

## Ο ΝΕΟΣ ΚΑΤΑΛΟΓΟΣ ΔΕΔΟΜΕΝΩΝ ΤΟΥ SODANET – Η ΜΕΤΑΒΑΣΗ ΑΠΟ ΤΟ NESSTAR ΣΤΟ DATAVERSE

Παραμετροποίηση, βελτιστοποίηση και επέκταση των λειτουργιών  
του Dataverse για τις ανάγκες των αποθετηρίων του SoDaNet

Απόστολος Λιναρδής  
Κωνσταντίνος Αλεξανδρής  
Νικόλας Κληρονόμος

### Περίληψη

*Ο κατάλογος δεδομένων που χρησιμοποιούσε μέχρι πρότινος η ερευνητική υποδομή SoDaNet για να διατηρεί και να διαχέει ερευνητικά δεδομένα ήταν το εμπορικό λογισμικό Nesstar. Το Nesstar διαχειριζόταν σύνολα δεδομένων με μεταβλητές, καθώς και ερωτήσεις από τις οποίες προήλθαν οι μεταβλητές, και υποστήριζε online αναλύσεις μεταβλητών και γεωγραφικές απεικονίσεις. Παρά την αποτελεσματική διαχείριση των μεταβλητών, το Nesstar αντιμετώπιζε σοβαρά προβλήματα στη διαχείριση χρηστών, στη διαχείριση αρχείων, στην αναζήτηση των ερευνών, στην πολυγλωσσικότητα και στην απόδοση μόνιμων ταυτοποιητών, με αποτέλεσμα να καθιστά βασικές λειτουργίες ενός αποθετηρίου προβληματικές.*

*Ο προβληματισμός σε σχέση με τις λειτουργίες του Nesstar, καθώς και το γεγονός ότι το NSD (το νορβηγικό αρχείο δεδομένων) σταμάτησε την υποστήριξη του λογισμικού, οδήγησε το SoDaNet στην αναζήτηση αξιόπιστων λύσεων. Η λύση που προέβρινε ήταν το Dataverse, ένα λογισμικό ανοιχτού κώδικα που αναπτύχθηκε από το Ινστιτούτο για τις ποσοτικές κοινωνικές επιστήμες του Χάρβαρντ και από συνεργάτες παγκοσμίως. Ισχυρά χαρακτηριστικά του Dataverse είναι η αποτελεσματική διαχείριση των χρηστών και των αρχείων, στην οποία το Nesstar υστερούσε. Από την άλλη το Dataverse υστερεί σε χαρακτηριστικά που ήταν τα δυνατά σημεία του Nesstar: «Αντιλαμβάνεται» μεν αντικείμενα όπως το Dataset και το File, που παραπέμπουν στο «Study Description» και στο «Other Study Material» του DDI, αλλά δεν «αντιλαμβάνεται» οντότητες όπως οι μεταβλητές και οι ερωτήσεις.*

Σε συνέχεια των ανωτέρω, και στο πλαίσιο της πράξης “SoDaNet in Action”, η απόφαση που λήφθηκε αφορούσε την υλοποίηση μιας υποδομής που βασίστηκε στην παραμετροποίηση, βελτιστοποίηση και επέκταση των λειτουργιών του Dataverse, ώστε να συμπεριλάβει και τα «ισχυρά» χαρακτηριστικά του Nesstar σε μια ολοκληρωμένη υποδομή. Στην παρούσα εργασία περιγράφονται συνοπτικά οι λειτουργίες του Nesstar και της έκδοσης 4.16 του Dataverse, καθώς και όλες οι βελτιστοποιήσεις που έγιναν στο Dataverse, ώστε ο νέος κατάλογος δεδομένων να συμπεριλάβει τα «δυνατά» χαρακτηριστικά και των δύο λογισμικών.

Λέξεις-κλειδιά: *Dataverse, Nesstar, Αποθετήριο, Διαχείριση Δεδομένων, Διαχείριση Μεταβλητών.*

### **Abstract**

*The SoDaNet research infrastructure used Nesstar software to store and disseminate research data. Nesstar managed datasets, variables and questions that formed the variables, and supported online variable analysis and geographic mapping. Despite efficiently managing variables, Nesstar ran into problems with user management, file management, persistent identifiers and search, making key repository functions problematic.*

*Concerns about Nesstar’s operations, as well as the fact that NSD (the Norwegian Centre for Research Data) stopped supporting the software, led SoDaNet to look for reliable solutions. The solution advocated by SoDaNet was Dataverse, an open source software developed by the Harvard Institute for Quantitative Social Sciences and partners worldwide.*

*A powerful feature of Dataverse is the efficient management of users and files, an area where Nesstar was lagging behind. Dataverse, on the other hand, lags behind in features that were Nesstar’s strengths. Dataverse “perceives” objects such as Dataset and File, which correspond to DDI’s “Study Description” and “Other Study Material”, but does not “perceive” entities such as variables and questions.*

*Following the aforementioned, the decision was made to implement an infrastructure based on the configuration, optimization and expansion of Dataverse functions, so as to include the “strong” features of Nesstar in an integrated system. This paper describes the strengths and weaknesses of Nesstar and Dataverse version 4.16, as well as all the optimizations made to Dataverse so that the new data catalogue can include the “strong” features of both softwares. The data catalogue was implemented in the context of the “SoDaNet in Action” project.*

**Keywords:** *Dataverse, Nesstar, Repository, Data Management, Variable Management.*

### **Εισαγωγή: Οι υπηρεσίες της υποδομής SoDaNet**

Η υποδομή SoDaNet αποτελεί την ελληνική ερευνητική υποδομή για τις κοινωνικές επιστήμες και η δικτύωσή της επιδιώκεται σε εθνικό και ευρωπαϊκό επίπεδο. Το εθνικό ερευνητικό δίκτυο SoDaNet συγκροτήθηκε το 2012 και αποτελείται από το Εθνικό Κέντρο Κοινωνικών Ερευνών (EKKE) και 6 πανεπιστημιακά τμήματα (<https://sodanet.gr/about/sodanet-members/>). Τον Ιούνιο του 2015 η Ελλάδα, με την ερευνητική υποδομή SoDaNet, γίνεται μέλος της Ευρωπαϊκής Κοινοπραξίας Αρχείων Δεδομένων των Κοινωνικών Επιστημών - CESSDA (Consortium of European Social Data Archives), και αποτελεί τον ελληνικό κόμβο της ευρωπαϊκής υποδομής. Η κοινοπραξία CESSDA αποτελείται σήμερα από 22 χώρες μέλη και μία χώρα παρατηρητή (observer). Αρκετές ευρωπαϊκές χώρες βρίσκονται στη διαδικασία να γίνουν μέλη ή παρατηρητές της CESSDA. Η CESSDA έχει επίσης 13 εταίρους (partners) σε διάφορες χώρες εκτός της κοινοπραξίας (<https://www.cessda.eu/About/Consortium>).

Οι υπηρεσίες της εθνικής ερευνητικής υποδομής SoDaNet αφορούν τα κάτωθι:

- παροχή on line πρόσβασης σε τρίτους σε έργα δεδομένων, όπως ποσοτικές έρευνες, ποιοτικές έρευνες, μεικτές έρευνες, κύβους, δείκτες και ταξινομήσεις, μεταδεδομένα ερευνών της ελληνικής στατιστικής αρχής και των στατιστικών φορέων, σώματα κειμένων και ρέπλικες αναλύσεων (για τον ορισμό των έργων δεδομένων δείτε στην ενότητα «Τα έργα δεδομένων και οι πόροι» της παρούσας εργασίας). Παράλληλα παρέχεται η δυνατότητα online οπτικοποιήσεων μέσω περιγραφικών στατιστικών και διαγραμμάτων και παραγωγής χαρτών για σύνολα δεδομένων που περιλαμβάνουν γεωγραφικές μεταβλητές
- παροχή online εκπαιδευτικών ενοτήτων για τη διαχείριση και την ανακάλυψη δεδομένων, τη μεθοδολογία της έρευνας και την ανάπτυξη και τον σχεδιασμό ερευνητικών υποδομών
- παροχή συμβουλευτικής σε φορείς και ιδιώτες για τη δημιουργία Σχεδίων Διαχείρισης Δεδομένων –Data Management Plans (DMPs)– σχεδίων που απαιτούνται πλέον από πολλούς χρηματοδοτικούς φορείς για την υποβολή προτάσεων που περιλαμβάνουν διαχείριση δεδομένων (περισσότερα για τα DMPs βρείτε στα: CESSDA Training Team, 2017-2019· Research Data Management and Sharing, Coursera, 2018· Research Data MANTRA, University of Edinburgh, 2018· Λιναρδής & Ιωαννίδης, 2022)
- κατάθεση έργων δεδομένων τρίτων στην ερευνητική υποδομή με στόχο τη μακροπρόθεσμη διαχείρισή τους, την απόδοση μόνιμων ταυτοποιητών και τη διάχυσή τους στην ερευνητική και ακαδημαϊκή κοινότητα

Εν τέλει, η ερευνητική υποδομή SoDaNet υποστηρίζει τη διεπιστημονική έρευνα στο πεδίο των κοινωνικών επιστημών και προωθεί την απόκτηση, ανταλλαγή, επεξεργασία και διάχυση των δεδομένων που προκύπτουν από –και αφορούν την– κοινωνική έρευνα.

Η βασικότερη όμως υπηρεσία του SoDaNet είναι η παροχή πρόσβασης σε δεδομένα και μεταδεδομένα για δευτερογενή χρήση και ανάλυση μέσω του καταλόγου δεδομένων. Μέχρι πρότινος το SoDaNet χρησιμοποιούσε το λογισμικό Nesstar για την τεκμηρίωση των δεδομένων και την πρόσβαση στον κατάλογο. Τα προβλήματα του Nesstar οδήγησαν το SoDaNet στην εξεύρεση αξιόπιστων λύσεων ώστε: α) να ικανοποιούνται οι εθνικές ανάγκες ως προς την τεκμηρίωση και διαχείριση των δεδομένων και β) παράλληλα να υπάρχει συμβατότητα με τα απαιτούμενα σε ευρωπαϊκό επίπεδο, ήτοι τη CESSDA, ώστε τα δεδομένα του SoDaNet να αναζητούνται από κοινού με άλλες χώρες μέσα από τον κεντρικό κατάλογο δεδομένων της CESSDA (Cessda Data Catalogue - CDC) που είναι διαθέσιμος στη διεύθυνση <https://datacatalogue.cessda.eu/>). Η συμβατότητα σε ευρωπαϊκό επίπεδο προϋποθέτει τα μεταδεδομένα των αποθετηρίων της CESSDA να ακολουθούν το πρότυπο μεταδεδομένων DDI (Data Documentation Initiative) και ειδικότερα μία από τις τρεις εκδόσεις αυτού: το DDI 1.2.2, το DDI 2.5 ή το DDI 3.2 (η CESSDA παρέχει εργαλείο επικύρωσης των ως άνω εκδόσεων του DDI στον σύνδεσμο: <https://cmv.cessda.eu/#!validation>).

Τα κατάλληλα μεταδεδομένα εκτίθενται από το εκάστοτε αποθετήριο στον κατάλογο δεδομένων της CESSDA (στην Εικόνα 1 παρουσιάζονται κάποια έργα δεδομένων του SoDaNet μέσα από τον κατάλογο της CESSDA), ώστε να είναι

The screenshot displays the CESSDA Data Catalogue interface. At the top, it identifies the Consortium of European Social Science Data Archives. The search bar is set to Greek, and the results show 65 studies found in Greek from a total of 33767. The search filters on the left include Topic, Collection years, Country, and Publisher. The Publisher filter is selected, showing 'SoDaNet - Greek Research Infrastructure for Social Science (65)'. The search results list two studies:

- To Διαδίκτυο στην Ελλάδα 2019**: Δεμιρτζής Νίκος (Εθνικό Κέντρο Κοινωνικών Ερευνών & Εθνικά και Καποδιστριακά Πανεπιστήμια Αθηνών); Τσέκερης Χαράλαμπος (Εθνικό Κέντρο Κοινωνικών Ερευνών). 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 μεγάλης κλίμακας έρευνα με κύριο πεδίο μελέτης την καταγραφή της διεύθυνσης και των πολλαπλών επιδράσεων των διαδικτυακών υποδομών στην ...
- To Διαδίκτυο στην Κύπρο 2008**: Δεμιρτζής Νίκος (Εθνικό Κέντρο Κοινωνικών Ερευνών, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών). Η έρευνα του World Internet Project για την Κύπρο διεξήχθη στα τέλη του 2008. Ένα δείγμα 1000 Κυπρίων ερωτήθηκαν σχετικά με τη χρήση και τις στάσεις τους στο Διαδίκτυο.

Εικόνα 1: Τα αποτελέσματα του καταλόγου δεδομένων της CESSDA, με κριτήριο «Publisher», το SoDaNet

εφικτή η από κοινού αναζήτηση δεδομένων. Κατόπιν ο κατάλογος της CESSDA οδηγεί τον χρήστη στο ίδιο το αποθετήριο δεδομένων, ώστε να επιτευχθεί/ζητηθεί η πρόσβαση στα ίδια τα δεδομένα. Για να είναι δυνατή η διαλειτουργικότητα του CDC και των αποθετηρίων, τα αποθετήρια θα πρέπει να έχουν ενεργοποιημένο στον κατάλόγό τους το πρωτόκολλο OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting - <https://www.openarchives.org/pmh/>), ώστε να εκθέτουν τα δομημένα μεταδεδομένα τους στον CDC και σε άλλες πύλες (portal).

Τον Σεπτέμβριο του 2019, στο πλαίσιο της πράξης «SoDaNet in Action», το SoDaNet αποφάσισε να αντικαταστήσει το λογισμικό Nesstar με την έκδοση 4.16 του Dataverse. Παρακάτω δίνεται μια πολύ σύντομη περιγραφή του προτύπου μεταδεδομένων DDI, ώστε να διευκολυνθεί η σύνδεση με τα επόμενα, και κατόπιν παρουσιάζονται τα πλεονεκτήματα και μειονεκτήματα του Dataverse και του Nesstar, καθώς και οι φορείς που τα χρησιμοποιούν. Στο τέλος της παρούσας εργασίας παρουσιάζονται οι υλοποιήσεις του SoDaNet σχετικά με την παραμετροποίηση, βελτιστοποίηση και επέκταση των λειτουργιών του Dataverse για τις ανάγκες των αποθετηρίων του Sodanet, καθώς και οι ανάγκες για μελλοντικές υλοποιήσεις.

