

Η αναγκαιότητα αυστηρής
και ποιοτικής τεκμηρίωσης
στις κοινωνικές έρευνες.
Το πρότυπο τεκμηρίωσης
Data Documentation Initiative (DDI).

Απόστολος Λιναρδής
Ερευνητής ΕΚΚΕ

Κείμενα Εργασίας 2006/14
Working Papers 2006/14

Περιεχόμενα

1.	Η αναγκαιότητα αυστηρής και ποιοτικής τεκμηρίωσης στις κοινωνικές έρευνες.....	4
2.	Τα Αρχεία Δεδομένων (ΑΔ).....	4
3.	Το πρότυπο DDI.	6
3.1.	Εισαγωγή.....	6
3.2.	Η δομή του DDI.....	7
3.3.	Το ιστορικό του DDI και των προτύπων τεκμηρίωσης.....	10
3.4.	Πλεονεκτήματα της έκδοσης DDI 2.0.....	10
3.5.	Μειονεκτήματα της έκδοσης DDI 2.0.....	11
3.6.	Η νέα έκδοση DDI 3.0.....	11
4.	Λογισμικό που χρησιμοποιεί το πρότυπο DDI.	12
4.1.	Nesstar Server.....	12
4.2.	Nesstar Publisher.....	13
4.3.	Nesstar Hierarchy Builder.....	14
4.4.	Nesstar Cube Builder.....	15
5.	Συμπεράσματα.....	17

1. Η αναγκαιότητα αυστηρής και ποιοτικής τεκμηρίωσης στις κοινωνικές έρευνες.

Η πληροφορία που παρέχεται για μεταβλητές ερευνών δεν είναι ιδιαίτερα χρήσιμη, αν δε συνοδεύεται από συμπληρωματική τεκμηρίωση, με την οποία να περιγράφεται ο πληθυσμός, το δείγμα και η μέθοδος δειγματοληψίας της έρευνας. Πολλές έρευνες χρησιμοποιούν σύνθετους σχεδιασμούς δείγματος ενώ άλλες χρησιμοποιούν βάρη που πρέπει να τεκμηριωθούν. Μερικά δεδομένα, όπως αυτά που προκύπτουν από έρευνες γνώμης, μπορεί να είναι εξαρτημένα από τον χρόνο και αυτό απαιτεί να ξέρουμε πότε συλλέχθηκαν. Οι διαχρονικές έρευνες πολλές φορές απαιτούν πολλά αρχεία των οποίων θα πρέπει να τεκμηριώνεται ο τρόπος σύνδεσης. Τα δεδομένα μπορεί να αλλοιώνονται κατά την αποθήκευση ή την μεταφορά · για αυτόν τον λόγο η καλή και αυστηρή τεκμηρίωση απαιτεί συχνότητες και περιγραφικά στατιστικά για όλες τις μεταβλητές. Η αυστηρή και καλής ποιότητας τεκμηρίωση απαιτείται γενικά για τους εξής λόγους:

- Για να καταλάβουν και για να χρησιμοποιήσουν τα δεδομένα άλλοι ερευνητές.
- Για να μπορεί ο αρχικός ερευνητής να επιστρέψει στα δεδομένα, από τη στιγμή που όλες οι λεπτομέρειες έχουν σβηστεί από τη μνήμη του.
- Για να παρέχει τη βάση για συστηματικό αθροιστικό χτίσιμο της προηγούμενης γνώσης.
- Για να είναι εφικτή η εκ των υστέρων ομογενοποίηση όπου μπορεί να εφαρμοστεί.

Ο βασικός λόγος είναι ότι δεδομένα χωρίς τεκμηρίωση καθίστανται ευτελή με την πάροδο του χρόνου. (Grant Blank, Karsten Boye Rasmussen 2004)

2. Τα Αρχεία Δεδομένων (ΑΔ).

Τα τελευταία χρόνια έχουν δημιουργηθεί οργανισμοί στην Ευρώπη και στην Αμερική, οι οποίοι ασχολούνται με τη συσσώρευση, τεκμηρίωση και διάχυση των δεδομένων. Αυτοί οι οργανισμοί ονομάζονται Αρχεία Δεδομένων (Data Archives – θα αναφερόμαστε πλέον σε αυτούς τους οργανισμούς με τη συντομογραφία ΑΔ, ώστε να αποφεύγεται η σύγχυση με τα αρχεία δεδομένων ενός υπολογιστή).

Τα ΑΔ υποστηρίζουν συνήθως την ποσοτική έρευνα και τη δευτερογενή ανάλυση. Διατηρούν συλλογές από αρχεία δεδομένων και από αρχεία τεκμηρίωσης που μπορούν να διαβάζονται από υπολογιστές, συντονίζουν την απόκτηση πρόσθετων δεδομένων από μία πληθώρα πηγών και παρέχουν πρόσβαση σε δημόσια διαθέσιμα δεδομένα. Τα ΑΔ αποτελούν κόμβους εθνικών δικτύων και διεθνών οργανισμών και διατηρούν επαφές με ερευνητές και προμηθευτές δεδομένων έτσι, ώστε αυτοί να είναι ενήμεροι των νέων συλλογών δεδομένων, τεχνικών διαχείρισης της πληροφορίας και νέων τεχνικών υπολογισμών. Χρησιμοποιούνται κυρίως από καθηγητές, φοιτητές, ερευνητές, ιδιώτες.

Ένας ευρωπαϊκός οργανισμός αρχείων είναι η CESSDA (Council of European Social Science Data Archives). Η CESSDA προωθεί την απόκτηση, αρχειοθέτηση και διάθεση ηλεκτρονικών δεδομένων για τη διδασκαλία και την έρευνα των κοινωνικών επιστημών στην Ευρώπη. Ενθαρρύνει την ανταλλαγή δεδομένων και τεχνολογίας και υποστηρίζει την ανάπτυξη νέων οργανισμών, οι οποίοι συμπλέουν με τους στόχους του συμβουλίου. Το συμβούλιο αναπτύσσει σχέσεις και συνεργάζεται με διεθνείς οργανισμούς, οι οποίοι υιοθετούν παρόμοιους στόχους. Κάποια από τα Ευρωπαϊκά ΑΔ της CESSDA είναι τα εξής:

A/A	Χώρα	Λογότυπο ΑΔ	Πλήρες Όνομα	URL
1	Αυστρία	WISDOM	Wiener Institut für Sozialwissenschaftliche Dokumentation und Methodik	http://www.wisdom.at
2	Βέλγιο	BASS	Belgian Archives for the Social Sciences	http://logi.rspo.ucl.ac.be/
3	Γαλλία		Riseau Quetelet	http://www.centre.quetelet.cnrs.fr/
4	Γερμανία	ZA	Zentralarchiv für Empirische Sozialforschung	http://www.gesis.org/ZA/
5	Δανία	DDA	Danish Data Archives	http://www.dda.dk/
6	Ελβετία	SIDOS	Swiss Information and Data Archive Service for the Social Sciences	http://www.sidos.ch/
7	Ελλάδα	GSDB	Greek Social Data Bank	http://www.gsdb.gr
8	Εσθονία	ESSDA	Estonian Social Science Data Archives	http://psych.ut.ee/esta/
9	Ηνωμένο Βασίλειο	UKDA	UK Data Archive	http://www.data-archive.ac.uk/
10	Ιρλανδία	ISSDA	Irish Social Science Data Archive	http://www.ucd.ie/issda/
11	Ισπανία	CIS	Centro de Investigaciones Sociológicas	http://www.cis.es/cis/openms/ES/index.html
12	Ιταλία	ADPSS	Archivio Dati e Programmi per le Scienze Sociali	http://www.sociologia.unimib.it/sociodata/
13	Λουξεμβούργο	CEPS	International Networks for Studies in Technology, Environment, Alternatives, Development	http://www.ceps.lu/
14	Νορβηγία	NSD	Norwegian Social Science Data Services	http://www.nsd.uib.no/
15	Ολλανδία	DANS	Data Archiving and Networked Services	http://www.niwi.knaw.nl/nl/
16	Ουγγαρία	TARKI	Social Research Informatics Center	http://www.tarki.hu/index-e.html
17	Ρουμανία	RODA	Romanian Social Data Archive	http://www.roda.ro/
18	Σλοβενία	ADP	Social Science Data Archives	http://www.adp.fdv.uni-lj.si/
19	Σουηδία	SSD	Swedish Social Science Data Services	http://www.ssd.gu.se/enghome.html
20	Τσεχία	SDA	Sociological Data Archive	http://archiv.soc.cas.cz/
21	Φινλανδία	FSD	Finnish Social Science Data Archive	http://www.fsd.uta.fi/

Επίσης ένα ιδιαίτερα αξιόλογο ΑΔ στην Αμερική είναι το Inter-university Consortium for Political and Social Research ([ICPSR](#)). Αυτό είναι ένα διεθνές ΑΔ ποσοτικών δεδομένων που εδρεύει στο Ινστιτούτο Κοινωνικής Έρευνας στο Πανεπιστήμιο του Michigan.

Τα καινούργια ΑΔ, όπως η GSDB (Greek Social Data Bank) θα πρέπει να υιοθετήσουν τεχνογνωσία από τα ήδη υπάρχοντα καθώς και τρόπους επικοινωνίας μεταξύ τους. Η στρατηγική που θα πρέπει να ακολουθηθεί περιλαμβάνει τις εξής αρχές: α) την αξιολόγηση πρώτα των προσπαθειών οι οποίες έχουν ήδη γίνει από τους προγενέστερους του χώρου και την απόφαση του ΑΔ για το τι θα υιοθετήσει και τι θα δημιουργήσει εξ' αρχής, ώστε να μπορεί να εξυπηρετεί και τις τοπικές ιδιαιτερότητες και β) την αρχή DRY (Don't Repeat Yourself) όπως αρθρώθηκε από τον Andrew Hunt και David Thomas (Hunt and Thomas 2000, σελ. 26-28), δηλ. να μην «ανακαλύψουμε τον τροχό από την αρχή».

Τα ΑΔ οφείλουν να διατηρούν τα δεδομένα αυστηρά και καλώς τεκμηριωμένα με βάση διεθνή πρότυπα, στοχεύοντας σε μια συγκριτική οπτική και παρέχοντας κατόπιν τη δυνατότητα να τα διαχέουν σε τρίτους. Ένα από τα πρότυπα που έχει επικρατήσει και που θα πρέπει να υιοθετηθεί από ένα ΑΔ, ώστε να καθίσταται δυνατή η ανταλλαγή πληροφορίας μεταξύ των ΑΔ, είναι το Data Documentation Initiative (DDI).

3. Το πρότυπο DDI.

3.1. Εισαγωγή.

Η ανταλλαγή δεδομένων μεταξύ επιστημόνων είναι μία από τις βασικές απαιτήσεις της επιστημονικής προόδου. Το κρίσιμο σε μία αποτελεσματική ανταλλαγή δεδομένων είναι η ύπαρξη τεκμηρίωσης και η πλήρης κατανόηση των δεδομένων χωρίς τη συμβολή του δημιουργού τους. Στις κοινωνικές επιστήμες ήταν ευρέως γνωστό ότι η τεκμηρίωση που συνόδευε τα σύνολα δεδομένων ήταν συνήθως ανεπαρκής.

Επιπροσθέτως η αδόμητη και μη συμβατή με υπολογιστές τεκμηρίωση ήταν αδύνατο να διαβαστεί από ηλεκτρονικούς υπολογιστές. Για να μοιραστεί και να ξαναχρησιμοποιηθεί η τεκμηρίωση των κοινωνικών δεδομένων από λογισμικό και οργανισμούς – και να γίνει μέρος του σημασιολογικού ιστού – πρέπει να χρησιμοποιεί μία τυποποιημένη γλώσσα και η πληροφορία θα πρέπει να είναι καλώς ορισμένη και δομημένη.

Παραδοσιακά οι αρχειοθέτες δεδομένων ήταν αυτοί που προσπαθούσαν να λύσουν τα προβλήματα επικοινωνίας που περιβάλλουν τη δευτερογενή ανάλυση. Ενώ οι δημιουργοί και οι πρώτοι χρήστες των στατιστικών έχουν άμεση γνώση των δεδομένων, οι δευτερογενείς χρήστες πρέπει να βασίζονται στην παρεχόμενη τεκμηρίωση. Τα μεταδεδομένα παρέχουν τον ουσιαστικό σύνδεσμο μεταξύ του πρωτογενούς υλικού και της δευτερογενούς χρήσης.

Τα ΑΔ εκτελούν επίσης συχνά «μεταφορά δεδομένων» και είναι σύνηθες το γεγονός τα δεδομένα να μη μεταφέρονται σωστά κατά τη διάδοση. Γι' αυτό τον λόγο

η τεκμηρίωση θα πρέπει να περιέχει αθροίσματα ελέγχου, συχνότητες των μεταβλητών και περιγραφικά στατιστικά για επιβεβαίωση της ακεραιότητας των δεδομένων.

Εντός της κοινότητας των κοινωνικών επιστημών υπάρχει ανάγκη για υψηλής ποιότητας τεκμηρίωσης, διότι:

- Οι δευτερογενείς ερευνητές πρέπει να καταλάβουν και να χρησιμοποιήσουν τα δεδομένα.
- Τα δεδομένα θα πρέπει να διατηρούνται και να μεταφέρονται με ακρίβεια.
- Τα δεδομένα θα πρέπει να είναι εύκολα προσβάσιμα.
- Οι κοινωνικές επιστήμες είναι αναγκαίο να γίνουν κομμάτι του σημασιολογικού ιστού .

3.2. Η δομή του DDI.

Το DDI είναι ένα διεθνές πρότυπο - βασισμένο σε XML - για την τεκμηρίωση συνόλων δεδομένων των κοινωνικών επιστημών. Είναι ένα καλά δομημένο πρότυπο για ηλεκτρονική τεκμηρίωση. Η τεκμηρίωση με μεταδεδομένα, δεδομένα για τα δεδομένα, αποτελεί την πληροφορία, ώστε να καθίσταται αποτελεσματική, επαρκής και ακριβής η χρήση αυτών των συνόλων δεδομένων.

