

## FAIR ΔΕΔΟΜΕΝΑ

### Δυνατότητες και προκλήσεις για ερευνητικές υποδομές και ερευνητικές κοινότητες

Δήμητρα Κονδύλη  
Νικόλαος Κληρονόμος

#### Περίληψη

*Η ανάδυση των διαφορετικών και ποικίλων δεδομένων, η πολυπλοκότητα των υπηρεσιών και οι αρχές της Ανοικτής Επιστήμης επηρέασαν τις μεθόδους και τα εργαλεία διαχείρισης και διάθεσης των ερευνητικών δεδομένων. Επηρέασαν επίσης τα περιβάλλοντα διάθεσης (δηλαδή τα Αποθετήρια και τις Ερευνητικές Υποδομές), ενώ εισήγαγαν τις έννοιες της εμπιστοσύνης (trust) αλλά και του FAIR, ακρωνύμιο με συνειρμική αναφορά στο επίθετο δίκαιος. Η ελεύθερη, γρήγορη, μαζική παραγωγή και διακίνηση δεδομένων επιβάλλει κανόνες και τεχνικές που προωθούν στους κόλπους της Ανοικτής Επιστήμης περισσότερο αξιόπιστα δεδομένα και υπηρεσίες, τουλάχιστον ως προς την τεχνική τους αρτιότητα. Θα διερευνήσουμε μέσα από μια σύντομη ιστορική αναδρομή αυτές τις εξελίξεις, τους περιορισμούς και τις μελλοντικές προκλήσεις, έχοντας ως μελέτη περίπτωσης την ευρωπαϊκή υποδομή CESSDA ERIC και την αντίστοιχη ελληνική SoDaNet, καθώς και τις προοπτικές που μπορεί να διανοίξει, όπως και τα διλήμματα που μπορεί να προκαλέσει, η υιοθέτηση των FAIR δεδομένων στις ερευνητικές υποδομές των κοινωνικών επιστημών.*

Λέξεις-κλειδιά: FAIR δεδομένα, Ανοικτή Επιστήμη, Ερευνητικές Υποδομές, Dataverse.

#### Abstract

*The emergence of various and diverse data, the complexity of services along with the principles of Open Science have influenced the methods and tools for managing research data and making it available. They have also influenced the environments*

*for its disposal (i.e. Repositories and/or Research Infrastructures), and introduced the concepts of trust and FAIR, an acronym with connotations of fairness and justice. The free, fast, mass production and circulation of data imposes rules and techniques that promote, within Open Science, more reliable data and services, at least in terms of their technical excellence, thus rendering FAIR data a sine qua non precondition. We will explore, through a brief historical review, the developments, limitations and future challenges involved, taking as a case study the European infrastructure CESSDA ERIC and its Greek member SoDaNet, as well as the prospects and the dilemmas that the adoption of fair data in social science research infrastructures may bring about.*

Keywords: *FAIR data, Open Science, Research Infrastructures, Dataverse.*

## Εισαγωγή

*Σκοπός της επιστήμης δεν είναι να ανοίξει την πόρτα στην απέραντη σοφία, αλλά να θέσει όρια στο απέραντο σφάλμα.*

Μπέρτολτ Μπρεχτ, *Η ζωή του Γαλιλαίου*, 1945

Οι τεχνολογικές και θεσμικές εξελίξεις των τελευταίων δεκαετιών επηρέασαν σε μεγάλο βαθμό την ερευνητική παραγωγή. Η αυξανόμενη τον τελευταίο καιρό επικράτηση των αρχών της ανοικτής πρόσβασης επιτάχυνε τις διαδικασίες κυκλοφορίας και πρόσβασης στα ερευνητικά πορίσματα και δεδομένα (Berlin, 2010). Σε πρόσφατο άρθρο του περιοδικού *Nature*, τα δεδομένα αναφέρονται ως το «άλφα και το ωμέγα της επιστημονικής και κοινωνικής έρευνας, πολύπλευρα και πολύτροπα αγαθά που υφίστανται ως πρωτογενές υλικό παραγωγής γνώσης και, μετά την επεξεργασία και ερμηνεία τους από τους ειδικούς, ως τελικό προϊόν της ερευνητικής άσκησης<sup>1</sup>». Η τελευταία μάλιστα διετία της τρέχουσας πανδημίας απέδειξε, πέραν πάσης αμφιβολίας, ότι η αντιμετώπιση φαινομένων που διαρρηγνύουν ή απειλούν την κανονικότητα χρήζει συλλογικών ενεργειών, οι οποίες τεκμηριώνονται στη βάση της έρευνας και της επιστήμης και τίθενται στη διάθεση όλων των κοινωνικών δρώντων, από τους πολιτειακούς θεσμούς που φέρουν το βάρος της υλοποίησης έως τον κάθε πολίτη που επιθυμεί να ενημερωθεί έγκαιρα ώστε να ανταποκριθεί στην πρωτόγνωρη σημερινή κατάσταση πραγμάτων. Η ανάγκη για αξιόπιστα δεδομένα και πληροφορίες ως αποτέλεσμα δημόσια χρηματοδοτούμενης έρευνας προς όφελος της κοινωνίας (OECD

1. «Data are the alpha and omega of scientific and social research. A versatile good, they exist both as raw material for producing knowledge and, when processed and interpreted with an expert eye, the end product of the exercise». Nature Publishing Group (13 Μαρτίου, 2018). Everyone needs a data-management plan. Nature News. Πρόσβαση στις 10 Ιανουαρίου, 2022, από <https://www.nature.com/articles/d41586-018-03065-z> Η μετάφραση είναι των συγγραφέων.

2007, 2017) και η εδραίωση των αρχών της Ανοικτής Επιστήμης μέσω της παραγωγής επιστημονικών δημοσιεύσεων, δεδομένων και μεταδεδομένων υποστηρίζουν τις συνέργειες μεταξύ θεσμών και κοινωνίας για την επίτευξη των παραπάνω στόχων. Στο σημείο αυτό θεωρούμε σκόπιμο να εισαγάγουμε τις βασικές/λειτουργικές έννοιες βάσει των οποίων αρθρώνεται το παρόν κείμενο.

Ως Ανοικτή Επιστήμη ορίζεται ένα νέο πρότυπο πρακτικών, μέσων και συνεργασίας για την παραγωγή και τη διάθεση επιστημονικού έργου και αποτελεσμάτων της έρευνας, με άμεσο επιστημονικό, οικονομικό και κοινωνικό αντίκτυπο, το οποίο αποτελεί κρίσιμη προτεραιότητα της Ευρωπαϊκής Ένωσης για τη βιώσιμη ανάπτυξη, παραγωγή και αξιοποίηση ερευνητικών αποτελεσμάτων και τη συνεργασία μεταξύ των επιστημονικών φορέων και των πολιτειακών θεσμών με απώτερο στόχο τη διάχυση και τον εκδημοκρατισμό της γνώσης (Εθνικό Σχέδιο για την Ανοικτή Επιστήμη, 2020). Δηλαδή η παραγωγή γνώσης μέσω της διάθεσης και αξιοποίησης δημόσιας χρηματοδότησης προϋποθέτει, ή τουλάχιστον θα έπρεπε να προϋποθέτει, δημόσια ευθύνη και λογοδοσία. Η συγκυρία της πανδημίας του Covid-19, καθώς και η ανταπόκριση της ερευνητικής/ακαδημαϊκής κοινότητας μέσω της διάθεσης ερευνητικών ανοικτών δεδομένων, ενίσχυσε την εδραίωση της Ανοικτής Επιστήμης και τη διάχυση και κυκλοφορία τεκμηριωμένων πληροφοριών.

Ως προς τη διαδικασία της τεκμηρίωσης και διαχείρισης, αναφερόμαστε σε δεδομένα και άλλους πόρους που έχουν σχεδιαστεί και παραχθεί βάσει συγκεκριμένων προτύπων, έχουν αξιολογηθεί από την ίδια την επιστημονική κοινότητα, υπάρχει δηλαδή συναίνεση όσον αφορά την εγκυρότητά τους. Μεταξύ άλλων, η Ανοικτή Επιστήμη προάγει την ερευνητική αριστεία –καθώς διαθέτει την παραγόμενη επιστημονική γνώση και τους διαμεσολαβητές αυτής, την ερευνητική κοινότητα– στην κοινωνία και την πολιτεία, επιτυγχάνοντας τα ακόλουθα: εμπιστοσύνη των πολιτών στην επιστημονική γνώση, επιστημονικά τεκμηριωμένο σχεδιασμό και εφαρμογή πολιτικών που άπτονται τόσο της λειτουργίας του κράτους όσο και της ζωής των πολιτών. Κατ' αυτό τον τρόπο προωθείται η εξωστρέφεια των ερευνητικών και επιστημονικών οργανισμών και βελτιώνεται η αποτελεσματικότητα των χρηματοδοτικών μηχανισμών, καθώς κεφαλαιοποιείται ο κοινωνικός τους αντίκτυπος. Σήμερα οι αρχές της Ανοικτής Επιστήμης, σε συνέχεια της υιοθέτησής τους από το Ευρωπαϊκό Συμβούλιο από κοινού με την αναγκαιότητα επανάχρησης των ερευνητικών δεδομένων, έχουν επίσης ενσωματωθεί στις εθνικές πολιτικές μεγάλου αριθμού κρατών της Ευρωπαϊκής Ένωσης, συνδιαμορφώνοντας και αλληλοεπιδρώντας στην ανάπτυξη και τη λειτουργία του ευρωπαϊκού οικοσυστήματος δεδομένων.

Στο πλαίσιο αυτής της εργασίας θα επιχειρηθεί να αποσαφηνιστεί η έννοια και σημασία των FAIR<sup>2</sup> δεδομένων και η σύνδεσή τους με τις αρχές της Ανοικτής

---

2. FAIR: findable, accessible, interoperable, reusable (εξερευνήσιμα, προσβάσιμα, διαλειτουργικά, επανα-χρησιμοποιήσιμα).