### Το πρότυπο DDI

Όπως αναφέρθηκε προηγουμένως, η συμβατότητα του λογισμικού με το πρότυπο μεταδεδομένων DDI αποτέλεσε βασικό κριτήριο για την επιλογή του λογισμικού που θα αντικαθιστούσε τον υφιστάμενο κατάλογο δεδομένων. Παρακάτω παρουσιάζονται οι πέντε βασικοί τομείς της δεύτερης έκδοσης του DDI<sup>1</sup> (Λιναρδής, 2006) που χρησιμοποιούνται για την τεκμηρίωση ερευνών (Η εμβόλιμη πληροφορία για το πρότυπο DDI, δίνεται ώστε ο αναγνώστης της εργασίας να κατανοήσει καλύτερα τις αποφάσεις και τις σχετικές υλοποιήσεις γύρω από τον κατάλογο δεδομένων):

- *Document Description* (docDscr - περιγραφή εγγράφου): περιέχει βιβλιογραφική πληροφορία του ίδιου του XML εγγράφου που περιγράφει την έρευνα (δηλαδή περιλαμβάνει βιβλιογραφικού τύπου μεταδεδομένα για τα μεταδεδομένα της έρευνας). Ορίζει έτσι τη βιβλιογραφική πληροφορία που περιγράφει όλο το έγγραφο, όμοια με την πληροφορία που προσθέτει ένας εκδότης σε ένα βιβλίο. Παραδείγματα πεδίων: title (περιλαμβάνει τον πλήρη τίτλο του XML εγγράφου τεκμηρίωσης, που συνήθως συμπίπτει με τον τίτλο της έρευνας με κωδικό 2.1.1.), Authoring Entity (το άτομο ή ο φορέας

---

1. Βρείτε την αναλυτική περιγραφή της δενδρικής δομής του DDI2 στη διεύθυνση: <http://www.ddialliance.org/Specification/DDI-Codebook/2.1/DTD/Documentation/DDI2-1-tree.html> Στο: <http://www.ddialliance.org/Specification/DDI-Codebook/2.1/DTD/Documentation/version2-1-all.html>, εκτός από τη δομή, εμπεριέχεται και η σημασιολογία των XML στοιχείων με παραδείγματα.

που είναι υπεύθυνος για το ουσιαστικό και νοητικό περιεχόμενο του XML εγγράφου), Producers (το άτομο ή ο φορέας που δημιούργησε το XML έγγραφο τεκμηρίωσης) κ.ά. Η περιγραφή του εγγράφου είναι προαιρετική και μπορεί να επαναλαμβάνεται (\*).

- *Study Description* (stdyDscr - περιγραφή έρευνας): κάποια ενδεικτικά μετα-δεδομένα αυτού του τομέα είναι τα ακόλουθα: ο επίσημος τίτλος έρευνας, η περίληψη της έρευνας, τα άτομα ή οι οργανισμοί που είναι υπεύθυνα για τη συλλογή των δεδομένων, τα άτομα ή οργανισμοί που είναι υπεύθυνα για τη διάχυση των δεδομένων, δείγμα και μέθοδοι δειγματοληψίας, τεκμηρίωση βαρών, χρόνος και τόπος συλλογής των δεδομένων, μονάδες ανάλυσης, περιορισμοί στη χρήση των δεδομένων, λέξεις κλειδιά, βιβλιογραφικές αναφορές κ.ά. Όσο πιο πλούσια είναι η πληροφορία σε αυτόν τον τομέα, τόσο πιο ολοκληρωμένη είναι η πληροφορία σχετικά με τον σχεδιασμό και τη συλλογή των δεδομένων της έρευνας. Η περιγραφή της έρευνας είναι υποχρεωτική και μπορεί να επαναλαμβάνεται (+).
- *Data File Description* (fileDscr - περιγραφή αρχείων δεδομένων): η περιγραφή των αρχείων δεδομένων αναφέρεται στα δεδομένα της έρευνας. Μια έρευνα μπορεί να περιέχει πολλά αρχεία δεδομένων και συνεπώς να υπάρχουν πολλές περιγραφές αρχείων δεδομένων στον συγκεκριμένο τομέα. Αποθηκεύεται πληροφορία, όπως η έκδοση του αρχείου δεδομένων, ο σύνδεσμος στο αρχείο δεδομένων, το πλήθος των μεταβλητών, το πλήθος των περιπτώσεων κ.ά. Το μεγαλύτερο μέρος της πληροφορίας δημιουργείται με αυτοματοποιημένο τρόπο κατά τη μεταφόρτωση, επεξεργασία και ενσωμάτωση (ingestion) του αρχείου, όταν χρησιμοποιείται κατάλληλο λογισμικό. Η περιγραφή των αρχείων δεδομένων (Data File Description) είναι προαιρετική και μπορεί να επαναλαμβάνεται (\*).
- *Variable Description* (dataDscr - περιγραφή μεταβλητών): σ' αυτό τον τομέα περιέχεται μια περιεκτική περιγραφή των μεταβλητών των αρχείων δεδομένων, όπως ονόματα μεταβλητών, ετικέτες, πληθυσμός στον οποίο αναφέρεται κάθε μεταβλητή, λεκτικό των ερωτήσεων, κατηγορίες/τιμές μεταβλητών, κατανομές και περιγραφικά στατιστικά των μεταβλητών, ελλείπουσες τιμές, λόγοι για τους οποίους δεν εισήχθησαν δεδομένα σε μια μεταβλητή, λεπτομέρειες για τον αλγόριθμο με τον οποίο δημιουργήθηκαν νέες μεταβλητές, οι ερωτήσεις από τις οποίες δημιουργήθηκαν οι μεταβλητές κ.ά. Είναι δυνατή επίσης η δημιουργία και η περιγραφή ομάδων μεταβλητών για το αρχείο δεδομένων. Οι ομάδες μεταβλητών περιέχουν μεταβλητές που μπορούν είτε να σχετίζονται με μια θεματική ή να αποτελούν έναν συγκεκριμένο τομέα ενός ερωτηματολογίου. Η περιγραφή των μεταβλητών είναι προαιρετική και μπορεί να επαναλαμβάνεται (\*).
- *Other Study Related Materials* (otherMat - άλλα τεκμήρια σχετικά με την

έρευνα): αυτός ο τομέας αποτελεί μια αποθήκη από ψηφιακά υλικά και περιλαμβάνει πρόσθετο υλικό για την τεκμηρίωση της έρευνας και των μεταβλητών, όπως ερωτηματολόγια, ταξινομικές λίστες, αποτελέσματα, αρχεία syntax, πληροφορία κωδικοποίησης, εγχειρίδια χρήσης, γλωσσάρια όρων, συμβουλές για τον συνεντευκτή, χάρτες, σχήματα βάσεων δεδομένων, κάρτες που χρησιμοποιήθηκαν κατά τη συνέντευξη, σχέδια συνέντευξης κ.ά. Η περιγραφή άλλων τεκμηρίων σχετικών με την έρευνα είναι προαιρετική και μπορεί να επαναλαμβάνεται (\*).

### Το λογισμικό NESSTAR

Το Nesstar είναι ένα εμπορικό λογισμικό που δημιουργήθηκε από το Νορβηγικό κέντρο για τα ερευνητικά δεδομένα, το NSD (<http://nesstar.com/>). Αποτελείται από τρία επιμέρους λογισμικά: α) το Nesstar Publisher, μια εφαρμογή που εγκαθίσταται σε προσωπικούς υπολογιστές και στην οποία ο χρήστης μπορεί να τεκμηριώσει τοπικά κυρίως ποσοτικές έρευνες βάσει του προτύπου DDI 1.2.2, να εισαγάγει τα γενικά μεταδεδομένα της έρευνας καθώς και τα ίδια τα δεδομένα από το SPSS ή από άλλα στατιστικά πακέτα, να τεκμηριώσει περαιτέρω τις μεταβλητές και τις ερωτήσεις μιας έρευνας και κατόπιν να δημοσιοποιήσει τα δεδομένα στον Nesstar Server, εφόσον διαθέτει τα κατάλληλα username και password, β) τον Nesstar Server, έναν web εξυπηρετητή που ευρετηριάζει τα δεδομένα και τα στέλνει σε web εφαρμογές, όπως το γ) Nesstar Webview, όπου ένας χρήστης που διεξάγει δευτερογενή ανάλυση μπορεί να αναζητήσει ερευνητικά δεδομένα, να τα αναλύσει online χρησιμοποιώντας περιγραφική στατιστική, πινακοποιήσεις, συσχετίσεις και παλινδρομήσεις, να δημιουργήσει online χάρτες και να κατεβάσει τοπικά τα δεδομένα και την τεκμηρίωση της έρευνας, εφόσον διαθέτει την κατάλληλη πρόσβαση (περισσότερα για το Nesstar βρείτε στα Nesstar, 2011· Nesstar, 2008). Η έκδοση 4.0 των λογισμικών είναι διαθέσιμη στον παρακάτω σύνδεσμο: <http://www.nesstar.com/software/download.html>. Επισημαίνεται ότι, μαζί με την εγκατάσταση του Nesstar Server σε κάποιον εξυπηρετητή, εγκαθίσταται και το Nesstar WebView, ενώ το Nesstar Publisher εγκαθίσταται τοπικά. Το Nesstar Publisher διατίθεται δωρεάν, ενώ ο Nesstar Server διατίθεται σε δοκιμαστική έκδοση 60 ημερών και εγκαθίσταται είτε σε λειτουργικό σύστημα Windows, είτε σε Linux.

Παρά την αποτελεσματική διαχείριση των μεταβλητών, το Nesstar αντιμετώπιζε σοβαρά προβλήματα στη διαχείριση χρηστών και ειδικότερα στη διαχείριση αρχείων, στην αναζήτηση των ερευνών, στην εξαγωγή της τεκμηρίωσης σε διάφορες γλώσσες και στην απόδοση μόνιμων ταυτοποιητών, με αποτέλεσμα να καθιστά βασικές λειτουργίες ενός αποθετηρίου προβληματικές. Χαρακτηριστικό παράδειγμα για τα προβλήματα που δημιουργεί η χρήση του Nesstar σε σχέση με τη

μη αποτελεσματική διαχείριση των πόρων-αρχείων μιας έρευνας αποτελεί η περίπτωση του Τσέχικου Αρχείου Κοινωνικών Δεδομένων (Czech Social Science Data Archive - CSDA) της CESSDA, που χρησιμοποιεί το λογισμικό Nesstar ως βασικό λογισμικό για τον κατάλογο δεδομένων (ο κατάλογος δεδομένων του CSDA είναι διαθέσιμος στη διεύθυνση: <http://nesstar.soc.cas.cz/webview/>), αλλά για τη διαχείριση των πόρων-αρχείων της έρευνας που αντιστοιχούν στον τομέα του «Other Study Materials» του DDI χρησιμοποιεί το λογισμικό Dspace, που είναι διαθέσιμο στη διεύθυνση <http://dspace.soc.cas.cz:8080/>.

Το Σλοβένικο Αρχείο από την άλλη (ADP) έχει υλοποιήσει τον βασικό του κατάλογο δεδομένων (<https://www.adp.fdv.uni-lj.si/orisi/>) μέσω προσαρμοσμένης εφαρμογής (custom), αλλά εξακολουθεί να χρησιμοποιεί παράλληλα το Nesstar για οντότητες και λειτουργίες που το Nesstar διαχειρίζεται πολύ ικανοποιητικά, όπως είναι οι μεταβλητές και η online διαχείριση και ανάλυση αυτών. Ενδεικτικά δείτε στην Εικόνα 2, τα σημεία με τα βελάκια που οδηγούν στο Nesstar. Είναι προφανές ότι το ADP έχει αντικαταστήσει μεν το Nesstar με άλλο λογισμικό για τον βασικό του κατάλογο, αλλά, αναγνωρίζοντας την αποτελεσματική διαχείριση των μεταβλητών και των αρχείων δεδομένων, διατηρεί ακόμα το Nesstar και οδηγεί τον χρήστη σε αυτό για την «περιγραφή των δεδομένων» (Data description) και την online ανάλυση (Nesstar Browser). Επιπλέον το σλοβένικο αποθετήριο παραθέτει το «Study description» και στο Nesstar, αλλά και στον βασικό του κατάλογο, με αποτέλεσμα να δημιουργείται σύγχυση στον χρήστη. Επιπροσθέτως, το Τσέχικο αλλά και το Σλοβένικο αποθετήριο δεν χρησιμοποιούν μόνιμους ταυτοποιητές, όπως τα DOI's, αφού το Nesstar δεν διαθέτει DOI μηχανισμούς παρά μόνο ad hoc μηχανισμούς ταυτοποίησης ανά έρευνα.

Από τα δύο παραπάνω παραδείγματα συμπεραίνουμε ότι το Nesstar δεν επαρκεί πλέον για την αποτελεσματική διαχείριση των πόρων και των μεταδεδομένων των αποθετηρίων και τα αποθετήρια καταφεύγουν σε λύσεις που χρησιμοποιούν δύο ή παραπάνω λογισμικά για τη διαχείριση βασικών λειτουργιών του καταλόγου τους. Επιπλέον, το γεγονός ότι το NSD σταμάτησε την υποστήριξη του λογισμικού οδήγησε τη CESSDA και τα μέλη της στην αναζήτηση αξιόπιστων λύσεων λογισμικού που θα αντικαταστήσουν το Nesstar. Τέλος, ένα πολύ βασικό μειονέκτημα αφορούσε το γεγονός ότι η δομή του εθνικού δικτύου, και κατ' επέκταση και της εθνικής ερευνητικής υποδομής, με επτά αποθετήρια (ένα για κάθε φορέα) –όπου τα δεδομένα αναζητούνται μεν από κοινού, αλλά κάθε αποθετήριο οφείλει να έχει πρόσβαση μόνο στα δικά του δεδομένα–, δεν υποστηριζόταν από το Nesstar, αφού ο χρήστης με ρόλο «Publisher» δεν γίνονταν να περιοριστεί μόνο στο δικό του αποθετήριο και είχε τη δυνατότητα να δημοσιεύει και να διαχειρίζεται έρευνες σε οποιοδήποτε αποθετήριο (φάκελο).

Καθώς η ερευνητική υποδομή SoDaNet διέθετε ήδη ένα μερικώς λειτουργικό αποθετήριο βασιζόμενο στο λογισμικό Nesstar –με γνωστά όμως μειονεκτήματα





The screenshot shows the ADP - Social Science Data Archives website. At the top, there is a logo with the letters 'adp' in a stylized font, followed by the text 'ADP - SOCIAL SCIENCE DATA ARCHIVES' and the tagline 'Analyze data! Deposit study! Promote science!'. Below this is a navigation bar with icons and text for 'USE DATA', 'DEPOSIT STUDY', 'LEARN ABOUT', and 'DISCOVER ADP'. The main content area shows the breadcrumb 'ADP Catalogue / / eb621' and the title 'EUROBAROMETER 62.1; OCTOBER - DECEMBER 2004: THE FUTURE OF THE EUROPEAN UNION, VOCATIONAL TRAINING, ENVIRONMENT, IT AT WORK, AND PUBLIC SERVICES'. There are tabs for 'Study description', 'Data description', 'Accompanying Materials', and 'Nesstar Browser'. The 'Basic Data File Description' section is visible, showing the title 'EB621 - Eurobarometer 62.1 [data file], 2004', the format '\*.por - SPSS PORTABLE', and the version 'August 2006'.

