



Ακρωνύμιο πρότασης:

**DataPopEU**

Τίτλος πρότασης:

**Καινοτόμες μέθοδοι και δεδομένα υψηλής ποιότητας για τη μελέτη του  
λαϊκισμού και του Ευρωσκεπτικισμού**

## **Παραδοτέο 6.1**

**Σχέδιο διαχείρισης και διαμοιρασμού των δεδομένων (πρώτη έκδοση)**

Μάρτιος 2020

## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΙΣΑΓΩΓΗ.....	3
ΣΥΝΟΨΗ ΕΡΓΟΥ.....	3
ΣΥΝΟΨΗ ΔΕΔΟΜΕΝΩΝ.....	4
ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΔΙΑΔΙΚΑΣΙΕΣ ΔΙΑΣΦΑΛΙΣΗΣ ΠΟΙΟΤΗΤΑΣ..	4
ΧΡΗΣΙΜΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ.....	5
ΓΕΝΙΚΟΣ ΚΑΝΟΝΙΣΜΟΣ ΠΡΟΣΤΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ (GDPR).....	5
ΠΡΟΣΤΑΣΙΑ ΔΕΔΟΜΕΝΩΝ.....	7
ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ ΔΙΑΤΗΡΗΣΗ ΔΕΔΟΜΕΝΩΝ.....	8
ΔΙΑΜΟΙΡΑΣΜΟΣ ΔΕΔΟΜΕΝΩΝ.....	9
ΠΑΡΑΡΤΗΜΑΤΑ.....	10
Παράρτημα Α: Συμφωνητικό στην ιστοσελίδα του FORS που καλούνται να αποδεχτούν οι ερευνητές προκειμένου να έχουν πρόσβαση στα δεδομένα της έρευνας Υποψηφίων Βουλευτών 2019.....	10
Παράρτημα Β: Βεβαίωση τήρησης Κώδικα Δεοντολογίας του ΑΠΘ για την Εθνική Μετεκλογική Έρευνα Ψηφοφόρων 2019.....	11

## ΕΙΣΑΓΩΓΗ

Η συγκεκριμένη αναφορά συνιστά μία επισκόπηση της στρατηγικής διαχείρισης των δεδομένων του έργου. Εστιάζοντας, λοιπόν, στην πολιτική διαχείρισης των δεδομένων (διαδικασίες συλλογής και διαμοιρασμού των δεδομένων), περιγράφεται i) η μεθοδολογία που θα εφαρμοστεί όσον αφορά στη συλλογή των δεδομένων, ii) η διαχείρισή τους κατά τη διάρκεια και μετά τη λήξη του έργου, iii) ο σκοπός συλλογής τους, iv) η χρησιμότητά τους, v) και οι διαδικασίες που θα ακολουθηθούν σχετικά με τον διαμοιρασμό, την προστασία και τη διατήρηση των δεδομένων μετά τη λήξη του έργου. Πρόκειται για την πρώτη έκδοση της στρατηγικής διαχείρισης των δεδομένων του DataPopEU, ενώ η τελική έκδοση θα υποβληθεί τον τελευταίο μήνα του έργου (M24).

## ΣΥΝΟΨΗ ΕΡΓΟΥ

Σκοπός του έργου DataPopEU είναι η ανάπτυξη καινοτόμων μεθόδων και τεχνικών για τη συλλογή, την επεξεργασία και ανάλυση μεγάλου όγκου δεδομένων που προέρχονται από διαφορετικές πηγές όπως διαδικτυακές έρευνες, μέσα κοινωνικής δικτύωσης και Τύπο. Επιπλέον, η δημιουργία της Πυξίδας για λαϊκισμό και Ευρωσκεπτικισμό (PopEUCompass) θα αποτελέσει ακόμα μια πηγή δεδομένων καθώς οι απαντήσεις των πολιτών στις ερωτήσεις της Πυξίδας θα παράξουν καινούργια δεδομένα. Όλα τα παραπάνω δεδομένα συγκεντρώνονται και αναλύονται κυρίως στο πλαίσιο δύο πολιτικών φαινομένων, του λαϊκισμού και του Ευρωσκεπτικισμού. Οι διαφορετικές πηγές δεδομένων επιτρέπουν τη μελέτη ομάδων των πολιτών με διαφορετικά χαρακτηριστικά και συνδέονται άμεσα με τους στόχους του συγκεκριμένου έργου.

Ο πρώτος στόχος του έργου είναι η παραγωγή δεδομένων υψηλής ποιότητας δεδομένων των διαδικτυακών ερευνών του έργου μέσω της δημιουργίας και εφαρμογής δεικτών ποιότητας απαντήσεων (response quality indicators) και μεθόδων καθαρισμού δεδομένων (data cleaning methods).

Δεύτερος στόχος του έργου είναι η ανάπτυξη και εφαρμογή καινοτόμων μεθόδων για τη συλλογή, την ανάλυση και την ερμηνεία των δεδομένων που προέρχονται από τα μέσα κοινωνικής δικτύωσης και σχετίζονται με τα δύο υπό εξέταση πολιτικά φαινόμενα.

Τρίτος στόχος του έργου είναι η δημιουργία σωμάτων κειμένου (corpora) ανοικτής πρόσβασης σχετικά με τον λαϊκισμό και τον Ευρωσκεπτικισμό προερχόμενα από άρθρα του Τύπου.