Το DDI εκτός από την τεκμηρίωση των μετρήσεων σε ένα σύνολο από δεδομένα, όπως ονόματα και ετικέτες μεταβλητών και τιμές μεταβλητών, παρέχει πληροφορία για τη μεθοδολογία της έρευνας και πληροφορία για τον προσδιορισμό της ακριβούς θέσης αποθήκευσης ενός συνόλου δεδομένων. Ειδικότερα καλύπτοντας πληροφορίες όπως: α) ο τρόπος συλλογής των δεδομένων, β) οι μέθοδοι δειγματοληψίας που χρησιμοποιήθηκαν, γ) ο πληθυσμός και οι μονάδες, δ) η γεωγραφική και χρονική έκταση της μελέτης, ε) οι θεματικές ταξινομήσεις και οι λέξεις κλειδιά, στ) οι οργανισμοί και τα άτομα που είναι υπεύθυνα για την παραγωγή του τελικού συνόλου δεδομένων, ζ) η έκδοση του συνόλου δεδομένων που αποθηκεύεται και η) οι συνθήκες πρόσβασης για τη χρήση του συνόλου δεδομένων.

Η γλώσσα στην οποία το DDI εκφράζει την πληροφορία είναι η XML. Ο έλεγχος ορθής εισαγωγής γίνεται βάσει ενός αρχείου DTD (Document Type Definition) που ορίζει όλα τα στοιχεία και τις ιδιότητες της τεχνικής τεκμηρίωσης και τις σχέσεις μεταξύ τους (<http://www.icpsr.umich.edu/DDI/Version2-1.dtd>). Συνοδεύεται επίσης από ένα αρχείο που δείχνει το δενδρικό σχήμα του DDI (<http://www.icpsr.umich.edu/DDI/Version2-1.xsd>). Το πρότυπο DDI απαρτίζεται από πέντε βασικά μέρη:

- docDscr (περιγραφή του ίδιου του XML εγγράφου).
- stdyDscr (Αναφορές στην έρευνα).
- fileDscr (Αναφορές στα φυσικά αρχεία δεδομένων).
- dataDscr (Μεταβλητές).
- othMat (Πρόσθετο υλικό για την τεκμηρίωση της έρευνας και των μεταβλητών).

Ας θεωρήσουμε το επόμενο παράδειγμα και να δούμε πως αυτό τεκμηριώνεται μέσω του προτύπου DDI σε επίπεδο μεταβλητής:

Έστω η ερώτηση ενός ερωτηματολογίου: Ποια η εργασιακή σας κατάσταση;

- 1.Εργάζεστε τώρα
- 2.Εργαστήκατε στο παρελθόν
- 3.Δεν εργαστήκατε ποτέ.

Από την ερώτηση προκύπτει η μεταβλητή με όνομα: job_status με ετικέτα (label): “Εργασιακή κατάσταση” και συχνότητες:

	Τιμές	Συχνότητα
1	Εργάζεστε τώρα	680
2	Εργαστήκατε στο παρελθόν	851
3	Δεν εργαστήκατε ποτέ	393
	Total	1924

Η απεικόνιση της παραπάνω πληροφορίας μέσω του προτύπου DDI είναι η εξής:

```
<var ID="V1" name="job_status" dcml="0" intrvl="discrete">
  <labl>
    Εργασιακή κατάσταση
  </labl>
  <qstn>
    <qstnLit>
      Ποια η εργασιακή σας κατάσταση;
    </qstnLit>
  </qstn>
  <sumStat type="vald">
    1924
  </sumStat>
  <sumStat type="invd">
    0
  </sumStat>
  <sumStat type="min">
    1
  </sumStat>
  <sumStat type="max">
    3
  </sumStat>
  <catgry><catValu>1</catValu> <labl>Εργάζεστε τώρα </labl>
    <catStat type="freq"> 680</catStat>
  </catgry>
  <catgry><catValu>2</catValu><labl>Εργάστηκατε στο παρελθόν</labl>
    <catStat type="freq"> 851</catStat>
  </catgry>
  <catgry> <catValu>3</catValu><labl>Δεν εργαστήκατε ποτέ</labl>
    <catStat type="freq">393</catStat>
  </catgry>
  <varFormat type="numeric" schema="other"/>
</var>
```

Κάθε ένα από τα παραπάνω στοιχεία μιας μεταβλητής έχει το δικό του ξεχωριστό tag. Τα ορισμένα tags που χρησιμοποιεί το DDI καθώς και η δομή τους καθορίζονται από το αρχείο DTD (Document Type Definition). Δεν πρέπει κάποιος να συμβουλευτείται συνεχώς το DTD, ώστε να δημιουργήσει το DDI XML codebook. Υπάρχουν ειδικοί XML συντάκτες (XML editors) που βοηθούν στη δημιουργία ενός έγκυρου και καλοσχεδιασμένου XML αρχείου και που παρέχουν στον χρήστη τα κατάλληλα στοιχεία, ανάλογα με το DTD, καθώς ο χρήστης συντάσσει το codebook.

Το πρότυπο επιτρέπει επίσης:

- Αναφορές σε συνοδευτικά τεκμήρια , όπως : δημοσιεύσεις, ερωτηματολόγια, ταξινομικές λίστες, αποτελέσματα, αρχεία syntax.
- Συνδέσμους και αναφορές εντός και εκτός αρχείου με τη χρήση των ιδιοτήτων (attributes) ID, IDRefs και URI.

Κάποια ενδεικτικά tags σε επίπεδο έρευνας είναι τα εξής:

- Κυριότερος ερευνητής ή οργανισμός.
- Επίσημος τίτλος του συνόλου δεδομένων.
- Πηγές χρηματοδότησης.
- Άτομα ή οργανισμοί υπεύθυνα για τη συλλογή των δεδομένων.
- Δείγμα και μέθοδοι δειγματοληψίας.
- Τεκμηρίωση βαρών.
- Βαθμός απόκρισης.
- Ημερομηνία και τόπος συλλογής των δεδομένων.
- Χρονική περίοδος συλλογής δεδομένων.
- Μονάδες ανάλυσης.
- Περιορισμοί στη χρήση των δεδομένων.
- Βιβλιογραφικές αναφορές .

Κάποια ενδεικτικά tags σε επίπεδο μεταβλητής είναι τα εξής:

- Ακριβές λεκτικό της ερώτησης.
- Αριθμός ερώτησης.
- Όνομα μεταβλητής που προκύπτει από την ερώτηση.
- Missing values.
- Πληροφορία για τους λόγους που δεν εισήχθησαν δεδομένα για κάθε μεταβλητή.
- Λεπτομέρειες για τον τρόπο που δημιουργήθηκαν νέες τεχνητές μεταβλητές.
- Ακριβές λεκτικό όλων των πιθανών απαντήσεων.
- Κατανομές και περιγραφικά στατιστικά των μεταβλητών (μην υπολογίζοντας τα βάρη τους) .