Εικόνα 2: Ο κατάλογος δεδομένων του Σλοβένικου αρχείου (ADP). Το Ευρωβαρόμετρο 62.1 διαθέσιμο στη διεύθυνση: <https://www.adp.fdv.uni-lj.si/opisi/eb621/>

όπως περιγράφηκαν παραπάνω– ξεκινήσαμε να διερευνούμε τις δυνατότητες του λογισμικού Dataverse. Σκοπός μας ήταν η ανεύρεση εναλλακτικών επιλογών που να καλύπτουν τις αποθετηριακές ανάγκες της υποδομής. Κατά τη διερεύνηση αυτή, εξετάσαμε ενδελεχώς τις λειτουργίες, τις δυνατότητες, τα πλεονεκτήματα αλλά και τα μειονεκτήματα του Dataverse σε σχέση τουλάχιστον με τις αντίστοιχες λειτουργίες που διαθέταμε ήδη μέσω του Nesstar.

### Τα πλεονεκτήματα του λογισμικού Dataverse

Το Dataverse είναι ένα λογισμικό ανοιχτού κώδικα που αναπτύχθηκε από το Ινστιτούτο για τις ποσοτικές κοινωνικές επιστήμες του Χάρβαρντ και από συνεργάτες παγκοσμίως. Αυτό σημαίνει ότι οποιοσδήποτε φορέας μπορεί να το εγκαταστήσει, να το παραμετροποιήσει ανάλογα με τις ανάγκες του αλλά και να το επανασχεδιάσει-επεκτείνει, από τη στιγμή που έχει στη διάθεσή του τον πηγαίο κώδικα της αποθετηριακής αυτής υποδομής. Σκοπός της δημιουργίας του λογισμικού είναι η επίλυση προβλημάτων που σχετίζονται με τη διάχυση των δεδομένων των κοινωνικών και πολιτικών επιστημών, όπως η ανάγκη αναγνώρισης των δημιουργών των δεδομένων, η μακροχρόνια διατήρησή τους και η δημόσια και ανοιχτή δημοσιοποίησή τους (King, 2007). Το Dataverse είναι υλοποιημένο με χρήση Java, Hibernate και JSF, ενώ για τον μηχανισμό αναζήτησης χρησιμοποιεί το λογισμικό Apache Solr. Η έκδοση Dataverse v4.16, που χρησιμοποιήθηκε για την ανάπτυξη του κα-

ταλόγου δεδομένων του SoDaNet (<https://datacatalogue.sodanet.gr/>), χρησιμοποιεί την έκδοση 7.3.x του Solr. Η βάση δεδομένων του είναι Postgresql. Τα μεταδεδομένα αναφοράς (citation), τα γεωχωρικά μεταδεδομένα (geospatial) και τα μεταδεδομένα που αφορούν τις κοινωνικές επιστήμες (social sciences and humanities metadata) είναι συμβατά με το DDI Lite και το DDI 2.5 (<https://guides.dataverse.org/en/latest/user/appendix.html>).

Το dataverse είναι γενικό αποθετηριακό λογισμικό, που υποστηρίζει πολλές επιστημονικές πειθαρχίες, όπως κοινωνικές επιστήμες, τέχνες και ανθρωπιστικές επιστήμες, περιβαλλοντικές επιστήμες, ιατρική, υγεία και επιστήμες της ζωής, νομική, γεωπονικές επιστήμες, πληροφορική, φυσική, μηχανική, αστρονομία και αστροφυσική, διοίκηση επιχειρήσεων, χημεία και μαθηματικές επιστήμες. Για κάθε μία από αυτές διαθέτει ξεχωριστό σύνολο μεταδεδομένων. Έχει τη φιλοσοφία της αυτοτεκμηρίωσης των ερευνών που φιλοξενεί. Με απλά λόγια, δίνει στον εκάστοτε χρήστη τη δυνατότητα να εγγραφεί και να δημιουργήσει το δικό του αποθετήριο με φακέλους και υποφακέλους (dataverses), μέσα στους οποίους μπορεί να εισάγει σύνολα δεδομένων με μεταδεδομένα (datasets), καθώς και αρχεία (files) εντός των συνόλων δεδομένων. Συνεπώς η ιεραρχία που υποστηρίζει η δομή του Dataverse είναι: Φάκελος > Σύνολα Δεδομένων > Αρχεία. Αυτό το χαρακτηριστικό έχει ως αποτέλεσμα την ύπαρξη ενός πολύ καλού συστήματος διαχείρισης δικαιωμάτων των χρηστών τόσο σε επίπεδο αρχείου (file), όσο και σε επίπεδο συνόλου δεδομένων (dataset) αλλά και φακέλου (dataverse). Πρακτικά, η λειτουργικότητα αυτή επιτυγχάνεται μέσω της χορήγησης ρόλων οι οποίοι καθορίζουν τις ενέργειες που επιτρέπεται να κάνουν οι χρήστες στους διάφορους φακέλους, σύνολα δεδομένων ή/και αρχεία. Κάθε ρόλος συνοδεύεται από ένα σύνολο δικαιωμάτων, τα οποία καθορίζουν τις συγκεκριμένες ενέργειες στις οποίες μπορούν να προβούν οι χρήστες. Ρόλοι και δικαιώματα μπορούν επίσης να χορηγηθούν σε ομάδες. Οι ομάδες μπορούν να οριστούν ως μια ομαδοποίηση λογαριασμών χρηστών του αποθετηρίου, μια συλλογή διευθύνσεων IP (π.χ. όλοι οι χρήστες των υπολογιστών μιας βιβλιοθήκης) ή ακόμα και μια ομαδοποίηση όλων των χρηστών, που συνδέονται χρησιμοποιώντας ένα συγκεκριμένο θεσμικό login (π.χ. όλοι όσοι συνδέονται μέσω του ιδρυματικού λογαριασμού ενός συγκεκριμένου πανεπιστημίου). Ενδεικτικά αναφέρουμε τα επίπεδα πρόσβασης κάθε ρόλου<sup>2</sup> που μπορεί να ανατεθεί και τις δυνατότητες αυτών στην έκδοση 4.16:

- *Διαχειριστής (Admin)*: στο επίπεδο αυτό ο χρήστης έχει πλήρη δικαιώματα σε έναν φάκελο και στα σύνολα δεδομένων και τα αρχεία που περιλαμβάνονται σε αυτά. Μπορεί να προσθέσει, να τροποποιήσει και να διαγράψει υποφακέλους (dataverses), σύνολα δεδομένων και μεμονωμένα αρχεία.

---

2. <https://guides.dataverse.org/en/4.16/user/dataverse-management.html>

Μπορεί να δημοσιεύσει τα σύνολα δεδομένων και τους φακέλους που είναι σε πρόχειρη (draft) μορφή και τέλος μπορεί να τροποποιήσει και να αναθέσει ο ίδιος άδειες για το φάκελο, τα σύνολα δεδομένων και τα αρχεία που περιλαμβάνονται σε αυτά.

- *Συντελεστής/συνεργάτης (contributor)*: ο χρήστης μπορεί να δει, να επεξεργαστεί και να διαγράψει πρόχειρες (drafts) εκδόσεις συνόλων δεδομένων, αλλά και να κατεβάσει αρχεία.
- *Επιμελητής (curator)*: μπορεί να προσθέσει νέους υποφακέλους και σύνολα δεδομένων, να βλέπει μη δημοσιευμένους φακέλους, να βλέπει και να δημοσιεύει μη δημοσιευμένα σύνολα δεδομένων, να τροποποιεί σύνολα δεδομένων, να κάνει λήψη αρχείων και τέλος να τροποποιεί τα επίπεδα πρόσβασης σε ένα σύνολο δεδομένων.
- *Δημιουργός έρευνας (dataset creator)*: μπορεί μόνο να προσθέσει ένα καινούριο σύνολο δεδομένων, χωρίς όμως να μπορεί να προβεί στη δημοσίευσή του (publish).
- *Δημιουργός υποφακέλου (dataverse creator)*: ο χρήστης διαθέτει μόνο δυνατότητα προσθήκης υποφακέλων, χωρίς όμως δυνατότητα δημοσιοποίησής τους και εισαγωγής δεδομένων σε αυτούς.
- *Δημιουργός έρευνας και υποφακέλου (dataset and dataverse creator)*: ο χρήστης διαθέτει τη δυνατότητα δημιουργίας υποφακέλων και συνόλων δεδομένων, χωρίς όμως τη δυνατότητα δημοσίευσής τους (publish).
- *Λήπτης αρχείων (file downloader)*: ο χρήστης διαθέτει τη δυνατότητα να κάνει λήψη αρχείων από τα σύνολα δεδομένων που εμπεριέχονται σε ένα φάκελο.
- *Μέλος (member)*: μπορεί να βλέπει φακέλους και σύνολα δεδομένων (datasets) που βρίσκονται σε πρόχειρη μορφή, αλλά και να κατεβάζει αρχεία.

Αξίζει να σημειώσουμε ότι το dataverse παρέχει τη δυνατότητα στον εκάστοτε διαχειριστή (admin) να τροποποιήσει τους παραπάνω ρόλους, να προσθαφαιρέσει δικαιώματα στους ήδη υπάρχοντες ρόλους και να δημιουργήσει καινούριους. Τα παραπάνω –δομή, ρόλοι και δικαιώματα– ήταν ιδανικά για την υλοποίηση της αποθετηριακής υποδομής του SoDaNet, που περιλαμβάνει επτά αποθετήρια δεδομένων με διακριτούς χρήστες στο καθένα, αλλά με δυνατότητες από κοινού αναζήτησης.

Επιπλέον το Dataverse διαθέτει εγγενείς μηχανισμούς για τη διαχείριση των παρακάτω λειτουργιών-κλειδιών ενός αποθετηρίου:

- Προσθήκη μεταδεδομένων και ελεγχόμενων λεξιλογίων σε επίπεδο συνόλου δεδομένων

- ΟΑΙ-ΡΜΗ στα υποστηριζόμενα πεδία από την εγκατάστασή του (<https://guides.dataverse.org/en/latest/admin/harvestserver.html>)
- Εξαγωγή σε DDI στα υποστηριζόμενα πεδία από την εγκατάστασή του
- Κατηγορίες πρόσβασης χρηστών (Access Categories)
- Μερική υποστήριξη πολυγλωσσικής διεπαφής χρήστη
- Μηχανισμό απόδοσης μόνιμου ταυτοποιητή DOI

Ένα ακόμα ελκυστικό χαρακτηριστικό Dataverse ήταν το φιλικό περιβάλλον χρήστη. Το στοιχείο αυτό είναι πολύ σημαντικό για το αποθετήριο, καθώς βελτιώνει σημαντικά την ορατότητα των δεδομένων στους εξωτερικούς χρήστες, αλλά ταυτόχρονα μειώνει τα λάθη και τον χρόνο που πρέπει να δαπανηθεί για την εκπαίδευση των χρηστών της υποδομής. Παράλληλα, αξίζει να αναφέρουμε και άλλα χαρακτηριστικά που κάνουν το Dataverse ελκυστικό:

- Τόσο η απλή όσο και η σύνθετη αναζήτηση είναι πολύ απλά εφαρμόσιμες και λειτουργικές.
- Είναι μια εφαρμογή εξ ολοκλήρου διαδικτυακή (web-based application). Αυτό σημαίνει ότι οποιοσδήποτε χρήστης μπορεί να έχει πρόσβαση στην ίδια εφαρμογή ανά πάσα στιγμή και από οποιοδήποτε μέρος, χωρίς να χρειάζεται να εγκαθιστά πρόσθετα λογισμικά στη συσκευή του. Το χαρακτηριστικό αυτό είναι πολύ σημαντικό, ειδικότερα για ένα αποθετήριο όπως αυτό του SoDaNet, αποτελούμενο από φορείς που βρίσκονται σε διαφορετικά μέρη της Ελλάδας. Με την είσοδό τους (login) στο σύστημα, οι εσωτερικοί χρήστες μπορούν να προσθέσουν δεδομένα και μεταδεδομένα, χωρίς τη χρήση άλλων λογισμικών πλην του φυλλομετρητή (browser). Σημειώνουμε ότι στο Nesstar ήταν απαραίτητη η εγκατάσταση του λογισμικού Nesstar Publisher στον τοπικό υπολογιστή, καθώς και η χρήση των κατάλληλων προτύπων τεκμηρίωσης (templates).
- Δεν υπάρχει κανένας περιορισμός σε σχέση με τους τύπους των αρχείων που μπορεί κανείς να εισαγάγει και να τεκμηριώσει. Το Dataverse μπορεί να φιλοξενήσει: βίντεο (.mpg, .avi κ.λπ.), φωτογραφίες (.bmp, .jpeg κ.λπ.), συμπιεσμένα αρχεία (.rar, .zip), μορφότυπους εξειδικευμένων λογισμικών όπως του NVivo, αρχεία δεδομένων (.xls, .csv, .por, .sav), pdf, docx κ.ά.

Η ευρεία χρήση του dataverse αποτελεί σίγουρα ένα από τα σημαντικότερα πλεονεκτήματά του. Με βάση τα επίσημα στοιχεία του [dataverse.org](http://dataverse.org), υπάρχουν 75 φορείς που χρησιμοποιούν το λογισμικό dataverse παγκοσμίως. Ανάμεσα σε αυτούς βρίσκουμε:

- Μεγάλα πανεπιστήμια, όπως το UCLA<sup>3</sup> στο Los Angeles, το Johns Hopkins<sup>4</sup> στην Βαλτιμόρη και το Harvard<sup>5</sup>, που είναι φυσικά και ο δημιουργός του λογισμικού
- Εθνικούς φορείς έρευνας, όπως το National Institute of Education<sup>6</sup> στη Σιγκαπούρη, η Science Po<sup>7</sup> στη Γαλλία, το Australian Data Archive<sup>8</sup> στην Αυστραλία κ.ά.
- Service Providers της CESSDA, όπως το DANS<sup>9</sup> στην Ολλανδία, το AUSSDA<sup>10</sup> στην Αυστρία, το CROSSDA<sup>11</sup> στην Κροατία, το SODHA<sup>12</sup> στο Βέλγιο, το LiDA<sup>13</sup> στην Λιθουανία κ.ά.
- Εξειδικευμένα αποθετήρια, όπως το Qualitative Data Repository<sup>14</sup> (QDR), το οποίο αποθηκεύει, συντηρεί και δημοσιεύει ψηφιακά δεδομένα που έχουν παραχθεί μέσω ποιοτικών μεθόδων έρευνας στις κοινωνικές επιστήμες
- Κοινοπραξίες πανεπιστημίων ή ερευνητικών φορέων όπως το Abacus Data Network<sup>15</sup> στην British Columbia του Καναδά, το Texas Data Repository<sup>16</sup> για τα πανεπιστήμια της πολιτείας του Τέξας στις Η.Π.Α. κ.ά.
- Τέλος, υπάρχουν περιπτώσεις ενσωμάτωσης του λογισμικού Dataverse σε ευρύτερα έργα και υλοποιήσεις, όπως η προσπάθεια που γίνεται από το Public Knowledge Project (PKP) για τη διάδραση του Dataverse με το Open Journal Systems (OJS), που έχει αναπτυχθεί από τα πανεπιστήμια Stanford και Simon Fraser. Από κοινού τα δύο αυτά λογισμικά ανοικτού κώδικα επιτρέπουν στους διάφορους ερευνητές να υποβάλλουν επιστημονικά άρθρα, αλλά και τα συνοδευτικά σύνολα δεδομένων, μέσω μιας υφιστάμενης διεπαφής διαχείρισης επιστημονικών περιοδικών (Castro & Garnett, 2014). Η προσπάθεια ενσωμάτωσης του Dataverse με το Archivematica στο Scholar's Portal<sup>17</sup> (Goodchild & Hurley, 2019) είναι ακόμη μια τέτοια περίπτωση.