Τέταρτος στόχος του έργου είναι η δημιουργία μιας Πυξίδας για λαϊκισμό και Ευρωσκεπτικισμό (PopEUCompass) στη λογική του HelpMeVote (Ηλεκτρονικός Σύμβουλος Ψήφου).

Πέμπτος στόχος του έργου είναι η δημιουργία αναθεωρημένων ερωτηματολογίων για την Ελληνική Εθνική Μετεκλογική Έρευνα Υποψηφίων Βουλευτών 2019 και την Ελληνική Εθνική Μετεκλογική Έρευνα Ψηφοφόρων 2019. Σε αυτόν συμπεριλαμβάνεται και η προσαρμογή και αναθεώρηση των ερωτήσεων που περιέχονται στο PopEUCompass μέχρι το τέλος του έργου.

Τελευταίος, αλλά εξίσου σημαντικός στόχος του έργου είναι η μελέτη των δύο υπό εξέταση πολιτικών φαινομένων και η διερεύνηση της σχέσης μεταξύ των δύο χρησιμοποιώντας τα δεδομένα που συλλέχθηκαν και τέλος, η παρακολούθηση της εξέλιξης του λαϊκισμού και του Ευρωσκεπτικισμού στα ελληνικά πολιτικά κόμματα από το 2015 μέχρι το 2019 μέσα από μία συγκριτική και συνδυαστική ανάλυση των δεδομένων που συλλέχθηκαν από τις μετεκλογικές έρευνες για τις τρεις εκλογικές αναμετρήσεις (εθνικές βουλευτικές εκλογές Ιανουαρίου και Σεπτεμβρίου 2015 και Ιουλίου 2019).

## ΣΥΝΟΨΗ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα του DataPopEU περιλαμβάνουν δεδομένα που συλλέγονται στα πλαίσια του εν λόγω έργου και δεδομένα από προηγούμενες έρευνες του ΕΛΝΕΣ (Μετεκλογικές Εθνικές Έρευνες). Ειδικότερα, περιλαμβάνει δεδομένα των ερευνών του 2015 (έρευνες υποψηφίων βουλευτών και ψηφοφόρων) όπως και δεδομένα του HelpMeVote του 2015 (Ηλεκτρονικός Σύμβουλος Ψήφου) που σχετίζονται με τον λαϊκισμό και τον Ευρωσκεπτικισμό.

Τα δεδομένα που συλλέγονται στα πλαίσια του παρόντος έργου σχετίζονται με τον λαϊκισμό και τον Ευρωσκεπτικισμό και ανήκουν σε τέσσερις κατηγορίες: α) δεδομένα διαδικτυακών ερευνών (έρευνες υποψηφίων βουλευτών, ψηφοφόρων, εμπειρογνομόνων), β) δεδομένα από τα μέσα κοινωνικής δικτύωσης, γ) δεδομένα από την Πυξίδα για λαϊκισμό και Ευρωσκεπτικισμό (PopEUCompass) που θα δημιουργηθεί τον δέκατο μήνα του έργου και δ) δεδομένα από τον Τύπο (δημοσιευμένα άρθρα τα οποία μετά από κατάλληλη επεξεργασία μετατρέπονται σε σώματα κειμένου).

Όλα τα δεδομένα του έργου θα είναι υπό τη μορφή ακέραιων (integers), τύπων δεδομένων αληθείας (booleans), χαρακτήρων (characters) και αλφαριθμητικών συμβολοσειρών (strings) ενώ ο μορφότυπος των δεδομένων (data format) θα είναι είτε σε αριθμητική μορφή π.χ σε αρχείο .csv και RData για τα δεδομένα διαδικτυακών ερευνών και PopEUCompass και .RData για τα δεδομένα από τα μέσα κοινωνικής δικτύωσης, είτε σε μορφή κειμένου σε αρχεία .txt (utf8) για τα δεδομένα από τον Τύπο (σώματα κειμένου προερχόμενα από άρθρα του Τύπου).

Όσον αφορά στο μέγεθος των δεδομένων, i) για τα δεδομένα της έρευνας ψηφοφόρων εκτιμάται ότι το μέγεθος του αρχείου θα είναι 200-250 KB, ii) για τα δεδομένα της έρευνας υποψηφίων βουλευτών το μέγεθος του αρχείου εκτιμάται ότι θα είναι 200-300 KB, iii) για τα δεδομένα της έρευνας εμπειρογνομόνων (expert survey) υπολογίζεται ότι το μέγεθος του αρχείου θα είναι 70-120 KB, iv) για τα δεδομένα της Πυξίδας (PopEUCompass) υπολογίζεται ότι το μέγεθος του αρχείου θα είναι 10-15 MB, v) για τα σώματα κειμένου που προκύπτουν από την επεξεργασία των άρθρων του Τύπου υπολογίζεται ότι το μέγεθος του αρχείου θα είναι λιγότερο από 1 GB και τέλος, vi) για τα δεδομένα από τα μέσα κοινωνικής δικτύωσης υπολογίζεται να είναι κατά προσέγγιση 800-1000 KB ανά ημέρα.

## ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΔΙΑΔΙΚΑΣΙΕΣ ΔΙΑΣΦΑΛΙΣΗΣ ΠΟΙΟΤΗΤΑΣ

Καθώς πρόκειται για διαφορετικούς τύπους δεδομένων, ο τρόπος με τον οποίο συλλέγονται διαφέρει αρκετά. Οι τρεις διαδικτυακές έρευνες (έρευνα ψηφοφόρων, έρευνα υποψηφίων βουλευτών και έρευνα εμπειρογνομόνων) διεξάγονται μέσω του LimeSurvey, ένα ανοικτής πρόσβασης λογισμικό για διαδικτυακές έρευνες. Από το ίδιο λογισμικό εξάγονται και τα δεδομένα των τριών ερευνών. Από τον αρχικό σχεδιασμό των ερευνών σε κάθε συμμετέχοντα μοιράζεται ένα συγκεκριμένο «κουπόνι» (token), έτσι ώστε να διασφαλιστεί ότι η έρευνα συμπληρώνεται από έναν και μόνο συμμετέχοντα.

Στη συνέχεια, αφού συγκεντρωθούν τα δεδομένα, θα εφαρμοστούν μέθοδοι καθαρισμού δεδομένων οι οποίες βασίζονται σε δείκτες ποιότητας απαντήσεων όπως είναι το μέγεθος της απάντησης στις ερωτήσεις ανοικτού τύπου, η παράλειψη απάντησης ερωτήσεων (skipping), ο χρόνος που αφιερώθηκε σε κάθε απάντηση κ.λπ. Πιο συγκεκριμένα, θα υπολογίσουμε την αναλογία των ερωτήσεων που δεν απαντήθηκαν για κάθε ερωτώμενο, την αναλογία των ενδιάμεσων απαντήσεων στις (5-βάθμιες) κλίμακες likert, τον αριθμό των χαρακτήρων των απαντήσεων στις ερωτήσεις ανοικτού τύπου, την αναλογία των ίδιων απαντήσεων σε ερωτήσεις διάταξης (grid) και την αναλογία των ακραίων απαντήσεων στις κλίμακες του ερωτηματολογίου.

Σχετικά με τα δεδομένα των μέσων κοινωνικής δικτύωσης, αυτά συλλέγονται μέσω της R. Τα δεδομένα που αφορούν τους υποψηφίους βουλευτές συλλέγονται αφού έχουν συγκεντρωθεί τα

ονόματα των λογαριασμών τους στο Twitter και το Facebook (Twitter and Facebook handles). Με βάση τα ονόματά τους στα μέσα κοινωνικής δικτύωσης συγκεντρώνονται οι δημοσιεύσεις τους και πλήθος μεταβλητών που περιέχει διάφορες πληροφορίες σχετικά με αυτές. Τα υπόλοιπα δεδομένα που μαζεύονται από το Twitter συγκεντρώνουν δημοσιεύσεις οι οποίες περιέχουν λέξεις κλειδιά, ανεξάρτητα με το ποιος τις έχει δημοσιεύσει. Και στις δύο περιπτώσεις, οι δημοσιεύσεις ελέγχονται ώστε να έχουν πραγματοποιηθεί στο χρονικό διάστημα το οποίο είναι υπό εξέταση. Καθώς η ανάλυση που θα πραγματοποιηθεί σε αυτά τα δεδομένα είναι αυτόματη ανάλυση περιεχομένου (automated content analysis), για να διασφαλιστεί η ποιότητα των δεδομένων αφαιρούνται από τα κείμενα των δημοσιεύσεων όλα τα σημεία στίξης, εικονίδια (emoji/emoticons), σύνδεσμοι κ.λπ.

Σχετικά με τα άρθρα που συγκεντρώνονται από τον Τύπο και μετατρέπονται σε σώματα κειμένου, για τη διασφάλιση της ποιότητας των σωμάτων κειμένου θα δοθεί ιδιαίτερη προσοχή στη διαδικασία καθαρισμού των κειμένων.

## ΧΡΗΣΙΜΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα του DataPopEU θα είναι χρήσιμα, τόσο στους επιστήμονες όσο και στην κοινωνία και τις επιχειρήσεις. Πιο συγκεκριμένα, όσον αφορά στα δεδομένα των διαδικτυακών ερευνών (υποψηφίων βουλευτών, ψηφοφόρων, εμπειρογνομόνων) και τα δεδομένα από τα μέσα κοινωνικής δικτύωσης, ο συνδυασμός αυτών των δύο τύπων δεδομένων μπορεί να βοηθήσει τους επιστήμονες αλλά και τις επιχειρήσεις που ασχολούνται με τις έρευνες με ερωτηματολόγια να μειώσουν το μέγεθος των ερωτηματολογίων χωρίς να απουσιάζουν οι πληροφορίες που θέλουν να συλλέξουν. Για την ακρίβεια, τα δεδομένα από τα social media μπορούν να υποκαταστήσουν σε ορισμένες περιπτώσεις ερωτήσεις διαδικτυακών ερευνών. Έτσι, με τον συνδυασμό των δύο μεθόδων μπορούμε να λάβουμε όλες τις απαραίτητες πληροφορίες ελαχιστοποιώντας ταυτόχρονα το μέγεθος των ερωτηματολογίων και ως επακόλουθο και τον χρόνο που θα χρειαστεί για την απάντησή τους. Ελαχιστοποιώντας, λοιπόν, το μέγεθος των ερωτηματολογίων και τον χρόνο απάντησής τους μπορούμε να λάβουμε περισσότερα και πιο ποιοτικά δεδομένα καθώς ο χρόνος που χρειάζεται για να απαντηθεί ένα ερωτηματολόγιο είναι ένας αρκετά σημαντικός παράγοντας που επηρεάζει την ποιότητα των δεδομένων που συλλέγονται.

