορίζεται μια μεταβλητή-στόχος (target variable), όπως αυτή προκύπτει από την επεξεργασία μίας ή περισσότερων πηγαίων μεταβλητών (source variables). Για κάθε μεταβλητή-στόχο καταγράφονται μεταδεδομένα, όπως το όνομα της μεταβλητής, η περιγραφή της μεταβλητής, οι κατηγορίες της και ο τύπος της (αν πρόκειται για δείκτη ή για ταξινόμηση), ενώ για κάθε πηγαία μεταβλητή προσδιορίζονται το όνομα, η περιγραφή της, οι κατηγορίες της, καθώς και ενδεικτικές ερωτήσεις που μπορεί να τεθούν ώστε να προσδιοριστεί ορθά η πηγαία μεταβλητή (βλέπε Εικόνα 8).

Η ως άνω τεκμηρίωση αφορά μεταδεδομένα σε επίπεδο έργου δεδομένων. Σε επίπεδο πόρου απαιτείται τουλάχιστον ο αλγόριθμος με τον υπολογισμό της μεταβλητής-στόχου από τις πηγαίες μεταβλητές. Ο αλγόριθμος μπορεί να γράφεται σε μορφή ελεύθερου κειμένου, αλλά ιδεατά ζητείται σε SPSS syntax ή σε κώδικα άλλου στατιστικού πακέτου. Επιπλέον μπορεί να δημιουργηθεί κάποιο αρχείο δεδομένων –ως παράδειγμα– στο οποίο να εφαρμόζεται άμεσα ο αλγόριθμος.

Η διαφορά μιας ταξινόμησης από ένα δείκτη είναι ότι στην ταξινόμηση προσδιορίζεται διαφορετική τιμή ταξινόμησης ανά μονάδα (π.χ. άτομο), ενώ ο δεί-

Indices & Classifications ▲	
Target Variable Name ⓘ	Unempl_rate
Target Variable Description ⓘ	The rate of unemployment
Target Variable Type ⓘ	Index
Surveys applied to ⓘ	Labour Force Survey
Comments ⓘ	The unemployment rate ranges from 0% to 100%
Source Variable ⓘ	Name: ILOSTAT Description: ILO working status Comments: The values of ILOSTAT variable are: 1--> Employed, 2 --> Unemployed

Εικόνα 8: Μεταβλητή-στόχος και πηγαίες μεταβλητές για τον προσδιορισμό του ποσοστού ανεργίας στο αποθετήριο του Δημοκρίτειου Πανεπιστημίου Θράκης, στο SoDaNet

κτης αναφέρεται συνήθως σε μία συγκεντρωτική τιμή για μια ομάδα μονάδων. Ενδεικτικά αναφέρονται ως ταξινομήσεις: η Ευρωπαϊκή Κοινωνικο-Οικονομική Ταξινόμηση (ESeC), η διεθνής ταξινόμηση των επαγγελμάτων (ISCO), η διεθνής ταξινόμηση της Εκπαίδευσης (ISCED), η Στατιστική ταξινόμηση των οικονομικών δραστηριοτήτων (NACE) κ.ά. Πολλοί δείκτες είναι διαθέσιμοι στη βάση δεδομένων της Eurostat στην ακόλουθη διεύθυνση: <https://ec.europa.eu/eurostat/en/web/main/data/database>. Ενδεικτικά αναφέρονται: το ποσοστό απασχόλησης, το προσδόκιμο ζωής, ο δείκτης γονιμότητας, το κατά κεφαλήν ΑΕΠ κ.λπ. Παρακάτω παρουσιάζεται ο αλγόριθμος σε SPSS Syntax για τον υπολογισμό του δείκτη του ποσοστού ανεργίας, όπως τέθηκε στο αποθετήριο του Δημοκρίτειου Πανεπιστημίου Θράκης (ΔΠΘ) στο SoDaNet. Ο αλγόριθμος προϋποθέτει την ύπαρξη μίας μόνο πηγαίας μεταβλητής με την κατάσταση εργασίας του ατόμου (ILOSTAT) με κατηγορίες 1 ➔ Εργαζόμενος και 2 ➔ Άνεργος.

Μετά υπολογίζουμε τον αριθμό των εργαζομένων και των ανέργων ως νέες μεταβλητές:

```
AGGREGATE
/OUTFILE=* MODE=ADDVARIABLES
/BREAK=
/Numb_empl 'Number employed'=CIN(ILOSTAT 1 1).
AGGREGATE
/OUTFILE=* MODE=ADDVARIABLES
/BREAK=
/Numb_unempl 'Number unemployed'=CIN(ILOSTAT 2 2).
```

Τέλος, υπολογίζουμε μια νέα μεταβλητή που μόνη τιμή της είναι το ποσοστό ανεργίας:

```
COMPUTE Unempl_rate= (Numb_unempl / (Numb_empl+Numb_unempl)) * 100.
VARIABLE LABELS Unempl_rate 'Unemployment Rate'.
EXECUTE.
```

Το σχετικό έργο δεδομένων για τον υπολογισμό του δείκτη του ποσοστού ανεργίας είναι διαθέσιμο στον σύνδεσμο: <https://doi.org/10.17903/FK2/KKZFMF>

Οι ρέπλικες αναλύσεων

Με τις ρέπλικες ποσοτικών αναλύσεων ο συγγραφέας παρέχει στον χρήστη που διεξάγει δευτερογενή ανάλυση όλα τα απαραίτητα εργαλεία ώστε να αναπαραγάγει απρόσκοπτα ολόκληρη την ποσοτική στατιστική ανάλυση που έχει εκπονήσει, στο πλαίσιο της διαφάνειας και της εγκυρότητας των αποτελεσμάτων που δημοσιεύει. Η ρέπλικα αποτελεί απαραίτητο συνοδευτικό τεκμήριο μιας δημοσίευσης και πλην του απαραίτητου αλγόριθμου απαιτείται επίσης το τμήμα των δεδομένων στο οποίο εφαρμόζεται ο αλγόριθμος καθώς και ο σύνδεσμος στη σχετική δημοσίευση (βλέπε Εικόνα 9).

Ενδεικτικά αναφέρονται οι οδηγίες για την αναπαραγωγή της ποσοτικής ανάλυσης, όπως τίθενται στο περιοδικό *Journal of Politics* (χ.η.) του Πανεπιστημίου του Σικάγο:

«... Το *Journal of Politics* εφαρμόζει αυστηρή πολιτική αναπαραγωγής δεδομένων με ισχύ από την 1η Ιανουαρίου 2021. Η πολιτική του JOP είναι να δημοσιεύει άρθρα μόνο εάν τα δεδομένα που χρησιμοποιήθηκαν στην ανάλυση είναι σαφώς τεκμηριωμένα και διαθέσιμα και εάν τα εμπειρικά ευρήματα μπορούν να αναπαραχθούν. Οι συγγραφείς των αποδεκτών άρθρων που περιέχουν εμπειρικές αναλύσεις, καθώς και αναλύσεις που βασίζονται σε προσομοιώσεις, υποχρεούνται να παρέχουν σύνολα δεδομένων, κώδικες και άλλες σχετικές πληροφορίες που είναι απαραίτητες για τη διευκόλυνση της αναπαραγωγής. Όλα τα άρθρα γίνονται δεκτά υπό την προϋπόθεση της δυνατότητας αναπαραγωγής τους, η οποία θα αξιολογηθεί από τους αναλυτές αναπαραγωγής του JOP που θα αναλάβουν την εργασία σας. Εργασίες για τις οποίες δεν είναι δυνατή η αναπαραγωγή δεν θα δημοσιεύονται. Το παρόν έγγραφο παρέχει κατευθυντήριες γραμμές σχετικά με τον τρόπο προετοιμασίας του υλικού αναπαραγωγής δεδομένων...»

Στην εικόνα 9 βρείτε μια ρέπλικα ανάλυσης στο αποθετήριο του EKKΕ για το άρθρο: «Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση». Το άρθρο αναφέρεται ως «Related Publication» στην τεκμηρίωση του έργου δεδομένων, ενώ οι πόροι του έργου δεδομένων αφορούν την αναπαραγωγή της ανάλυσης σε SPSS Syntax, το αρχείο δεδομένων σε SPSS, στο οποίο εφαρμόζεται ο αλγόριθμος, καθώς και σχέδιο εργασίας του άρθρου σε αρχική μορφή (βλέπε Εικόνα 9).

Το σχετικό έργο δεδομένων για την ως άνω ρέπλικα είναι διαθέσιμο στον σύνδεσμο: <https://doi.org/10.17903/FK2/OX9UYM>

Για περισσότερα σχετικά με τη σύνδεση μεταξύ άρθρων περιοδικών και ερευνητικών δεδομένων δείτε στο Castro & Garnett, 2014.

Τεκμήρια για την αναπαραγωγή ανάλυσης του άρθρου : "Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση"

Share it:    

Λιναρδής, Απόστολος, Νησιώτης, Κωνσταντίνος - Συμείων; Κληρονόμος, Νικόλαος, 2022, "Τεκμήρια για την αναπαραγωγή ανάλυσης του άρθρου : "Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση"", <https://doi.org/10.17903/FK2/OX9UYM>, Κατάλογος Δεδομένων SoDaNet, Πρώτη έκδοση, UNF:6:YDZv4xzBmARYPcW6ofricg== [fileUNF]

Cite Data Project Learn about Data Citation Standards. Contact Owner Share

Data Project Metrics 1 | 1 Download

Abstract

Τεκμήρια για την αναπαραγωγή της ποσοτικής ανάλυσης για το άρθρο: "Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση" που δημοσιεύθηκε το 2021 στον συλλογικό τόμο «Αξίες, Πολιτισμικά Πρότυπα & Κοινωνικοί Προσανατολισμοί στη Σύγχρονη Ελλάδα» σε επιμέλεια του Σωκράτη Κοινούδου και από τις εκδόσεις "Gutenberg".