---

3. <https://dataverse.ucla.edu>

4. <https://archive.data.jhu.edu>

5. <https://dataverse.harvard.edu>

6. <https://researchdata.nie.edu.sg>

7. <https://data.sciencespo.fr>

8. <https://dataverse.ada.edu.au>

9. <https://dataverse.nl>

10. <https://data.aussda.at>

11. <https://data.crossda.hr>

12. <https://www.sodha.be>

13. <https://lida.dataverse.lt>

14. <https://data.qdr.syr.edu>

15. <https://abacus.library.ubc.ca>

16. <https://dataverse.tdl.org>

17. <https://dataverse.scholarsportal.info>

### Τα μειονεκτήματα του Dataverse 4.16

Το Dataverse ως ένα λογισμικό αποθετηρίου γενικού σκοπού μπορεί να διαχειριστεί όλων των ειδών τα αρχεία, από εικόνες και βίντεο μέχρι απλά αρχεία κειμένου, προγραμματιστικού κώδικα κ.ά. Δεν υπάρχει όμως ιδιαίτερη μέριμνα για κάποιους τύπους αρχείων που κυριαρχούν στις ποσοτικές κοινωνικές έρευνες, όπως είναι τα πινακοποιημένα δεδομένα (tabular data) στα διάφορα δημοφιλή μορφότυπά τους (.xlsx, .sav, .por, .dta κ.α.). Κατά τη μεταφόρτωση, επεξεργασία και ενσωμάτωση των αρχείων (ingestion), δημιουργούνται αυτοματοποιημένα κάποια τεχνικά μεταδεδομένα του αρχείου, και στον τεκμηριωτή δίνεται η δυνατότητα ορισμού κάποιων λίγων μεταδεδομένων που αφορούν τη χρήση του αρχείου. Όταν όμως το πινακοποιημένο αρχείο περιέχει μεταβλητές ή/και μεταδεδομένα αυτών, το Dataverse δεν αναγνωρίζει τις μεταβλητές ως ξεχωριστές οντότητες, και στην πράξη ο εξωτερικός χρήστης πρέπει να κατεβάσει το αρχείο τοπικά για να μπορέσει να τις εντοπίσει. Βάσει των εθνικών αναγκών για την ορθή και αποτελεσματική διαχείριση ποσοτικών δεδομένων, η έκδοση 4.16 του Dataverse δεν διαθέτει τις ακόλουθες δυνατότητες:

- Δεν υπάρχει η δυνατότητα τεκμηρίωσης και αναζήτησης των μεταβλητών που εμπεριέχονται σε κάποιο αρχείο πινακοποιημένων δεδομένων.
- Δεν υπάρχει η δυνατότητα τεκμηρίωσης και αναζήτησης των ερωτήσεων από τις οποίες πηγάζουν οι μεταβλητές.
- Δεν υπάρχει η δυνατότητα ορισμού ομάδων μεταβλητών (variable groups). Μέσω αυτής της λειτουργίας, οι ερευνητές μπορούν να ομαδοποιήσουν τις μεταβλητές με διάφορα θεματικά κριτήρια και κατ' αυτό τον τρόπο να δώσουν μια καλύτερη εμπειρία πλοήγησης στα δεδομένα μιας έρευνας.
- Δεν υπάρχει η δυνατότητα δημιουργίας online στατιστικής ανάλυσης και σχετικών γραφημάτων των αρχείων εκείνων που περιέχουν μεταβλητές<sup>18</sup>.
- Αντίστοιχα, δεν υπάρχει μέριμνα για οπτικοποίηση δεδομένων με βάση γεωγραφικές μεταβλητές μέσω ενός συστήματος χαρτογράφησης γεωγραφικών πληροφοριών (GIS)<sup>19</sup>.

Άλλα μειονεκτήματα του Dataverse που αφορούν την υλοποίηση της ελληνικής ερευνητικής υποδομής είναι τα παρακάτω:

---

18. Θα πρέπει να σημειώσουμε εδώ ότι είχε ξεκινήσει από το Dataverse η προσπάθεια να καταστεί δυνατό κάτι τέτοιο, με τη χρήση του λογισμικού TwoRavens. Παρ' όλα αυτά η υλοποίηση μέχρι την έκδοση 4.16 δεν βρισκόταν σε ικανοποιητικό επίπεδο, οπότε και εγκαταλείφθηκε.

19. Το Dataverse επίσης είχε χρησιμοποιήσει τις εφαρμογές Geoconnect και WorldMap για GIS, αλλά και αυτή η υλοποίηση δεν βρισκόταν σε ικανοποιητικό επίπεδο.

- Το περιβάλλον χρήστη υποστηρίζει μεν και άλλες γλώσσες πλην των αγγλικών, αλλά όχι την ελληνική. Στα πλαίσια των ευρωπαϊκών αποθετηρίων συνηθίζεται η δίγλωσση (ή και πολύγλωσση) διάθεση του περιβάλλοντος χρήστη και των πεδίων μεταδεδομένων από τις ερευνητικές υποδομές. Επισημαίνεται πως, παρά το γεγονός ότι το Dataverse εγγενώς υποστηρίζει στο σύνολό του τα αγγλικά, εντούτοις διαθέτει εύκολα εφαρμόσιμους μηχανισμούς μετάφρασης του μεγαλύτερου τμήματος του περιβάλλοντος χρήστη για την υποστήριξη της πολυγλωσσικότητας (για παράδειγμα, δεν υποστηρίζεται η πολυγλωσσικότητα για τον μηχανισμό διόδων αναζήτησης - facets).
- Παρά το γεγονός ότι υποστηρίζεται το πρότυπο μεταδεδομένων DDI 2.5, δεν υπάρχει πλήρης συμφωνία με το πρότυπο. Υπάρχουν αρκετές παραβιάσεις που αφορούν περισσότερο τη δομή του προτύπου (το xsd αρχείο του DDI 2.5) και λιγότερο το περιεχόμενο.
- Τα πεδία δεν ταυτίζονται με το Core Metadata Model (CMM), που έχει συμφωνηθεί από τα μέλη της CESSDA (<https://zenodo.org/record/3238175>). Εντούτοις, σε αυτή την περίπτωση παρέχει μηχανισμούς για αλλαγές και προσθήκες στα πεδία μεταδεδομένων.
- Δεν υποστηρίζει τον ορισμό περιόδου εμπάργκο για τα αρχεία μιας έρευνας, με αποτέλεσμα οι υπεύθυνοι έργων που δεν έχουν ολοκληρωθεί να είναι διστακτικοί ως προς την κατάθεση της έρευνας πριν από τη δημοσιοποίηση των σχετικών αποτελεσμάτων.

### Τα έργα δεδομένων και οι πόροι

Μία από τις απαιτήσεις των χρηστών και μελών της ελληνικής υποδομής ήταν να υπάρχει διάκριση μεταξύ των συνόλων δεδομένων που φιλοξενούνται στην υποδομή. Η διάκριση αυτή αφορούσε τα κάτωθι οκτώ αντικείμενα (βλέπε εικόνα 3):

- Ποσοτικές έρευνες (Quantitative Studies)
- Ποιοτικές έρευνες (Qualitative Studies)
- Μεικτές έρευνες (Mixed Studies)
- Μεταδεδομένα Στατιστικών Ερευνών (Statistical data)
- Κύβιοι (Cubes)
- Δείκτες και Ταξινομήσεις (Indices and Classifications)
- Ρέπλικες για την αναπαραγωγή στατιστικών αναλύσεων (Replicas)
- Σώματα κειμένων (Corpora)

Από την άλλη, τα παραπάνω αντικείμενα δεν αφορούν κατ' ανάγκη Data-Sets (σύνολα δεδομένων) αλλά και σύνολα μεταδεδομένων (π.χ. Δείκτες και Ταξινομήσεις, Μεταδεδομένα Στατιστικών Ερευνών). Γι' αυτό ο όρος «Datasets» στην ελληνική

ερευνητική υποδομή αντικαταστάθηκε από τον ευρύτερο όρο «Data Projects», ήτοι έργα δεδομένων που στη γενική περίπτωση μπορεί να αφορούν είτε έργα δεδομένων, είτε έργα μεταδεδομένων. Επίσης αντικαταστήσαμε τον όρο File –που παραπέμπει σε τεχνικό όρο– με τον όρο Resource (Πόρος), που παραπέμπει περισσότερο σε αποθετηριακό όρο και αναφέρεται σε περιεχόμενο, που περιλαμβάνει, εκτός από δεδομένα, και εργαλεία (π.χ. ερωτηματολόγια, βιβλία κωδικοποίησης κ.λπ.), αλλά και σε μεταδεδομένα, που είναι απαραίτητα για την υλοποίηση ενός έργου παραγωγής δεδομένων (Κάλλας & Παπαριστείδη, 2022).



Εικόνα 3: Οι κατηγορίες των Έργων Δεδομένων (Data Projects) που υποστηρίζει το SoDaNet Dataverse μετά την επέκτασή του

Επιπλέον το Dataverse διέθετε εγγενείς ρόλους που εξυπηρετούσαν τη δομή του δικτύου και της ερευνητικής υποδομής SoDaNet. Δημιουργήθηκαν επτά βασικοί φάκελοι (Dataverses), ένας για κάθε αποθετήριο. Οι γενικοί διαχειριστές είχαν δικαιώματα διαχειριστή σε όλους τους φακέλους, και δημιουργήθηκαν τοπικοί διαχειριστές που είχαν δικαιώματα διαχειριστή μόνο στο αποθετήριο όπου ανήκαν. Στους τεκμηριωτές δόθηκε ο ρόλος του επιμελητή (curator) και μόνο για το αποθετήριο όπου ανήκαν. Οι χρήστες που εγγράφονται στο SoDaNet δεν λαμβάνουν κάποιο ρόλο και τα δικαιώματά τους αφορούν το να βλέπουν και να κατεβάζουν τοπικά όλα τα μεταδεδομένα και τα δεδομένα που δεν είναι περιορισμένης πρόσβασης.

Οι νέες επεκτάσεις/παραμετροποιήσεις/βελτιστοποιήσεις που υλοποιήθηκαν στο πλαίσιο της πράξης «SoDaNet in action» ήταν οι παρακάτω:

- Επέκταση του σχήματος μεταδεδομένων Έργου Δεδομένων με:
  - Γενικά Μεταδεδομένα.
  - Ελεγχόμενα Λεξιλόγια.
- Επέκταση του σχήματος μεταδεδομένων πόρου με:
  - Γενικά Μεταδεδομένα.
  - Ελεγχόμενα Λεξιλόγια.
- Μεταδεδομένα Μεταβλητών και Ερωτήσεων



- Επέκταση του υποσυστήματος διαχείρισης της πρόσβασης χρηστών σε πόρους
- Παραμετροποίηση και επέκταση του ΟΑΙ-ΡΜΗ
- Εξαγωγή μεταδεδομένων σε DDI 2.5
- Επαναχρησιμοποίηση εσωτερικών και εξωτερικών πόρων
- Εικαστικός και λειτουργικός επανασχεδιασμός του περιβάλλοντος χρήστη
- Πολυγλωσσικότητα
- Επέκταση του μηχανισμού αναζήτησης
- Επέκταση του μηχανισμού απόδοσης μόνιμου ταυτοποιητή DOI (Digital Object Identifier)
- Δημιουργία online εφαρμογής για τη διεξαγωγή στατιστικών αναλύσεων και χαρτογραφικών απεικονίσεων

Παρακάτω βρείτε μια αναλυτική περιγραφή όλων των ως άνω υλοποιήσεων.

### **Η προσαρμογή του Dataverse στο πλαίσιο του SoDaNet in action**

#### *Επέκταση του σχήματος μεταδεδομένων Έργου Δεδομένων*

Στο πλαίσιο του έργου δεδομένων, το προεπιλεγμένο (default) σχήμα μεταδεδομένων του Dataverse επεκτάθηκε και προσαρμόστηκε κατάλληλα ώστε να εξυπηρετήσει τις νέες ανάγκες. Για τον λόγο αυτό:

- Προστέθηκαν οι κατηγορίες των έργων δεδομένων της εικόνας 3.
- Προστέθηκαν αρκετά νέα πεδία στις υφιστάμενες κατηγορίες μεταδεδομένων.
- Δημιουργήθηκαν τρεις νέες κατηγορίες μεταδεδομένων για την τεκμηρίωση α) των Δεικτών & Ταξινομήσεων, β) των μεταδεδομένων των Στατιστικών Ερευνών, γ) των Εξωτερικών Πόρων (αφορά πόρους, αλλά για τεχνικούς λόγους η υλοποίησή του «πακέτου μεταδεδομένων» κρίθηκε ευκολότερο να τεθεί σε επίπεδο έργου δεδομένων, αλλά να εμφανίζεται ως πόρος).

Συνολικά προστέθηκαν πάνω από 150 νέα πεδία μεταδεδομένων. Οι τύποι των πεδίων μεταδεδομένων που προστέθηκαν/τροποποιήθηκαν διακρίνονται σε:

- πεδία ελεύθερου κειμένου (είτε απλού κειμένου, είτε html)
- ελεγχόμενα λεξιλόγια (Controlled Vocabularies - CV), δηλαδή λίστες τιμών με δυνατότητα μονής ή πολλαπλής επιλογής
- σύνθετα πεδία μεταδεδομένων που περιλαμβάνουν περισσότερα του ενός υποπεδία

Για την επέκταση του σχήματος μεταδεδομένων του έργου δεδομένων χρησιμοποιήθηκε και επεκτάθηκε ο εγγενής μηχανισμός του Dataverse με την ανάρτηση tsν αρχείων<sup>20</sup>.

Σε υφιστάμενα ή νέα πεδία τέθηκαν ελεγχόμενα λεξιλόγια βάσει των πρότυπων CV της CESSDA (βλέπε τα CV της CESSDA εδώ: <https://vocabularies.cessda.eu/>). Ενδεικτικά αναφέρονται ελεγχόμενα λεξιλόγια για τη μονάδα ανάλυσης, τον εκδότη (ένας από τους εκδότες είναι και το SoDaNet), τον τύπο των δεδομένων, τη διαδικασία δειγματοληψίας, τη χρονική διάσταση της συλλογής δεδομένων κ.ά.