Συνολικά, όλα τα δεδομένα του έργου και η ανάλυσή τους θα συμβάλλουν στο να διατεθεί πολύτιμη επιστημονική γνώση σχετικά με τα δύο υπό εξέταση πολιτικά φαινόμενα στην κοινωνία. Αυτή η επιστημονική γνώση θα απλοποιηθεί και θα δημοσιοποιηθεί στην ιστοσελίδα του έργου προκειμένου να γίνει κατανοητή στο ευρύ κοινό (π.χ. μέσα από τη χρήση της Πυξίδας για λαϊκισμό και Ευρωσκεπτικισμό). Η διάχυση της επιστημονικής γνώσης προς την κοινωνία και ειδικότερα πληροφοριών που σχετίζονται με δύο σημαντικά πολιτικά φαινόμενα μπορεί να συμβάλει στην καλύτερη αντιπροσώπευση των πολιτών, οι οποίοι, γνωρίζοντας καλύτερα τις στάσεις των πολιτικών κομμάτων και των υποψηφίων σε διάφορα πολιτικά ζητήματα, μπορούν να επιλέξουν το κόμμα ή τον υποψήφιο του οποίου οι απόψεις ταιριάζουν περισσότερο με τις δικές τους. Επιπλέον, μιας και τα δεδομένα θα είναι ελεύθερης πρόσβασης προς επιστήμονες, ενισχύεται και ενθαρρύνεται η επιστημονική έρευνα. Πιο συγκεκριμένα, τα δεδομένα του DataPopEU μπορούν να χρησιμοποιηθούν από κοινωνικούς επιστήμονες, από επιστήμονες δεδομένων (data scientists), από όσους -επιστήμονες και μη- ασχολούνται με τις διαδικτυακές έρευνες κ.λπ.

## ΓΕΝΙΚΟΣ ΚΑΝΟΝΙΣΜΟΣ ΠΡΟΣΤΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ (GDPR)

Ο Κανονισμός 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016 καθορίζει τον νέο Γενικό Κανονισμό Προστασίας Δεδομένων στην Ευρωπαϊκή Ένωση. Ο

νέος κανονισμός θεσπίζει αυστηρούς κανόνες όσον αφορά στη συγκέντρωση, επεξεργασία και αποθήκευση δεδομένων προσωπικού χαρακτήρα πολιτών της Ευρωπαϊκής Ένωσης. Σε αυτούς συμπεριλαμβάνονται μεμονωμένα άτομα, εταιρείες, δημόσιοι ή μη κυβερνητικοί οργανισμοί.

Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) εγκρίθηκε στις 27 Απριλίου 2016 και τέθηκε σε εφαρμογή στις 25 Μαΐου 2018 έπειτα από μια μεταβατική περίοδο δύο ετών. Αντικατέστησε την προηγούμενη οδηγία 95/46/EK του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 24ης Οκτωβρίου 1995 για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα και για την ελεύθερη κυκλοφορία των δεδομένων αυτών. Ωστόσο, μιας και ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) αποτελεί κανονισμό και όχι οδηγία δεν χρειάζεται ψήφιση από τα εθνικά κοινοβούλια των κρατών-μελών της Ευρωπαϊκής Ένωσης και είναι δεσμευτικός για όλα τα κράτη-μέλη.

Ωστόσο, η νομική ερμηνεία του όρου «προσωπικά δεδομένα» είναι αρκετά ευρεία καθώς πρόκειται για δεδομένα που σχετίζονται με το άτομο και αφορούν πληροφορίες, τόσο για την προσωπική όσο και την επαγγελματική και δημόσια ζωή του, οι οποίες αν συγκεντρωθούν όλες μαζί μπορούν να οδηγήσουν σε ταυτοποίησή του. Για παράδειγμα, τέτοιου είδους πληροφορίες είναι το όνομα, η διεύθυνση κατοικίας, η διεύθυνση ηλεκτρονικού ταχυδρομείου, η διεύθυνση διαδικτυακού πρωτοκόλλου (IP) κ.ο.κ.

Σύμφωνα με τον Γενικό Κανονισμό Προστασίας Δεδομένων, η συλλογή δεδομένων μπορεί να ξεκινήσει μόνο αφού ο/η ερωτώμενος/η έχει σαφώς εκφράσει τη συγκατάθεσή του/της. Οι ερωτώμενοι/ες πρέπει επίσης να ενημερώνονται με σαφήνεια για τα δεδομένα που συλλέγονται και για ποιο σκοπό. Αφού ενημερωθούν για το είδος και τον σκοπό των δεδομένων, τον χρόνο διατήρησής τους και τον τρόπο που μπορούν να ασκήσουν τα δικαιώματά τους, πρέπει να συμφωνήσουν ρητά με τη συλλογή των δεδομένων. Αυτό διασφαλίζεται μέσω της συμπλήρωσης μιας φόρμας συγκατάθεσης στην αρχή κάθε έρευνας. Μόνο εάν δώσουν τη συγκατάθεσή τους, μπορούν να συνεχίσουν την έρευνα. Σε περίπτωση που δεν συμφωνούν, η έρευνα θα πρέπει να τερματιστεί αμέσως και θα πρέπει να ενημερωθούν ότι δεν μπορούν να ολοκληρώσουν την έρευνα λόγω της διαφωνίας τους με τη συλλογή των δεδομένων τους.