Η ρέπλικα περιλαμβάνει:

1. Το τρίμηνο των δεδομένων του 7ου γύρου της Παγκόσμιας Έρευνας Αξιών για την Ελλάδα που χρησιμοποιήθηκε, ώστε να απαντηθούν τα ερευνητικά ερωτήματα που τέθηκαν στο άρθρο (σε SPSS μορφή).
2. Τον αλγόριθμο για την αναπαραγωγή των αναλύσεων σε SPSS Syntax.
3. Το ίδιο το άρθρο με περιεχόμενο και μορφή μερικώς τροποποιημένη, σε σχέση με το άρθρο που δημοσιεύθηκε.

Related Publication

Λιναρδής Α., Νησιώτης, Κ.Σ., Κληρονόμος Ν. (2021). Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση, στο Κοινούδου Σ. (επιμ.), «Αξίες, Πολιτισμικά Πρότυπα & Κοινωνικοί Προσανατολισμοί στη Σύγχρονη Ελλάδα»

Metadata Resources Terms

Search this data project...

Filter by

Sort

1 to 3 of 3 Resources

	<p>Σχέδιο εργασίας (draft paper) για το άρθρο: "Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση"</p> <p>Description: Η παρούσα μελέτη χρησιμοποιεί τα δεδομένα του 7ου γύρου της Παγκόσμιας Έρευνας Αξιών για την Ελλάδα, ώστε να εξεταστεί η διαμόρφωση του φαινομένου του «παράδοξου της δημοκρατίας» για την ελληνική κοινωνία, τη συσχέτιση μεταξύ χειραφετητικών αξιών (emancipative values) και των εννοιών/εκφάνσεων της δημοκρατίας (notions of democracy), ενώ παράλληλα θα γίνει προσπάθεια συσχέτισης των παραπάνω, με την πολιτική αυτοτοποθέτηση στον άξονα Αριστεράς – Δεξιάς. Οδηγός στην προσπάθεια μας αυτή, είναι η αντίστοιχη δουλειά που έχει γίνει από τον Welzel στο 8ο και 10ο κεφάλαιο του Freedom Rising (2013) για τα συνολικά παγκόσμια δεδομένα του 5ου κύματος της Παγκόσμιας Έρευνας Αξιών.</p> <p>Resource Category: Research Outcomes: Other</p> <p>Filename: Paradox_of_Democracy_Linaradis_Nisiotis_Klironomos.pdf</p> <p>Other Information: Adobe PDF - 522.2 KB - Jan 27, 2022 - 1 Download</p>	<input type="button" value="Access"/>
	<p>Αλγόριθμος σε SPSS Syntax για την αναπαραγωγή της ανάλυσης στο άρθρο "Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση"</p> <p>Resource Category: Tool: Algorithm in SPSS Syntax</p> <p>Filename: SPSS_Syntax_Democracy's_Paradox_Replicate.sps</p> <p>Other Information: SPSS Syntax - 26.6 KB - Jan 27, 2022 - 0 Downloads</p>	<input type="button" value="Access"/>
	<p>Υποσύνολο δεδομένων του κύματος WV57_GR για τις ανάγκες αναπαραγωγής της ανάλυσης στο άρθρο "Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση"</p> <p>Description: Στο παρόν υποσύνολο δεδομένων εφαρμόζεται το SPSS Syntax για την αναπαραγωγή της στατιστικής ανάλυσης στο άρθρο "Το Παράδοξο της Δημοκρατίας στην Ελλάδα: Χειραφετητικές αξίες, εκφάνσεις της δημοκρατίας και πολιτική αυτοτοποθέτηση".</p> <p>Resource Category: Data: Microdata</p> <p>Filename: wv57_gr_replica_paradox.tab</p> <p>Data Ingestion Details: 33 Number of Variables, 1200 Number of Cases - UNF:6:YDZv4xzBmARYPcW6ofricg==</p> <p>Other Information: Tabular Data - 85.2 KB - Jan 27, 2022 - 0 Downloads</p>	<input type="button" value="Explore upon login"/> <input type="button" value="Access"/>






Εικόνα 9: Ρέπλικα ανάλυσης στο αποθετήριο δεδομένων του ΕΚΚΕ στο SoDaNet

Μεταδεδομένα στατιστικών ερευνών

Τέλος, μια νέα λειτουργία του SoDaNet αφορά την τεκμηρίωση ερευνών που προέρχονται από την ΕΛΣΤΑΤ ή από φορείς του Ελληνικού Στατιστικού Συστήματος. Οι ως άνω φορείς, όταν εκπονούν έρευνες, οφείλουν να καταθέτουν στην Eurostat μεταδεδομένα βάσει της Ενιαίας Ολοκληρωμένης Δομής Μεταδεδομένων SIMS (Single Integrated Metadata Structure) (European Commission, 2014).

Στην ερευνητική υποδομή SoDaNet αναπαράγονται οι τεκμηριώσεις κάποιων ερευνών της Ελληνικής Στατιστικής Αρχής και των Στατιστικών Φορέων βάσει των μεταδεδομένων SIMS. Στα μεταδεδομένα αναφέρεται ρητά η διαδικασία με την οποία επιτυγχάνεται η πρόσβαση στα στατιστικά δεδομένα.

Το σχετικό έργο δεδομένων που παρουσιάζεται στην Εικόνα 10 για την έρευνα της ΕΛΣΤΑΤ με τίτλο «Έρευνα Εκπαίδευσης Ενηλίκων» είναι διαθέσιμο στον σύνδεσμο: <https://doi.org/10.17903/FK2/SLR6KM>

Statistical Data 	
Contact 	<p>Organization: Ελληνική Στατιστική Αρχή Organization unit: Στατιστικών Πληθυσμού και Αγοράς Εργασίας Name: Ζαχαρίου, Σ. Mail address: Πειραιώς 46 και Επονητών, ΤΚ 18510 Πειραιάς. E-mail address: s.zachariou@statistics.gr Phone number: (+30) 2133152173 Fax number: (+30) 2133152948</p>
Introduction 	<p>Η έρευνα απευθύνθηκε σε άτομα ηλικίας 18-64 ετών και είχε ως στόχο να συγκεντρώσει πληροφορίες για τη συμμετοχή των ενηλίκων σε εκπαιδευτικά προγράμματα, τα χαρακτηριστικά των προγραμμάτων που παρακολούθησαν τους τελευταίους 12 μήνες, τους παράγοντες που επηρέασαν τη συμμετοχή τους σε εκπαιδευτικές διαδικασίες, το επίπεδο γνώσης ξένων γλωσσών, και βασικά δημογραφικά και κοινωνικοοικονομικά χαρακτηριστικά των συμμετεχόντων στην έρευνα, όπως τη κατάσταση απασχόλησης, την υπηκοότητα και τη χώρα γέννησης, το επίπεδο εκπαίδευσης και η χώρα γέννησης των γονέων τους καθώς και το φύλο, την ηλικία και την οικογενειακή κατάσταση.</p>
Metadata Update 	<p>Last certified: Φεβρουάριος 2018 Last posted: Φεβρουάριος 2018 Last updated: Φεβρουάριος 2018</p>
Statistical Presentation 	<p>Data description: Η έρευνα απευθύνθηκε σε άτομα ηλικίας 18-64 ετών και είχε ως στόχο να συγκεντρώσει πληροφορίες για τη συμμετοχή των ενηλίκων σε εκπαιδευτικά προγράμματα, τα χαρακτηριστικά των προγραμμάτων που παρακολούθησαν τους τελευταίους 12 μήνες, τους παράγοντες που επηρέασαν τη συμμετοχή τους σε εκπαιδευτικές διαδικασίες, το επίπεδο γνώσης ξένων γλωσσών, και βασικά δημογραφικά και κοινωνικοοικονομικά χαρακτηριστικά των συμμετεχόντων στην έρευνα, όπως τη κατάσταση απασχόλησης, την υπηκοότητα και τη χώρα γέννησης, το επίπεδο εκπαίδευσης και η χώρα γέννησης των γονέων τους καθώς και το φύλο, την ηλικία και την οικογενειακή κατάσταση.</p>

Εικόνα 10: Μεταδεδομένα της έρευνας της ΕΛΣΤΑΤ για την «Εκπαίδευση Ενηλίκων» στο αποθετήριο του Πανεπιστημίου Πελοποννήσου, στο SoDaNet

Η επίδραση των αποθετηρίων στην οργάνωση της ερευνητικής διαδικασίας

Παρακάτω θα αναφερθούμε σε εκείνες τις διαδικασίες που οφείλουν να ακολουθήσουν α) οι καταθέτες για την υποβολή ποσοτικών δεδομένων και ερευνών σε αποθετήρια δεδομένων, β) τα αποθετήρια για τη διαχείριση των δεδομένων που κατα-

τέθηκαν στην υποδομή τους και γ) οι χρήστες που διεξάγουν δευτερογενή ανάλυση για την πρόσβαση και την εξερεύνηση αυτών των δεδομένων. Όπως αναφέρθηκε και στην εισαγωγή, η ως άνω διαδικασία αποτελεί μεθοδολογική παρέμβαση στη μέχρι πρότινος ερευνητική διαδικασία που περιλάμβανε τέσσερα στάδια: τον σχεδιασμό, τη συλλογή και την ανάλυση των δεδομένων και τη δημοσιοποίηση των αποτελεσμάτων (Κάλλας, Λιναρδής 2004). Τα τρία νέα στάδια της ερευνητικής διαδικασίας απεικονίζονται στη Εικόνα 11 και θα επεξηγηθούν αναλυτικά στη συνέχεια.