Επιστήμης, να αναδειχθούν οι επιμέρους ιδιότητες και τα προαπαιτούμενα συγκρότησης της έννοιας του FAIR, όπως και η σύνδεση με ερευνητικά περιβάλλοντα που προάγουν τα FAIR, η συζήτηση για τις δεξιότητες και η εξειδίκευση των επιστημόνων που εμπλέκονται στη διαχείρισή τους, καθώς και οι προοπτικές και τα διλήμματα από την επίδρασή τους στην παραγωγή δεδομένων ποσοτικών ερευνών ή ερευνών προκαθορισμένου σχεδίου (Κάλλας, 2008: 32)<sup>3</sup> και σε νέους τύπους δεδομένων που ανέκυσαν με την ευρεία χρήση των τεχνολογιών πληροφορίας και επικοινωνιών.

### **FAIR: ένα ακρωνύμιο που μεταβάλλει το οικοσύστημα των ερευνητικών δεδομένων**

Τα τελευταία τριάντα χρόνια παρατηρείται αναμφισβήτητα ένας «καταιγισμός δεδομένων και πληροφοριών», αποκαλούμενος και «επανάσταση των δεδομένων»<sup>4</sup>, που αυξάνει με γεωμετρική πρόοδο τόσο το ποσοστό όσο και τον όγκο των ερευνητικών δεδομένων τα οποία διατίθενται δυνητικά προς επανάχρηση. Η αύξηση σε όλο το φάσμα των επιστημών δημιούργησε ευκαιρίες αλλά και προκλήσεις ως προς την αξιοποίησή τους χάρη στην παράλληλη τεχνολογική και μεθοδολογική εξέλιξη. Έγινε επίσης αντιληπτό ότι η αξιοποίησή τους προϋποθέτει κοινά σχήματα, τεκμηρίωση και εργαλεία, εκ παραλλήλου με τη διαμόρφωση στρατηγικής σε επίπεδο θεσμικών φορέων που θα προάγουν αυτές τις προσπάθειες. Ως εκ τούτου, οι πρώτοι προβληματισμοί σχετικά με την αξιοποίηση και την επανάχρηση των ερευνητικών δεδομένων διατυπώθηκαν από τον OECD το 2007 σε ένα κείμενο που επηρέασε την πρόσβαση στα ερευνητικά δεδομένα δημόσια χρηματοδοτούμενης έρευνας, το οποίο το 2013 διαμορφώθηκε ως Αρχές και Συστάσεις από την Ευρωπαϊκή Επιτροπή κατά τη διαμόρφωση του χρηματοδοτικού πλαισίου Horizon 2014-2020<sup>5</sup>.

Η απαρχή της επικράτησης των αρχών FAIR τοποθετείται σε μια συνάντηση τον Ιανουάριο του 2014 στο Lorentz Center, με τη συμμετοχή διαφορετικών δρώντων, που συμπεριελάμβαναν εκπροσώπους της ερευνητικής και πανεπιστημιακής κοινότητας, επιστήμονες των δεδομένων, εταιρείες και οργανισμούς του δημόσιου και ιδιωτικού τομέα, προκειμένου να αναστοχαστούν σχετικά με τον εντοπισμό

---

3. Στην εμπειρική κοινωνική έρευνα, οι έρευνες προκαθορισμένου σχεδίου βασίζονται σε μεθόδους ποσοτικής ανάλυσης και στη στατιστική οντολογία για την παραγωγή ενός συνόλου δεδομένων που αποτελεί την προνομαϊκή μορφή τεκμηρίου της ποσοτικής έρευνας.

4. Data revolution.

5. European Commission (2020). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European strategy for data*. COM (2020) 66 final. Brussels: EURLex, 19.02.2020. (Αποθήκευση: 28 Αυγούστου 2020)

και την επανάχρηση των ερευνητικών δεδομένων και την περαιτέρω ενδυνάμωση του σύνθετου πλέγματος *ερευνητική κοινότητα - επιστήμονες των δεδομένων - παραγωγοί και θεσμικοί χρήστες δημόσιου και ιδιωτικού τομέα*<sup>6</sup>. Το λογοπαίγνιο της Ομάδας FORCE 11 άντλησε έμπνευση από τις έννοιες της δικαιοσύνης και της ίσης πρόσβασης στη γνώση, με την έννοια ότι τα βασικά χαρακτηριστικά των δεδομένων που είναι FAIR αποσκοπούν στην ευρύτερη διάθεση, διάδοση και χρήση τους και εν τέλει σε μια σχετική ισονομία ως προς την προώθηση της πληροφόρησης και γνώσης μεταξύ των πολιτών.

Ως FAIR ορίζεται ένα σύστημα βασικών κατευθυντήριων κανόνων και αρχών που θα πρέπει να διέπουν τα δεδομένα και που διασφαλίζουν ότι οι διάφοροι ερευνητικοί πόροι μπορούν να επαναχρησιμοποιηθούν και να είναι όσο το δυνατόν πιο χρήσιμοι στην ερευνητική και την ευρύτερη κοινότητα χρηστών, παρέχοντας ένα συνεχές σύστημα αυξανόμενης επανάχρησης των ερευνητικών δεδομένων (Mons et al., 2017). Με βάση τα παραπάνω, πρόθεσή μας στη συγγραφή του παρόντος κειμένου δεν είναι η έμφαση στα τεχνικά προαπαιτούμενα που θα πρέπει να πληρούνται, αλλά κυρίως η διερεύνηση και, ελπίζουμε, η ανάδειξη της «φιλοσοφίας» τους και μιας κουλτούρας τελικού διαμοιρασμού και κοινοχρησίας που θα διέπει τα διάφορα στάδια της ερευνητικής διαδικασίας. Η δική μας προσέγγιση των FAIR data είναι συμπεριληπτική, αφορά τόσο τον σχεδιασμό μιας έρευνας όσο και τη διαχείριση των δεδομένων εντός των Αποθετηρίων.

Ως FAIR ορίζονται τα δεδομένα που πληρούν τις αρχές της Ευρεσιμότητας (Findability), της Προσβασιμότητας (Accessibility), της Διαλειτουργικότητας (Interoperability) και της Επαναχρησιμότητας (Reusability), όπως ορίστηκαν από τους δημιουργούς του ακρωνυμίου σε άρθρο τους το 2016 στο περιοδικό Scientific Data (Wilkinson et al., 2016). Στο ίδιο άρθρο ορίζονται πιο αναλυτικά οι κατευθυντήριες αρχές των FAIR για το κάθε επιμέρους στοιχείο:

Κατευθυντήριες αρχές των FAIR δεδομένων	
Να είναι ευρέσιμα (findable):	
F1.	Στα δεδομένα ή/και μεταδεδομένα αποδίδεται ένα παγκοσμίως μοναδικό και μόνιμο αναγνωριστικό, που λειτουργεί ως μόνιμος ταυτοποιητής (Persistent Identifier), π.χ. DOI, Handle, ORCID κ.ά.
F2.	Τα δεδομένα περιγράφονται με πλούσια και πλήρη τεκμηρίωση μεταδεδομένων, όπως ορίζεται από το R1 παρακάτω.

6. Οι διαφορετικοί εταίροι ανταποκρίθηκαν στο κάλεσμα του Ολλανδικού Κέντρου eScience Center καθώς και του Ολλανδικού Techcentre για τις Επιστήμες της Ζωής (Dutch Techcentre for the Life Sciences (DTL) στους χώρους του Κέντρου Lorentz στην πόλη Λέιντεν της Ολλανδίας. Η συγκεκριμένη Ομάδα εμπειρογνομόνων και θεσμών έγινε γνωστή ως FORCE11. Περισσότερα στο <https://www.force11.org/group/fairgroup>

Κατευθυντήριες αρχές των FAIR δεδομένων	
F3.	Τα μεταδεδομένα περιλαμβάνουν σαφώς και ρητά το μοναδικό αναγνωριστικό (PID) των δεδομένων που περιγράφουν.
F4.	Τα δεδομένα ή/και μεταδεδομένα είναι καταχωρημένα και μπορούν να αναζητηθούν, π.χ. μέσω των Αποθετηρίων δεδομένων.
Να είναι προσβάσιμα (accessible):	
A1.	Τα δεδομένα ή/και μεταδεδομένα είναι ανακτήσιμα με βάση τον ταυτοποιητή τους (PID), χρησιμοποιώντας ένα τυποποιημένο πρωτόκολλο επικοινωνίας.
A1.1.	Το προαναφερθέν πρωτόκολλο είναι ανοιχτό, ελεύθερο και καθολικά εφαρμόσιμο από όλους τους χρήστες χωρίς περιορισμούς.
A1.2.	Το προαναφερθέν πρωτόκολλο επιτρέπει μια διαδικασία πιστοποίησης και εξουσιοδότησης, όπου είναι απαραίτητο.
A2.	Τα μεταδεδομένα είναι διαθέσιμα ακόμα και στην περίπτωση όπου τα δεδομένα δεν είναι πλέον προσβάσιμα.
Να είναι διαλειτουργικά (interoperable):	
I1.	Τα μεταδεδομένα ή/και τα δεδομένα είναι διαθέσιμα σε μια επίσημη, προσβάσιμη, κοινή και ευρέως χρησιμοποιούμενη γλώσσα.
I2.	Τα μεταδεδομένα ή/και τα δεδομένα χρησιμοποιούν λεξικά (vocabularies) που ακολουθούν τις αρχές FAIR.
I3.	Τα δεδομένα ή/και τα μεταδεδομένα περιλαμβάνουν εξειδικευμένες αναφορές σε άλλα δεδομένα ή/και μεταδεδομένα.
Να είναι επαναχρησιμοποιήσιμα (reusable):	
R1.	Τα δεδομένα ή/και μεταδεδομένα περιγράφονται επαρκώς παρέχοντας τα απαραίτητα έγκυρα και σχετικά στοιχεία.
R1.1.	Τα δεδομένα ή/και μεταδεδομένα δημοσιεύονται με σαφώς προσδιορισμένη και επιτρεπτής πρόσβασης άδεια χρήσης δεδομένων (data usage licence).
R1.2.	Τα δεδομένα ή/και μεταδεδομένα συνδέονται με λεπτομερή στοιχεία της προέλευσής τους.
R1.3.	Τα δεδομένα ή/και μεταδεδομένα πληρούν τα σχετικά με τον επιστημονικό τομέα πρότυπα.

Στις επόμενες ενότητες θα επιχειρηθεί η περαιτέρω ανάλυση των παραπάνω αρχών.