Είναι προφανές ότι για κάποιες έρευνες θα υπάρχουν tags τα οποία θα είναι κενά, διότι το πρότυπο έχει δημιουργηθεί, ώστε να υποστηρίζει το σύνολο των ερευνών. Παρ' όλα αυτά υπάρχει μία λίστα (DDI – Lite) με τα υποχρεωτικά καθώς και με τα συστηνόμενα πεδία στην ακόλουθη διεύθυνση:

<http://www.icpsr.umich.edu/DDI/dtd/lite/ddi-lite.html>

Αναλυτική πληροφορία για τα DDI tags καθώς και για την ερμηνεία τους υπάρχει στην παρακάτω διεύθυνση:

<http://webapp.icpsr.umich.edu/cocoon/DDI-LIBRARY/Version2-1.xsd?section=all>

3.3. Το ιστορικό του DDI και των προτύπων τεκμηρίωσης.

Το χρονικό του DDI καθώς και των γνωστών προτύπων τεκμηρίωσης που χρησιμοποιήθηκαν από τα ΑΔ συνοψίζεται στον παρακάτω πίνακα:

Χρόνος	Όνομα πρότυπου τεκμηρίωσης	Οργανισμός	Παρατηρήσεις
1970	OSIRIS	Inter-university Consortium for Political and Social Research (ICPSR) and other archives.	Τεκμηρίωση σε επίπεδο μεταβλητής (όνομα, ετικέτα, πεδίο τιμών, missing values). Περιορισμένη τεκμηρίωση για την έρευνα.
Μάρτιος 2000	DDI Έκδοση 1.0	The International Association for Social Science Information Service and Technology (IASSIST) Codebook Action Group	Πλήρης τεκμηρίωση μιας έρευνας.
Ιούλιος 2003	DDI Έκδοση 2.0	The IASSIST Codebook Action Group	Καλύτερη έκδοση του DDI.
Μάρτιος 2007	DDI Έκδοση 3.0	DDI Alliance	Το DDI 3.0 φιλοδοξεί να τεκμηριώνει μία έρευνα από το στάδιο της σχεδίασης και να μην εστιάζεται μόνο στο τελικό σύνολο δεδομένων της έρευνας.

Τώρα πλέον η εξέλιξη του DDI είναι θέμα μιας συμμαχίας που δημιουργήθηκε από 27 οργανισμούς στη Βόρεια Αμερική και Ευρώπη, την DDI Alliance. Η νέα έκδοση του DDI 3.0 υπολογίζεται ότι θα είναι διαθέσιμη τον Μάρτιο του 2007.

3.4. Πλεονεκτήματα της έκδοσης DDI 2.0.

Οι παρατηρήσεις που λαμβάνονται από μία εμπειρική έρευνα μπορεί να αποτελούνται από μία μέτρηση σε κάποιο συγκεκριμένο χρονικό και χωρικό σημείο ή από επαναλαμβανόμενες μετρήσεις στον χώρο και στον χρόνο. Το DDI έχει σχεδιαστεί έτσι, ώστε να περιλαμβάνει όλα αυτά τα είδη των δεδομένων και να παρέχει στον αναλυτή όλη τη σχετική πληροφορία. Γενικά τα πλεονεκτήματα της έκδοσης 2.0 του DDI είναι τα παρακάτω:

- Η δομή με tags ενός DDI XML αρχείου είναι διαχειρίσιμη από υπολογιστή και μπορεί να παρουσιαστεί με διάφορα style sheets.
- Η εκτέλεση αναζητήσεων είναι μεγάλης ακρίβειας.
- Η λειτουργικότητα μεταξύ συστημάτων και οργανισμών γίνεται πιο εύκολη. Τα codebooks που παράγονται από το DDI μπορεί να ανταλλάσσονται και να μεταφέρονται .

-
- Η παροχή πληρέστερης εικόνας των δεδομένων στον αναλυτή.
 - Τα αρχεία DDI που παράγονται εισάγονται ευκολότερα σε συστήματα on line ανάλυσης.
 - Η συγκριτική ανάλυση γίνεται πολύ πιο εφικτή λόγω της ενσωμάτωσης συνδέσμων στο DDI XML.
 - Η προοπτική της δημιουργίας open source λογισμικού λόγω της διάδοσης του DDI ως ενός ευρέως αποδεχόμενου προτύπου. Αυτό μπορεί επίσης να βοηθήσει στη μείωση του κόστους για την παραγωγή τεκμηρίωσης καλής ποιότητας.

3.5. Μειονεκτήματα της έκδοσης DDI 2.0.

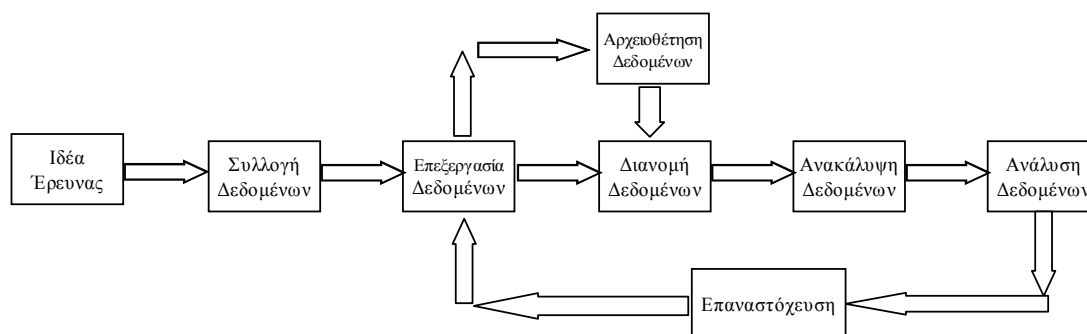
Τα μειονεκτήματα της έκδοσης 2.0 του DDI - κάποια από τα οποία θα προσπαθήσει να λύσει η καινούργια έκδοση 3.0 – είναι τα εξής:

- Δεν υπάρχει μοντέλο που να περιγράφει τις σχέσεις μεταξύ των οντοτήτων.
- Η προσέγγιση της τεκμηρίωσης μέσω ενός codebook υποφέρει από την έλλειψη μιας ξεκάθαρης ιδέας για το τι περιλαμβάνει ένα codebook.
- Δεν υπάρχει τεκμηρίωση στο στάδιο της σχεδίασης μιας έρευνας αλλά εστιάζεται μόνο στο τελικό σύνολο δεδομένων της.
- Η οντότητα της έρευνας (study) περιλαμβάνει πληροφορία που λογικά αρμόζει σε άλλες οντότητες που θα πρέπει να δημιουργηθούν [στιγμιότυπο έρευνας, έργο (η επιχειρησιακή πληροφορία μιας έρευνας)].
- Δεν υποστηρίζει τεκμηρίωση μιας έρευνας που γίνεται με τη μέθοδο της εκ των υστέρων ομογενοποίησης.
- Οι τοπικές ανάγκες δεν προσαρμόζονται σε ένα γενικό πρότυπο.
- Η εφευρετική πρωτοβουλία πρέπει να καθυστερείται έτσι ώστε να προωθείται η πληροφορία που είναι αποθηκευμένη στην υπάρχουσα δομή. Οι αναθεωρήσεις ενός προτύπου πρέπει να είναι αργές και συμφωνημένες από το σύνολο των χρηστών και επίσης να έχουμε στο μυαλό τη συμβατότητα με τις προηγούμενες εκδόσεις. Αυτός είναι ο λόγος που πολλά ΑΔ έχουν δικά τους συστήματα τεκμηρίωσης και εισάγουν ή εξάγουν τις έρευνές τους στο πρότυπο DDI, ώστε να είναι συμβατά με τα υπόλοιπα ΑΔ.