The screenshot shows the 'Eξωτερικοί Πόροι' (External Sources) section in the Dataverse metadata editor. It features a 'Τίτλος' (Title) field with the value 'Δεδομένα για Flash Ευρωβαρόμετρ'. To the right of the title field are '+' and '-' buttons. Below the title is a 'Σημειώσεις' (Notes) field containing the text: 'Ακολουθώντας τον σύνδεσμο, ο χρήστης θα πρέπει να κάνει είσοδο με τον λογαριασμό του για να έχει πρόσβαση στα δεδομένα. Τα δεδομένα διατίθενται σε μορφές: .sav, .por και .dta'. Below the notes are two fields: 'URL' with the value 'https://search.gesis.org/research\_d' and 'Κατηγορία Πόρου' (Source Category) with a dropdown menu showing 'Δεδομένα: Μικροδεδομένα'. At the bottom is an empty 'Περιγραφή' (Description) field.

Εικόνα 4: Σύνθετα πεδία μεταδεδομένων στο Dataverse

Τα σύνθετα πεδία είναι πεδία μεταδεδομένων που ανήκουν σε μια γενικότερη ομαδοποίηση. Συμπλήρωση ενός σύνθετου πεδίου μεταδεδομένων σημαίνει συμπλήρωση τουλάχιστον ενός από τα επιμέρους υποπεδία που το απαρτίζουν. Παράδειγμα σύνθετου πεδίου είναι ο εξωτερικός πόρος ενός έργου δεδομένων, όπως παρουσιάζεται στην Εικόνα 4. Ένας εξωτερικός πόρος αποτελείται από πέντε υποπεδία: τίτλο, σημειώσεις, URL, κατηγορία πόρου και περιγραφή. Το πεδίο αυτό έχει πολλαπλότητα, δίνοντας στο χρήστη την επιλογή να προσθέσει όσες φορές χρειάζεται το σύνθετο αυτό πεδίο (η προσθήκη επιτυγχάνεται από το κουμπί «+» της εικόνας 4).

20. <https://guides.dataverse.org/en/latest/admin/metadatatcustomization.html>

### *Επέκταση του σχήματος μεταδεδομένων Πόρου*

Το περιβάλλον επεξεργασίας/τεκμηρίωσης πόρου επεκτάθηκε ώστε να περιλαμβάνει διαφορετικές λειτουργικότητες, μέσω τεσσάρων διαφορετικών καρτελών, που είναι οι κάτωθι:

- Γενικές πληροφορίες πόρου.
- Περιγραφή μεταβλητών.
- Ομάδες μεταβλητών.
- Ερωτήσεις.
- Διάταξη (layout) κύβου.

Η «περιγραφή μεταβλητών» και «οι ομάδες μεταβλητών» αφορούν πόρους με πινακοποιημένα δεδομένα με μεταβλητές. Εφόσον τα πινακοποιημένα δεδομένα περιλαμβάνουν μικροδεδομένα ερευνών, τότε έχει νόημα η τεκμηρίωση και των ερωτήσεων από τις οποίες προήλθαν οι εκάστοτε μεταβλητές, στην καρτέλα «Ερωτήσεις».

### *Γενικές πληροφορίες Πόρου*

Από την εγκατάστασή του, το Dataverse περιλαμβάνει μόνο τρία πεδία για την περιγραφή/τεκμηρίωση πόρων. Αυτά είναι: το όνομα του αρχείου (File Name), το path του αρχείου (File Path) και η περιγραφή του (Description). Για τον λόγο αυτό, και για να καλυφθούν οι ανάγκες του έργου, το περιβάλλον διαχείρισης των πόρων ανασχεδιάστηκε και υλοποιήθηκε εξαρχής Για να γίνει αυτό προστέθηκαν άλλα 9 νέα πεδία μεταδεδομένων σε επίπεδο πόρου (συνολικά 12 πεδία). Δεδομένου ότι δεν υπάρχει εγγενής μηχανισμός του Dataverse για τη διαχείριση πόρων, χρειάστηκε να επεκταθεί κατάλληλα τόσο το λογισμικό όσο και η βάση δεδομένων. Όπως και σε επίπεδο έργου δεδομένων, έτσι και στο επίπεδο του πόρου τα πεδία μεταδεδομένων διαχωρίζονται σε: α) πεδία ελεύθερου κειμένου και β) ελεγχόμενα λεξιλόγια. Στην περίπτωση των ελεγχόμενων λεξιλογίων, όπου υπάρχουν κοινές λίστες με αυτές του έργου δεδομένων, η σχεδίαση έγινε με τέτοιο τρόπο ώστε η ανάκτηση της λίστας να γίνεται με χρήση κοινού μηχανισμού, έτσι ώστε αν αλλάξει μία τιμή στη λίστα (τροποποίηση, προσθήκη ή διαγραφή), να γίνεται αυτόματα ορατή τόσο στο περιβάλλον τεκμηρίωσης του έργου δεδομένων, όσο και σε αυτό του πόρου. Τα εννέα νέα πεδία, που προστέθηκαν ήταν τα εξής:

- Τίτλος πόρου (Το Dataverse δεν υποστήριζε τον τίτλο του πόρου! Αντί για τίτλο χρησιμοποιούσε το «File Name». Όλο το νέο περιβάλλον χρήστη ανασχεδιάστηκε, ώστε να προτάσσει τον τίτλο του πόρου.)
- Γλώσσα (επιλογή από CV)
- DOI / URL

- Τύπος πόρου (επιλογή από CV)
- Σημειώσεις
- Πηγή δεδομένων
- Πρόσβαση μετά από σύνδεση/login (Yes/No)
- Σε περίοδο εμπάργκο έως ... (πεδίο ημερομηνίας όπου δηλώνεται η ημερομηνία λήξης της περιόδου εμπάργκο για τον εκάστοτε πόρο)
- Οδηγίες για την οπτικοποίηση (που εξυπηρετεί τις ανάγκες του pivoting στους κύβους)

### *Περιγραφή μεταβλητών*

Η συγκεκριμένη λειτουργικότητα εξυπηρετεί την ανάγκη επεξεργασίας και διαχείρισης των μεταβλητών ενός πόρου τον οποίο το Dataverse αντιλαμβάνεται εγγενώς ως «tabular». Συγκεκριμένα, επεκτάθηκε ο μηχανισμός μεταφόρτωσης, επεξεργασίας και ενσωμάτωσης (ingestion) του Dataverse, ώστε ο χρήστης/τεκμηριωτής να μπορεί να έχει πρόσβαση τόσο στη λίστα των μεταβλητών ενός π.χ. spss ή csv αρχείου, όσο και στην επεξεργασία τους. Ειδικότερα, επεκτάθηκε η βάση δεδομένων ώστε να μπορεί να αποθηκευτεί η περιγραφή μιας μεταβλητής με τις παρακάτω ιδιότητες (Εικόνα 5):

- Ετικέτα
- Όνομα
- Τύπος (επιλογή από CV)
- Γεωγραφική (επιλογή από CV: αφορά στη σύνδεση με συγκεκριμένους χάρτες που παρέχονται από τη γεωγραφική εφαρμογή του SoDaNet)
- Ρόλος (επιλογή από CV: ορίζεται αν αφορά στάθμιση ή ταυτοποίηση)
- Προσθετικότητα (Stock, Flow, Non Additive): ιδιότητα που αφορά κυρίως τη διαχείριση των κύβων
- Μέτρηση (επιλογή από CV, όπως: Ονομαστική/Nominal, Διατακτική/Ordinal, Κλίμακας/Scale)
- Κατηγορίες - Τιμές: καθορισμός κατηγοριών μεταβλητών με δυνατότητα καθορισμού ελλειπουσών τιμών (missing)

Κάθε μία από τις παραπάνω ιδιότητες αποτελεί μια νέα στήλη στη βάση δεδομένων και πιο συγκεκριμένα στον πίνακα datavariabale.

Για την ευκολότερη διαχείριση των μεταβλητών, ο πίνακας προβολής της λίστας μεταβλητών παρέχει τις παρακάτω δυνατότητες (Εικόνα 5):

- Αναζήτηση μεταβλητών με εισαγωγή λέξης-κλειδιού
- Φιλτράρισμα μεταβλητών ανά ιδιότητα

<span>Metadata</span> <span>Variable Description</span> <span>Variable Groups</span> <span>Questions</span> <span>Versions</span>							
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 5: Η περιγραφή των μεταβλητών στο frontend της επέκτασης της εφαρμογής

- Σελιδοποίηση με δυνατότητα ορισμού πλήθους προβαλλόμενων μεταβλητών ανά σελίδα

### *Ομάδες Μεταβλητών*

Μέσα από το κατάλληλα διαμορφωμένο διαχειριστικό περιβάλλον, ο πιστοποιημένος χρήστης μπορεί να προσθέσει, να διαγράψει και να επεξεργαστεί ομάδες μεταβλητών. Μια ομάδα μεταβλητών περιλαμβάνει (Εικόνα 6):

- Κωδικό: μοναδικό ταυτοποιητή της κάθε ομάδας (identifier)
- Ετικέτα της ομάδας μεταβλητών
- Περιγραφή: μία σύντομη περιγραφή της ομάδας
- Μεταβλητές: το σύνολο των μεταβλητών που ανήκουν στη συγκεκριμένη ομάδα. Ο χρήστης έχει τη δυνατότητα επιλογής μίας ή περισσότερων μεταβλητών από το σύνολο των μεταβλητών του αντίστοιχου πόρου. Το σύνολο των μεταβλητών δημιουργείται σε μορφή λίστας πολλαπλών επιλογών με δυνατότητα φιλτραρίσματος (variable name, variable label). Οι μεταβλητές ανακτώνται δυναμικά από τη βάση, όπως αυτές έχουν αποθηκευτεί κατά τη διαδικασία της μεταφόρτωσης, επεξεργασίας και ενσωμάτωσης (ingestion), μετά την ανάρτηση του πόρου.

Metadata	Variable Description	Variable Groups	Questions	Versions
	<ul style="list-style-type: none"> <li>▶ Δημογραφικά στοιχεία</li> <li>▶ Γενική Πρόσβαση και Χρήση</li> <li>▶ Άσχημη Διαδικτυακή Εμπειρία</li> <li>▶ Πολιτική Επάρκεια</li> <li>▶ Διαδικτυακές σχέσεις και κοινωνικό κεφάλαιο</li> <li>▶ Κοινωνική Εμπιστοσύνη και Κοινωνική Ζωή εκτός Διαδικτύου</li> <li>▼ Online Δραστηριότητες <ul style="list-style-type: none"> <li>Συχνότητα ελέγχου e-mail (q26_1)</li> <li>Γνωρίζω πώς να ρυθμίζω το με ποιους μοιράζομαι περιεχόμενο στο διαδίκτυο (π.χ. φωτογραφίες, κείμενα) (q25_3)</li> <li>Γνωρίζω πώς να δημιουργώ περιεχόμενο και να το ανεβάζω στο διαδίκτυο (q25_4)</li> <li>Συχνότητα που παίζετε παιχνίδια (q28_1)</li> <li>Συχνότητα που κατεβάζετε ή ακούτε μουσική (q28_2)</li> <li>Συχνότητα που ψάχνετε για τον ορισμό μιας λέξης (q30_1)</li> </ul> </li> </ul>			

Εικόνα 6: Ο ορισμός των ομάδων μεταβλητών στο frontend της επέκτασης της εφαρμογής

### Ερωτήσεις

Αντίστοιχα, το περιβάλλον ερωτήσεων δίνει τη δυνατότητα εισαγωγής, διαγραφής και επεξεργασίας ερωτήσεων. Η κάθε ερώτηση αποτελείται από έξι βασικές ιδιότητες (Εικόνα 7):

- Κωδικό: μοναδικό ταυτοποιητή της κάθε ερώτησης (identifier)
- Προ-ερώτηση (prequestion): πεδίο ελεύθερου κειμένου που αφορά το κείμενο που υπάρχει πριν από την κύρια ερώτηση

Metadata	Variable Description	Variable Groups	Questions	Versions
	<ul style="list-style-type: none"> <li>▶ Q01) Πόσα άτομα ζουν σε αυτό το σπίτι και είναι 15 ετών και άνω;...</li> <li>▶ Q02) Από αυτά τα άτομα (15 και άνω), πόσα είναι στο σπίτι αυτή τη στιγμή;...</li> <li>▶ Q03) Είστε μαθητής ή φοιτητής;...</li> <li>▶ Q04) Εργάζεστε;...</li> <li>▶ Q05) Ποιο είναι ακριβώς το επάγγελμά σας;...</li> <li>▶ Q06) Είστε μερικής ή πλήρους απασχόλησης;...</li> <li>▶ Q07) Είστε:...</li> <li>▶ Q07_other) Άλλο ... προσδιορίστε...</li> <li>▼ Q08) Τον καιρό αυτό χρησιμοποιείτε το διαδίκτυο;... <ul style="list-style-type: none"> <li><i>Literal:</i> Τον καιρό αυτό χρησιμοποιείτε το διαδίκτυο;</li> <li><i>Instructions:</i> ΑΠΑΝΤΑΝΕ ΟΛΟΙ</li> <li><i>Variable:</i> Χρήση Διαδικτύου (q8)</li> </ul> </li> </ul>			

Εικόνα 7: Ο ορισμός των ερωτήσεων στο frontend της επέκτασης της εφαρμογής

- Κύρια ερώτηση (literal question)
- Κείμενο μετά την κύρια ερώτηση (postquestion)
- Οδηγίες που δίνονται για την ορθή συμπλήρωση της ερώτησης (instructions)
- Τη μεταβλητή που δημιουργήθηκε από την ερώτηση. Ο χρήστης έχει τη δυνατότητα επιλογής μίας μόνο μεταβλητής (σχέση 1 προς 1 μεταξύ ερώτησης και μεταβλητής) από το σύνολο των μεταβλητών του αντίστοιχου πόρου. Το σύνολο των μεταβλητών δημιουργείται σε μορφή λίστας απλής επιλογής με δυνατότητα φιλτραρίσματος ως προς κριτήρια όπως το όνομα και η ετικέτα της μεταβλητής. Οι μεταβλητές ανακτώνται δυναμικά από τη βάση, όπως αυτές έχουν αποθηκευτεί κατά τη διαδικασία μεταφόρτωσης, επεξεργασίας και ενσωμάτωσης (ingestion), μετά την ανάρτηση του πόρου.

### *Διάταξη κύβου (Cube Layout)*

Για δεδομένα που περιλαμβάνουν μεν μεταβλητές αλλά αφορούν αθροιστικά δεδομένα (aggregated), η τεκμηρίωση επεκτάθηκε ώστε σε επίπεδο πόρου να καθορίζεται η αρχική διάταξη ενός κύβου, προσδιορίζοντας τις γραμμές του πίνακα (rows), τις στήλες του πίνακα (columns) και τις μετρήσεις (measures). Επισημαίνεται ότι η επιλογή των μεταβλητών γίνεται από τη δεξαμενή μεταβλητών που δημιουργείται μετά από τη μεταφόρτωση, επεξεργασία και ενσωμάτωση (ingestion) του αρχείου δεδομένων. Κατόπιν είναι εφικτή η αναπαράσταση του κύβου μέσω του κατάλληλου εργαλείου pivoting.