Εικόνα 11: Τα τρία νέα στάδια της ερευνητικής διαδικασίας

Ο ρόλος των Αποθετηρίων Δεδομένων

Τα τελευταία χρόνια έχουν δημιουργηθεί οργανισμοί ή τμήματα οργανισμών, κυρίως στην Ευρώπη και στην Αμερική, που ασχολούνται με τη συσσώρευση, τεκμηρίωση και διάχυση των μικροδεδομένων. Αυτοί οι οργανισμοί ονομάζονται Αποθετήρια Δεδομένων (ΑΔ – Data Repositories). Ενίοτε για τα ΑΔ χρησιμοποιούνται εναλλακτικά οι όροι: βιβλιοθήκες δεδομένων (Data Libraries) ή αρχεία δεδομένων (Data Archives). Τα ΑΔ υποστηρίζουν την τεκμηρίωση και τη συσσώρευση εμπειρικών ερευνών και δεδομένων καθώς και τη δευτερογενή ανάλυση και χρήση μέσω της διάχυσης των δεδομένων και δρουν ως διαμεσολαβητές μεταξύ των παραγωγών και των αναλυτών. Διατηρούν συλλογές από έργα δεδομένων και από πόρους, συντονίζουν την απόκτηση πρόσθετων δεδομένων από πληθώρα πηγών και παρέχουν πρόσβαση σε δημόσια διαθέσιμα δεδομένα. Τα ΑΔ αποτελούν κόμβους εθνικών δικτύων και διεθνών οργανισμών, διασφαλίζοντας την επικοινωνία μεταξύ ερευνητών και προμηθευτών δεδομένων, έτσι ώστε αυτοί να είναι ενήμεροι περί των νέων συλλογών δεδομένων και των τεχνικών διαχείρισης της πληροφορίας. Χρησιμοποιούνται κυρίως από την ερευνητική και ακαδημαϊκή κοινότητα, καθώς και από φοιτητές και

ιδιώτες. Τα ΑΔ σπάνια εμπλέκονται στην παραγωγή των πρωτογενών δεδομένων και χρησιμεύουν ως διαμεσολαβητές μεταξύ των προμηθευτών δεδομένων και της ερευνητικής και ακαδημαϊκής κοινότητας. Τα δεδομένα προέρχονται είτε από τον δημόσιο τομέα (στατιστικές υπηρεσίες, ερευνητικά κέντρα, πανεπιστήμια κ.λπ.), είτε από τον ιδιωτικό τομέα (εταιρείες έρευνας αγοράς και δημοσκοπήσεων). Τα ΑΔ όχι μόνο συντηρούν τα δεδομένα για μελλοντική χρήση, αλλά δίνουν και μια προστιθέμενη αξία στις συλλογές δεδομένων, διότι (Ryssevick, & Musgrave, 2001):

- Τα δεδομένα που παραλαμβάνονται από τα ΑΔ υποβάλλονται σε ποικίλους ελέγχους και διαδικασίες καθαρισμού, ώστε να εξασφαλιστεί η ακεραιότητά τους.
- Αφαιρούνται οποιεσδήποτε εξαρτήσεις από συστήματα ή λογισμικό, ώστε τα δεδομένα να μπορούν να διαβάζονται οποιαδήποτε στιγμή στο μέλλον, ανεξάρτητα από τις εκάστοτε τεχνολογικές εξελίξεις.
- Τα δεδομένα καταλογογραφούνται και γίνονται διαθέσιμα μέσω πληροφοριακών συστημάτων αναζήτησης και ανάκτησης.
- Δημιουργούνται σειρές και συλλογές δεδομένων.

Τα ΑΔ οφείλουν να παρέχουν τεκμηρίωση και διαδικασίες αναζήτησης που να ικανοποιούν εν πολλοίς όλους τους δυνητικούς χρήστες, είτε πρόκειται για ερευνητές και ακαδημαϊκούς, είτε για ιδιώτες. Ένα γενικό χαρακτηριστικό των χρηστών ενός ΑΔ που διεξάγουν δευτερογενή ανάλυση είναι ότι οι περισσότεροι από αυτούς δεν έχουν ασχοληθεί με τη δημιουργία και την παραγωγή ενός συνόλου δεδομένων. Συνεπώς δεν κατανοούν εις βάθος όλες τις έννοιες που αφορούν την ερευνητική διαδικασία και γι' αυτό η τεκμηρίωση πρέπει να είναι απλή και κατανοητή, αλλά ταυτόχρονα να διατηρεί τον επιστημονικό λόγο και να μην εκλαϊκεύει. Σε σχέση με την περίπτωση ερευνητών και ακαδημαϊκών χρηστών –οι οποίοι συνήθως συγκρίνουν και συνδυάζουν δεδομένα από μια ευρεία σειρά πηγών σε χωροχρονικό επίπεδο και ασχολούνται επίσης με τους ορισμούς των εννοιών που χρησιμοποιούνται– το πληροφοριακό σύστημα του ΑΔ θα πρέπει να επιτρέπει τη χωροχρονική αναζήτηση βάσει λέξεων-κλειδιών, καθώς και τη σύνθεση των δεδομένων, ενώ στην περίπτωση ιδιωτών που ενδιαφέρονται άμεσα για κάποια νούμερα, π.χ. δημοσιογράφων, θα πρέπει να επιτρέπει την εύκολη δημιουργία πινάκων, διαγραμμάτων και απλών στατιστικών. Κοινός παρονομαστής των παραπάνω χαρακτηριστικών είναι η σχετική απόσταση μεταξύ των τελικών χρηστών ενός συνόλου δεδομένων και της διαδικασίας παραγωγής του. Ενώ οι δημιουργοί και οι αρχικοί χρήστες των συνόλων δεδομένων μπορούν να κατέχουν «ατεκμηρίωτη» και άτυπη γνώση, που τους καθοδηγεί κατά τη διαδικασία ανάλυσης, οι χρήστες που διεξάγουν δευτερογενή ανάλυση πρέπει να στηριχθούν στα μεταδεδομένα των δεδομένων.

Η τεκμηρίωση βασίζεται σε ένα πρόσθετο σώμα δεδομένων που περιγράφει τόσο τα δεδομένα όσο και όλες τις φάσεις της ερευνητικής διαδικασίας. Το πρό-

σθετο σώμα δεδομένων ονομάζεται «μεταδεδομένα». Τα μεταδεδομένα παρέχουν τον ουσιαστικό σύνδεσμο μεταξύ του πρωτογενούς υλικού και της δευτερογενούς χρήσης του. Ενώ δηλαδή τα δεδομένα είναι το τελικό προϊόν της εμπειρικής κοινωνικής έρευνας, για να μπορέσουν να αξιοποιηθούν από τρίτους, θα πρέπει αυτά να συνοδεύονται από ένα πρόσθετο σώμα δεδομένων που ονομάζουμε «μεταδεδομένα» (Λιναρδής, 2006). Συνεπώς τα κοινωνικά δεδομένα γίνονται προσβάσιμα μέσω των μεταδεδομένων τους. Τα μεταδεδομένα παρέχουν τη γέφυρα μεταξύ των παραγωγών των δεδομένων και των χρηστών τους και μεταβιβάζουν ουσιαστικές πληροφορίες στους χρήστες που διεξάγουν δευτερογενή ανάλυση. Ένα ΑΔ παρέχει καλής ποιότητας τεκμηρίωση όταν οι χρήστες αντιλαμβάνονται τις λεπτομέρειες που κρύβουν τα δεδομένα χωρίς παρέμβαση και πρόσθετη πληροφόρηση από τον αρχικό δημιουργό τους. Οι παραγωγοί δεδομένων –μέχρι πρότινος– δεν είχαν κάποιο ιδιαίτερο κίνητρο ώστε να τεκμηριώνουν αυστηρά και ποιοτικά ένα έργο δεδομένων, και τις περισσότερες φορές η τεκμηρίωση των δεδομένων που κατέθεταν στα ΑΔ ήταν ελλιπής. Στις κοινωνικές επιστήμες ήταν ευρέως γνωστό ότι η τεκμηρίωση που συνόδευε τα σύνολα δεδομένων ήταν συνήθως ανεπαρκής. Πολλές φορές η μοναδική τεκμηρίωση που συνόδευε ένα σύνολο δεδομένων δεν ήταν παρά μια ελλιπής περιγραφή της καρτέλας των μεταβλητών. Παραδοσιακά οι αρχειοθέτες δεδομένων ήταν αυτοί που προσπαθούσαν να λύσουν τα προβλήματα επικοινωνίας και ανταλλαγής δεδομένων ανάμεσα στους παραγωγούς των πρωτογενών δεδομένων και την κοινότητα των χρηστών. Ενώ οι παραγωγοί των πρωτογενών δεδομένων έχουν άμεση γνώση των δεδομένων, οι χρήστες-αναλυτές πρέπει να βασίζονται στην παρεχόμενη τεκμηρίωση. Πλέον –όπως θα διαπιστωθεί παρακάτω– συντρέχουν αρκετοί λόγοι ώστε τα ΑΔ να γίνονται *αρκετά πειστικά* στους παραγωγούς ως προς την παρεχόμενη τεκμηρίωση και ποιότητα δεδομένων και μεταδεδομένων.