3.6. Η νέα έκδοση DDI 3.0.

Μέχρι τώρα το DDI 2.0 είχε εστιαστεί στις ανάγκες αρχειοθέτησης των συνόλων δεδομένων. Το DDI 3.0 φιλοδοξεί να τεκμηριώνει μία έρευνα από το στάδιο της σχεδίασης και να μην εστιάζεται μόνο στο τελικό σύνολο δεδομένων της έρευνας, δηλ. από την ιδέα της έρευνας μέχρι την ανάλυση και την επαναχρησιμοποίηση των

δεδομένων για δευτερογενή ανάλυση ή επαναστόχευση. Κατ' αυτόν τον τρόπο είναι δυνατό να καταγραφεί πληροφορία για όλες τις φάσεις της έρευνας. Η καινούργια έκδοση του DDI έχει υιοθετήσει πολλές από τις ιδέες του έργου Metadater. Στο παρακάτω σχήμα φαίνεται όλος ο κύκλος ζωής ενός συνόλου δεδομένων που καλείται να τεκμηριώσει η νέα έκδοση του DDI:



Εικόνα 1: Κύκλος ζωής δεδομένων έρευνας σύμφωνα με το πρότυπο DDI 3.0

4. Λογισμικό που χρησιμοποιεί το πρότυπο DDI.

Το λογισμικό που δημιουργήθηκε για να υποστηρίζει το πρότυπο DDI είναι το Nesstar. Προφανώς ένα έγγραφο DDI μπορεί να δημιουργηθεί και μέσω ενός XML συντάκτη δίχως όμως τις ευκολίες που παρέχει το Nesstar. Το Nesstar έκδοσης 3.0.1 λογισμικό αποτελείται από τα παρακάτω μέρη:

- Nesstar Server (Nesstar WebView και User Management Tool)
- Nesstar Publisher
- Nesstar Hierarchy Builder
- Nesstar Cube Builder

Το Nesstar λογισμικό είναι διαθέσιμο χωρίς χρέωση μόνο προς τα ΑΔ που είναι μέλη της CESSDA. Εκτός από το Nesstar, δημιουργήθηκε και το MADDIE λογισμικό ώστε να υποστηρίζει το πρότυπο DDI, χωρίς όμως ιδιαίτερη επιτυχία. Πάντως, η καινούργια τάση η οποία κατευθύνεται από το DDI Alliance – το συμβούλιο που καθορίζει την εξέλιξη του DDI – είναι το DDI να υποστηρίζεται από open source εργαλεία ώστε να μειώνεται το κόστος.

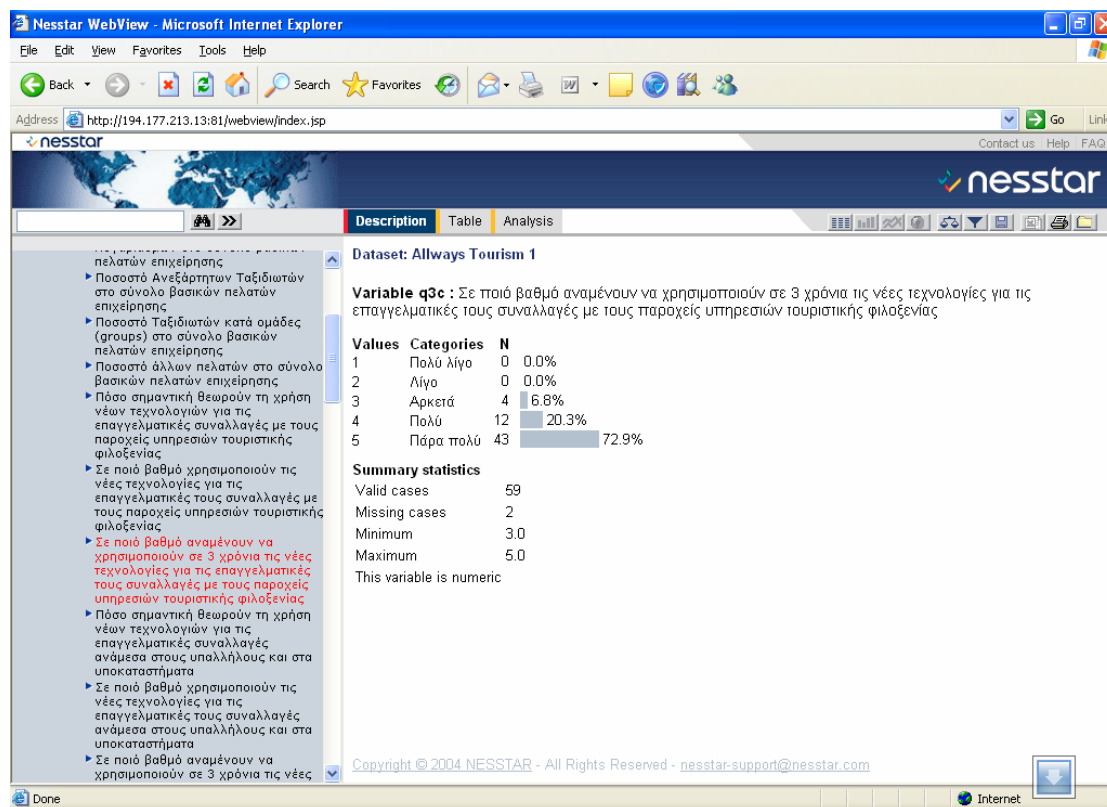
4.1. Nesstar Server

Ο Nesstar Server δημιουργήθηκε σαν επέκταση ενός κανονικού web server. Παρέχει όλες τις συνηθισμένες ευκολίες για δημοσιοποίηση web περιεχομένου και επιπροσθέτως είναι δυνατό για κάποιον να δημοσιεύσει στατιστική πληροφορία που μπορεί να αναλυθεί, να αναζητηθεί και να γίνει download από χρήστες. Αυτό γίνεται χρησιμοποιώντας έναν από τους γνωστούς browser καθώς και το Nesstar Webview.

Επίσης, το περιβάλλον του λογισμικού Nesstar παρέχει έναν τρόπο, ώστε να κάνει τα στατιστικά δεδομένα εύκολα στη δημοσιοποίηση στο web, προσβάσιμα και

προστατευμένα από χρήστες με μη εξουσιοδοτημένη πρόσβαση. Αυτό ρυθμίζεται από το περιβάλλον «User Management Tool» του Nesstar Server.

Το προϊόν Nesstar αναπαριστά ένα σύστημα αρχιτεκτονικής λογισμικού που κάνει εύκολη τη δημιουργία, αποθήκευση και πρόσβαση σε μεταδεδομένα και δεδομένα που βρίσκονται σε απομακρυσμένα συστήματα.