### *Εξερενήσιμα (findable) δεδομένα και μεταδεδομένα*

Με την επέλαση του διαδικτύου τις προηγούμενες τρεις δεκαετίες, υπήρξε ταχύτατη μεταβολή στην ανθρώπινη εμπειρία. Ολόκληροι κλάδοι προϊόντων και υπηρεσιών μετοίκησαν στα αναδυόμενα ψηφιακά περιβάλλοντα και οι διαστάσεις του χώρου και του χρόνου σχετικοποιήθηκαν. Η επιστημονική έρευνα έγινε πιο ανοικτή, δίνο-

ντας την δυνατότητα ευρύτερης διάχυσης των ερευνητικών δεδομένων σε κάθε ενδιαφερόμενο/η χρήστη μέσω του διαδικτύου. Ταυτόχρονα όμως νέες ανάγκες και προβληματισμοί ανέκυψαν έναντι αυτού του ταχέως μεταβαλλόμενου περιβάλλοντος. Μια πρόκληση αυτού του χαρακτήρα αποτέλεσε η «μεταβλητότητα» των ψηφιακών διευθύνσεων σε σχέση με τις φυσικές διευθύνσεις, γεγονός που, όπως είναι εύλογο, δημιούργησε προβληματισμό στην επιστημονική κοινότητα, καθώς γι' αυτήν είναι αναγκαία η δυνατότητα παραπομπής στα δεδομένα από τα οποία εν τέλει προέρχεται η ίδια η επιστημονική γνώση. Προς αυτή την κατεύθυνση τείνει η χρήση μόνιμων και μοναδικών ταυτοποιητών (PIDs) τόσο σε επίπεδο δεδομένων και δημοσιεύσεων, όπως οι υπηρεσίες που προσφέρουν το DOI<sup>7</sup> (Digital Object Identifier) ή το Handle<sup>8</sup>, όσο και σε επίπεδο μεταδεδομένων, όπως είναι οι ποικίλες φυσικές ή μη οντότητες, οι επιστήμονες που συμμετέχουν στην παραγωγή μιας δημοσίευσης ή ενός ερευνητικού έργου με υπηρεσίες όπως αυτές που προσφέρει το ORCID<sup>9</sup>. Η λογική πάνω στην οποία εδράζεται η παραπάνω λειτουργία είναι αρκετά απλή και δημιουργεί επί της ουσίας μια βάση δεδομένων με μοναδικό και μη επαναλήψιμο αναγνωριστικό (ID) για το εκάστοτε αντικείμενο, είτε είναι δημοσίευση, όπως στην περίπτωση του DOI, είτε είναι άτομο, όπως στην περίπτωση του ORCID.

Παράλληλα, υπάρχει η ανάγκη για πλήρη τεκμηρίωση των μεταδεδομένων που περιγράφουν ένα ψηφιακό αντικείμενο. Ο κύριος λόγος έγκειται στο γεγονός ότι, όταν γίνεται αναζήτηση για δεδομένα, ο/η εκάστοτε χρήστης –φυσικό πρόσωπο ή ακόμα και μηχανή– αρχικά θα έρθει σε επαφή με κάποια από τα μεταδεδομένα, καθώς αυτά είναι ευρετηριασμένα (indexed) και αναζητήσιμα μαζί με τα δεδομένα. Καθώς στα μεταδεδομένα περιλαμβάνονται κυρίως τα περιγραφικά μεταδεδομένα μιας έρευνας<sup>10</sup>, είναι εκείνα από τα οποία ο/η εκάστοτε χρήστης θα μπορεί να διαγνώσει τη συνάφεια των αποτελεσμάτων με την αναζήτησή του. Επίσης, θα πρέπει ο μόνιμος και μοναδικός ταυτοποιητής (PID) να αναφέρεται ρητά μέσα στο σύνολο των μεταδεδομένων για έναν επιπλέον λόγο: τα μεταδεδομένα που βρίσκονται ευρετηριασμένα στα διάφορα επιμέρους αποθετήρια και καταλόγους συχνά συλλέγονται αυτοματοποιημένα (harvested) από ευρύτερους καταλόγους, αποκτώντας έτσι μεγαλύτερη ορατότητα. Εντούτοις, κατά τη διαδικασία της συλλογής, πολλές φορές δεν περιλαμβάνονται και τα δεδομένα της έρευνας αλλά μόνο τα μεταδεδομένα, και έτσι οποιοσ/αδήποτε χρήστης ψάχνει σε έναν από τους μεγάλους καταλόγους θα πρέπει να μπορεί να έχει πρόσβαση στα δεδομένα που τον/

---

7. Digital Object Identifier System. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://www.doi.org/>

8. Handle.net registry. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://www.handle.net/>

9. ORCID. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://orcid.org/>

10. Τα μεταδεδομένα δεν περιορίζονται πάντα μόνο στα περιγραφικά δεδομένα, αλλά διακρίνονται σε τέσσερις κατηγορίες μεταδεδομένων: α) ορισμού β) επιχειρησιακής πληροφόρησης γ) ταξινόμησης και δ) ελέγχου (Κάλλας, 2002: 129-130).



την ενδιαφέρουν και ως εκ τούτου αυτή η διαδικασία μπορεί να επιτευχθεί μέσω της αναφοράς του μοναδικού ταυτοποιητή στα μεταδεδομένα. Τέλος, με βάση τα προαναφερθέντα, γίνεται αντιληπτό ότι τόσο τα δεδομένα όσο και τα μεταδεδομένα θα πρέπει να φιλοξενούνται σε αποθετήρια τα οποία να είναι ευρετηριασμένα και να μπορούν να διασυνδεθούν με ευρύτερους καταλόγους για την όσο το δυνατόν μεγαλύτερη ορατότητα και ευρεσιμότητά τους. Συνάγεται λοιπόν ότι ad hoc ψηφιακοί τόποι, όπως είναι ο ιστότοπος ενός ερευνητικού έργου που φιλοξενεί τα ερευνητικά αποτελέσματα του συγκεκριμένου έργου, αφενός δύσκολα θα μπορέσουν να γίνουν ορατοί στην πλειονότητα της ερευνητικής κοινότητας, αφετέρου, μετά τη λήξη του έργου και αν δεν έχει εξασφαλισθεί η συντήρησή του, σε ένα πεπερασμένο χρονικό διάστημα (όχι και τόσο μεγάλο) μπορεί να μη λειτουργεί ο σχετικός ιστότοπος.

### *Προσβάσιμα (accessible) δεδομένα και μεταδεδομένα*

Μια δεύτερη κατευθυντήρια αρχή των δεδομένων FAIR έγκειται στο να είναι προσβάσιμα. Ακολουθώντας τη διαδικασία που περιγράφηκε παραπάνω, εφόσον έχει γίνει η εύρεση είτε των δεδομένων είτε των μεταδεδομένων, θα πρέπει οι εξωτερικοί/ές χρήστες να μπορούν να έχουν πρόσβαση σε αυτά μέσω της χρήσης τυποποιημένων μηχανισμών και πρωτοκόλλων. Πρωτόκολλο είναι ο τεχνικός όρος για το πρότυπο του τρόπου με τον οποίο μεταφέρονται τα δεδομένα μέσω ενός δικτύου. Τυποποιημένοι μηχανισμοί είναι για παράδειγμα τα σχετικά ευρέως χρησιμοποιούμενα πρωτόκολλα στις διαδικτυακές περιηγήσεις, όπως τα πρωτόκολλα TCP/IP<sup>11</sup>, http(s)<sup>12</sup> ή FTP<sup>13</sup>, σε αξιόπιστα αποθετήρια και καταλόγους δεδομένων.

Υπάρχουν όμως αρκετές περιπτώσεις πρόσβασης ενός ατόμου στα δεδομένα ή μεταδεδομένα για τις οποίες απαιτείται η μεσολάβηση πρόσθετων διαδικασιών, όπως η συμπλήρωση συμφώνου εμπιστευτικότητας κ.ά. Ακόμα και αυτές οι περιπτώσεις, όπου φαίνεται σαν τα δεδομένα να είναι «κλειστά» ή να διατίθενται υπό όρους, δεν παρεκκλίνουν από τις αρχές των δεδομένων FAIR εφόσον στα μεταδεδομένα διατυπώνονται με σαφή και ρητή αναφορά οι όροι και οι μηχανισμοί πρόσβασης στα δεδομένα αυτών. Για παράδειγμα, θα μπορούσαν να εμπεριέχουν τηλεφωνικούς αριθμούς, συνδέσμους διαδικτυακών «εντύπων» επικοινωνίας, διευθύνσεις ηλεκτρονικού ταχυδρομείου ή ακόμη και φυσικές διευθύνσεις –με όρους εμπορικούς και όχι προστασίας– μαζί με τις διαδικασίες και τις προϋποθέσεις που θα πρέπει να πληρούνται για να παρέχεται πρόσβαση στα δεδομένα. Τα πρωτόκολλα που αναφέρθηκαν παραπάνω θα πρέπει να είναι ανοικτά, ελεύθερα και κα-

11. Transmission Control Protocol/Internet Protocol.

12. Hypertext Transfer Protocol και Hypertext Transfer Protocol Secure.

13. File Transfer Protocol.



θολικά εφαρμόσιμα υπό την έννοια ότι κάθε άτομο μπορεί να τα χρησιμοποιήσει χωρίς να χρειάζεται να καταβάλει κάποιου είδους συνδρομή ή τέλος χρήσης. Η επιλογή περιοριστικών στη χρήση πρωτοκόλλων δημιουργεί προσκόμματα σε κάθε εξωτερικό/ή χρήστη καθιστώντας έτσι δύσκολη, ή και αδύνατη, την εύρεση και επανάχρηση των δημοσιευμένων δεδομένων.

Παράλληλα, θα πρέπει τόσο τα προαναφερθέντα πρωτόκολλα, όσο και η δομή του Αποθετηρίου ή καταλόγου όπου φιλοξενούνται τα μεταδεδομένα και δεδομένα να υποστηρίζουν διαδικασίες ταυτοποίησης και εξουσιοδότησης εφόσον, εννοείται, κρίνονται απαραίτητες. Θα μπορούσε, για παράδειγμα, ένα αποθετήριο δεδομένων να διαθέτει μια πολιτική «ανοικτών» μεταδεδομένων για όλους, χωρίς ταυτοποίηση και διαδικασίες εισόδου και εξουσιοδότησης πρόσβασης σε οποιον/αδήποτε εξωτερικό/ή χρήστη, ενώ ταυτόχρονα να υπάρχει απαίτηση αυτών για την πρόσβαση στα δεδομένα. Στην περίπτωση όπου, κατά την πρόσβαση στα δεδομένα ενός αποθετηρίου, οι διαδικασίες είναι μη αυτοματοποιημένες και ζητείται η ρητή συναίνεση του κατόχου των δεδομένων σχετικά με την παροχή τους, το αποθετήριο ή το σύστημα αποθήκευσης θα πρέπει να είναι σε θέση να επικοινωνήσει με τον κάτοχο ώστε να επιτραπεί η πρόσβαση.