Το υποκείμενο των δεδομένων έχει δικαίωμα να ανακαλέσει την συγκατάθεσή του ανά πάσα στιγμή. Η ανάκληση της συγκατάθεσης δεν θίγει τη νομιμότητα της επεξεργασίας που βασίστηκε στη συγκατάθεση προ της ανάκλησής της. Σε κάθε περίπτωση και σε οποιοδήποτε χρόνο, το υποκείμενο των δεδομένων έχει δικαίωμα στην πρόσβαση, διόρθωση και διαγραφή (δικαίωμα στη λήθη) των δεδομένων προσωπικού χαρακτήρα, στον περιορισμό της επεξεργασίας τους, στην αντίταξη της επεξεργασίας τους και στην φορητότητά τους.

Επιπλέον, σύμφωνα με τις προβλέψεις του κανονισμού (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016 (άρθρο 8, παράγραφος 1) η προσφορά υπηρεσιών της κοινωνίας των πληροφοριών απευθείας σε παιδιά, και η επεξεργασία δεδομένων προσωπικού χαρακτήρα παιδιού είναι σύννομη εάν το παιδί είναι τουλάχιστον 16 ετών. Εάν το παιδί είναι ηλικίας κάτω των 16 ετών, η επεξεργασία αυτή είναι σύννομη μόνο εάν και στον βαθμό που η εν λόγω συγκατάθεση παρέχεται ή εγκρίνεται από το πρόσωπο που έχει τη γονική μέριμνα του παιδιού.

Όταν συλλέγονται δεδομένα προσωπικού χαρακτήρα, ο υπεύθυνος επεξεργασίας παρέχει στο υποκείμενο τις ακόλουθες πληροφορίες: την ταυτότητα και τα στοιχεία επικοινωνίας του υπευθύνου επεξεργασίας (και του υπευθύνου προστασίας προσωπικών δεδομένων, αν υπάρχει), τους σκοπούς της επεξεργασίας για τους οποίους προορίζονται τα δεδομένα προσωπικού χαρακτήρα και τον τρόπο άσκησης των δικαιωμάτων του.

## ΠΡΟΣΤΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα που θα συλλεχθούν κατά τη διάρκεια του έργου έχουν δύο μορφές: α) δεδομένα που είναι διαθέσιμα δημοσίως και β) προσωπικά δεδομένα (δεδομένα προσωπικού χαρακτήρα). Όσον αφορά στη συλλογή προσωπικών δεδομένων θα υπάρχει πλήρης συμμόρφωση με τον Γενικό Κανονισμό Προστασίας Δεδομένων (GDPR). Πιο συγκεκριμένα, σε όλες τις έρευνες που θα διεξαχθούν από την ερευνητική ομάδα του έργου -τόσο διαδικτυακές όσο και τηλεφωνικές- στην πρώτη σελίδα της έρευνας θα υπάρχει φόρμα συγκατάθεσης η οποία θα περιλαμβάνει πληροφορίες σχετικά με τον σκοπό της έρευνας, τη διασφάλιση της ανωνυμίας και τα δικαιώματα των συμμετεχόντων ενώ θα τους ενημερώνει πως η συμμετοχή τους στην έρευνα είναι προαιρετική και μπορούν να τη διακόψουν σε οποιοδήποτε στάδιο διεξαγωγής της έρευνας. Για να συμμετέχει κανείς στις εν λόγω έρευνες είναι απαραίτητο να δηλώσει ρητά τη συγκατάθεσή του για καταγραφή, αποθήκευση και στατιστική επεξεργασία των απαντήσεών του.

Επιπλέον, στις έρευνες που απευθύνονται στους ψηφοφόρους, θα υπάρχει ερώτηση η οποία θα εμφανίζεται εφόσον ο ερωτώμενος έχει δηλώσει τη συγκατάθεσή του για καταγραφή, αποθήκευση και στατιστική επεξεργασία των απαντήσεών του και θα εξετάζει το ενδεχόμενο να είναι κάτω των 16 ετών. Αν ο ερωτώμενος απαντήσει πως είναι κάτω από 16 ετών αυτομάτως αποκλείεται από την έρευνα καθώς σύμφωνα με το άρθρο 8 παράγραφος 1 του κανονισμού (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016 δεν μπορεί να συμμετέχει σε έρευνες χωρίς τη συγκατάθεση κηδεμόνα.