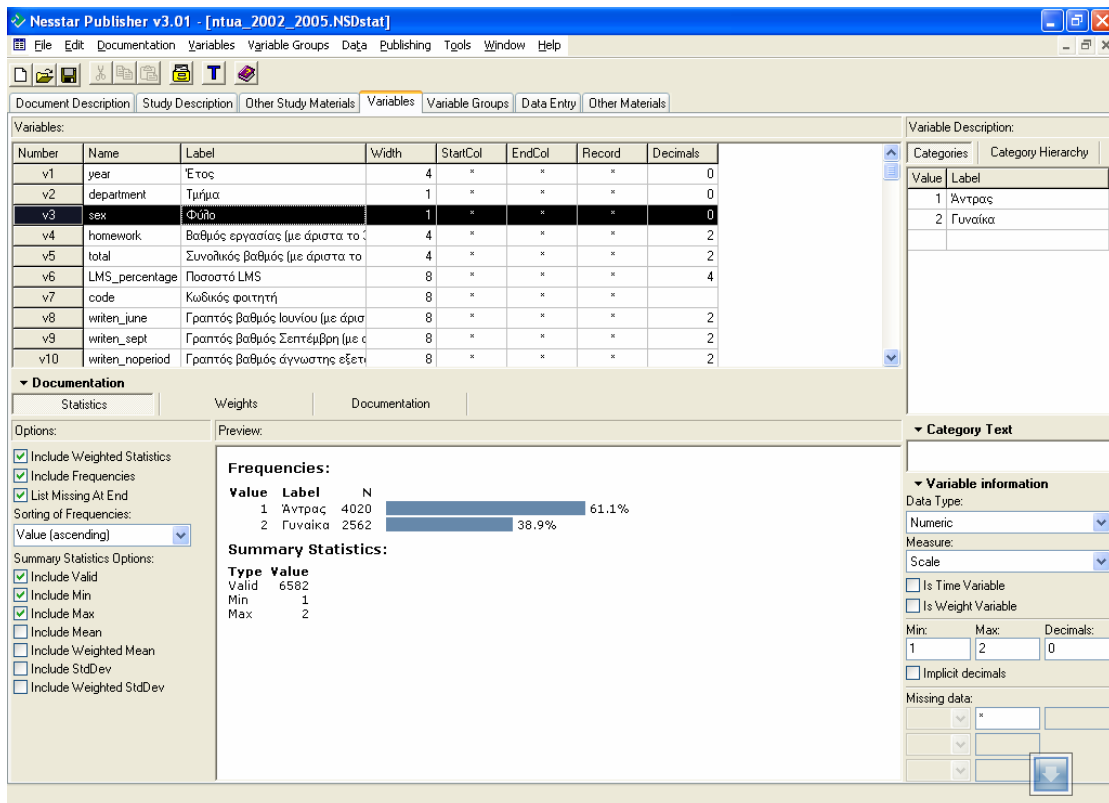


Εικόνα 2: Nesstar Server (WebView)

4.2. Nesstar Publisher

Το Nesstar Publisher χρησιμοποιείται για να φορτώσει δεδομένα και μεταδεδομένα στον Nesstar server. Η πρόσβαση σε αυτές τις πηγές μπορεί κατόπιν να είναι ελεύθερη σε όλον τον κόσμο ή μπορεί να περιορίζεται για συγκεκριμένους χρήστες και ανάλογα με τις απαιτήσεις του παροχέα των δεδομένων. Τότε όλα τα υλικά είναι έτοιμα να χρησιμοποιηθούν και να αναζητηθούν χρησιμοποιώντας το Nesstar WebView.

Ένα από τα ισχυρά χαρακτηριστικά του Nesstar Publisher είναι ότι μπορεί να εισάγει αυτόματα μεταδεδομένα και απλά στατιστικά από αρχεία δεδομένων στατιστικών πακέτων ή βάσεων δεδομένων όπως: SPSS (sav, por, sps), SAS (sp1), STATA (dta – STATA 7 και STATA 8), Statistica (Sta), dBase (dbf), Text αρχεία (txt, csv). Κατ' αυτόν τον τρόπο ο χρήστης αποφορτίζεται από μεγάλο όγκο εργασίας. (Τα μεταδεδομένα της παραγράφου 3.2 παράχθηκαν όλα αυτόματα από το Nesstar Publisher, αφότου έγινε εισαγωγή αρχείου SPSS, πλην του λεκτικού της ερώτησης).



Εικόνα 3: Nesstar Publisher

Το Nesstar Publisher είναι ένα εργαλείο συγγραφής δεδομένων – μεταδεδομένων που προετοιμάζει και δημοσιεύει έρευνες και κύβους. Ο χρήστης δεν απαιτείται να έχει ιδιαίτερη εξοικείωση με τη γλώσσα XML ώστε να εξάγει μία έρευνα στο πρότυπο DDI. Το Nesstar Publisher χρησιμοποιείται ώστε:

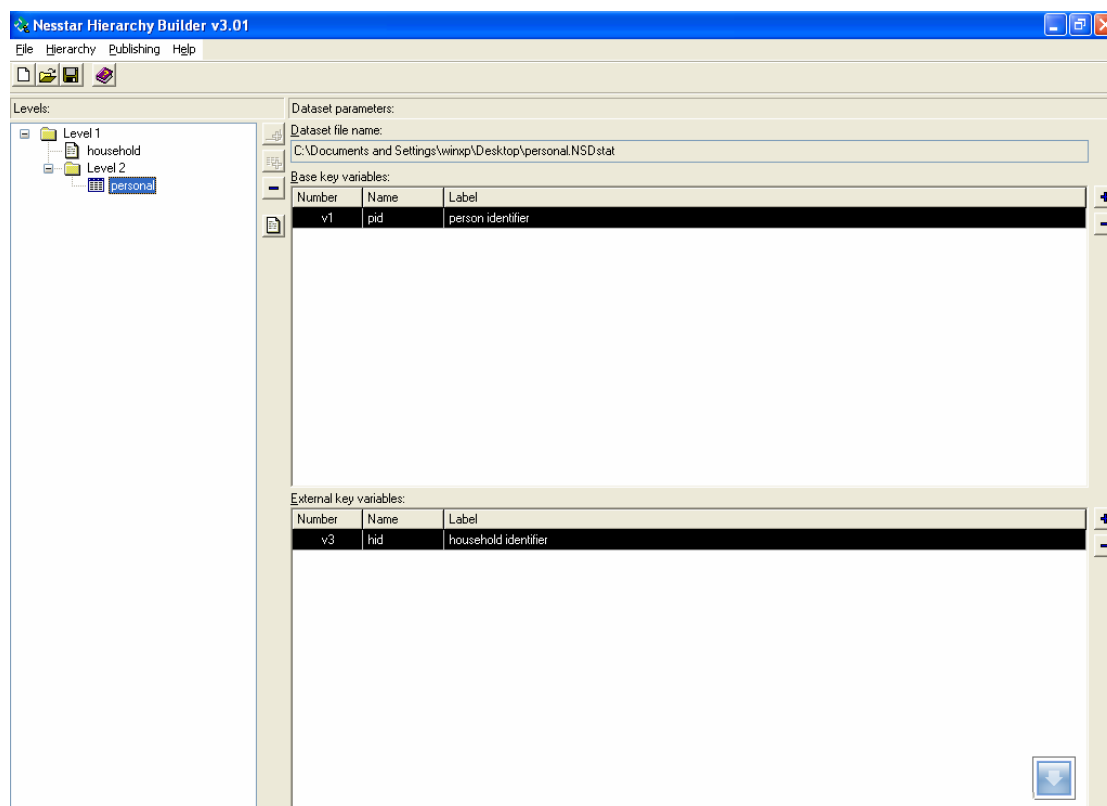
- Να εξάγει την τεκμηρίωση μιας έρευνας σε πρότυπο DDI.
- Να προσθέσει μεταδεδομένα σε μία έρευνα ή κύβο.
- Να δημιουργήσει κύβους.
- Να φορτώσει ένα σύνολο δεδομένων σε έναν κατάλογο του Nesstar Server.
- Να δημοσιεύσει ένα σύνολο δεδομένων σε έναν Nesstar WebView client.