### **Επέκταση υποσυστήματος διαχείρισης πρόσβασης χρηστών σε πόρους**

Όπως αναφέρθηκε παραπάνω, στις γενικές πληροφορίες του πόρου προστέθηκαν δύο πεδία που εξειδικεύουν την πρόσβαση του χρήστη στον εκάστοτε πόρο. Αυτά είναι τα εξής:

- Πρόσβαση μετά από σύνδεση/ login (Yes/No)
- Σε περίοδο εμπάργκο έως ... (πεδίο ημερομηνίας όπου δηλώνεται η ημερομηνία λήξης της περιόδου εμπάργκο)

### *Πρόσβαση μετά από σύνδεση*

Η λειτουργικότητα αυτή σχεδιάστηκε και αναπτύχθηκε ώστε σε συγκεκριμένους αναρτημένους πόρους να έχουν πρόσβαση μόνο συνδεδεμένοι χρήστες, δηλαδή χρήστες που έχουν την κατάλληλη διαβάθμιση διασύνδεσης στον κατάλογο δε-

The screenshot shows a web interface with a navigation bar at the top containing three tabs: 'Μεταδομένα', 'Πόροι', and 'Όροι'. Below the navigation bar is a search bar with the text 'Αναζήτηση σε αυτό το έργο δεδομένων ...' and a search icon. To the right of the search bar is a dropdown menu for sorting, labeled 'Ταξινόμηση'. Below the search bar is a filter section with the text 'Φιλτράρετε κατά'. The main content area displays a list of documents, with the first four items visible. Each item includes a document icon, a title, a category, a filename, and a date. The titles are: 'Ανωνυμοποιημένη απομαγνητοφώνηση 1ης συνέντευξης', 'Ανωνυμοποιημένη απομαγνητοφώνηση 2ης συνέντευξης', 'Ανωνυμοποιημένη απομαγνητοφώνηση 3ης συνέντευξης', and 'Ανωνυμοποιημένη απομαγνητοφώνηση 4ης συνέντευξης'. The categories are 'Δεδομένα: Συνέντευξη' for the first three and 'Δεδομένα: Μικροδοδεμένα' for the fourth. The filenames are 'R\_I\_01.docx', 'R\_I\_02.docx', 'R\_I\_03.docx', and 'R\_I\_04.docx'. The dates are '16 Ιουν 2021 - 1 Αύγουστος', '16 Ιουν 2021 - 1 Αύγουστος', '16 Ιουν 2021 - 0 Αύγουστος', and '16 Ιουν 2021 - 0 Αύγουστος'. The file sizes are '23.3 KB', '15.3 KB', '15.4 KB', and '15.1 KB' respectively.

Εικόνα 8: Περιορισμένη πρόσβαση (κλειδώμα) σε πόρους για μη συνδεδεμένους χρήστες

δομένων μέσα από το Single-Sign-On (SSO) σύστημα αυθεντικοποίησης χρηστών του Sodanet. Για την ενεργοποίηση του περιορισμού ενός πόρου μόνο για συνδεδεμένους χρήστες, το άτομο που αναρτά τον πόρο πρέπει να επιλέξει στη σελίδα επεξεργασίας μεταδομένων του πόρου το πεδίο «Πρόσβαση κατόπιν Σύνδεσης». Με την επιλογή του αυτή, ο πόρος «κλειδώνεται» για τους μη συνδεδεμένους (anonymous) χρήστες, με αντίστοιχη ένδειξη στη λίστα των πόρων (βλέπε Εικόνα 8).

### *Embargo*

Η λειτουργία ορισμού ημερομηνίας embargo εξυπηρετεί την ανάγκη εμφάνισης/λήψης αναρτηθέντων πόρων σε διαφορετική (μεταγενέστερη) ημερομηνία από αυτή της ανάρτησής τους. Έτσι το άτομο που αναρτά τον πόρο μπορεί μεν να προβεί στην ανάρτηση οποιαδήποτε χρονική στιγμή, αλλά με ορισμό της ημερομηνίας embargo ο πόρος είναι διαθέσιμος για λήψη από τους τελικούς χρήστες μόνο αφού περάσει η ημερομηνία αυτή. Ο ορισμός της ημερομηνίας embargo γίνεται μέσα από το περιβάλλον εισαγωγής και επεξεργασίας πόρου, με τη χρήση κατάλληλα διαμορφωμένου πεδίου ημερομηνίας. Για το λόγο αυτό έχει γίνει η επέκταση της βάσης δεδομένων και του ανοιχτού κώδικα, ώστε να υποστηριχθεί η αποθήκευση και διαχείριση της πληροφορίας αυτής τόσο σε διαχειριστικό περιβάλλον, όσο και σε περιβάλλον πρόσβασης.

Επισημαίνεται ότι η δυνατότητα αυτή υλοποιήθηκε από το SoDaNet στην έκδοση v4.16 του Dataverse, τον Αύγουστο του 2019, ενώ η κοινότητα του Dataverse



την υλοποίησε στην έκδοση 5.8 τον Νοέμβριο του 2021 (<https://dataverse.org/blog/dataverse-software-58-release>).

Μετά τις ως άνω παρεμβάσεις- επεκτάσεις, και σε συνδυασμό με τα εγγενή επίπεδα πρόσβασης του Dataverse, υποστηρίζονται τα εξής 4 επίπεδα πρόσβασης ανά πόρο:

- 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)

Επισημαίνεται ότι για τις τρεις πρώτες κατηγορίες πρόσβασης ο χρήστης μπορεί συνδυαστικά να θέσει περίοδο εμπάργκο.

### *OAI-PMH/ Export DDI*

Η προσθήκη πολλών νέων πεδίων μεταδεδομένων έκρινε επιτακτική τη βελτίωση και επέκταση του μηχανισμού δημιουργίας xml εγγραφών τόσο σε σχέση με το πρωτόκολλο OAI-PMH, όσο και με τον εγγενή μηχανισμό δημιουργίας και εξαγωγής αρχείων xml με βάση το σχήμα DDI. Για τον λόγο αυτό, ο μηχανισμός δημιουργίας της αναπαράστασης DDI μέσω δομής xml έχει επεκταθεί ώστε να περιλαμβάνει:

- Όλα τα νέα πεδία μεταδεδομένων που έχουν προστεθεί
- Την ενσωμάτωση επιπλέον ιδιοτήτων (attributes), όπως αυτή της γλώσσας σε συγκεκριμένα πεδία. Η έλλειψη του προσδιορισμού της γλώσσας αποτελούσε τη συνηθέστερη παραβίαση συμβατότητας με το DDI 2.5.
- Αλλαγές ώστε το σχήμα που εκτίθεται μέσω του μηχανισμού OAI-PMH να είναι πλήρως συμβατό με τους κανόνες της CESSDA, βάσει του σχετικού εργαλείου επικύρωσης που διατίθεται στον σύνδεσμο: <https://cmv.cessda.eu/> και για το βασικό προφίλ του DDI 2.5

### **Επαναχρησιμοποίηση εσωτερικών και εξωτερικών πόρων**

Η αρχική εγκατάσταση του λογισμικού Dataverse δεν έχει καμία προνόηση για την επαναχρησιμοποίηση πόρων. Απαραίτητη ήταν η σχεδίαση και ανάπτυξη μηχανισμών δημιουργίας αποθήκευσης και προβολής συσχετίσεων είτε για εσωτερικούς είτε για εξωτερικούς πόρους (που υφίστανται εκτός των αποθετηρίων του SoDaNet). Και στις δύο παραπάνω περιπτώσεις επεκτάθηκε το σχήμα της βάσης,

το περιβάλλον τεκμηρίωσης έργων δεδομένων, καθώς και το περιβάλλον προβολής μεταδεδομένων του έργου δεδομένων.

### *Εσωτερικοί πόροι*

Συγκεκριμένα, για την επαναχρησιμοποίηση εσωτερικών πόρων έχει δημιουργηθεί ένα νέο πεδίο στην κατηγορία «Citation Metadata», όπου ο χρήστης μπορεί να εισαγάγει πολλαπλούς κωδικούς εσωτερικών πόρων (Comma Separated Resource Ids) (Εικόνα 9).

Εσωτερικοί πόροι 

Εικόνα 9: Καθορισμός κωδικών εσωτερικών πόρων για επαναχρησιμοποίηση

Για την εύρεση του κωδικού του κάθε πόρου, ο χρήστης χρειάζεται να ανατρέξει στην καρτέλα «Τεχνικά Μεταδεδομένα» της σελίδας του πόρου και απλώς να αντιγράψει την τιμή του πεδίου «Κωδικός Καταλόγου Δεδομένων» (Εικόνα 10).

Τεχνικά Μεταδεδομένα 	
UNF του Αρχείου	UNF:6:vwYBaUMeTS30mvcz7mIKIQ==
Αρχική Μορφή Αρχείου	SPSS Binary
Τύπος Αρχείου	Tab-Delimited
Αρχικό Αρχείο MD5	83cda388ee5ba137a59d14e6a4cc6eec
Μέγεθος	338.4 KB
Κωδικός Καταλόγου Δεδομένων	966
Λήψη Διεύθυνσης URL	Use the Download URL in a Wget command or a download manager to avoid interrupted downloads, time outs or other failures. <a href="#">User Guide - Downloading via URL</a> <a href="https://datacatalogue.sodanet.gr/api/access/datafile/966">https://datacatalogue.sodanet.gr/api/access/datafile/966</a>

Εικόνα 10: Η εξέυρεση του κωδικού για την επαναχρησιμοποίηση του πόρου

Με την αποθήκευση της διασύνδεσης εσωτερικών πόρων, ενημερώνεται άμεσα και το περιβάλλον μεταδεδομένων του έργου δεδομένων, ώστε η καρτέλα “Πόροι” να περιλαμβάνει και τους εσωτερικούς πόρους που έχουν συσχετιστεί (Εικόνα 11). Προφανώς οι συσχετισμένοι εσωτερικοί πόροι έχουν ακριβώς τα ίδια δικαιώματα πρόσβασης με αυτά που έχουν οριστεί στο έργο δεδομένων, που είναι φυσικά αναρτημένα.

The screenshot shows the SoDaNet interface with the following elements:

- Navigation:** Metadata, Resources, Terms tabs.
- Change View:** Table, Tree buttons.
- Search:** Search this data project... Search button.
- Filter by:** Filter by dropdown.
- Sort:** Sort dropdown.
- Resources List:**
  - Item 1:** Αρχείο με τα δεδομένα του πρώτου κύματος της Ελλάδας για την έρευνα VIC. Description: Το αρχείο περιλαμβάνει τα δεδομένα, όπως προέρχονται από τις ερωτήσεις του ερωτηματολογίου χωρίς τον υπολογισμό πρόσθετων δεκτών ή μεταβλητών. Resource Category: Data, Microdata. Filename: VIC-GR\_data\_1540\_15062021\_GR.tab. Data Ingestion Details: 106 Number of Variables, 1540 Number of Cases - UNF:6k2aH4Aρν7c5G2. Other Information: Tabular Data - 327.8 KB - Jun 16, 2021 - 1 Download.
  - Item 2:** Ερωτηματολόγιο πρώτου κύματος VIC στα Αγγλικά (docx). Description: Πρόκειται για master ερωτηματολόγιο της έρευνας "Αξίς στην Κρίση" για την Ελλάδα στα αγγλικά σε μορφή docx. Resource Category: Tool, Questionnaire: Structured. Filename: VIC-GR\_Questionnaire\_EN.docx. Other Information: MS Word - 62.7 KB - Jun 14, 2021 - 0 Downloads.
  - Item 3:** Ερωτηματολόγιο πρώτου κύματος VIC στα Αγγλικά (pdf). Description: Είναι το ερωτηματολόγιο της έρευνας "Αξίς στην Κρίση" για την Ελλάδα στα αγγλικά σε μορφή pdf. Resource Category: Tool, Questionnaire: Structured. Filename: VIC-GR\_Questionnaire\_EN.pdf. Other Information: Adobe PDF - 309.4 KB - Jun 14, 2021 - 0 Downloads.
  - Item 4:** Ερωτηματολόγιο πρώτου κύματος VIC στα Ελληνικά (docx). Description: Είναι το ερωτηματολόγιο της έρευνας "Αξίς στην Κρίση" για την Ελλάδα στα ελληνικά σε μορφή docx. Resource Category: Tool, Questionnaire: Structured. Filename: VIC-GR\_Questionnaire\_GR.docx. Other Information: MS Word - 192.8 KB - Jun 14, 2021 - 0 Downloads.
  - Item 5:** Ερωτηματολόγιο πρώτου κύματος VIC στα Ελληνικά (pdf). Description: Είναι το ερωτηματολόγιο της έρευνας "Αξίς στην Κρίση" για την Ελλάδα στα ελληνικά σε μορφή pdf. Resource Category: Tool, Questionnaire: Structured. Filename: VIC-GR\_Questionnaire\_GR.pdf. Other Information: Adobe PDF - 442.5 KB - Jun 14, 2021 - 0 Downloads.
- Other Resources:** 1 Resource from other Sodanet data projects.
  - Item 6:** Αρχείο με τα δεδομένα του πρώτου κύματος της Ελλάδας για την έρευνα VIC. Abstract: Το αρχείο περιλαμβάνει τα δεδομένα, όπως προέρχονται από τις ερωτήσεις του ερωτηματολογίου χωρίς τον υπολογισμό πρόσθετων δεκτών ή μεταβλητών. Resource Category: Data, Microdata. Filename: VIC-GR\_data\_1540\_15062021\_GR.tab. Data Ingestion details: 106 Number of Variables, 1540 Number of Cases - UNF:6k2aH4Aρν7c5G2. Other Information: Tabular Data - 327.8 KB - Jun 16, 2021 - 1 Download.

Εικόνα 11: Επαναχρησιμοποίηση εσωτερικού πόρου του SoDaNet, σε έργο δεδομένων

### Εξωτερικοί πόροι

Για την υλοποίηση συσχετίσεων με εξωτερικούς πόρους έχει δημιουργηθεί νέα κατηγοριοποίηση μεταδεδομένων με τίτλο «Εξωτερικοί πόροι». Όπως παρουσιάστηκε στην Εικόνα 4, η κατηγορία αυτή περιλαμβάνει ένα σύνθετο πεδίο με 5 επιμέρους υποπεδία. Το σύνθετο αυτό πεδίο είναι υλοποιημένο με πολλαπλότητα, ώστε ο χρήστης να μπορεί να εισαγάγει περισσότερους του ενός εξωτερικούς πόρους.

Κατά την αποθήκευση ενός εξωτερικού πόρου, μια νέα καρτέλα δημιουργείται στη σελίδα μεταδεδομένων της έρευνας με τίτλο «Εξωτερικοί Πόροι». Η παρουσίαση της λίστας των εξωτερικών πόρων προσομοιάζει με αυτή της καρτέλας των "Πόρων" για την καλύτερη δυνατή εμπειρία χρήση (Εικόνα 12).