Ουσιαστικά, μέσω των παραπάνω διαδικασιών οι κάτοχοι των δεδομένων μπορούν να ελέγχουν πού και πώς διατίθενται τα δεδομένα τους, κάτι που είναι σε απόλυτη σύμπτωση με τις αρχές FAIR καθώς, όπως θα δούμε και αναλυτικότερα παρακάτω, οι έννοιες των Ανοικτών Δεδομένων (Open Data) δεν ταυτίζονται με τις αρχές FAIR, παρά το γεγονός ότι συχνά συγχέονται. Τέλος, δεδομένου του κόστους συντήρησης των δεδομένων, είναι συχνό το φαινόμενο να αποσύρονται ερευνητικά δεδομένα από την ψηφιακή τους κυκλοφορία. Στην ατυχή αυτή περίπτωση, θα πρέπει να είναι σε θέση τα διάφορα αποθετήρια και συστήματα αποθήκευσης να διατηρούν τουλάχιστον τα μεταδεδομένα, ενημερώνοντας αρχικά για την παρελθοντική ύπαρξη των δεδομένων και τη μη διαθεσιμότητά τους, ενώ ταυτόχρονα θα παρέχουν πληροφορίες σχετικά με το πλαίσιο, τους συγγραφείς, τους φορείς καθώς και όλα τα στοιχεία που ίσως είναι χρήσιμα στους/στις εξωτερικούς/ές χρήστες για την αναζήτηση περισσότερων λεπτομερειών. Με βάση όλα τα παραπάνω, γίνεται αντιληπτό ότι οι περισσότερες σύγχρονες και αξιόπιστες πλατφόρμες αποθήκευσης και δημοσίευσης δεδομένων προσφέρουν αυτές τις υπηρεσίες στους κατόχους των δεδομένων, διασφαλίζοντας ότι συμμορφώνονται με αυτές τις αρχές.

### *Διαλειτουργικά (interoperable) δεδομένα και μεταδεδομένα*

Άλλη μια κατευθυντήρια αρχή των δεδομένων FAIR είναι να καθίστανται τα διαμοιραζόμενα δεδομένα διαλειτουργικά, καθώς οι άνθρωποι αλλά κυρίως οι μηχανές μπορεί να έχουν πρόβλημα στην «ερμηνεία» των δεδομένων λόγω των διαφορούμενων αλλά και σημασιολογικά διαφορετικών εννοιών που μπορεί να έχει

κάποια λέξη ή τιμή, και φυσικά της πληθώρας διαφορετικών γλωσσών στις οποίες είναι αποτυπωμένα τα δεδομένα. Πρακτικά, συνιστάται η χρήση κοινών προτύπων δεδομένων για την αναπαράστασή τους, καθώς και η χρήση λεξικών (vocabularies) και οντολογιών (ontologies) που επιδιώκουν την ταξινόμηση και επεξήγηση οντοτήτων, την αναπαράσταση τιμών και τη σήμανση των δεδομένων. Τα λεξικά και οι οντολογίες συχνά ορίζονται στο πλαίσιο των διαφόρων ερευνητικών κοινοτήτων και συλλογικοτήτων και αποτελούν έναν αδιαμφισβήτητο τρόπο προσθήκης σημασιολογικού νοήματος στα δεδομένα και τα μεταδεδομένα.

Για παράδειγμα, μέσω των Controlled Vocabularies<sup>14</sup> (CV) της CESSDA ERIC<sup>15</sup>, τα διάφορα αποθετήρια κοινωνικών δεδομένων ανά την Ευρώπη μπορούν να τυποποιηθούν σε ένα βαθμό την διαδικασία τεκμηρίωσης των δεδομένων τους δίνοντάς τους την ίδια νοηματοδότηση: π.χ. χρησιμοποιώντας την ίδια ταξινόμηση θεματικών μέσω του Topic Classification Controlled Vocabulary<sup>16</sup> –μιας τυπολογίας των κύριων θεματικών ή του αντικειμένου στο οποίο αναφέρονται τα δεδομένα– και επιτρέποντας στον/στην εξωτερικό/ή χρήστη που αναζητά μια θεματική στον Κατάλογο Δεδομένων της CESSDA (CDC)<sup>17</sup> να έχει πρόσβαση σε μεγάλο πλήθος διαφορετικών αποτελεσμάτων, τα οποία όμως «συνομιλούν» μεταξύ τους μέσω των κοινών προτύπων και θεματικών. Ομοίως, τα λεξικά θα πρέπει να διέπονται από κάποιες αρχές: για παράδειγμα, να τυγχάνουν αποδοχής από την επιστημονική ή ευρύτερη κοινότητα από την οποία προέρχονται και στην οποία απευθύνονται, να είναι καλά τεκμηριωμένα, να συντηρούνται, να μπορεί να γίνει παραπομπή και αναφορά σε αυτά μέσω μόνιμων αναγνωριστικών κ.ά.

Επιπλέον, στο επίπεδο των μεταδεδομένων που τεκμηριώνουν ένα σύνολο δεδομένων, θα πρέπει να δίνεται η απαραίτητη αναγνώριση, συσχέτιση και παραπομπή σε άλλα σύνολα δεδομένων ή δημοσιεύσεις που σχετίζονται με το τεκμηριωμένο σύνολο δεδομένων. Η δυνατότητα αυτή επιτυγχάνεται μέσω εξειδικευμένων πεδίων μεταδεδομένων, όπως είναι οι «συναφείς δημοσιεύσεις» (related publications), τα «συναφή έργα παραγωγής δεδομένων» (related data projects) ή το «συναφές υλικό» (related material), τα οποία εντάσσονται στα δημοφιλή σχήματα μεταδεδομένων

14. CESSDA Vocabulary Service (CVS). Για περισσότερες πληροφορίες μπορείτε να ανατρέξετε στη σελίδα του CVS: <https://vocabularies.CESSDA.eu/>

15. Consortium of European Social Science Data Archives - European Research Infrastructure Consortium. Για περισσότερες πληροφορίες μπορείτε να ανατρέξετε στη σελίδα της CESSDA: <https://www.CESSDA.eu>

16. CESSDA Controlled Vocabulary for CESSDA Topic Classification. Για περισσότερες πληροφορίες μπορείτε να ανατρέξετε στη σελίδα του CESSDA Controlled Vocabulary for CESSDA Topic Classification: <https://vocabularies.CESSDA.eu/vocabulary/TopicClassification?lang=en>

17. CESSDA data catalogue. Για περισσότερες πληροφορίες μπορείτε να ανατρέξετε στη σελίδα του CDC: <https://datacatalogue.CESSDA.eu/> 39.885 σύνολα δεδομένων και μελέτες στις 10 Ιανουαρίου 2022.

όπως είναι, μεταξύ άλλων, το DDI<sup>18</sup> ή το Dublin Core<sup>19</sup>. Μέσω της διαδικασίας αυτής, πέρα από την διαλειτουργικότητα των δεδομένων και μεταδεδομένων, διασφαλίζεται η ορθή αναφορά και αναγνώριση της διανοητικής ιδιοκτησίας στα άτομα και τους οργανισμούς που παρήγαγαν και κατέστησαν διαθέσιμα τα δεδομένα τα οποία αποτέλεσαν τη βάση για την παραγωγή των δικών μας δεδομένων.

Τέλος, ένας ακόμα παράγοντας διαλειτουργικότητας των δεδομένων, είναι η δυνατότητα και προσπάθεια διάθεσής τους σε μορφές (data format) ευρέως αποδεκτές και χρησιμοποιούμενες από την κοινότητα, οι οποίες υποστηρίζονται σε διάφορους τύπους υπολογιστικών ή άλλων συστημάτων. Για παράδειγμα, η διάθεση ερευνητικών δεδομένων και σε τύπους αρχείων που είναι «ανοικτούί» –υπό την έννοια ότι ο καθένας μπορεί να τα χρησιμοποιήσει χωρίς να χρειάζεται να καταβάλει κάποιου είδους συνδρομή ή τέλος χρήσης– προάγει αναμφίβολα τη διαλειτουργικότητα των δεδομένων και, εντέλει, την επανάχρησή τους.

#### *Επαναχρησιμοποιήσιμα (reusable) δεδομένα και μεταδεδομένα*

Όπως ήδη αναφέραμε στην ενότητα περί εξερευνησιμότητας των δεδομένων και μεταδεδομένων, η διαδικασία της επαρκούς τεκμηρίωσης των δεδομένων από την πλευρά των καταθετών των δεδομένων επιτρέπει σε εξωτερικούς/ές χρήστες να αναζητήσουν και να εντοπίσουν τα αποτελέσματα που τους είναι χρήσιμα. Παράλληλα όμως βοηθά τόσο τους ανθρώπους, όσο και τις μηχανές να κατανοήσουν το πλαίσιο μέσα στο οποίο δημιουργήθηκαν τα δεδομένα, τους σκοπούς τους και φυσικά το περιεχόμενό τους. Ως εκ τούτου, όσο περισσότερες πληροφορίες παρέχουμε, κυρίως με τη μορφή μεταδεδομένων, τόσο περισσότερο μπορούν οι εξωτερικοί/ές χρήστες να κατανοήσουν το περιεχόμενο των δεδομένων μας και επομένως να κρίνουν αν τα δεδομένα μας μπορεί να είναι χρήσιμα και άρα επαναχρησιμοποιήσιμα για περαιτέρω δική τους ανάλυση. Πληροφορίες που μπορεί αρχικά να μας φαίνονται μη σχετικές, μπορεί να είναι τμήμα της επιλογής (filtering) και αναζήτησης δεδομένων για άλλους χρήστες. Στην κατεύθυνση αυτή, η χρήση ελεγχόμενων λεξικών (controlled vocabularies) και οντολογιών γενικότερα (όπως περιγράφηκε αναλυτικά πιο πάνω) μπορεί να είναι μείζονος σημασίας, καθιστώντας την περιγραφή του περιεχομένου των δεδομένων όσο το δυνατόν πιο αδιαμφισβήτητη.