Τον Ιούνιο του 2015 η Ελλάδα –με την ερευνητική υποδομή SoDaNet– γίνεται μέλος της Ευρωπαϊκής Κοινοπραξίας Αρχείων Δεδομένων των Κοινωνικών Επιστημών CESSDA και αποτελεί τον ελληνικό κόμβο της ευρωπαϊκής υποδομής. Η κοινοπραξία CESSDA αποτελείται σήμερα από 22 χώρες μέλη και μία χώρα παρατηρητή (observer). Η CESSDA προωθεί την απόκτηση, αρχειοθέτηση και διάθεση ψηφιακών δεδομένων των κοινωνικών επιστημών για ερευνητικούς και διδακτικούς σκοπούς σε όλη την Ευρώπη. Μεγάλου βεληνεκούς αποθετήρια που συμμετέχουν σε αυτήν είναι: το γερμανικό ΑΔ GESIS (χ.η.), το νορβηγικό ΑΔ NSD (χ.η.) και του Ηνωμένου Βασιλείου UK Data Service (χ.η.). Ο ελληνικός κόμβος είναι διαθέσιμος στη διεύθυνση: <https://sodanet.gr/>.

Η κατάθεση των δεδομένων και των μεταδεδομένων στα ΑΔ

Τα ΑΔ λοιπόν διαθέτουν διαδικασίες τις οποίες γνωστοποιούν στους υποψήφιους καταθέτες δεδομένων μέσω των ιστοτόπων τους. Πιο συγκεκριμένα στον ιστότοπο

της ερευνητικής υποδομής SoDaNet αναφέρονται τα παρακάτω βήματα για την κατάθεση δεδομένων στα αποθετήρια της υποδομής (<https://sodanet.gr/data-services/data-deposition-process>):

- Εκδήλωση ενδιαφέροντος για κατάθεση των δεδομένων μέσω σχετικής φόρμας (<https://sodanet.gr/data/deposit-data>).
- Επικοινωνία του αποθετηρίου με τον καταθέτη για διευκρινίσεις.
- Υποβολή φόρμας μεταδεδομένων (https://sodanet.gr/storage/submission-forms/deposit_data.xlsx) και των σχετικών δεδομένων από τον καταθέτη στο αποθετήριο. Στη σχετική φόρμα ο καταθέτης προσδιορίζει επίσης το επίπεδο πρόσβασης ανά πόρο/αρχείο που καταθέτει. Τα 4 επίπεδα πρόσβασης ανά πόρο, που υποστηρίζονται από την υποδομή, είναι τα εξής:
 - Πρόσβαση χωρίς κάποιο περιορισμό (Unrestricted).
 - Πρόσβαση μετά από σύνδεση (Access upon login).
 - Περιορισμένη πρόσβαση – Διαθέσιμος πόρος κατόπιν αιτήματος (Restricted – Available upon Request).
 - Καθολικά περιορισμένη πρόσβαση (Restricted).

Σε όλες τις παραπάνω κατηγορίες εφαρμόζεται περίοδος εμπάργκο.

- Επιπλέον καθορίζονται οι άδειες χρήσης των δεδομένων βάσει των έξι διαφορετικών αδειών Creative Commons (χ.η.).
- Τέλος, όταν το ΑΔ δημοσιοποιεί το έργο δεδομένων, γνωστοποιεί στον καταθέτη το DOI και την προτεινόμενη παραπομπή (citation).

Επισημαίνεται ότι η φόρμα μεταδεδομένων είναι συμβατή με το πρότυπο τεκμηρίωσης Data Documentation Initiative 2.5 (DDI) και επιπλέον χρησιμοποιεί ελεγχόμενα λεξιλόγια (controlled vocabularies) βάσει των προτεινόμενων ελεγχόμενων λεξιλογίων από τη CESSDA (<https://vocabularies.cessda.eu/>).

Η διαχείριση των δεδομένων από τα ΑΔ

Αυτή αφορά τις κάτωθι ενέργειες:

- Διενέργεια λογικών και δομικών ελέγχων και ελέγχων ανωνυμοποίησης των τεκμηρίων και ειδικότερα των δεδομένων
- Τεκμηρίωση του έργου δεδομένων και των πόρων βάσει προτύπων μεταδεδομένων, όπως το DDI
- Τεκμηρίωση των αρχείων δεδομένων σε επίπεδο μεταβλητών και ερωτήσεων βάσει του DDI
- Αποθήκευση του έργου δεδομένων και πρόνοια για αντίγραφα ασφαλείας των τεκμηρίων

- Καθορισμό επιπέδων πρόσβασης και αδειών χρήσης ανά πόρο
- Ενεργοποίηση μηχανισμών απόδοσης μόνιμων ταυτοποιητών, όπως τα DOI (Digital Object Identifier).
- Δημοσιοποίηση των έργων δεδομένων στο πληροφοριακό σύστημα της υποδομής.
- Ενεργοποίηση του πρωτοκόλλου OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting – <https://www.openarchives.org/pmh/>), ώστε να είναι εφικτή η έκθεση των μεταδεδομένων σε άλλες πύλες (portal).

Οι έλεγχοι των τεκμηρίων

Η διενέργεια δομικών και λογικών ελέγχων των δεδομένων αφορά κυρίως τα κάτωθι:

- Ελέγχους κωδικών και τιμών εκτός αποδεκτού εύρους. Για παράδειγμα, ενώ η μεταβλητή ILOSTATUS δέχεται τιμές 1 à Εργαζόμενος και 2 à Άνεργος, η κατανομή συχνοτήτων μέσω της διαδικασίας «Frequencies» ενδέχεται να αναδείξει τιμές εκτός των δύο αυτών κωδικών, που πρέπει εκ των υστέρων να διορθωθούν. Επίσης η κατανομή συχνοτήτων μπορεί να αναδείξει και λάθη «εκτός του αποδεκτού εύρους». Για παράδειγμα, νοικοκυριό με 99 παιδιά.
- Ελέγχους συνοχής και ειδικότερα σε μεταβλητές που προέρχονται από ερωτήσεις-φίλτρο. Για παράδειγμα, ένας ερευνώμενος δηλώνει ότι είναι άνεργος, ενώ ταυτόχρονα υπάρχει πληροφορία για το τρέχον επάγγελμά του και την οικονομική δραστηριότητα της επιχείρησης στην οποία εργάζεται. Ο έλεγχος αυτός πραγματοποιείται με τη διασταύρωση της μεταβλητής που προέρχεται από την ερώτηση-φίλτρο με τις υπόλοιπες μεταβλητές που επηρεάζονται από αυτή.
- Λογικούς ελέγχους που προκύπτουν από τη διασταύρωση κάποιων μεταβλητών. Για παράδειγμα, η διασταύρωση του ανώτατου επιπέδου εκπαίδευσης με την ηλικία μπορεί να αναδείξει άτομα μικρής ηλικίας με διδακτορικό!

Προτού τα δεδομένα διαμοιραστούν σε τρίτους, πρέπει να υποστούν ελέγχους ανωνυμοποίησης έτσι ώστε να μην μπορούν να εντοπιστούν τα άτομα, οι οργανισμοί ή οι επιχειρήσεις που έλαβαν μέρος στην έρευνα. Ταυτόχρονα όμως πρέπει να διατηρηθούν τα δεδομένα με όσο το δυνατόν περισσότερη ουσιαστική πληροφορία. Η ανωνυμοποίηση απαιτείται για λόγους δεοντολογίας, 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 ούτε από διασταυρώσεις των μεταβλητών της έρευνας, εκτός αν ένας ερευνώμενος έχει δώσει γραπτώς ρητή συγκατάθεση. Η ταυτότητα ενός ατόμου μπορεί να αποκαλυφθεί από (UK Data Service – Anonymization, χ.η):

- άμεσα αναγνωριστικά (direct identifiers) όπως: ονόματα, διευθύνσεις, αριθμοί τηλεφώνου, εικόνες, αριθμοί ταυτότητας, που επιτρέπουν την άμεση αναγνώριση του ερωτώμενου, ενώ 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 και ελέγχου της έρευνας.
- έμμεσα αναγνωριστικά (indirect identifiers), δηλαδή 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 με ιδιαίτερα χαρακτηριστικά σε μία συγκεκριμένη περιοχή. Τα έμμεσα αναγνωριστικά είναι όμως απαραίτητα για ερευνητικούς σκοπούς.

Οι τεχνικές ανωνυμοποίησης αφορούν τη διαγραφή μεταβλητών, την ομαδοποίηση τιμών και κατηγοριών των μεταβλητών ή τη μείωση της ακρίβειας και της λεπτομέρειας της κειμενικής πληροφορίας μιας μεταβλητής. Πιο συγκεκριμένα, οι τεχνικές αφορούν τα κάτωθι (UK Data Service – Anonymization quantitative, χ.η):

- Αφαίρεση άμεσων αναγνωριστικών (direct identifiers). Για παράδειγμα, οι μεταβλητές στις οποίες αποθηκεύονται τα ονόματα των ερευνώμενων πρέπει να διαγραφούν ή να αντικατασταθούν από κάποιον κωδικό. Στοιχεία επικοινωνίας όπως διευθύνσεις, τηλέφωνα, email κ.ά. πρέπει να διαγράφονται.
- Ομαδοποίηση κατηγοριών ή μείωση της ακρίβειας μιας μεταβλητής, όπως η ηλικία, η περιοχή κατοικίας ή το επάγγελμα. Κατά γενικό κανόνα, αναφέρεται το χαμηλότερο επίπεδο γεωαναφοράς που δεν αποκαλύπτει στοιχεία, στάσεις και αντιλήψεις κάποιου ερευνώμενου. Το ακριβές επίπεδο εξαρτάται από το είδος των δεδομένων που συλλέγονται, αλλά λεπτομερείς γεωαναφορές όπως ταχυδρομικοί κώδικες ή ονόματα μικρών πόλεων ή χωριών είναι συνήθως προβληματικές. Στην περίπτωση που εντοπιστεί πρόβλημα αποκάλυψης ταυτοτήτων, η λύση έγκειται στην επανακωδικοποίηση των μεταβλητών σε ευρύτερες κατηγορίες: για παράδειγμα, η καταγραφή της γέννησης ενός ατόμου να γίνεται με το έτος και όχι με ακρίβεια ημέρας, η καταγραφή του ταχυδρομικού κώδικα να γίνεται μόνο με τα 2-3 πρώτα ψηφία του κώδικα και όχι τα πλήρη, η ταξινόμηση των επαγγελματιών ISCO να γίνεται με μονοψήφιο ή διψήφιο κωδικό επαγγελματιών και όχι με τριψήφιο ή τετραψήφιο κωδικό.
- Γενίκευση του νοήματος κειμενικής πληροφορίας, στην περίπτωση που ο ερωτώμενος αναφέρεται σε λεπτομέρειες που μπορεί να αποκαλύψουν την ταυτότητά του.