The screenshot shows a web interface with a navigation bar at the top containing 'Metadata', 'Resources', 'External Resources', and 'Terms'. The 'External Resources' tab is active. Below the navigation bar, there is a list of resources, each with a 'URL' icon and a title. The first resource is 'Δεδομένα για Flash Ευρωβαρόμετρο 211' with a note about file formats (.sav, .por, .dta) and a 'Resource Category' of '-'. The following resources are questionnaires in various languages: 'Βασικό Ερωτηματολόγιο στα αγγλικά και στα γαλλικά', 'Ερωτηματολόγιο Αυστρίας', 'Ερωτηματολόγιο Βελγίου στα γαλλικά', 'Ερωτηματολόγιο Βελγίου στα ολλανδικά', 'Ερωτηματολόγιο Βουλγαρίας', 'Ερωτηματολόγιο Κύπρου', and 'Ερωτηματολόγιο Τσεχίας'. All these questionnaires have a 'Resource Category' of 'Tool: Questionnaire: Structured'.

Εικόνα 12: Εξωτερικοί πόροι από το Γερμανικό αποθετήριο GESIS που χρησιμοποιούνται για την τεκμηρίωση της ελληνικής μετάφρασης του Flash Ευρωβαρόμετρο 211

### *Περιβάλλον Χρήστη*

Το περιβάλλον χρήστη έχει ανασχεδιαστεί πλήρως –εικαστικά και λειτουργικά– ώστε να προσφέρει μια καλύτερη εμπειρία χρήστη. Για τον λόγο αυτό, έχει επεκταθεί όλος ο frontend μηχανισμός, χρησιμοποιώντας Bootstrap & JQuery τεχνολογίες, ώστε να μπορεί να υποστηρίξει:

- Την προβολή της πλατφόρμας σε όλες τις οθόνες
- Την παροχή λειτουργιών για ΑΜΕΑ
- Την ενημέρωση για τη χρήση cookies
- Την πολυγλωσσικότητα

Ειδικότερα για την υποστήριξη της πολυγλωσσικότητας, το Dataverse διαθέτει ενσωματωμένο μηχανισμό με χρήση των αρχείων ιδιοτήτων (property files). Συ-

**SOCIAL DATA NETWORK DATA CATALOGUE**

ΥΠΟΣΤΗΡΙΞΗ ΕΠΙΚΟΙΝΩΝΙΑ ΕΛΛΗΝΙΚΑ

ΑΠΟΣΤΟΛΟΣ ΛΙΝΑΡΙΔΗΣ

**ΚΑΛΩΣ ΗΛΘΑΤΕ ΣΤΟ SODANET DATA CATALOGUE**

Η ερευνητική υποδομή SoDaNet αποτελεί την Εθνική Ερευνητική Υποδομή για τις Κοινωνικές Επιστήμες.

ΑΝΑΖΗΤΗΣΤΕ ΣΤΟΝ ΚΑΤΑΛΟΓΟ...

ΣΥΝΘΕΤΗ ΑΝΑΖΗΤΗΣΗ

**ΚΑΤΗΓΟΡΙΕΣ ΕΡΓΩΝ ΔΕΔΟΜΕΝΩΝ**

Ο Κατάλογος Δεδομένων του SoDaNet περιλαμβάνει τις ακόλουθες οκτώ διαφορετικές κατηγορίες Έργων Δεδομένων, : Ποσοτικές έρευνες, Ποιοτικές έρευνες, Μεικτές έρευνες, Κύβους, Δείκτες & Ταξινομήσεις, Μεταδεδομένα Στατιστικών Ερευνών, Σώματα Κειμένων και Ρέπλικες Ποσοτικών Αναλύσεων. Επιλέξτε την κατηγορία Έργου Δεδομένων που επιθυμείτε.

ΠΟΣΟΤΙΚΕΣ ΕΡΕΥΝΕΣ ΚΥΒΟΙ ΣΤΑΤΙΣΤΙΚΕΣ ΕΡΕΥΝΕΣ ΔΕΙΚΤΕΣ & ΤΑΞΙΝΟΜΗΣΕΙΣ ΠΟΙΟΤΙΚΕΣ ΕΡΕΥΝΕΣ

Εικόνα 13: Τμήμα της αρχικής σελίδα του νέου καταλόγου δεδομένων στο <https://datacatalogue.sodanet.gr/>

γκεκριμένα, το Dataverse υποστηρίζει στο μεγαλύτερο τμήμα του την πολυγλωσσικότητα, αλλά όχι την ελληνική γλώσσα. Για το λόγο αυτό έγιναν δύο βασικές ενέργειες:

- Μεταφράστηκε η αγγλική διεπαφή στα ελληνικά μέσω του αρχείου Bundle.properties.
- Επεκτάθηκε ο μηχανισμός πολυγλωσσικότητας του ανοιχτού λογισμικού, ώστε να μπορούν να μεταφραστούν μηνύματα που υποστηρίζονταν μόνο στα αγγλικά (π.χ. δίοδοι αναζήτησης).

Ενδεικτικά βρείτε τη σελίδα σύνθετης αναζήτησης στον σύνδεσμο: <https://datacatalogue.sodanet.gr/dataverse/root/search>, μία σελίδα ενός έργου δεδομένων στον σύνδεσμο: <https://datacatalogue.sodanet.gr/dataset.xhtml?persistentId=doi:10.17903/>

FK2/DZCJTC, και μία σελίδα πόρου με μικροδεδομένα και μεταβλητές στον σύνδεσμο: <https://datacatalogue.sodanet.gr/file.xhtml?fileId=966&version=4.0>

### *Μηχανισμός Αναζήτησης*

Η έκδοση 4.16 του Dataverse, στην οποία βασίστηκε η δημιουργία του καταλόγου δεδομένων, χρησιμοποιεί ως βασικό μηχανισμό ευρετηρίασης και αναζήτησης επί των δεδομένων και μεταδεδομένων το Solr. Το Solr είναι το πιο δημοφιλές λογισμικό στην κατηγορία των μηχανών αναζήτησης ανοιχτού κώδικα (Java). Η έκδοση 7.3 του Solr παρουσιάζει μια πολύ σημαντική αδυναμία/κενό ασφαλείας<sup>21</sup>, με αποτέλεσμα την αναγκαιότητα αναβάθμισής του στην πιο πρόσφατη έκδοση (v8.9). Η αναβάθμιση του Solr:

- Έλυσε τα θέματα ασφαλείας που είχαν εντοπιστεί.
- Βελτίωσε τους χρόνους απόκρισης της μηχανής αναζήτησης.
- Διευκόλυνε την ενσωμάτωση προηγμένων αλγορίθμων αναζήτησης, όπως α) αναζήτηση με χρήση τόνων, β) αναζήτηση με χρήση stemming, γ) προηγμένη αναζήτηση σε ελληνικές εγγραφές (λαμβάνοντας υπόψη γένη, πτώσεις και αριθμούς) κ.λπ.

Οι παραπάνω επεκτάσεις στην αναζήτηση πραγματοποιήθηκαν ώστε να καλύπτουν τόσο τα πεδία μεταδεδομένων ενός έργου δεδομένων αλλά μεταβλητών.

### *DOI - Datacite*

Η επιχειρησιακή ανάγκη διασύνδεσης του καταλόγου δεδομένων με την πλατφόρμα παραγωγής και διάθεσης DOIs έκανε επιτακτική την παραμετροποίηση και επέκταση του Dataverse, ώστε:

- Να στέλνονται τα κατάλληλα μεταδεδομένα στην πλατφόρμα του Datacite μέσω του Dataverse API
- Σε δοκιμαστικό περιβάλλον (staging), να υποστηρίζεται η παραγωγή τόσο ψευδών (fake) όσο και πραγματικών DOIs στο δοκιμαστικό περιβάλλον της πλατφόρμας Datacite
- Σε παραγωγικό περιβάλλον, να υποστηρίζεται η παραγωγή πραγματικών DOIs στην παραγωγική πλατφόρμα Datacite

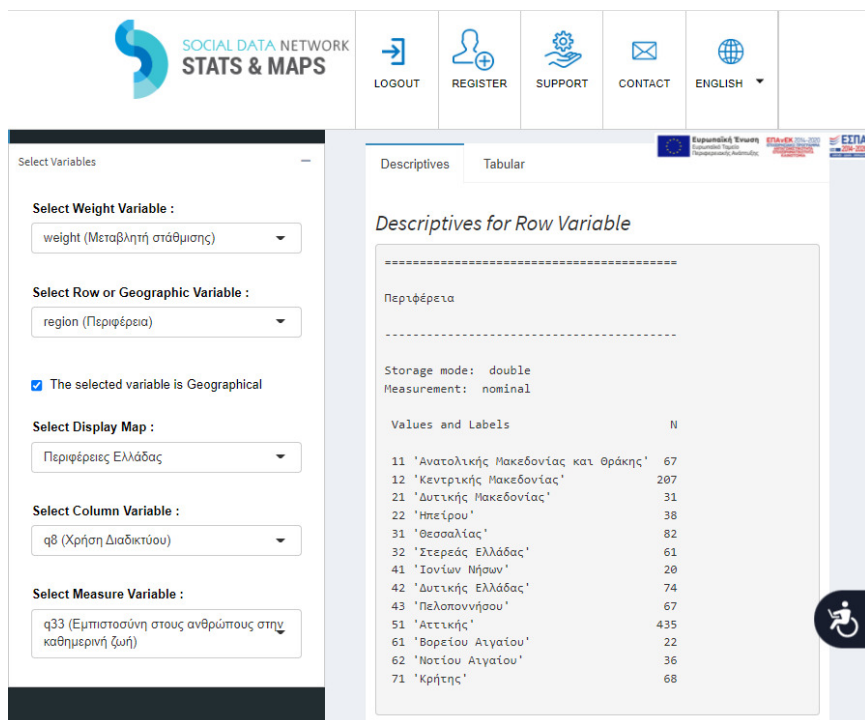
Η παραγωγή των DOIs γίνεται έπειτα από αυτοματοποιημένη διαδικασία αυθεντικοποίησης στην πλατφόρμα Datacite, με χρήση του API του Dataverse.

---

21. <https://github.com/amulcse/solr-kinsing-malware>

### Εφαρμογές Stats and Maps & Pivoting

Τέλος, υλοποιήθηκε εφαρμογή με τη χρήση των λογισμικών R-Server, R-Shiny και Geoserver για τη δημιουργία online στατιστικών αναλύσεων και χαρτογραφικών απεικονίσεων.



The screenshot shows the 'StatsAndMaps' application interface. At the top, there is a navigation bar with the 'SOCIAL DATA NETWORK STATS & MAPS' logo and buttons for LOGOUT, REGISTER, SUPPORT, CONTACT, and ENGLISH. Below the navigation bar, there are logos for the European Union, the Ministry of Education and Religious Affairs, and the NSRF (ΕΣΠΑ).

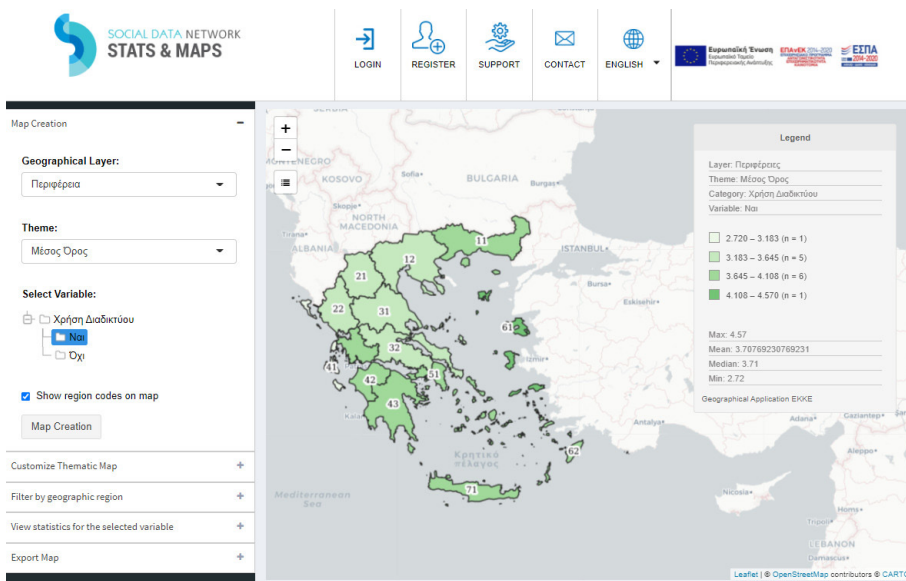
The interface is split into two main sections:

- Select Variables (Left Sidebar):**
  - Select Weight Variable:** weight (Μεταβλητή στάθμησης)
  - Select Row or Geographic Variable:** region (Περιφέρεια)
  - The selected variable is Geographical
  - Select Display Map:** Περιφέρειες Ελλάδας
  - Select Column Variable:** q8 (Χρήση Διαδικτύου)
  - Select Measure Variable:** q33 (Εμπιστοσύνη στους ανθρώπους στην καθημερινή ζωή)
- Descriptives (Main Content Area):**
  - Tabular view selected.
  - Descriptives for Row Variable**
  - Storage mode: double
  - Measurement: nominal
  - Values and Labels**

Values and Labels	N
11 'Ανατολικής Μακεδονίας και Θράκης'	67
12 'Κεντρικής Μακεδονίας'	207
21 'Δυτικής Μακεδονίας'	31
22 'Ηπείρου'	38
31 'Θεσσαλίας'	82
32 'Στερεάς Ελλάδας'	61
41 'Ιονίων Νήσων'	20
42 'Δυτικής Ελλάδας'	74
43 'Πελοποννήσου'	67
51 'Αττικής'	435
61 'Βορείου Αιγαίου'	22
62 'Νοτίου Αιγαίου'	36
71 'Κρήτης'	68

Εικόνα 14: StatsAndMaps: Online εφαρμογή για την παραγωγή περιγραφικών στατιστικών με συνδυασμό έως τριών μεταβλητών

Για αρχεία δεδομένων με μεταβλητές, είναι δυνατή πλέον η online στατιστική ανάλυση, με τη χρήση ή μη μεταβλητής στάθμησης, για την εκπόνηση περιγραφικών στατιστικών και διαγραμμάτων. Πιο συγκεκριμένα, δίνεται η δυνατότητα στον χρήστη να υπολογίσει μέτρα κεντρικής τάσης, μέτρα διασποράς, συχνότητες και σχετικές συχνότητες, τα σχετικά διαγράμματα, καθώς και διασταυρώσεις μεταβλητών (crosstabulations) για έως 3 μεταβλητές (Εικόνα 14). Για μεταβλητές που είναι γεωγραφικές και έχουν διασυνδεθεί εκ των προτέρων, από τις ιδιότητες των μεταβλητών με προφορτωμένους χάρτες, στον GeoServer, υπάρχει η δυνατότητα online χαρτογραφικών απεικονίσεων (Εικόνα 15). Κατόπιν ο χρήστης έχει τη δυνατότητα να κατεβάσει τοπικά τον χάρτη σε μορφή εικόνας καθώς και τα σχετικά δεδομένα που αναπαριστούνται στον χάρτη.