Άλλη σημαντική πτυχή των αρχών FAIR είναι η παροχή αδειών χρήσης (licences of use) τόσο σε επίπεδο δεδομένων όσο και μεταδεδομένων, καθώς, όπως ήδη έχουμε αναφέρει και θα δούμε αναλυτικότερα πιο κάτω, οι αρχές FAIR και

---

18. Data Documentation Initiative. Για περισσότερες πληροφορίες μπορείτε να ανατρέξετε στη σελίδα του DDI: <https://ddialliance.org>

19. Dublin Core Metadata Initiative, Για περισσότερες πληροφορίες μπορείτε να ανατρέξετε στη σελίδα του Dublin Core: <https://dublincore.org>

εν προκειμένω η αδειοδότηση των δεδομένων μπορούν να εφαρμοστούν τόσο σε άδειες που περιορίζουν, μερικώς ή ακόμα και ολικώς, τα δικαιώματα χρήσης ως προς την πρόσβαση στα δεδομένα, όσο και σε ανοικτά διατιθέμενα δεδομένα (Open Data). Οι άδειες χρήσης είναι τυποποιημένες δηλώσεις, αναγνώσιμες τόσο από μηχανήματα όσο και από ανθρώπους, και θέτουν τους όρους και τις προϋποθέσεις με τις οποίες μπορεί να γίνει η χρήση των δεδομένων και μεταδεδομένων. Διατίθενται πολλές διαφορετικές άδειες για δεδομένα – μεταξύ των δημοφιλέστερων και φιλικών προς τον χρήστη είναι η γκάμα αδειών Creative Commons<sup>20</sup>.

Στις άδειες Creative Commons δηλώνεται ρητά ο τρόπος παραπομπής τους κατά την επανάχρηση και αν είναι αναγκαία η παραπομπή, παρέχονται πιθανές επιλογές επανάχρησης των έργων δεδομένων κ.ά. Κατά την κατάθεση δεδομένων σε κάποιο πιστοποιημένο αποθετήριο, δίνεται συνήθως η επιλογή εφαρμογής αδειών χρήσης στα κατατιθέμενα μεταδεδομένα και δεδομένα, με τη δυνατότητα να επιλεχθούν πέραν της μίας άδειες για τα διαφορετικά αρχεία δεδομένων και αντικείμενα που αποτελούν τα συνοδευτικά έγγραφα ενός έργου δεδομένων<sup>21</sup>. Μεγάλο μέρος της αξίας των δεδομένων συνιστά η ικανότητα μιας μηχανής ή ενός ανθρώπου να κρίνει την προέλευσή τους` έτσι αξιολογεί συχνά κατά πόσον τα δεδομένα είναι επαναχρησιμοποιήσιμα σε ένα νέο πλαίσιο, αξιολόγηση που περιλαμβάνει την ικανότητα να γνωρίζει κανείς πώς δημιουργήθηκαν τα δεδομένα και από ποιον, να ενημερώνεται για τη χρήση των σχετικών εργαλείων κ.ά. Άλλες πληροφορίες σχετίζονται με την επεξεργασία των δεδομένων: αν έχουν υποστεί επεξεργασία ή είναι ακατέργαστα, ποιες ενέργειες έχουν υποστεί κατά την επεξεργασία και ούτω καθεξής. Οι αρχές FAIR ενθαρρύνουν τη χρήση αρχείων προέλευσης (provenance files), όπου οι υπεύθυνοι των δεδομένων μπορούν να ορίσουν με σαφήνεια όλες τις παραπάνω πληροφορίες, βοηθώντας με αυτό τον τρόπο τους/τις εξωτερικούς/ές χρήστες να αξιολογήσουν και εν τέλει να επαναχρησιμοποιήσουν τα εν λόγω δεδομένα.

Η εργασία με σύνολα δεδομένων από ποικίλες πηγές θα ήταν απλούστερη αν μπορούσε να υπάρξει συναίνεση σε έναν ορισμένο, τυποποιημένο τρόπο οργάνωσης και περιγραφής των δεδομένων. Αυτός είναι και ο λόγος για τον οποίο πολλοί επιστημονικοί κλάδοι έχουν δημιουργήσει πρότυπα μεταδεδομένων για την περιγραφή των δεδομένων, καταλόγους συνιστώμενων μορφών αρχείων κ.ά. Η τήρηση αυτών των προτύπων θα επιτρέψει την ενσωμάτωση των νέων δεδομένων στο όλο οικοσύστημα των δεδομένων που είναι εύκολα και κατάλληλα για επανάχρηση από άλλους. Καθώς όμως δεν είναι δυνατή ακόμη η απόλυτη τυποποίηση, ενώ ταυτόχρονα σε πολλές επιστημονικές πειθαρχίες ανοίγονται νέα ερευνητικά μονοπάτια

20. Για περισσότερες πληροφορίες για τις άδειες Creative Commons μπορείτε να επισκεφτείτε τον ιστότοπο: <https://creativecommons.org>.

21. Ως Έργο παραγωγής Δεδομένων (data project) ορίζεται ένα οικείο σύνολο δεδομένων εμπλουτισμένο με διαφορετικού τύπου τεκμήρια ή πόρους που συνοδεύουν το σύνολο δεδομένων (την έρευνα per se).

χωρίς πρότερη ύπαρξη προτύπων, οι προκλήσεις για τους επιστήμονες των δεδομένων προϋποθέτουν διαρκή εγρήγορση, περιέργεια και εκπαίδευση, έτσι ώστε οι παρεχόμενες υπηρεσίες τους να εναρμονίζονται με τις αναδυόμενες τάσεις και ανάγκες σε ό,τι αφορά την ορθή τεκμηρίωση, διατήρηση και διάχυση των δεδομένων.

### *Αποσαφηνίζοντας ποια δεδομένα δεν είναι FAIR*

Αρχικά, όπως ήδη αναφέραμε, οι αρχές FAIR είναι ουσιαστικά κατευθύνσεις στην προσπάθεια να γίνουν τα δεδομένα και οι προσφερόμενες υπηρεσίες όσο το δυνατόν χρησιμότερες και να υπάρχει πραγματική επανάχρησή τους από την ερευνητική και ευρύτερη κοινότητα` δεν αποτελούν κάποιου είδους πρότυπο που καθορίζει συγκεκριμένη μορφή ή τεχνολογία η οποία θα πρέπει να χρησιμοποιηθεί ή να αναπτυχθεί, αλλά επιτρέπουν διαφορετικές προσεγγίσεις ως προς την επίτευξή τους. Κατά αυτό τον τρόπο, οι αρχές FAIR δεν αποτελούν πρότυπα που θα «ξεπεραστούν» κάποια στιγμή από κάποια άλλη τεχνολογία, αλλά μια συνεχή και πολύμορφη διαδικασία. Αυτό σημαίνει ότι δεν μπορούν να ταυτιστούν με κάποια ήδη υπάρχουσα τεχνολογία ή εφαρμογή, όπως το Resource Description Framework (RDF)<sup>22</sup> ή τα Διασυνδεδεμένα Δεδομένα (Linked Data)<sup>23</sup>, παρά το γεγονός ότι τούτα χρησιμοποιούνται ευρέως αυτή την στιγμή καθώς ικανοποιούν τις «απαιτήσεις» των δεδομένων FAIR (Mons, 2018, σ. 57).

Οι αρχές FAIR φιλοδοξούν να παραμείνουν αρχές και να μπορούν να υλοποιηθούν μέσα από ένα φάσμα τεχνολογιών και προτύπων (formats). Παράλληλα οι αρχές FAIR δεν περιορίζονται μόνο σε κάποιον ή κάποιους συγκεκριμένους επιστημονικούς τομείς, αλλά μπορούν να εφαρμοστούν σε όλες τις επιστήμες και στα παραγόμενα δεδομένα και μεταδεδωμένα των ερευνών αυτών. Τέλος, θα πρέπει να καταστεί σαφές ότι οι αρχές των FAIR δεδομένων δεν απευθύνονται μόνο σε φυσικά πρόσωπα που επιθυμούν να εντοπίσουν δεδομένα, να έχουν πρόσβαση σε αυτά και τελικά να τα ξαναχρησιμοποιήσουν. Αντίθετα, εφόσον οι αρχές FAIR γίνουν καθολικές, οι μηχανές και οι αλγόριθμοι θα μπορούν αυτοματοποιημένα

---

22. Για περισσότερες πληροφορίες για τα RDF: <https://www.w3.org/standards/semanticweb/data>.

23. Τα Διασυνδεδεμένα Δεδομένα συνδέουν διαφορετικούς τύπους δεδομένων: για παράδειγμα, στις κοινωνικές επιστήμες, δεδομένα των μέσων κοινωνικής δικτύωσης με δεδομένα που έχουν παραχθεί στο πλαίσιο συγκεκριμένης έρευνας και δειγματοληψίας. Αυτός μπορεί να είναι ένας τρόπος να συνδυαστούν τα μοναδικά πλεονεκτήματα και να αντιμετωπιστούν ορισμένοι από τους αντίστοιχους περιορισμούς αυτών των δύο τύπων δεδομένων. Παρά το ότι τα αποθετήρια επιχειρούν να αξιοποιήσουν διαφορετικές τεχνικές και εργαλεία και κυρίως να παράσχουν υπηρεσίες που θα διασφαλίζουν την αποθήκευση και συντήρηση τέτοιων νέων τύπων δεδομένων, μια σημαντική δυσκολία αφορά την προβλεπόμενη εκ του ΓΚΠΔ ενήμερη συγκατάθεση των υποκειμένων της έρευνας, που θα λαμβάνει υπόψη τα προαπαιτούμενα προστασίας των υποκειμένων αυτών (Breuer et al., 2021) .

να λαμβάνουν, να επεξεργάζονται και να επαναχρησιμοποιούν τα δεδομένα, εξοικονομώντας τεράστιους πόρους, ενώ παράλληλα η επανάχρηση των δεδομένων αυτών μπορεί να οδηγήσει σε αποκαλύψεις μοτίβων και συνδέσεων που δεν ήταν δυνατές έως τώρα.