Όσον αφορά στη διασφάλιση της ανωνυμίας, τα δεδομένα της έρευνας ψηφοφόρων δεν μπορούν εύκολα να οδηγήσουν στην ταυτοποίηση των συμμετεχόντων καθώς σε αρκετές επιλογές απαντήσεων έχει υιοθετηθεί η τεχνική της γενίκευσης (π.χ. στην ερώτηση που αφορά το εισόδημα του ερωτώμενου η απάντηση ζητείται να δοθεί κατά προσέγγιση δηλ. μέχρι 10.000€, από 10.001€ - 15.000€ κ.ο.κ.). Επιπρόσθετα, μιας και οι προσκλήσεις για συμμετοχή στην έρευνα ψηφοφόρων αποστέλλονται σε αριθμούς κινητών τηλεφώνων είναι αναγκαίο να διευκρινιστεί πως η επιλογή των αριθμών γίνεται τυχαία μέσω ηλεκτρονικού υπολογιστή (RDD – Random Digit Dialing) και επομένως δεν είναι δυνατή η συσχέτιση ενός αριθμού με τον κάτοχό του. Την αποστολή των προσκλήσεων για συμμετοχή στην έρευνα σε τυχαία επιλεγμένους αριθμούς έχει αναλάβει μία ιδιωτική εταιρεία. Αξίζει να σημειωθεί πως δεν θα υπάρχει καμία σύνδεση μεταξύ του αριθμού τηλεφώνου και των απαντήσεων των συμμετεχόντων. Τα προσωπικά στοιχεία τηρούνται χωριστά από τις απαντήσεις, ώστε κανείς να μην μπορεί να ταυτίσει απαντήσεις με αριθμό τηλεφώνου. Τα προσωπικά στοιχεία των συμμετεχόντων μαζί με το ειδικό κουπόνι-κωδικό (token) -αποτελεί μέρος του link που συμπεριλαμβάνεται στην πρόσκληση- θα βρίσκονται στον διακομιστή (server) της εταιρείας που θα αναλάβει την αποστολή των SMS και θα διαγράφονται αμέσως μετά τη διαδικασία της αποστολής. Επιπλέον, θα βρίσκονται -από κοινού με τις απαντήσεις στις ερωτήσεις της έρευνας- στον διακομιστή (server) του επιστημονικού υπευθύνου σε κρυπτογραφημένη μορφή. Πρόκειται για servers υψηλής ασφαλείας και ο μόνος τρόπος για να αποσπαστούν πληροφορίες που θα ταυτίσουν αριθμό κινητού τηλεφώνου με συγκεκριμένες απαντήσεις στην έρευνα είναι αν πέσουν θύματα ηλεκτρονικής πειρατείας (hacking). Ωστόσο, αν προκύψει οποιοδήποτε πρόβλημα σε έναν από τους δύο servers ενημερώνεται αυτόματα ο άλλος και λαμβάνονται ισχυρότερα μέτρα ασφαλείας. Επιπλέον, η έρευνα ψηφοφόρων πληροί τις προϋποθέσεις του ισχύοντος Κώδικα Δεοντολογίας του ΑΠΘ όπως φαίνεται στο Παράρτημα Β.

Επιπλέον, όλοι οι συμμετέχοντες στην έρευνα ψηφοφόρων ενημερώνονται σχετικά με τους σκοπούς της έρευνας, καθώς αποστέλλεται σε όλους ο σχετικός σύνδεσμος (<http://epolls.gr/elnes2019>). Μέσα από την πλοήγηση στη συγκεκριμένη ιστοσελίδα μπορεί κανείς να βρει πληροφορίες σχετικά με την πολιτική προστασίας προσωπικών δεδομένων, τα δικαιώματα των συμμετεχόντων και τον τρόπο άσκησής τους, όπως και τις πληροφορίες επικοινωνίας του

υπευθύνου επεξεργασίας των δεδομένων και επιστημονικά υπευθύνου του έργου (<http://epolls.gr/node/29>).

Όσον αφορά στη διασφάλιση της ανωνυμίας στην έρευνα που αφορά τους υποψήφιους βουλευτές καθώς και στα δεδομένα από τα μέσα κοινωνικής δικτύωσης που αφορούν δημοσιεύσεις (posts) υποψηφίων βουλευτών στο Twitter ή/και το Facebook, ισχύουν όσα αναφέρονται στο άρθρο 28 του Νόμου υπ' αριθμ. 4624/2019 σχετικά με τον συμβιβασμό του δικαιώματος στην προστασία των δεδομένων προσωπικού χαρακτήρα με το δικαίωμα στην ελευθερία της έκφρασης και πληροφόρησης, συμπεριλαμβανομένης της επεξεργασίας για δημοσιογραφικούς σκοπούς και για σκοπούς ακαδημαϊκής, καλλιτεχνικής ή λογοτεχνικής έκφρασης. Πιο συγκεκριμένα, σύμφωνα με αυτόν τον νόμο η επεξεργασία δεδομένων προσωπικού χαρακτήρα επιτρέπεται όταν υπερέχει το δικαίωμα στην ελευθερία της έκφρασης και το δικαίωμα της πληροφόρησης έναντι του δικαιώματος προστασίας των δεδομένων προσωπικού χαρακτήρα του υποκειμένου, ιδίως για θέματα γενικότερου ενδιαφέροντος ή όταν αφορά δεδομένα προσωπικού χαρακτήρα δημοσίων προσώπων. Ωστόσο, σε περιπτώσεις που είναι εύκολη η ταυτοποίηση του υποψηφίου βουλευτή είναι δυνατή η απόκρυψη απαντήσεων που θα βοηθούσαν προς την ταυτοποίησή του. Τα δεδομένα της συγκεκριμένης έρευνας θα διατεθούν στο FORS (μία αποθήκη δεδομένων) η οποία διασφαλίζει την ανωνυμία και τη σωστή χρήση τους από τους ενδιαφερόμενους (βλ. Παράρτημα Α, Συμφωνητικό FORS).