- Περιορισμό του εύρους μιας συνεχούς μεταβλητής ώστε να μην αποκλύπτονται οι ακραίες τιμές – είτε πολύ μεγάλες είτε πολύ μικρές. Οι ασυνήθιστα μεγάλες ή μικρές τιμές μπορεί να επανακωδικοποιηθούν σε νέες κατηγορίες. Για παράδειγμα, το ετήσιο εισόδημα θα μπορούσε να είναι κωδικοποιημένο για τα υψηλόμισθα άτομα ως εξής: «50.000 ευρώ και άνω».

Η τεκμηρίωση

Όπως αναφέρθηκε προηγουμένως, η τεκμηρίωση αφορά τρία επίπεδα: το έργο δεδομένων, τον πόρο και τις μεταβλητές – σε περίπτωση που ο πόρος ενέχει μεταβλητές. Το έργο δεδομένων αντιστοιχεί στον τομέα «Study Description» του DDI, οι πόροι στον τομέα του DDI που αναφέρεται ως «Other Study Materials», ο πόρος που περιλαμβάνει μεταβλητές στον τομέα «File Description», ενώ οι μεταβλητές αντιστοιχούν στο «Data Description» του DDI (Λιναρδής, 2006). Η τεκμηρίωση σε επίπεδο έργου δεδομένων περιλαμβάνει την παρακάτω πληροφορία (UK Data Service – Study, χ.η):

- Τον τίτλο και το περίγραμμα σχεδιασμού και συλλογής δεδομένων: περίληψη, στόχοι, ερευνητικές υποθέσεις, εμπλεκόμενοι φορείς, ερευνητές και χρηματοδότες
- Τις μεθόδους συλλογής δεδομένων: πρωτόκολλα συλλογής δεδομένων, σχεδιασμός δειγματοληψίας, ροές εργασιών, όργανα μέτρησης που χρησιμοποιήθηκαν, λογισμικό και υλικό που χρησιμοποιήθηκε, γεωγραφική και χρονική κάλυψη
- Τις διαδικασίες επικύρωσης, ελέγχου, καθαρισμού δεδομένων στο πλαίσιο των ποιοτικών ελέγχων που διεξήχθησαν
- Την καταγραφή των αλλαγών που διενεργήθηκαν στα δεδομένα από τη στιγμή που δημοσιοποιήθηκαν, παρέχοντας ταυτοποίηση και αναγνώριση των διαφορετικών εκδόσεων των συνόλων δεδομένων
- Την καταγραφή αλλαγών σε χρονοσειρές ή διαχρονικές έρευνες, που επήλθαν στη μεθοδολογία, στο περιεχόμενο των μεταβλητών, στο κείμενο της ερώτησης, στις ετικέτες των μεταβλητών, στις μετρήσεις και στη δειγματοληψία από κύμα σε κύμα. Στην περίπτωση ερευνών panel, καταγράφεται ο τρόπος με τον οποίο έγινε η διαχείριση των ατόμων στον χρόνο και μεταξύ των κυμάτων.
- Την καταγραφή γενικής πληροφορίας σε σχέση με την πρόσβαση και τη χρήση των δεδομένων

Η τεκμηρίωση σε επίπεδο πόρου αφορά πόρους όπως ερωτηματολόγια, βιβλίο κωδικοποίησης, δημοσιεύσεις, παρουσιάσεις, άλλα ερευνητικά αποτελέσματα, δευ-

τερογενείς πηγές δεδομένων που χρησιμοποιήθηκαν στο έργο δεδομένων κ.ά. Για κάθε πόρο προσδιορίζονται τα παρακάτω πεδία:

- Ο τίτλος του πόρου
- Το όνομα του αρχείου και τεχνικά χαρακτηριστικά
- Η γλώσσα συγγραφής
- Το DOI / URL
- Ο τύπος πόρου (επιλέγεται από ελεγχόμενο λεξιλόγιο)
- Οι σημειώσεις
- Η πηγή δεδομένων
- Το επίπεδο πρόσβασης και η άδεια χρήσης των δεδομένων
- Η περίοδος εμπάργκο

Εφόσον ο πόρος περιέχει μεταβλητές, καταγράφεται επιπλέον σε επίπεδο πόρου το πλήθος των μεταβλητών και των περιπτώσεων και σε επίπεδο μεταβλητών η κάτωθι πληροφορία:

- Το όνομα, η ετικέτα και η περιγραφή της μεταβλητής. Οι ετικέτες των μεταβλητών θα πρέπει να είναι σύντομες και, όπου αυτό είναι εφικτό, θα πρέπει να καταγράφεται στην ετικέτα ή στο όνομα η αναφορά στον αριθμό της ερώτησης του ερωτηματολογίου ώστε να είναι άμεση η μετάβαση από την ερώτηση στη μεταβλητή (κατ' αυτό τον τρόπο δίνεται όλη η πληροφορία που ενέχει ένα βιβλίο κωδικοποίησης και δεν απαιτείται ξεχωριστό αρχείο για τη μετάβαση από το ερωτηματολόγιο στο σύνολο δεδομένων).
- Ο καθορισμός του πληθυσμού που αποκλίνει από τον γενικό πληθυσμό της έρευνας, π.χ. σε ερωτήσεις που καλούνται να απαντήσουν μόνο οι εργαζόμενοι ή οι άνεργοι
- Οι κωδικοί και οι ετικέτες κατηγοριών
- Οι κωδικοί και οι ετικέτες κατηγοριών για ελλείπουσες τιμές (missing values), όπως «Δεν γνωρίζω», «Δεν απαντώ», «Μη εφαρμόσιμο» κ.ά.
- Η κωδικοποίηση βάσει ταξινομητικών σχημάτων. Για παράδειγμα, κωδικοί για την ταξινόμηση των επαγγελματιών ISCO 2008, της οικονομικής δραστηριότητας NACE rev. 2 κ.ά. Παρατηρήστε ότι κάποιες από τις ταξινομήσεις αλλάζουν και φέρουν διαφορετικές εκδόσεις στον χρόνο.
- Για μεταβλητές που παρήχθησαν μέσω ενός σύνθετου αλγόριθμου, να συμπεριλαμβάνεται ο ίδιος ο αλγόριθμος ως πρόσθετη τεκμηρίωση.
- Για μεταβλητές στάθμισης, να καταγράφεται η διαδικασία με την οποία δημιουργήθηκαν καθώς και οδηγίες χρήσης τους.
- Οι ομαδοποιήσεις μεταβλητών βάσει των τομέων του ερωτηματολογίου ή άλλων θεματικών

- Οι ερωτήσεις από τις οποίες προήλθαν οι μεταβλητές και ο σύνδεσμος ερώτησης και μεταβλητής

Η δευτερογενής χρήση και ανάλυση

Η δευτερογενής ανάλυση είναι μια μέθοδος αξιοποίησης πρωτογενών δεδομένων η οποία εντάσσεται σε ερευνητικούς στόχους πέραν των πρωταρχικών που τέθηκαν από τους αρχικούς ερευνητές (Hakim, 1982· Vartanian, 2010). Η δευτερογενής ανάλυση δεν είναι απλώς μια επανεπεξεργασία των δεδομένων που έχουν ήδη αναλυθεί για έναν πρωτεύοντα σκοπό, αλλά έχει αξιοσημείωτη δυναμική πρωτοτυπίας και ελέγχου θεωρητικών όψεων. Η Hakim υποστηρίζει επίσης ότι η δευτερογενής ανάλυση έχει αφεαυτής τη δυναμική «να σπάσει τα μονοπάλια» στην κοινωνική έρευνα, αφού οι κρατικοί οργανισμοί δεν είναι πλέον σε θέση να διατηρούν για τους ίδιους μόνο δεδομένα που έχουν συγκεντρωθεί με δημόσια δαπάνη. Διαθέτει σημαντική δυναμική αφού νέα ερευνητικά ερωτήματα μπορεί να απαντώνται από τα υφιστάμενα δεδομένα. Δεν είναι τυχαίο ότι επιστήμονες της NASA ανακάλυψαν νέους πλανήτες από επεξεργασία παλαιότερων δεδομένων. Αποτελεί σχετικά ανέξοδη μέθοδο και εξασφαλίζει εξοικονόμηση πόρων. Η δευτερογενής ανάλυση σειρών δεδομένων επιτρέπει τη διαχρονική διεξαγωγή συγκρίσεων.

Τα προφανή οφέλη της δευτερογενούς χρήσης και ανάλυσης είναι ότι δεν προϋποθέτει τη φάση της συλλογής δεδομένων, με αποτέλεσμα την οικονομία σε κόστος και χρόνο. Επίσης, τα αποτελέσματα που προκύπτουν μπορεί να χρησιμοποιηθούν συγκριτικά με αυτά της πρωτογενούς ανάλυσης και ενδεχομένως να εφαρμόζουν πιο προηγμένες στατιστικές τεχνικές από αυτές της πρωτογενούς ανάλυσης.