### *Ανοικτά δεδομένα και δεδομένα FAIR*

Οι έννοιες και ιδιότητες των FAIR και ανοικτών δεδομένων δεν αποτελούν de facto αναπόσπαστες ή αλληλοσυμπληρούμενες έννοιες<sup>24</sup> τα FAIR δεδομένα δηλαδή μπορεί να μην είναι ανοικτά και να διατίθενται υπό προϋποθέσεις (EC Expert Group on Fair Data, 2018). Θα θέλαμε να καταστεί σαφές ότι οι ιδιότητες του FAIR προσδίδουν προστιθέμενη αξία στα δεδομένα και μπορούν να επαναχρησιμοποιηθούν ανεξαρτήτως του τρόπου πρόσβασης σε αυτά, όπως στην περίπτωση προσωπικών και ευαίσθητων δεδομένων ή δεδομένων που έχουν συλλεχθεί για εμπορικούς σκοπούς. Σε μια ευρύτερη ερμηνεία της Ανοικτής Επιστήμης, θα μπορούσε επίσης να υποστηριχθεί ότι η στόχευση είναι η πλήρης διάθεση των δεδομένων, εντός όμως σαφώς προσδιορισμένων πλαισίων διαβαθμισμένης πρόσβασης ή σε συμμόρφωση με την εκάστοτε εθνική νομοθεσία που διέπει την κυκλοφορία. Οι τεχνικές ανωνυμοποίησης, οι συμφωνίες με τους παρόχους δεδομένων, καθώς και οι ασφαλείς χώροι διάθεσης αυτών όπως τα Αποθετήρια αποτελούν τις κύριες λύσεις στην αντιμετώπιση της διάκρισης ή και της επιλογής μεταξύ FAIR και ανοικτών δεδομένων.

Στους κόλπους της ερευνητικής κοινότητας των Κοινωνικών Επιστημών, καθώς και στα περιβάλλοντα των ψηφιακών υποδοχέων, είτε Αποθετήρια είτε Αρχεία Κοινωνικών Δεδομένων είναι αυτοί, όπως η ευρωπαϊκή υποδομή CESSDA ERIC, έχει υιοθετηθεί η προσέγγιση «ανοικτά στο μέτρο του δυνατού, κλειστά στο μέτρο του αναγκαίου<sup>24</sup>». Ως προέκταση αυτής της προσέγγισης, μπορεί να διατυπωθεί αντίστοιχα το «ανοικτά στο μέτρο του δυνατού και FAIR στο μέτρο του αναγκαίου» (EC Expert Group on FAIR Data, 2018). Και οι δύο προαναφερόμενες προσεγγίσεις αφορούν κυρίως τη δημόσια χρηματοδοτούμενη έρευνα και τα παραγόμενα δεδομένα και παρουσιάζονται σε διατυπώσεις μελετών και εκθέσεων της Ευρωπαϊκής Επιτροπής.

### **Ο ρόλος των Αποθετηρίων στα δεδομένα FAIR**

Σημείοντα ρόλο μεταξύ των διαύλων μεταβίβασης της επιστημονικής παραγωγής κατέχουν συγκροτημένοι μηχανισμοί αποθήκευσης και διάθεσης των δεδομένων FAIR, που είναι τα Αποθετήρια Δεδομένων. Τα πιστοποιημένα Αποθετήρια Δεδομένων, μέσω του σκοπού και της αποστολής τους, μπορούν να συμβάλουν στην

24. As open as possible, as closed as necessary.



προαγωγή των FAIR δεδομένων με στενότερη ή χαλαρότερη διασύνδεση προς τις αρχές της ανοικτότητας των υπηρεσιών τους. Επιπλέον, δεδομένης της μαζικής αύξησης της παραγωγής δεδομένων, ο ρόλος τους αναδεικνύεται σημαντικός όσον αφορά την εδραίωση των αρχών των δεδομένων FAIR, αλλά και τη συγκρότηση κριτηρίων προτεραιοποίησης και επιλογής ως προς τα σύνολα δεδομένων που θα διατεθούν προς επανάχρηση σε σχέση με άλλα που δεν θα πληρούν τα κριτήρια. Οι ερευνητικές υποδομές θα μπορούσαν λοιπόν να συμβάλουν στην οργάνωση αυτού του «αδιαμφισβήτητου καταιγισμού δεδομένων και πληροφοριών εν τη γενέσει του», έτσι ώστε να κεφαλαιοποιηθεί η αξία τους, προς το συμφέρον μιας συνεκτικής και αναστοχαστικής κοινωνίας.

Τα αποθετήρια γνώρισαν ιδιαίτερη άνθηση τα τελευταία χρόνια. Η αύξηση της δημιουργίας αποθετηρίων οφείλει πολλά στην «Ανοικτή Επιστήμη», που πρόωθησε την ανάγκη συγκέντρωσης δεδομένων και διάχυσής τους, λόγω της αυξανόμενης ζήτησης για διάχυση και «ανοικτότητα». Χρησιμοποιώντας την περίπτωση της ανάπτυξης της ερευνητικής υποδομής SoDaNet, θα αναδείξουμε μέσα από συγκεκριμένα παραδείγματα την δυναμική της επικοινωνίας στους κόλπους μιας διεθνούς ερευνητικής κοινότητας δεδομένων και την κεφαλαιοποίηση των δεδομένων FAIR σε μια ευρωπαϊκή χώρα, συνυπολογίζοντας την περατότητα της χρηματοδότησης και των ανθρώπινων πόρων. Η σημερινή ανταλλαγή και συνομιλία μιας εθνικής, ευρωπαϊκής και διεθνούς κοινότητας επιτρέπει στα κενά και τις ελλείψεις να καλύπτονται μέσα από ένα συνεχές ταξίδι γνώσης.

Η προσπάθεια για τη δημιουργία μιας εθνικής υποδομής για την υποστήριξη της έρευνας στις κοινωνικές επιστήμες στη χώρα μας ξεκίνησε με την δημιουργία της Τράπεζας Κοινωνικών Δεδομένων από το ΕΚΚΕ στα τέλη της δεκαετίας του 1990 και σταθεροποιήθηκε χάρη στην έγκαιρη προσχώρησή της στο τότε Ευρωπαϊκό Συμβούλιο Κοινωνικών Αρχείων Δεδομένων CESSDA, με στόχο την ικανοποίηση των αναγκών της εθνικής ερευνητικής κοινότητας, σε συνομιλία με την ευρωπαϊκή κοινοπραξία για την ανάπτυξη συνεργασιών και την κεφαλαιοποίηση τεχνογνωσίας. Η μεγάλη δυσκολία της εποχής εκείνης συνίστατο σε τεχνολογικές εξελίξεις που απαιτούσαν την υιοθέτηση τεχνολογικών συστημάτων κλειστού κώδικα, ανταποκρινόμενων σε συγκεκριμένες ανάγκες. Οι εξελίξεις της ανοικτής επιστήμης και η συνδιαλλαγή και συνδιαμόρφωση πολιτικών για την ανάπτυξη των ερευνητικών υποδομών σε ευρωπαϊκό επίπεδο<sup>25</sup> συνέβαλαν στον εκσυγχρονισμό της εθνικής υποδομής στο πλαίσιο μιας ευρωπαϊκής κοινοπραξίας και υπόσχονται τη μελλοντική βιωσιμότητά της. Οι πόροι από ευρωπαϊκά έργα τεχνογνωσίας και κατάρτισης του προσωπικού που ενεπλάκη στην υπόθεση των δεδομένων συνέβαλαν επίσης στην περαιτέρω ανάπτυξή της.

Έτσι στα είκοσι χρόνια λειτουργίας της, χάρη και στη γενικότερη στρατηγική

---

25. ESFRI, EI RoadMaps.



σε επίπεδο πολιτείας και ερευνητικής κοινότητας, η ελληνική Τράπεζα Κοινωνικών Δεδομένων του ΕΚΚΕ μετουσιώθηκε στο εθνικό δίκτυο SoDaNet, που στοχεύει στον σχεδιασμό, την ανάπτυξη και τη βιώσιμη λειτουργία της εθνικής ερευνητικής υποδομής SoDaNet για την υποστήριξη της εμπειρικής έρευνας στις κοινωνικές επιστήμες στη χώρα μας. Η υποδομή SoDaNet επεδίωξε και πέτυχε την ορατότητά της, αυξάνοντας τα μέλη της στους κόλπους της ακαδημαϊκής κοινότητας με το δίκτυο χρηστών που τα μέλη της προσέδωσαν στην υποδομή. Σήμερα η εθνική υποδομή, με τη υιοθέτηση και τροποποίηση του συστήματος Dataverse ως βασικού αποθετηριακού εργαλείου, επιδιώκει την συνεχή συνομιλία και συνεργασία με μια διεθνή κοινότητα που προάγει συστήματα διαχείρισης δεδομένων ανοικτού κώδικα και διασφαλίζει τη διατήρηση και περαιτέρω ανάπτυξή της με τη ματιά στραμμένη στο «γίγνεσθαι» της διεθνούς κοινότητας δεδομένων. Παράλληλα οι εθνικές συνεργασίες και δικτύώσεις επιτρέπουν την ανταπόκριση στις ανάγκες της εθνικής κοινότητας και την παροχή υπηρεσιών που στηρίζουν σε ατομικό και συλλογικό επίπεδο υψηλής ποιότητας δεδομένα και εν γένει ψηφιακούς πόρους.