4.3. Nesstar Hierarchy Builder

Το Nesstar Hierarchy Builder μπορεί να χρησιμοποιηθεί σε συνδυασμό με το Nesstar Publisher έτσι, ώστε να προσδιορίσει μία ιεραρχική σχέση μεταξύ των αρχείων δεδομένων μιας έρευνας. Τα αρχεία δεδομένων μπορεί να σχετίζονται μεταξύ τους όπως π.χ. σε μεγάλες έρευνες νοικοκυριού, όπου οι ερωτήσεις δεν ερωτώνται μόνο σε επίπεδο νοικοκυριού αλλά και σε επίπεδο ατόμου.

Το Nesstar Hierarchy Builder επιτρέπει να προστεθούν αρχεία σε διαφορετικά επίπεδα και να οριστούν οι σχέσεις μεταξύ τους. Τα αρχεία δεδομένων μπορεί να συγχωνευτούν χρησιμοποιώντας αυτές τις μεταβλητές που είναι ορισμένες ως κλειδιά. Πρέπει πρώτα να οριστεί η ιεραρχία μεταξύ των αρχείων δεδομένων (π.χ. το αρχείο νοικοκυριού είναι σε υψηλότερο επίπεδο ιεραρχίας από το αρχείο των ατόμων) και κατόπιν να δημιουργηθεί ο σύνδεσμος. Ο σύνδεσμος μεταξύ των αρχείων γίνεται με τις βασικές και εξωτερικές μεταβλητές κλειδιά. Βασικές

μεταβλητές κλειδιά ονομάζουμε τις μεταβλητές που προσδιορίζουν μοναδικά μία περίπτωση σε ένα αρχείο δεδομένων. Οι εξωτερικές μεταβλητές κλειδιά είναι οι μεταβλητές εκείνες που είναι βασικά κλειδιά σε αρχείο δεδομένων υψηλότερου επιπέδου αλλά όχι βασικό κλειδί στο τρέχον αρχείο δεδομένων.