Από την άλλη πλευρά υπάρχουν τρεις τύποι προβλημάτων/περιορισμών που σχετίζονται με τη δευτερογενή ανάλυση: τα τεχνικά, τα θεσμικά και τα επιστημολογικά.

Τα τεχνικά προβλήματα αφορούν τον βαθμό εξοικείωσης του χρήστη με τα δεδομένα, την ύπαρξη σχετικής τεκμηρίωσης και ειδικότερα την ύπαρξη σχετικής τεκμηρίωσης σε γλώσσα που γνωρίζει ο δευτερογενής χρήστης. Οι χρήστες που διεξάγουν δευτερογενή ανάλυση υποστηρίζουν ότι αφιερώνουν πολύ χρόνο στην κατανόηση της δομής και των εννοιολογικών ορισμών των δεδομένων πριν από τη δευτερογενή ανάλυση. Επίσης, ότι απαιτείται πολύς χρόνος για την κατανόηση και εξέταση του θεωρητικού κορμού της έρευνας. Κάποιοι υποστηρίζουν ότι έργα που χρησιμοποιούν δευτερογενή ανάλυση υποεκτιμούν το θεωρητικό πλαίσιο, αφού παρακάμπτεται η φάση σχεδίασης του ερωτηματολογίου, στο οποίο αποτυπώνεται το θεωρητικό πλαίσιο. Υπάρχει επίσης ο κίνδυνος για τους χρήστες που διεξάγουν δευτερογενή ανάλυση να δίνουν μεγαλύτερη έμφαση στην ανάλυση των δεδομένων παρά στο θεωρητικό πλαίσιο που διέπει τις ερωτήσεις. Επιπλέον, τα τεχνικά προβλήματα έχουν να κάνουν και με τον τύπο των δεδομένων στα οποία έχει κα-

νείς πρόσβαση. Ενδεικτικά αναφέρεται ότι οι χρήστες απογραφών έχουν πρόσβαση σε κύβους, παρά στα μικροδεδομένα της απογραφής, με αποτέλεσμα πιθανώς να μην είναι υλοποιήσιμες όλες οι διασταυρώσεις που ενέχουν τα ερευνητικά τους ερωτήματα. Τέλος, τα δεδομένα μπορεί να περιέχουν δομικά λάθη ή να είναι λανθασμένα παρά τους πολλαπλούς ελέγχους ή και να μην είναι διαθέσιμα για τον χωροχρόνο που επιθυμεί ο δευτερογενής αναλυτής.

Τα θεσμικά προβλήματα αφορούν την πρόσβαση στα ίδια τα δεδομένα. Τα δεδομένα μπορεί να είναι σε περίοδο εμπάργκο ή να είναι διαθέσιμα κατόπιν αιτήματος ή ακόμα και κατόπιν πληρωμής. Επίσης, σε μερικές χώρες οι ξένοι ερευνητές πρέπει να διαπραγματευτούν με τα ΑΔ ή τις στατιστικές υπηρεσίες ή τα ερευνητικά κέντρα σε διαφορετική βάση απ' ό,τι οι γηγενείς. Σε άλλες περιπτώσεις γίνεται ή δεν γίνεται διάκριση μεταξύ ιδιωτών και ερευνητών/ακαδημαϊκών.

Εν τέλει τα επιστημολογικά προβλήματα αφορούν τη δεδομένη φύση των δεδομένων. Τα επιστημολογικά προβλήματα συνδέονται με το γεγονός ότι τα πρωτογενή δεδομένα παρήχθησαν κάτω από ένα συγκεκριμένο θεωρητικό πλαίσιο και σε σχέση με συγκεκριμένη προοπτική. Ο δευτερογενής αναλυτής δύναται να μην αφουγκράζεται το θεωρητικό πλαίσιο, όπως έχει τεθεί στην πρωτογενή συλλογή. Επιπλέον μπορεί να υφίστανται προβλήματα εφαρμογής σε σχέση, για παράδειγμα, με τις κατηγορίες που χρησιμοποιούνται σε κατηγορικές μεταβλητές.

Σε διεθνές επίπεδο τα προβλήματα είναι ακόμα εντονότερα, δεδομένου ότι ένας δευτερογενής αναλυτής που επιθυμεί να διεξαγάγει διεθνική έρευνα πρέπει να προσπαθήσει να κατανοήσει τις κοινωνικές, πολιτικές και πολιτισμικές ιδιαιτερότητες των δεδομένων των χωρών που επιθυμεί να συμμετάσχουν στην ανάλυσή του. Είναι δύσκολο να κατανοηθεί η φύση των κοινωνικοπολιτικών συμφωνιών που απεικονίζονται στη συλλογή των πρωτογενών δεδομένων. Συνεπώς, αρκετές φορές, για τη διεξαγωγή μιας μεθοδολογικά άρτιας διεθνικής δευτερογενούς ανάλυσης, απαραίτητη προϋπόθεση είναι η επαφή με «πρόσωπα-κλειδιά» προερχόμενα από την εκάστοτε χώρα.

Η εξερεύνηση και ανακάλυψη των δεδομένων

Η εξερεύνηση, ανακάλυψη και λήψη των κατάλληλων διαθέσιμων δεδομένων αποτελεί μια σύνθετη διαδικασία αποτελούμενη από πέντε στάδια (CESSDA ERIC, 2019). Μετά τον προσδιορισμό των αναγκών των ερευνητών σε δεδομένα θα πρέπει να εντοπιστούν και να διερευνηθούν οι διαθέσιμες βάσεις δεδομένων και να εντοπιστούν εντός τους τα δεδομένα που πιθανώς ταιριάζουν με τους σκοπούς της σχεδιαζόμενης έρευνας. Τα δεδομένα που θα βρεθούν θα πρέπει να αξιολογηθούν ως προς την ποιότητα και την αξιοπιστία τους και κυρίως ως προς τη χρησιμότητά τους για την πραγματοποίηση της σχεδιαζόμενης επιστημονικής μελέτης.

Το πρώτο στάδιο αφορά τη διαδικασία προσδιορισμού των αναγκών σε δεδομένα, βάσει της σχεδιαζόμενης έρευνας, των ερευνητικών ερωτημάτων, του θεωρητικού πλαισίου, του είδους της σχεδιαζόμενης έρευνας (πρωτότυπη, αναπαραγωγή-έλεγχος, έρευνα για εκπαιδευτικούς σκοπούς), των χαρακτηριστικών των δεδομένων (ποσοτικά ή ποιοτικά, πληθυσμός, περιοχές, χρονικά διαστήματα), και άλλων χαρακτηριστικών τους, όπως η ηλεκτρονική μορφή των αποθηκευμένων αρχείων και η αμεσότητα πρόσβασης.

Αφού ολοκληρωθεί η παραπάνω διαδικασία, μπορεί να εκκινήσει η διαδικασία της έρευνας για την εύρεση των δεδομένων. Η γνώση αξιόπιστων πηγών δεδομένων είναι ένα σημαντικό βήμα αυτής της διερεύνησης. Υπάρχουν πολλές πηγές και αποθετήρια κοινωνικών δεδομένων, ενώ μεταξύ τους σημαίνοντα ρόλο έχει ο κατάλογος δεδομένων της CESSDA DC (χ.η.), ως πύλη των αποθετηρίων που λειτουργούν ως πάροχοι υπηρεσιών δεδομένων (Service Providers) του CESSDA-ERIC. Το SoDaNet αποτελεί τον ελληνικό κόμβο στη CESSDA. Ωστόσο στην Ευρώπη υπάρχουν και αρκετές άλλες αξιόπιστες πηγές δεδομένων, διαφόρων μεγεθών, από μεγάλες ακαδημαϊκές έρευνες και κυβερνητικές πηγές μέχρι μικρές ερευνητικές ομάδες ή μεμονωμένους ερευνητές. Αντίστοιχα αρχεία-πηγές δεδομένων υπάρχουν και σε αρκετές άλλες χώρες του κόσμου. Επίσης υπάρχουν και τα αποθετήρια μεγάλων οργανισμών, όπως οι βάσεις δεδομένων της Eurostat (Eurostat Database, χ.η), του ΟΟΣΑ (OECD Data, χ.η), του ΔΝΤ (IMF Data, χ.η) και της Παγκόσμιας Τράπεζας (World Bank Data, χ.η). Πέραν αυτών, υπάρχουν μεγάλου βεληνεκούς διεθνείς ή ευρωπαϊκές έρευνες που διαθέτουν τα δεδομένα τους, όπως το Ευροβαρόμετρο (Eurobarometer Data Service, χ.η), το International Social Survey Programme (ISSP, χ.η), η European Social Survey (ESS, χ.η), η έρευνα Survey of Health, Ageing and Retirement Europe (SHARE, χ.η) κ.ά. (βρείτε όλες τις ως άνω πηγές δεδομένων στο: <https://sodanet.gr/related-links>).

Όπως ειπώθηκε νωρίτερα, τα ΑΔ αποτελούν σημαντική πηγή δεδομένων. Διακρίνονται σε αποθετήρια που εστιάζουν σε συγκεκριμένα επιστημονικά πεδία και σε γενικού χαρακτήρα αποθετήρια. Για τη διευκόλυνση της αναζήτησης στις βάσεις δεδομένων τους, δημιουργήθηκε το Μητρώο των Αποθετηρίων Ερευνητικών Δεδομένων (re3data, χ.η), που επιτρέπει την εύκολη εύρεση αποθετηρίων με ερευνητικά δεδομένα μεταξύ 2.500 και πλέον εγγραφών.