### *Οι αρχές των δεδομένων FAIR στην υποδομή SoDaNet*

Στο πλαίσιο του έργου SoDaNet in Action, το δίκτυο SoDaNet υιοθέτησε το Dataverse<sup>26</sup> ως βασικό αποθετηριακό λογισμικό για την υποδομή<sup>27</sup>, παραμετροποιώντας το σύμφωνα με τις ανάγκες του δικτύου. Έτσι, εκτός από τα διάφορα πλεονεκτήματα που είχε ως εφαρμογή σε σχέση με τα προγενέστερα λογισμικά της υποδομής (Λιναρδής & κ.ά., σε αυτό τον τόμο), μέσω του Dataverse ικανοποιούνταν και οι κατευθυντήριες αρχές των δεδομένων FAIR. Ενδεικτικά θα αναφέρουμε κάποια από τα τεχνικά χαρακτηριστικά του Dataverse που ικανοποιούν ορισμένες από τις αρχές FAIR:

1. Αρχικά, υποστηρίζει τα «ανοικτά» πρωτόκολλα του διαδικτύου TCP/IP και καθένας μπορεί να περιηγηθεί στο αποθετήριο με μια σύνδεση στο διαδίκτυο μέσω ενός διαδικτυακού περιηγητή (browser). Παράλληλα όμως, υποστηρίζει και την ταυτοποίηση των εγγεγραμμένων χρηστών δίνοντάς τους επίπεδα εξουσιοδότησης ως προς τις δυνατότητες πρόσβασης στα δεδομένα που φιλοξενούνται, ικανοποιώντας με αυτούς τους τρόπους την προσβασιμότητα σε δεδομένα και μεταδεδομένα (accessibility) που ορίζουν οι αρχές FAIR.

26. Το Dataverse είναι μια διαδικτυακή εφαρμογή ανοικτού κώδικα (open-source) για την κοινή χρήση, τη διατήρηση, την παράθεση, τη διερεύνηση και την ανάλυση ερευνητικών δεδομένων. Για περισσότερες πληροφορίες: <https://dataverse.org>

27. Για πρόσβαση στον κατάλογο δεδομένων του SoDaNet, ακολουθήστε τον σύνδεσμο: <https://datacatalogue.sodanet.gr>

2. Τα μεταδεδομένα προσφέρονται ελεύθερα σε όλους/ες –εγγεγραμμένους/ες και μη– χρήστες, ακόμα και στις περιπτώσεις που τα δεδομένα είναι περιορισμένης πρόσβασης ή σε κατάσταση embargo<sup>28</sup>. Παρέχονται μεταδεδομένα σε επίπεδο έργου δεδομένων, που περιλαμβάνουν τα μεταδεδομένα παραπομπής (citation metadata), αλλά και τα διάφορα εξειδικευμένα πακέτα μεταδεδομένων, όπως τα γεωχωρικά (geo-spatial), τα μεταδεδομένα κοινωνικών και ανθρωπιστικών επιστημών (social sciences and humanities metadata), και εκείνα που σχετίζονται με τους διαφορετικούς τύπους έργων δεδομένων που υποστηρίζει το αποθετήριο, όπως οι Δείκτες και οι Ταξινομήσεις (Indices and Classification metadata) ή τα μεταδεδομένα SIMS<sup>29</sup>, που χρησιμοποιούνται από την Eurostat – ενώ παρέχονται και μεταδεδομένα σε επίπεδο αρχείου (file). Όπως είδαμε και παραπάνω, η επαρκής και έγκυρη τεκμηρίωση και παροχή μεταδεδομένων, τόσο σε επίπεδο έργου δεδομένων όσο και σε επίπεδο αρχείου, διευκολύνει την ευρεσιμότητα, την διαλειτουργικότητα και εν τέλει την επανάχρηση των δεδομένων .
3. Παρέχεται μόνιμος ταυτοποιητής τύπου DOI σε κάθε έργο δεδομένων όταν αυτό δημοσιοποιηθεί, βοηθώντας στην ευρεσιμότητα των δεδομένων και μεταδεδομένων.
4. Τόσο σε επίπεδο έργου δεδομένων όσο και σε επίπεδο μοναδικού αρχείου (file) μπορούν να οριστούν διαφορετικού τύπου άδειες χρήσης και επίπεδα πρόσβασης στους διάφορους τύπους χρηστών, διευκολύνοντας έτσι την προσβασιμότητα των δεδομένων.
5. Τα πινακοποιημένα δεδομένα (tabular files) είναι διαθέσιμα και στον «ανοιχτό» μορφότυπο .tab –παρά το γεγονός ότι τα αυθεντικά αρχεία μπορεί να είναι «κλειστού» τύπου, όπως είναι οι διάφοροι μορφότυποι στατιστικών πακέτων: π.χ. .sav αρχεία του στατιστικού πακέτου SPSS<sup>30</sup>– συμβάλλοντας έτσι στην διαλειτουργικότητα και επανάχρηση των δεδομένων, όπως, για παράδειγμα, η εφαρμογή του Λεξικού Κοινωνικών Όρων<sup>31</sup> της υποδομής .

---

28. Ως embargo ορίζεται η κατάσταση κατά την οποία ορισμένα δεδομένα τελούν υπό απαγόρευση πρόσβασης για ορισμένο χρονικό διάστημα. Έτσι, παρά το γεγονός ότι έχουν υποβληθεί σε δημόσια προσβάσιμα αποθετήρια, δεν είναι διαθέσιμα για λήψη για το ρητά καθορισμένο χρονικό διάστημα που αποφασίζουν οι παραγωγοί των δεδομένων αυτών.

29. Single Integrated Metadata Structure. Για περισσότερες πληροφορίες: <https://ec.europa.eu/eurostat/documents/64157/4373903/03-Single-Integrated-Metadata-Structure-and-its-Technical-Manual.pdf>

30. Statistical Package for the Social Sciences, λογισμικό της IBM.

31. Για αναζήτηση στο Λεξικό Κοινωνικών Όρων του SoDaNet: <https://www.sodanet.gr/data-services/socialterms/search>

*Ο ρόλος των επιστημόνων των δεδομένων και  
οι δεξιότητες στη διαχείριση των δεδομένων*

Οι τεχνολογικές εξελίξεις προϋποθέτουν την ανάπτυξη δεξιοτήτων και γνώσεων του ανθρώπινου δυναμικού που δραστηριοποιείται στη διαχείριση των δεδομένων υπό την ευρεία έννοια. Η εκτίμησή μας είναι ότι, στο πέρασμα των χρόνων, οι δεξιότητες αυτές, που αφορούν τον σχεδιασμό, τη συλλογή/παραγωγή δεδομένων, καθώς και την περαιτέρω ανάλυση και διαχείριση και εν τέλει επανάχρηση αυτών, συνιστούν εγκάρσιες τομές που εντάσσονται σταδιακά σε ένα ενιαίο πλέγμα, όπου θα πρέπει να συνομιλούν και να ενσκήπτουν η μία μέσα στην άλλη. Αναφερόμαστε δηλαδή σε κοινωνικούς επιστήμονες που αναστοχάζονται την κοινωνική πραγματικότητα και παράγουν νοητικά σχήματα και εμπειρικά αποτελέσματα με το ανθρώπινο δυναμικό των ερευνητικών υποδομών και αποθετηρίων (Κονδύλη & Λιναρδής, 2020). Η ανάπτυξη αυτών των δεξιοτήτων και η συνεχής εκπαίδευση σε νέους τύπους δεδομένων και μεθόδους συνυπάρχουν σε κάποιο βαθμό και με την έλλειψη αυτών. Τα δεδομένα FAIR ή τα ανοικτά δεδομένα, όπως και η αξιοποίηση των μεγάλων δεδομένων (Big Data), αποτελούν κι αυτά τομείς αλληλόδρασης και συνεργασίας.

Έτσι, από τη μία πλευρά, οι κοινωνικοί επιστήμονες καλούνται να αναστοχαστούν τα μεθοδολογικά κεκτημένα της επιστημονικής παράδοσης (έρευνες, ερωτηματολόγια, συγκεκριμένοι πληθυσμοί δείγματος), που αποτυπώνουν δεδομένες κοινωνικές αναπαραστάσεις, και να αναπαραγάγουν αναλύσεις αυτών, πέραν της νομιμοποιητικής δικαιοδοσίας των επιστημών τους. Δηλαδή, άλλη μια συνέπεια της παραγωγής μεγάλων δεδομένων και της επανάχρησης αυτών συνίσταται στο γεγονός ότι οι κοινωνικοί επιστήμονες αρχίζουν να μελετούν τα κοινωνικά φαινόμενα με τη χρήση δεδομένων που αρχικά δεν συλλέχθηκαν για έρευνα, όπως είναι τα δεδομένα κοινωνικής δικτύωσης, καθώς και δεδομένα που προέρχονται από την εξόρυξη (data mining) διαδικτυακών πληροφοριών, ή από ψηφιακούς αισθητήρες, αλλά και από διοικητικά αρχεία (OECD, 2013). Θα μπορούσαμε επομένως να υποστηρίξουμε ότι οι σύγχρονες τεχνολογίες με τις έξυπνες συσκευές και τα συλλεγόμενα μέσω αυτών δεδομένα νοηματοδοτούν τον 21ο αιώνα, που σημαίνει ότι οι κοινωνικές επιστήμες είναι πλέον στην καρδιά όχι μόνο της γνώσης αλλά και της ευρύτερης διεπιστημονικής παραγωγής δεδομένων.

Στο άμεσο μέλλον, οι βασικές προκλήσεις θα συνίστανται στην ικανότητα του ερευνητικού δυναμικού να εργάζεται διεπιστημονικά, λαμβάνοντας υπόψιν τα ηθικά και δεοντολογικά ζητήματα διεξαγωγής έρευνας με τους νέους τύπους δεδομένων. Οι ερευνητικές υποδομές θα πρέπει να αναπτύσσουν και να βελτιώνουν τις παρεχόμενες υπηρεσίες και τα εργαλεία προς χρήση και αξιοποίηση από τους δυνητικούς χρήστες τους. Το ερευνητικό οικοσύστημα χρειάζεται το δίπολο αυτών των δρώντων. Μέσα από τις κοινές προσπάθειες των ερευνητικών επιστημών και των διεπιστημονικών ερευνητικών έργων –όπως τα πρόσφατα διεπιστημονικά

έργα που υλοποιούνται για την αξιοποίηση δεδομένων-, τις ανάγκες τεκμηρίωσης και γνώσης που δημιουργήσε η συνθήκη της πανδημίας Covid-19<sup>32</sup>, τις αρχές της Ανοικτής Επιστήμης και άλλες διεθνείς πρωτοβουλίες, τα πρώτα βήματα έχουν ήδη συντελεστεί σε εθνικό, ευρωπαϊκό και διεθνές επίπεδο.