Όσον αφορά στα δεδομένα από τα μέσα κοινωνικής δικτύωσης, πρόκειται για δημοσιεύσεις (posts) και πληροφορίες σχετικά με τις δημοσιεύσεις (π.χ. όνομα χρήστη, αριθμός likes, retweets κ.λπ.) που είναι ανοιχτές στο ευρύ κοινό και άρα δεν μπορούν να θεωρηθούν προσωπικά δεδομένα εφόσον δημοσιεύονται από τον ίδιο τον υποψήφιο βουλευτή.

Τέλος, όσον αφορά στα δεδομένα των σώματων κειμένου που δημιουργούνται από τα άρθρα του Τύπου, αλλά και τα δεδομένα της έρευνας εμπειρογνομόνων δεν αποτελούν δεδομένα προσωπικού χαρακτήρα.

Σχετικά με τα δεδομένα που προέρχονται από τον Τύπο δεν τίθεται θέμα παραβίασης πνευματικών δικαιωμάτων. Ο Οργανισμός Συλλογικής Διαχείρισης Έργων του Λόγου (ΟΣΔΕΛ), ο οποίος έχει ως σκοπό τη συλλογική διαχείριση και προστασία των δικαιωμάτων πνευματικής ιδιοκτησίας στα έργα του λόγου, εκπροσωπώντας τη συντριπτική πλειοψηφία των εκδοτών Τύπου, έχει αδειοδοτήσει όσες εταιρείες αποδελτίωσης έχουν καταβάλει σε αυτόν το υποχρεωτικό τέλος, να ψηφιοποιούν και να επεξεργάζονται το περιεχόμενο των εντύπων. Από την εταιρεία αποδελτίωσης θα παραλάβουμε το σύνολο των δημοσιευμάτων καθώς και το αρχείο που θα αποτελέσει την πηγή του corpus που θα αναλύσουμε. Καθώς δεν μας ενδιαφέρει ένα ένα ξεχωριστά κάθε άρθρο, ούτε θα προβούμε σε αναπαραγωγή ή διαμοιρασμό ως τέτοιο (ως άρθρο) αλλά μόνο οι μεμονωμένες λέξεις που συναποτελούν το corpus και μικρές φράσεις ή αποσπάσματα (συμφραστικοί πίνακες) δεν απαιτείται περαιτέρω άδεια. Σε περίπτωση που κριθεί απαραίτητο το θέμα των πνευματικών δικαιωμάτων θα αντιμετωπιστεί μέσω της πλατφόρμας φιλοξενίας του corpus CLARIN:EL (<https://www.clarin.gr/>).

## ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ ΔΙΑΤΗΡΗΣΗ ΔΕΔΟΜΕΝΩΝ

Όλα τα δεδομένα του παρόντος έργου θα αποθηκευτούν σε κρυπτογραφημένη μορφή στον διακομιστή (server) του επιστημονικού υπευθύνου και θα διατηρηθούν για 7 χρόνια. Παράλληλα, τα δεδομένα θα δημοσιευθούν στην ιστοσελίδα του έργου (<http://datapopeu.gr/>) όπου θα παραμείνουν διαθέσιμα για το ίδιο χρονικό διάστημα. Τα δεδομένα των ερευνών (Υποψηφίων Βουλευτών και Ψηφοφόρων 2019) θα αποθηκευτούν σε διαδικτυακά καταθετήρια δεδομένων ανοικτής πρόσβασης όπως είναι λόγου χάρη το FORS.

Επιπλέον, τα σώματα κειμένου θα αποθηκευτούν στην Υποδομή CLARIN:EL, το ελληνικό σκέλος της Ευρωπαϊκής Υποδομής CLARIN ERIC, που υλοποιείται από ένα πανευρωπαϊκό δίκτυο



πανεπιστημίων και ερευνητικών κέντρων, στο οποίο συμμετέχουν φορείς με ενεργό ερευνητική δράση και αναπτυξιακή δραστηριότητα στο χώρο των Γλωσσικών Πόρων και της Γλωσσικής Τεχνολογίας. Το CLARIN:EL είναι η εθνική Υποδομή Γλωσσικών πόρων και Τεχνολογιών στην Ελλάδα. Μέσω της Ερευνητικής Υποδομής CLARIN:EL συγκεντρώνονται, τεκμηριώνονται, συντηρούνται και διαμοιράζονται ψηφιακοί γλωσσικοί πόροι, εργαλεία γλωσσικής τεχνολογίας καθώς και πιστοποιημένες διαδικτυακές υπηρεσίες γλωσσικής επεξεργασίας.