Η διαδικασία της αναζήτησης περιλαμβάνει αναζήτηση δεδομένων βάσει διαφόρων κριτηρίων. Τα προφανή κριτήρια τα οποία τίθενται στα πληροφοριακά συστήματα των ΑΔ ώστε να διευκολύνουν την αναζήτηση στοιχείων για δευτερογενή ανάλυση είναι ο χωροχρόνος και η θεματική. Για παράδειγμα, ο κατάλογος δεδομένων της CESSDA (CESSDA Data Catalogue - CDC) επιτρέπει την αναζήτηση μεταξύ των ΑΔ των μελών του δικτύου με τη βοήθεια κριτηρίων όπως: θεματική, χώρα, έτη συλλογής δεδομένων και εκδότης. Κάποιοι πάροχοι δεδομένων χρησιμοποιούν το σύστημα λογισμικού NESSTAR (<http://www.nesstar.com/>) που είναι συμβατό με το

πρότυπο DDI 1.2.2, άλλοι έχουν δημιουργήσει τα δικά τους custom συστήματα, ενώ κάποιοι –λίγοι– χρησιμοποιούν την πλατφόρμα Colectica (<https://www.colectica.com/>), που είναι συμβατή με το DDI3 και με το DDI2. Η πλατφόρμα που προτιμάται τελευταία από πολλά ΑΔ είναι η πλατφόρμα Dataverse (<https://dataverse.org/>), που έχει δημιουργηθεί από το πανεπιστήμιο του Harvard και από κοινότητα προγραμματιστών παγκοσμίως, και πρόκειται για λογισμικό ανοικτού κώδικα.

Αφού εντοπιστούν τα πιθανά δεδομένα που είναι κατάλληλα για τη σχεδιαζόμενη έρευνα, θα πρέπει να ακολουθήσει η αξιολόγησή τους. Σε αυτή τη διαδικασία είναι κρίσιμη η ύπαρξη επαρκών μεταδεδομένων και άλλων πόρων, όπως ερωτηματολόγια, οδηγοί χρήσης κ.λπ. Σημαντικά κριτήρια επίσης είναι ο λόγος δημιουργίας των δεδομένων, ο τρόπος και ο χρόνος συλλογής τους, από ποιους πραγματοποιήθηκε η συλλογή, ποιοι είναι οι επιστημονικοί υπεύθυνοι της έρευνας ή των δεδομένων, εάν χρησιμοποιήθηκαν διαδικασίες πιστοποίησης ποιότητας, η εφαρμοζόμενη δειγματοληψία. Θα πρέπει επίσης τα εντοπισμένα δεδομένα να αξιολογηθούν και ως προς το αν καλύπτουν και σε ποιο βαθμό τις ανάγκες της έρευνας, έτσι όπως αυτές προσδιορίστηκαν στο αρχικό στάδιο. Χρειάζεται επίσης μια εκτίμηση του κόστους και του χρόνου πρόσβασης και ο συνυπολογισμός της ηλεκτρονικής μορφής αποθήκευσης στην οποία παρέχονται τα δεδομένα και τα μεταδεδομένα.

Αφού ολοκληρωθεί το προηγούμενο στάδιο και εντοπιστούν τα επιθυμητά και επαρκή για την έρευνα δεδομένα, ακολουθεί το στάδιο της λήψης τους. Για τη διευκόλυνση της πρόσβασης είναι σημαντικός ο έλεγχος των όρων πρόσβασης και των πιθανών διαφορετικών τρόπων πρόσβασης (ελεύθερη πρόσβαση, με εγγραφή, με λήψη κωδικού ή με λήψη ειδικής άδειας). Μόλις ολοκληρωθούν οι διαδικασίες αυτές, έχει έρθει η στιγμή της λήψης των δεδομένων. Η λήψη μπορεί να γίνει με διάφορους τρόπους, ανάλογα με το είδος των δεδομένων και τον τρόπο με τον οποίο παρέχονται σε κάθε αποθετήριο ή πηγή.

Δεν πρέπει να παραλείψουμε φυσικά και το τελευταίο στάδιο της διαδικασίας, που είναι η αναφορά στην πηγή των δεδομένων που έχουν χρησιμοποιηθεί. Οι λόγοι είναι γνωστοί και σημαντικοί (αξιοπιστία, δυνατότητα ελέγχου, απόδοση επαίνου στον πάροχο). Ο τρόπος πρέπει να εξασφαλίζει ότι η εύρεσή τους θα είναι δυνατή από τον αναγνώστη, και μπορεί να χρησιμοποιηθούν οι κανόνες του Datacite.org ή άλλων ανάλογων. Όπου υπάρχει αναγνωριστικός κωδικός ψηφιακού αντικειμένου (DOI), πρέπει να αναφέρεται στη σχετική παραπομπή.

Τα σχέδια διαχείρισης δεδομένων

Οι χρηματοδότες της έρευνας ζητούν πλέον ΣΔΔ από τους εν δυνάμει δικαιούχους, στα οποία σχέδια περιγράφονται με λεπτομέρεια οι ενέργειες για τη διαχείριση των δεδομένων, με στόχο τα δεδομένα να είναι FAIR (Findable, Accessible,

Interoperable, and Reusable), δηλαδή να μπορούν να αναζητηθούν, να είναι προσβάσιμα, διαλειτουργικά και επαναχρησιμοποιήσιμα. Σε πολλές περιπτώσεις δε, η κατάθεση ενός ΣΔΔ είναι υποχρεωτική ακόμα και για την υποβολή της πρότασης. Συνηθέστερα όμως είναι απαιτητή κατά τη διάρκεια του έργου. Ενδεικτικά αναφέρονται χρηματοδότες φορείς σε διάφορες χώρες που απαιτούν ή συστήνουν την υποβολή ενός ΣΔΔ, όπως: Research council of Norway, SkatteFUNN στη Νορβηγία, NWO, KNAW, ZonMw στην Ολλανδία, Wellcome Trust, Economic and Social Research Council, Cancer Research UK, Department for International Development στο Ηνωμένο Βασίλειο κ.ά. Σε σχέση με τα έργα Horizon 2020 της Ευρωπαϊκής Επιτροπής αναφερόταν ότι η κατάθεση του σχεδίου διαχείρισης των δεδομένων ήταν απαιτητή για όλα τα έργα στη θεματική προτεραιότητα «Open Research Data Pilot», ενώ ήταν προαιρετική για έργα άλλων θεματικών (European Commission, 2016).

Το ΣΔΔ αποτελεί ένα επίσημο έγγραφο που παρέχει ένα πλαίσιο για τη διαχείριση των δεδομένων κατά τη διάρκεια και μετά το πέρας ενός ερευνητικού έργου. Είναι ένα «ζωντανό» έγγραφο που μεταβάλλεται, εφόσον οι ανάγκες του έργου το επιτάσσουν. Ενημερώνεται σε όλο το έργο, ώστε να παρακολουθούνται οι αλλαγές σε σχέση με τη διαχείριση των δεδομένων και η τελευταία έκδοση αυτού αντικατοπτρίζει την τρέχουσα κατάσταση του έργου σε σχέση με τα δεδομένα. Τα ΣΔΔ δεν ακολουθούν κάποιο συγκεκριμένο πρότυπο, αλλά λίγο-πολύ όλες οι διαφορετικές εκδόχες πραγματεύονται τις ίδιες θεματικές. Το ΣΔΔ περιγράφει τον κύκλο ζωής-διαχείρισης των δεδομένων για όλα τα σύνολα δεδομένων που θα συλλεχθούν, θα δημιουργηθούν ή θα υποβληθούν σε επεξεργασία σε ένα έργο. Παρακάτω αναφέρονται οι βασικές θεματικές ενός ΣΔΔ, όπως τίθενται στον οδηγό για τη διαχείριση των δεδομένων της CESSDA (CESSDA Training Team, 2017-2019):

- **Επισκόπηση:** περιλαμβάνει στοιχεία όπως τον τίτλο του έργου και την περιγραφή, τις πηγές των δεδομένων που θα χρησιμοποιηθούν, τους επιστημονικούς υπεύθυνους του έργου, τους συνεργάτες, τους χρηματοδότες, το χρονοδιάγραμμα, τις προβλεπόμενες δαπάνες για τη διαχείριση των δεδομένων (δαπάνες για λογισμικό ή υλικό, για την αποθήκευση των δεδομένων και για αντίγραφα ασφαλείας, για την αρχειοθέτησή τους και τη διαχείρισή τους).
- **Οργάνωση δεδομένων και τεκμηρίωση:** Πώς θα γίνει η συλλογή και η οργάνωση των δεδομένων (ονοματοδοσία αρχείων, δομή φακέλων, έλεγχος εκδόσεων); Πώς θα πραγματοποιηθεί η τεκμηρίωση ώστε τα δεδομένα να καθίστανται επαναχρησιμοποιήσιμα, να γίνονται κατανοητά δίχως την παρέμβαση τρίτων και να διευκολύνουν την αναζήτηση; Είναι τα μεταδεδομένα κατανοητά από μηχανήματα; Χρησιμοποιούνται πρότυπα μεταδεδομένων;
- **Επεξεργασία των δεδομένων:** διαμόρφωση εκδόσεων, διαλειτουργικότητα, ποιοτικός έλεγχος δεδομένων