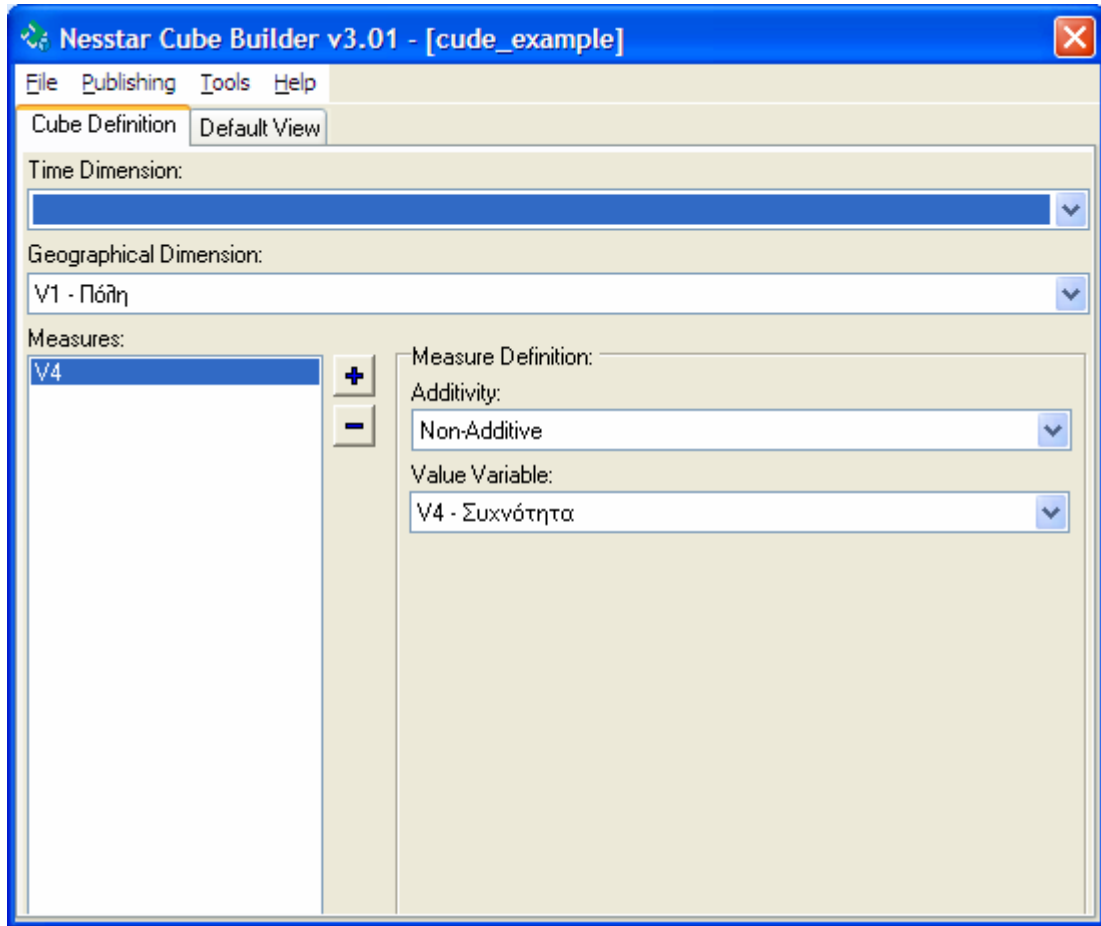


Εικόνα 4: Nesstar Hierarchy Builder

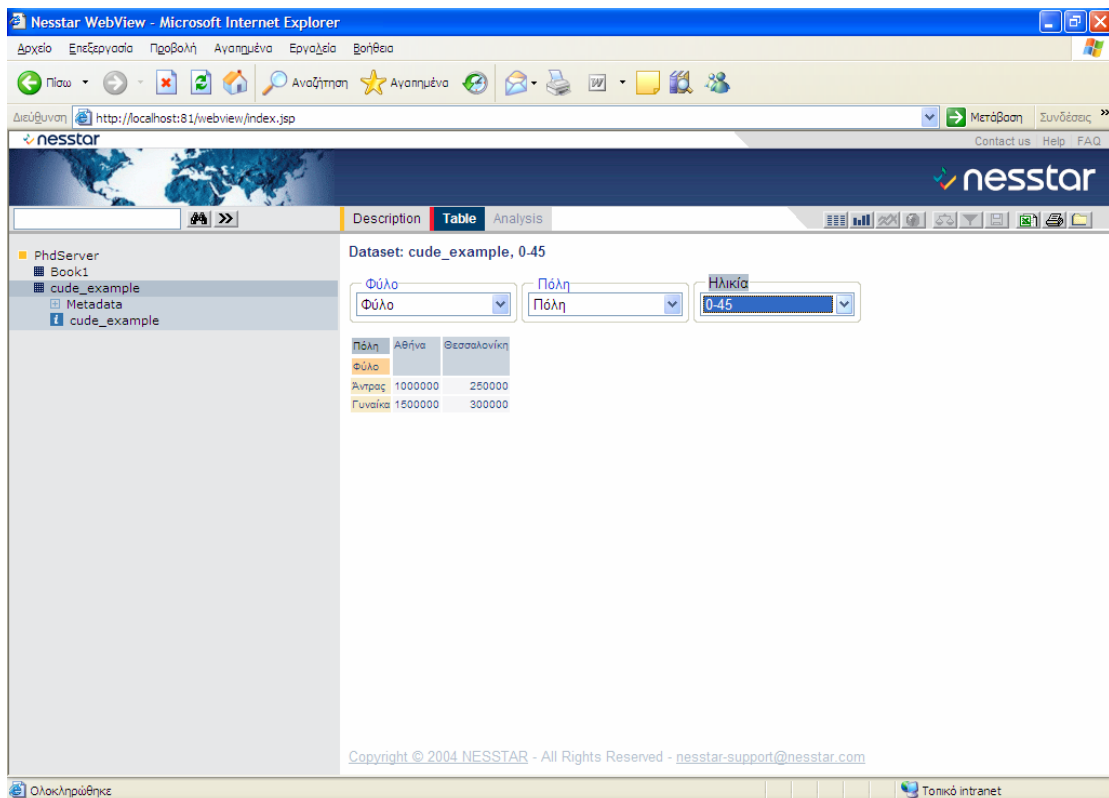
4.4. Nesstar Cube Builder

Ένας κύβος είναι ένας πολυδιάστατος πίνακας που περιέχει αθροιστικά δεδομένα. Τα δεδομένα ορίζονται από τις «διαστάσεις» και τις «μετρήσεις». Οι μεταβλητές που εκφράζουν διαστάσεις περιγράφουν τα δεδομένα και αποτελούνται από κατηγορίες π.χ. φύλο (άντρας-γυναίκα). Οι μεταβλητές που εκφράζουν «μετρήσεις» αναπαριστούν τα δεδομένα ή τις τιμές που υπάρχουν στα κελιά του πίνακα. Θα πρέπει να υπάρχει τουλάχιστο μία μεταβλητή μέτρησης, ώστε να δημιουργηθεί ένας κύβος. Κάθε διάσταση μπορεί να οριστεί και ιεραρχικά. Γεωγραφικές διαστάσεις μπορεί να συνδεθούν με χάρτες.

Τα μεταδεδομένα ενός κύβου αρχικά εισάγονται στο Nesstar Publisher. Μέσω του Nesstar Publisher πρέπει να οριστούν «ονόματα επιπέδων» (level names) για όλες τις διαστάσεις, διότι διαφορετικά δε θα μπορεί να δημιουργηθεί ο κύβος στο Nesstar Cube Builder. Τα δύο περιβάλλοντα συνδέονται άμεσα μεταξύ τους και ο χρήστης μπορεί εύκολα να μεταβαίνει από το ένα περιβάλλον στο άλλο. Στην εικόνα 5 φαίνεται το περιβάλλον Cube Builder όπου ορίζεται ο κύβος και η όψη του. Στην εικόνα 6 φαίνεται ο κύβος μετά την εισαγωγή του στον Nesstar Server.



Εικόνα 5: Nesstar Cube Builder



Εικόνα 6: Κύβος μέσα από το περιβάλλον Nesstar Webview

5. Συμπεράσματα

Η αναγκαιότητα αυστηρής και ποιοτικής τεκμηρίωσης στις κοινωνικές έρευνες κρίνεται επιτακτική, καθώς τα δεδομένα καθίστανται ευτελή με την πάροδο του χρόνου. Τα ΑΔ, προσπαθώντας να βρουν τρόπους για την εύκολη διάχυση και αυστηρή τεκμηρίωση των δεδομένων, αποφάσισαν να δημιουργήσουν ένα πρότυπο τεκμηρίωσης των κοινωνικών ερευνών. Η ανάπτυξη της τεχνολογίας στάθηκε αρωγός σε αυτήν την προσπάθεια. Το πιο αξιόλογο πρότυπο τεκμηρίωσης που δημιουργήθηκε είναι το DDI βασισμένο στην γλώσσα XML. Το DDI παρουσιάζει αρκετά πλεονεκτήματα αλλά και κάποια μειονεκτήματα. Ένα από τα σοβαρά μειονεκτήματα της τρέχουσας έκδοση 2.0 του προτύπου είναι ότι τεκμηριώνει μία έρευνα μόνο κατά το τελικό στάδιό της. Η έκδοση DDI 3.0 φιλοδοξεί να τεκμηριώνει μία έρευνα από το στάδιο της σχεδίασης και να μην εστιάζεται μόνο στο τελικό σύνολο δεδομένων της έρευνας.

Το λογισμικό που είναι ευρέως διαδεδομένο για την υποστήριξη του DDI είναι το λογισμικό Nesstar. Το λογισμικό Nesstar διατίθεται προς το παρόν ελεύθερα στα ΑΔ που είναι μέλη της CESSDA. Πάντως, η καινούργια τάση η οποία κατευθύνεται από το DDI Alliance – το συμβούλιο που καθορίζει την εξέλιξη του DDI – είναι το DDI να υποστηρίζεται από open source εργαλεία, ώστε να μειώνεται το κόστος.

Αναφορές

Blank G and Rasmussen K.B., “The Data Documentation Initiative: The Value and Significance of a Worldwide Standard”, Social Science Computer Review 22, no.3 (August 2004):307-318.

CESSDA Homepage, <http://www.nsd.uib.no/cessda/> .

Data Documentation Initiative. A Project of a Social Science Community. (2002) <http://www.icpsr.umich.edu/DDI>

Kallas J, Frydakis M, Linardis A, The Metadata Model of the Greek Social Data Bank, “The Node for Secondary Processing, a comparative research infrastructure”, ISBN:960-7093-88-7.

Miller K and Vardigan M, “How initiative Benefits the Research community – the Data Documentation Initiative”.

Leighton V, “Developing a new Data Archive in a Time of Maturing Standards.” IASSIST Quarterly 26, no. 1 (spring 2002):5-9.

Linardis A, A general description of the system “Node for Secondary Processing, “The Node for Secondary Processing, a comparative research infrastructure”, ISBN:960-7093-88-7.

Metadata Management and Production System for surveys in Empirical Socio-economic Research, <http://www.metadater.org/> .

Nesstar Software, <http://www.nesstar.com/> .

Ryssevik J and Musgrane S, “The Social Science Dream Machine”, Social Science Computer Review 19, no.2 (Summer 2001):163-174.

Stebe J, Experience with the Data Documentation Initiative Codebook Standard in Social Science Data Archive (ADP).