Εικόνα 15: StatsAndMaps:Χαρτογραφικές απεικονίσεις μεταβλητών

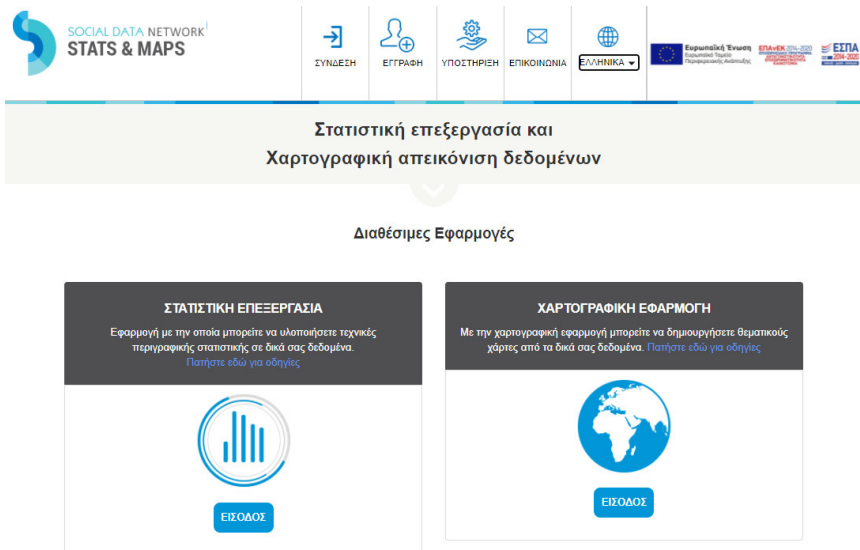
Η παραπάνω εφαρμογή ονομάστηκε «Stats and Maps» και μπορεί να επικοινωνήσει με το Dataverse, λαμβάνοντας τις κατάλληλες μεταβλητές και πληροφορία από αυτό, αλλά λειτουργεί και ως αυθύπαρκτη, αυτόνομη εφαρμογή στον σύνδεσμο <https://statsandmaps.sodanet.gr/>. Σε αυτή την περίπτωση ο χρήστης φορτώνει το δικό του σύνολο δεδομένων στην εφαρμογή και μπορεί να παραγάγει online στατιστικά και χαρτογραφικές απεικονίσεις (Εικόνα 16).

Για τους κύβους έχει δημιουργηθεί εργαλείο online pivoting (Εικόνα 17). Μέσω του εργαλείου, ο χρήστης μπορεί να έχει πρόσβαση στο pivoting της διάταξης του κύβου που έχει προταθεί από τον τεκμηριωτή, αλλά μπορεί να παρέμβει online σε αυτή τη διάταξη για να την τροποποιήσει και να δημιουργήσει τον πίνακα και τη διάταξη που επιθυμεί. Επίσης, στην περίπτωση που ο πίνακας διαθέτει στις γραμμές μεταβλητή που έχει συνδεθεί με προφορωμένους χάρτες, τότε είναι δυνατή η αναπαράσταση του κύβου σε χάρτη. Επισημαίνεται ότι στον Geoserver υπάρχουν προφορωμένοι οι παρακάτω χάρτες:

- Περιφέρειες Ελλάδας
- Περιφερειακές Ενώτητες Ελλάδας
- Δήμοι Ελλάδας
- Χάρτης Αποκεντρωμένης Διοίκησης
- Ιστορικός Χάρτης 46-64 (Γεωγραφικά διαμερίσματα, Εκλογικές Περιφέρειες, Επαρχίες, Νομοί, Δήμοι Αθηναίων και Πειραιά)



- Ευρωπαϊκός Χάρτης Χωρών
- Παγκόσμιος Χάρτης Χωρών



Εικόνα 16: StatsAndMaps:ως αυτόνομη εφαρμογή online απεικονίσεων μεταβλητών

Name	Label
Region	ΓΕΩΓΡΑΦΙΚΟ ΔΙΑΜΕΡΙΣΜΑ
Prefecture	ΝΟΜΟΣ
Province	ΕΠΑΡΧΙΑ
Gender	ΦΥΛΟ
Population	Πληθυσμός 1881

Table: Population

Sum: Population

Gender: Gender

Region	Prefecture	Province	Gender	Λήδρος	Γυναίκα	Totals	
ΘΕΣΣΑΛΙΑ ΚΑΙ ΑΡΤΑ	Άρτης	Άρτης		7.492.00	7.124.00	14.616.00	
		Τσομμερέων		8.579.00	7.983.00	16.562.00	
		Αγνιάς		6.613.00	5.805.00	12.418.00	
	Λαρίσης	Αλμυρού		5.389.00	4.864.00	10.253.00	
		Βόλου		29.376.00	27.124.00	56.500.00	
		Δομοκού και Φαραίων		9.586.00	8.133.00	17.719.00	
		Λαρίσης		18.419.00	15.853.00	34.272.00	
		Τυρνάβου		7.700.00	6.844.00	14.544.00	
	Τριυλίων	Καρδίτσας		11.121.00	9.629.00	20.750.00	
		Καρδίτσας		30.261.00	29.049.00	59.310.00	
		Τρικάλων		19.254.00	17.795.00	37.049.00	
	Totals				153.790.00	140.203.00	293.993.00

Εικόνα 17: Pivoting Κύβου: Απογραφή πληθυσμού Ελλάδας κατά το έτος 1881. Περιέχει δεδομένα για τις προσαρτηθείσες στο ελληνικό κράτος επαρχίες Ηπείρου και Θεσσαλίας από την Απογραφή πληθυσμού του Σεπτεμβρίου 1881.

## Προβλήματα και μελλοντικές υλοποιήσεις

### *Αναβάθμιση υποδομής σε επίπεδο υλικού*

Το βασικότερο πρόβλημα στην υφιστάμενη υποδομή αφορά την εγκατάσταση όλων των απαραίτητων λογισμικών σε μία μόνο εικονική μηχανή. Πιο συγκεκριμένα, η παρούσα εγκατάσταση του Dataverse έχει γίνει σε μια εικονική μηχανή που παρέχεται από την ΕΔΥΤΕ με τα παρακάτω χαρακτηριστικά:

- Λειτουργικό σύστημα: Centos [centos-release-7-9.2009.1.el7.centos.x86\_64]
- Μνήμη: 16GB
- Δίσκος: 500GB

Στην εικονική αυτή μηχανή έχουν εγκατασταθεί όλα τα απαιτούμενα λογισμικά ώστε να μπορεί να είναι παραγωγικό το αποθετήριο. Συγκεκριμένα έχουν εγκατασταθεί:

- Το ανοιχτό λογισμικό Dataverse με όλες τις επεκτάσεις του
- Εξυπηρετητής εφαρμογών Glassfish 4.x
- Βάση δεδομένων Postgres
- Μηχανή αναζήτησης Solr
- Εξυπηρετητής διαδικτύου Apache2

Το ίδιο πρόβλημα υφίσταται και για τα Geoserver και R-Server, όπου τα δύο λογισμικά έχουν εγκατασταθεί στην ίδια εικονική μηχανή. Για τον λόγο αυτό, προτείνεται μελλοντικά η κατανεμημένη εγκατάσταση ως εξής:

- Εγκατάσταση του Dataverse, με όλες τις επεκτάσεις του σε μία εικονική μηχανή
- Εγκατάσταση της βάσης δεδομένων (postgres) και του Solr σε άλλη εικονική μηχανή
- Εγκατάσταση υποδομής caching (κυρίως για στατικό περιεχόμενο και σελίδες)
- Λειτουργία υποδομής με δυνατότητες χειρισμού failover στρατηγικών
- Εγκατάσταση του R-Server σε διαφορετική εικονική μηχανή από τον Geoserver

### *Ενημέρωση θεμάτων ασφαλείας και επίλυση*

Δεδομένου ότι η δημιουργία, παραμετροποίηση και επέκταση του καταλόγου δεδομένων βασίστηκε στην έκδοση 4.16 του ανοιχτού λογισμικού, θα ήταν χρήσιμο

σε μελλοντική υλοποίηση να ενσωματωθούν στον κώδικα διαθέσιμες επιλεγμένες ενημερώσεις λογισμικού από μεταγενέστερες εκδόσεις του λογισμικού (>v4.16).

### *Επέκταση του περιβάλλοντος τεκμηρίωσης Έργου Δεδομένων και Πόρου*

Προσθήκη νέων πεδίων μεταδεδομένων σε επίπεδο έργου δεδομένων και πόρου, ώστε τα αποθετήρια του SoDaNet να είναι συμβατά με αυστηρότερες πύλες του CESSDA Validator (basic plus, standard, extended, strict) για το πρότυπο DDI 2.5. Προτείνεται ειδικότερη διαχείριση σε επίπεδο μεταδεδομένων ερωτήσεων, ώστε το αποθετήριο να είναι συμβατό και με το «European Question Bank (EQB) DDI 2.5 Profile», έτσι ώστε να είναι αναζητήσιμες στην «Ευρωπαϊκή Βάση Δεδομένων Ερωτήσεων» και οι ερωτήσεις από τις οποίες δημιουργούνται οι μεταβλητές.

### *Μετάφραση περιβάλλοντος χρήστη σε περισσότερες γλώσσες*

Υποστήριξη περισσότερων γλωσσών στους τελικούς χρήστες. Αυτή τη στιγμή το περιβάλλον διατίθεται μόνο στα ελληνικά και στα αγγλικά.

### *Προσαρμογή σε αυστηρότερους κανόνες προσβασιμότητας*

Επέκταση και προσαρμογή του αποθετηρίου, ώστε να υπακούει σε αυστηρότερους κανόνες προσβασιμότητας σε AMEA (ενδεικτικά AA και AAA, όπως ορίζεται από το πρότυπο προσβασιμότητας W3C<sup>22</sup>)

### *Διασύνδεση με τρίτα συστήματα*

- Επέκταση του μηχανισμού διαλειτουργικότητας OAI-PMH για υποστήριξη περισσότερων σχημάτων μεταδεδομένων. Αυτή τη στιγμή υποστηρίζεται πλήρως το πρότυπο DDI μέχρι και την έκδοση 2.5.
- Επέκταση του API για διασύνδεση με άλλα/τρίτα συστήματα (π.χ. Drupal, Wordpress)

## **Συμπεράσματα**

Εν κατακλείδει, στο πλαίσιο της πράξης «SoDaNet in Action», η απόφαση που λήφθηκε για την υλοποίηση του βασικού καταλόγου δεδομένων του SoDaNet ήταν η νέα εφαρμογή καταλόγου να βασιστεί στην παραμετροποίηση, βελτιστοποίηση

---

22. <https://www.w3.org/WAI/WCAG21/quickref/?showtechniques=125#text-alternatives>

και επέκταση των λειτουργιών του Dataverse. Οι επεκτάσεις, παραμετροποιήσεις και βελτιώσεις που υλοποιήθηκαν αφορούσαν: α) λειτουργίες παρόμοιες με του Nesstar για τη διαχείριση και ανάλυση των μεταβλητών και των ερωτήσεων, β) τη δυνατότητα επαναχρησιμοποίησης εσωτερικών και εξωτερικών πόρων στην τεκμηρίωση ενός έργου δεδομένων καθώς και γ) παραμετροποιήσεις και βελτιώσεις εγγενών μηχανισμών του Dataverse, όπως: το OAI-PMH, η απόδοση μόνιμων ταυτοποιητών DOI's, η προσθήκη και τροποποίηση νέων μεταδεδομένων και ελεγχόμενων λεξιλογίων και η πολυγλωσσική διεπαφή χρήστη. Θεωρούμε, ότι καταφέραμε να συμπεριλάβουμε στον νέο κατάλογο τα «ισχυρά» χαρακτηριστικά και των δύο λογισμικών καταλόγων: Dataverse και Nesstar. Ο νέος κατάλογος είναι διαθέσιμος στον σύνδεσμο: <https://datacatalogue.sodanet.gr/> Παράλληλα αναπτύξαμε το λογισμικό «Stats & Maps» για τη στατιστική και χαρτογραφική απεικόνιση των μεταβλητών βασιζόμενοι στα λογισμικά ανοιχτού κώδικα: R-Server, R-Shiny και Geoserver. Το λογισμικό αυτό επικοινωνεί με τον κατάλογο δεδομένων, ώστε να παράξει online απεικονίσεις μεταβλητών προερχόμενες από τον κατάλογο δεδομένων αλλά λειτουργεί και ως αυτόνομη εφαρμογή στη διεύθυνση: <https://statsandmaps.sodanet.gr/>.

### 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 Training Team. (2017 – 2019). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. DOI: 10.5281/zenodo.3820473. Ανακτήθηκε στις 14 Δεκεμβρίου 2021 από τη διεύθυνση: <https://www.cessda.eu/DMGuide>
- Goodchild, M., Hurley, G. (2019). Integrating Dataverse and Archivematica for research data preservation. 16th International Conference on Digital Preservation iPRES 2019, Amsterdam, The Netherlands. [https://ipres2019.org/static/pdf/iPres2019\\_paper\\_147.pdf](https://ipres2019.org/static/pdf/iPres2019_paper_147.pdf)
- Κάλλας, Γ. & Παπαριστείδη, Μ. (2022). Η ιδιαιτερότητα των υποδομών τεκμηρίωσης των κοινωνικών εμπειρικών ερευνών. Στο Γ., Κάλλας κ.α. (Επιμ.). *Ανάπτυξη υποδομών για την παραγωγή και τη διαχείριση δεδομένων στις κοινωνικές επιστήμες*.
- 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., Ιωαννίδης Α. (2022). Επεκτείνοντας τον κύκλο ζωής ποσοτικών ερευνών και δεδομένων. Ο ρόλος των αποθετηρίων δεδομένων. Στο επιμ. Κάλλας κ.α. «Τεκμηρίωση και διαχείριση εμπειρικών ερευνών. Ερευνητικές υποδομές στις κοινωνικές επιστήμες».
- King, G. (2007). An introduction to the Dataverse network as an infrastructure for data sharing. *Sociological Methods & Research*, 36(2), 173–199. <https://doi.org/10.1177/0049124107306660>
- Nesstar. (2011). User Guide of NESSTAR Publisher v4.0. Ανακτήθηκε στις 14 Δεκεμβρίου

ου 2021 από τη διεύθυνση: [http://nesstar.com/help/4.0/publisher/download\\_resources/Publisher\\_UserGuide\\_v4.0.pdf](http://nesstar.com/help/4.0/publisher/download_resources/Publisher_UserGuide_v4.0.pdf)

Nesstar. (2008). User Guide of NESSTAR Publisher v3.54. Ανακτήθηκε στις 14 Δεκεμβρίου 2021 από τη διεύθυνση: [http://nesstar.com/export/sites/default/doc/Publisher\\_UserGuide\\_v3.54.pdf](http://nesstar.com/export/sites/default/doc/Publisher_UserGuide_v3.54.pdf)

Research Data Management and Sharing. Coursera. (2018). Ανακτήθηκε στις 14 Δεκεμβρίου 2021 από τη διεύθυνση: <https://www.coursera.org/learn/data-management>.

Research Data MANTRA. University of Edinburgh. (2018). Ανακτήθηκε στις 14 Δεκεμβρίου 2021 από τη διεύθυνση: <https://mantra.edina.ac.uk/>