- Αποθήκευση δεδομένων και μεταδεδομένων: Ποιες λύσεις θα χρησιμοποιηθούν για την αποθήκευση των δεδομένων και των μεταδεδομένων και για τα αντίγραφα ασφαλείας (backup); Πώς θα ανακτηθούν τα δεδομένα σε περίπτωση απώλειας; Πώς θα προστατευτούν τα ευαίσθητα δεδομένα;
- Προστασία δεδομένων: έλεγχος δεοντολογίας, ενήμερη συναίνεση, πρόσβαση σε προσωπικά δεδομένα, πνευματικά δικαιώματα, συμφωνίες και περιορισμοί στη διαχείριση των δεδομένων
- Αρχαιοθέτηση και δημοσιοποίηση: Πώς και πού θα αποθηκευτούν τα δεδομένα μετά την ολοκλήρωση του έργου; Θα αρχειοθετηθούν τα δεδομένα σε ένα αξιόπιστο αποθετήριο (trusted repository); Θα υφίστανται μόνιμοι ταυτοποιητές για τα δεδομένα; Ποια μορφότυπα θα χρησιμοποιηθούν για την αρχειοθέτηση; Θα είναι τα δεδομένα ανοικτής πρόσβασης; Τι άδειες χρήσης απαιτούνται για τα δεδομένα; Πώς θα γίνεται η παραπομπή στα δεδομένα; Θα τεθεί περίοδος εμπάργκο; Υπάρχουν νομικοί περιορισμοί που εμποδίζουν τη δημοσίευση όλου του υλικού;
- Ανακάλυψη δεδομένων: προσδιορισμός αναγκών για χρήση υπαρχόντων δεδομένων, σκοπός χρήσης, πηγές όπου βρίσκονται αποθηκευμένα τα δεδομένα, τρόποι αναζήτησης των δεδομένων, ελάχιστη απαιτούμενη ποιότητα των δεδομένων, προσδιορισμός της διαδικασίας για την απόκτηση πρόσβασης στα δεδομένα και αναμενόμενο κόστος για την πρόσβαση στα δεδομένα και τη χρήση τους.

Όπως γίνεται κατανοητό απ' όλες τις θεματικές που πραγματεύεται ένα ΣΔΔ, αλλά ειδικότερα από τη θεματική «Αρχειοθέτηση και δημοσιοποίηση», ζητούνται ρητώς τα ΑΔ όπου θα αποθηκευτούν τα δεδομένα, τα κόστη και οι δαπάνες για την αρχειοθέτηση και προσδιορίζεται αν το ΑΔ όπου θα φιλοξενηθούν τα δεδομένα είναι αξιόπιστο.

Συμπεράσματα

Συμπερασματικά, τα ΑΔ εισέρχονται δυναμικά στη μεθοδολογική παρέμβαση της ερευνητικής διαδικασίας που επιτάσσουν χρηματοδοτικοί μηχανισμοί και επιστημονικά περιοδικά (Εικόνα 12). Πιο συγκεκριμένα:

- Η χρηματοδότηση της έρευνας επιτάσσει πολιτικές διαφάνειας στα δεδομένα ή/και στον τρόπο παραγωγής αυτών, καθώς και πρόσβαση σε αυτά από τρίτους.
- Οι πολιτικές διαφάνειας προδιαγράφουν την ορθή διαχείριση των δεδομένων είτε στα ΣΔΔ είτε στους κανόνες που θέτουν τα επιστημονικά περιοδικά για τη δημοσίευση εργασιών.

- Στα ΣΔΔ ή στις πολιτικές που καθορίζουν τα επιστημονικά περιοδικά προσδιορίζεται η αποτελεσματική διαχείριση των δεδομένων, που αφορά την κατάθεση δεδομένων, μεταδεδομένων ή/και ρέπλικων σε ΑΔ.
- Τα ΑΔ διαθέτουν υπό συνθήκες δεδομένα και μεταδεδομένα σε τρίτους για δευτερογενή έρευνα και χρήση.
- Τα δεδομένα για να παραχθούν απαιτούν πρωτογενή συλλογή ή δευτερογενή έρευνα.
- Τέλος, η έρευνα για να πραγματοποιηθεί απαιτεί πόρους, που πηγάζουν από τη χρηματοδότησή της.



Εικόνα 12: Τα αποθετήρια δεδομένων εισέρχονται στον κύκλο χρηματοδότησης της έρευνας και τα δεδομένα επεκτείνουν τον κύκλο ζωής τους

Καθίσταται προφανές, ότι ο κύκλος ζωής των ερευνητικών δεδομένων –και ειδικότερα των ποσοτικών– πλέον επεκτείνεται μέσω των ΑΔ. Η μέχρι πρότινος προσδοκία των συγγραφέων ήταν ότι η επέκταση αυτή θα πραγματοποιούνταν περισσότερο λόγω αλλαγής στην ερευνητική κουλτούρα του διαμοιρασμού των δεδομένων. Πλέον η επέκταση αυτή πραγματοποιείται. Λιγότερο όμως λόγω αλλαγής στην ερευνητική κουλτούρα του διαμοιρασμού των δεδομένων και περισσότερο λόγω των προϋποθέσεων που επιβάλλουν χρηματοδοτικοί μηχανισμοί έρευνας και επιστημονικά περιοδικά.

Βιβλιογραφικές αναφορές

- Castro, E., Garnett, A. (2014). Building a bridge between journal articles and Research Data: The PKP-dataverse integration project. *International Journal of Digital Curation*, 9(1), 176–184. <https://doi.org/10.2218/ijdc.v9i1.311>
- CESSDA (Consortium of European Social Science Data Archives). (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.cessda.eu/>
- CESSDA DC Data Catalogue. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://datacatalogue.cessda.eu/>
- CESSDA ERIC. (2019). Data discovery. CESSDA Train the Trainers Workshop, Αθήνα, Μάιος 2019.
- CESSDA Training Team.(2017 – 2019). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. DOI: 10.5281/zenodo.3820473. Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.cessda.eu/DMGuide>
- Creative Commons. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://creativecommons.org/>
- ESS. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.europeansocialsurvey.org/>
- European Commission. (2014). Technical Manual of the Single Integrated Metadata Structure (SIMS). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://ec.europa.eu/eurostat/documents/64157/4373903/03-Single-Integrated-Metadata-Structure-and-its-Technical-Manual.pdf>
- European Commission. (2016). Guidelines on Data Management in Horizon 2020. Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Eurostat Database. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://ec.europa.eu/eurostat/en/web/main/data/database>
- Eurobarometer Data Service. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.gesis.org/en/eurobarometer-data-service/search-data-access/data-access>
- GESIS. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.gesis.org/>
- Hakim C. 1982. *Secondary Analysis in social Research*. London: Allen & Unwin.
- IMF Data. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://data.imf.org/>
- ISSP. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.gesis.org/en/issp/home>
- Journal of Politics. (χ.η.). Guidelines for Data Replication. Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.journals.uchicago.edu/journals/jorp/data-replication>
- Κάλλας Γ., Λιναρδής Α. (2004). Η επίδραση των ερευνητικών υποδομών στη μεθοδολογία των κοινωνικών επιστημών, *Επιθεώρηση Κοινωνικών Ερευνών*, Τεύχος 115 Γ,139-155. <https://doi.org/10.12681/grsr.9348>.
- Λιναρδής Α. (2006). Η αναγκαιότητα αυστηρής και ποιοτικής τεκμηρίωσης στις κοινωνικές έρευνες. Το πρότυπο τεκμηρίωσης Data Documentation Initiative. Κείμενα εργασίας 2006/14, ΕΚΚΕ.
- Λιναρδής Α., Αλεξανδρής Κ., Κληρονόμος Ν. (2022). Ο νέος κατάλογος δεδομένων του SoDaNet: παραμετροποίηση, βελτιστοποίηση και επέκταση των λειτουργιών του Dataverse για τις ανάγκες των αποθετηρίων του Sodanet. Στο Κάλλας κ.ά. (επιμ.), *Τεκμηρίωση και διαχείριση εμπειρικών ερευνών. Ερευνητικές υποδομές στις κοινωνικές επιστήμες*.

- Nesstar. (2008). User Guide of NESSTAR Publisher v3.54. Ανακτήθηκε στις 7 Ιανουαρίου 2022 από τη διεύθυνση: http://nesstar.com/export/sites/default/doc/Publisher_UserGuide_v3.54.pdf
- Nesstar. (2011). User Guide of NESSTAR Publisher v4.0. Ανακτήθηκε στις 14 Δεκεμβρίου 2021 από τη διεύθυνση: http://nesstar.com/help/4.0/publisher/download_resources/Publisher_UserGuide_v4.0.pdf
- NSD. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.nsd.no/en>
- OECD Data. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://data.oecd.org/searchresults/?r=+f/type/datasets>
- re3data. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://www.re3data.org/>
- Ryssevik, J., & Musgrave, S. (2001). The social science dream machine: Resource discovery, analysis, and delivery on the Web. *Social science computer review*, 19(2), 163-174.
- Share. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <http://www.share-project.org/>
- UK Data Service. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://ukdataservice.ac.uk/>
- UK Data Service – Anonymization quantitative. (χ.η.). Anonymization quantitative data. Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://ukdataservice.ac.uk/learning-hub/research-data-management/anonymisation/anonymising-qualitative-data/>
- UK Data Service – Anonymization. (χ.η.). Anonymization step by step. Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://ukdataservice.ac.uk/learning-hub/research-data-management/anonymisation/anonymisation-step-by-step/>
- UK Data Service – Study. (χ.η.). Study level documentation. Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://ukdataservice.ac.uk/learning-hub/research-data-management/document-your-data/study-level-documentation/>
- Vartanian, T. P. 2010. *Secondary data analysis*. Oxford University Press.
- World Bank Data. (χ.η.). Ανακτήθηκε στις 5 Φεβρουαρίου, 2022 από τη διεύθυνση: <https://data.worldbank.org/>