## ΔΙΑΜΟΙΡΑΣΜΟΣ ΔΕΔΟΜΕΝΩΝ

Όλα τα δεδομένα του DataPopEU είναι δεδομένα ανοικτής πρόσβασης. Τα δεδομένα των διαδικτυακών ερευνών και ιδίως της έρευνας ψηφοφόρων και της έρευνας υποψηφίων βουλευτών θα διατεθούν σε καταθετήρια δεδομένων (π.χ. FORS: <https://forscenter.ch/>). Πρόκειται για πλατφόρμες ανοικτής πρόσβασης όπου, ωστόσο, χρειάζεται εγγραφή προκειμένου ένας επιστήμονας να κατεβάσει τα δεδομένα. Συγκεκριμένα, όσον αφορά στο FORS, ο ενδιαφερόμενος για τη λήψη των δεδομένων καλείται να αποδεχθεί τους όρους που θέτει η πλατφόρμα (το πλήρες συμφωνητικό παρέχεται στο Παράρτημα Α). Ο χρόνος παραμονής των δεδομένων στα καταθετήρια δεδομένων δεν μπορεί να προβλεφθεί καθώς οι συγκεκριμένες πλατφόρμες δεν ορίζουν συγκεκριμένο χρόνο διατήρησης των δεδομένων, αλλά ο στόχος τους είναι η διατήρηση των επιστημονικών δεδομένων το σύνολο του χρονικού διαστήματος που αυτές οι υπηρεσίες είναι εν λειτουργία. Όσον αφορά στα δεδομένα του Τύπου (σώματα κειμένου που προέρχονται από επεξεργασία άρθρων του Τύπου), θα διατεθούν στο κοινό μέσω του CLARIN:EL (<https://www.clarin.gr/>) και θα διατηρηθούν στη συγκεκριμένη υποδομή

Οι προαναφερθείσες «αποθήκες» δεδομένων χρησιμοποιούνται από πλήθος κοινωνικών επιστημόνων και είναι αρκετά εύχρηστες. Τα δεδομένα του έργου θα είναι εύκολο να βρεθούν χρησιμοποιώντας λέξεις κλειδιά όπως «populism», «Euroscpticism», «candidates», «voters» κ.λπ.

## ΠΑΡΑΡΤΗΜΑΤΑ

**Παράρτημα Α: Συμφωνητικό στην ιστοσελίδα του FORS που καλούνται να αποδεχτούν οι ερευνητές προκειμένου να έχουν πρόσβαση στα δεδομένα της έρευνας Υποψηφίων Βουλευτών 2019**



### User contract

I agree to the following:

1. to use the data cited above and related instruments only for the scientific research and/or academic teaching outlined in my description in the present contract, and for no other purpose;
2. to use the data with respect to Swiss federal law and the applicable standard norms of data protection;
3. not to undertake to attribute data to specific individuals; and to communicate my results only in a way that does not identify individual cases (i.e., persons, households, institutions, etc.);
4. to respect above all individual confidentiality, and not to divulge any information that might compromise this confidentiality;
5. to use the data in a conscientious and informed manner, notably by consulting the related documentation, and to respect scientific ethical rules of conduct;
6. to cite used data and documents in conformity with scientific standards (using the citation above);
7. to inform FORS of all publications based on these data;
8. not to transmit these data to third parties, whether in original or modified form;
9. to store the data in a way such that no third party can gain access to them;
10. to destroy the data at the latest upon expiry of the present contract, and to confirm this to FORS.

## Παράρτημα Β: Βεβαίωση τήρησης Κώδικα Δεοντολογίας του ΑΠΘ για την Εθνική Μετεκλογική Έρευνα Ψηφοφόρων 2019



ΕΙΔΙΚΟΣ ΛΟΓΑΡΙΑΣΜΟΣ ΚΟΝΔΥΛΙΩΝ ΕΡΕΥΝΑΣ  
ΑΡΙΣΤΟΤΕΛΕΙΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΘΕΣΣΑΛΟΝΙΚΗΣ

### ΕΠΙΤΡΟΠΗ ΗΘΙΚΗΣ & ΔΕΟΝΤΟΛΟΓΙΑΣ ΤΗΣ ΕΡΕΥΝΑΣ ΑΠΘ

Διεύθυνση: ΜΟΔΥ – ΕΛΚΕ ΑΠΘ  
Πληροφορίες: Νίκη Λώλου  
Τηλ: 2310.994051  
Φαξ: 2310.200392  
e-mail: ethics@rc.auth.gr  
Αρθ. Φακέλου: 92727

Θεσσαλονίκη, 28/11/2019  
Αρ. Πρωτ.: 242550/2019

**Προς κάθε ενδιαφερόμενο**

**Θέμα:** Βεβαίωση τήρησης Κώδικα Δεοντολογίας του ΑΠΘ

Σας πληροφορούμε ότι η Επιτροπή Ηθικής και Δεοντολογίας της Έρευνας ΑΠΘ, στη Συνεδρίασή της, 08/28-11-2019, ενέκρινε το υπ' αρ. πρωτ. 195449/2019 αίτημα του κ. Ιωάννη Ανδρέαδη, Αναπληρωτή Καθηγητή του Τμήματος Πολιτικών Επιστημών ΑΠΘ, για την έγκριση διεξαγωγής της έρευνας με τίτλο «Εθνική Μετεκλογική Έρευνα Ψηφοφόρων 2019 με χρήση διαδικτυακού ερωτηματολογίου», καθώς πληροί τις προϋποθέσεις του ισχύοντος Κώδικα Δεοντολογίας του ΑΠΘ.

Με εκτίμηση  
Ο Πρόεδρος της Επιτροπής Ηθικής και  
Δεοντολογίας Έρευνας



Φίλιππος Αραβανόπουλος  
Καθηγητής Τμήματος Δασολογίας & Φυσικού  
Περιβάλλοντος

3ης Σεπτεμβρίου – Πανεπιστημιούπολη  
546 36 Θεσσαλονίκη  
Τ: 2310 996745 F: 2310 200392  
E: research@rc.auth.gr  
www.rc.auth.gr