### **Οι δρόμοι που ανοίγονται: διλήμματα και προοπτικές**

Σε μια εμπνευσμένη εισήγηση προς την κοινότητα των επιστημόνων που ασχολούνται με τα δεδομένα, ο Mathew Wollard (2016: 29) μας επαναφέρει στις βασικές αρχές σε μια προσπάθεια πρόβλεψης των τάσεων της επόμενης δεκαετίας. Επανέρχεται στις αρχές που επιδιώκουν τα δεδομένα FAIR και η Ανοικτή Επιστήμη, με αρωγό την τεχνολογική εξέλιξη στους σημερινούς ψηφιακούς σχηματισμούς. Μας επαναφέρει σε μια έννοια-κλειδί, την εδραίωση της εμπιστοσύνης: «Πώς μπορούμε να προστατεύουμε και την ίδια στιγμή να σεβαστούμε με τον καλύτερο δυνατό τρόπο το υποκείμενο των δεδομένων (data subject), μεγιστοποιώντας παράλληλα τις δυνατότητες χρήσης στην έρευνα και την προστιθέμενη αξία που η έρευνα μπορεί να αποφέρει. Η αίσθησή μου είναι ότι όταν εμείς –ως κοινότητα υπηρεσιών δεδομένων– επιδεικνύουμε τη δέσμευσή μας στο σεβασμό των δεδομένων και την προστασία των εν λόγω υποκειμένων των δεδομένων, η πρόσβαση σε πολλές από τις νέες μορφές δεδομένων θα γίνει ευκολότερη». Θα προσθέταμε επίσης ότι η εμπιστοσύνη εγείρει ταυτόχρονα και ένα ζήτημα αξιοπιστίας ως προς την επανάχρηση των δεδομένων, που αφορά τόσο την άσκηση συνειδητού ελέγχου όσο και την επιβεβαίωση της ορθής χρήσης τους μέσα από διαδικασίες δοκιμής και σφάλματος.

Η Ανοικτή Επιστήμη και η σταδιακή εδραίωση των δεδομένων FAIR μέσω των μηχανισμών λειτουργίας τους διαμεσολαβούν μια νέα εποχή, η έκβαση της οποίας εξαρτάται εν πολλοίς από τις κοινότητες των εμπειρογνομόνων και των ειδικών.

Εν κατακλείδι, μία από τις προκλήσεις του καταγίσιμου δεδομένων και πληροφοριών δεδομένων που θα κληθούμε να αντιμετωπίσουμε στο άμεσο μέλλον έγκειται στην ανάπτυξη και εδραίωση αφενός της εμπιστοσύνης και αφετέρου της διευκόλυνσης κάθε ενδιαφερόμενου πολίτη προκειμένου να κάνει χρήση των υπη-

---

32. Ένα τέτοιο ενδεικτικό παράδειγμα είναι η υλοποίηση του ερευνητικού έργου με τίτλο Beyond COVID, στο οποίο προβλέπεται η συλλογή και διάθεση δεδομένων FAIR διαφορετικών επιστημονικών κλάδων και διεπιστημονικών υποδομών σε μια κοινή πλατφόρμα, με ενιαία σχήματα χρήσης δεδομένων Covid-19, καθώς και πιθανών άλλων επιδημιών, δίνοντας, μεταξύ άλλων, έμφαση στην εκπαίδευση και ανάπτυξη ικανοτήτων των επιστημόνων και δι-αφορετικών κοινοτήτων χρηστών και στην εξοικείωση με μεθοδολογικά εργαλεία που επι-τρέπουν την ανάλυση δεδομένων FAIR. Η εθνική ερευνητική υποδομή SoDaNet συμμετέχει μέσω του ΕΚΚΕ, εθνικού Service Provider και Τρίτου Συμβαλλόμενου μέρους της CESSDA ERIC. Συντονιστής είναι το European Molecular Biology Laboratory, με τη συμμετοχή 53 ακα-δημαϊκών και ερευνητικών οργανισμών ανά την Ευρώπη. Περισσότερες πληροφορίες <https://by-covid.org/>

ρεσιών των Αποθετηρίων και γενικότερα των Ερευνητικών Υποδομών. Η αξιοπιστία και χρήση των υπηρεσιών μας θεμελιώνεται στην πρωθύστερη εδραίωση της εμπιστοσύνης και επικοινωνίας των ερευνητικών υποδομών με την πανεπιστημιακή και ερευνητική κοινότητα, τα ερευνητικά δίκτυα ή ανεξάρτητους ερευνητές ως παραγωγούς δεδομένων και τεκμηρίων, που προάγουν την αποτύπωση όψεων των εκάστοτε κοινωνικών μετασχηματισμών. Ο ρόλος των Αποθετηρίων έγκειται στο να «επιστρέφει» στον κάθε πολίτη τα δεδομένα και, εν γένει, τα παραγόμενα τεκμήρια ή/και τους ψηφιακούς πόρους, υπακούοντας στις αρχές των δεδομένων FAIR και προστατεύοντας μέσα από τις λειτουργίες τους τη διαφάνεια και την προστασία των υποκειμένων της έρευνας.

Τα επόμενα χρόνια, η κουλτούρα και οι πρακτικές ως προς τα δεδομένα θα μεταβάλλονται διαρκώς. Οι παγκοσμιοποιημένες ερευνητικές κοινότητες αρχίζουν να εργάζονται διεπιστημονικά όσον αφορά την μελέτη των φαινομένων *per se*, αλλά και την διαλειτουργικότητα και επανάχρηση των ίδιων των δεδομένων. Οι διεθνείς, ευρωπαϊκοί και εθνικοί οργανισμοί που ασχολούνται με τα δεδομένα θα διαδραματίσουν σημαίνοντα ρόλο μέσα από την επιρροή τους στην ερευνητική κουλτούρα ανταλλαγής δεδομένων διεπιστημονικών εγχειρημάτων. Η σταδιακά αυξανόμενη εδραίωση των FAIR δεδομένων υπηρετεί αυτήν τη μεταβολή και κατεύθυνση.

### Βιβλιογραφικές αναφορές

- Breuer, J., Al Baghal, T., Sloan, L., Bishop, L., Kondyli, D., & Linardis, A. (2021). Informed consent for linking survey and social media data, *IASSIST Quarterly*, 45(1), 1-27. doi <https://doi.org/10.29173/iq988>
- European Commission (2020). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European strategy for data*. COM(2020) 66 final. Brussels: EURLex, 19.02.2020.
- European Commission Expert Group on FAIR Data (2018). *Turning FAIR into reality*. Final Report and Action Plan from the European Commission Expert Group on FAIR data. European Commission. Directorate General for Research and Innovation
- Κάλλας, Γ. (2008) Τα κοινωνικά φαινόμενα και η παρατήρησή τους στις κοινωνικές επιστήμες. Στο Γ. Κάλλας, Δ. Κονδύλη & Γ. Καραγιάννης (επιμ.), *Μεθοδολογικά ζητήματα και ερευνητικές υποδομές των κοινωνικών επιστημών* (17-37). Αθήνα: Ποταμός.
- Κάλλας, Γ. (2002). *Ζητήματα σχεδιασμού εμπειρικών ερευνών: Αξιοποίηση μεθόδων της πληροφορικής τεχνολογίας*. Αθήνα: ΕΚΚΕ & Νεφέλη
- Κονδύλη, Δ. & Λιναρδής, Α., (2021). Νέοι τύποι δεδομένων, νέοι ρόλοι των ερευνητικών υποδομών; Στο Ν. Ναγόπουλος (επιμ.), *Οι Κοινωνικές Επιστήμες σήμερα. Διλήμματα και προοπτικές πέρα από την κρίση*. Μυτιλήνη: Σχολή Κοινωνικών Επιστημών Πανεπιστήμιο Αιγαίου.
- Mons, B. (2018). *Data Stewardship for Open Science: Implementing FAIR Principles* (1η έκδ.). Boca Raton: Chapman and Hall/CRC. <https://doi.org/10.1201/9781315380711>
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O., & Wilkinson, M.

- D. (2017). Cloudy, increasingly fair; revisiting the Fair Data Guiding Principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49–56. doi: 10.3233/isu-170824
- OECD (2017). Business models for sustainable research data repositories. *OECD Science, Technology and Industry Policy Papers*, 47, OECD Publishing, Paris, <https://doi.org/10.1787/302b12bb-en>
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Woollard, M. (2017). Embracing the ‘Data Revolution’: Opportunities and Challenges for Research. *IASSIST Quarterly*, 40(2), 28. <https://doi.org/10.29173/iq784>.

### Διαδικτυακές πηγές

- A FAIRy tale. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://forskningsdata.dk/fairytales/>
- BY. Covid. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://by-covid.org/>
- CESSDA. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://www.cessda.eu/>
- CESSDA Controlled Vocabulary for CESSDA Topic Classification. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://vocabularies.cessda.eu/vocabulary/TopicClassification?lang=en>
- CESSDA Data Catalogue. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://datacatalogue.cessda.eu/>
- CESSDA Vocabulary Service (CVS). Πρόσβαση στις 10 Ιανουαρίου 2022, από: <https://vocabularies.cessda.eu/>
- Digital Object Identifier System. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://www.doi.org/>
- Document, discover and interoperate*. Welcome to the Data Documentation Initiative, Data Documentation Initiative. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://ddialliance.org/>
- Dublin Core*. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://dublincore.org>
- Handle.net registry. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://www.handle.net/>
- How to fair*. FAIR. Πρόσβαση στις 10 Ιανουαρίου, 2022, από <http://www.howtofair.dk/>
- Miller, J., Kemp, J., & Brown, J. (2021, 16 Δεκεμβρίου). FORCE11. Πρόσβαση στις 10 Ιανουαρίου, 2022, από <https://www.force11.org/>
- Nature Publishing Group. (2018, 13 Μαρτίου). Everyone needs a data-management plan. Nature News. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://www.nature.com/articles/d41586-018-03065-z>
- Social Data Network. Social Terms Dictionary. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://www.sodanet.gr/data-services/socialterms/search>
- ORCID. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://orcid.org/>
- The dataverse project*. The Dataverse Project - Dataverse.org. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://dataverse.org/>
- W3C. SEMANTIC WEB: LINKED DATA. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://www.w3.org/standards/semanticweb/data>
- When we share, everyone wins*. Creative Commons. (2021, 24 Μαΐου). Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://creativecommons.org/>
- Κατάλογος Δεδομένων sodanet*. Πρόσβαση στις 10 Ιανουαρίου 2022, από <https://datacatalogue.sodanet.gr/>