

**Έργο:** «ΘΑΛΗΣ: Ενίσχυση της Διεπιστημονικής ή και Διδρυματικής έρευνας και καινοτομίας με δυνατότητα προσέλκυσης ερευνητών υψηλού επιπέδου από το εξωτερικό μέσω της διενέργειας βασικής και εφαρμοσμένης έρευνας αριστείας»

**Τίτλος** «ΕΙΚΟΣ»: Θεωρητική και αλγοριθμική θεμελίωση για

**Υποέργου:** Προσωποκεντρικά Συνεργατικά Πληροφοριακά Συστήματα

## Παραδοτέο Π.1.1

### Μοντελοποίηση μη-παραδοσιακών δεδομένων

Σεπτέμβριος 2015



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ  
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ  
επένδυση στην κοινωνία της γνώσης

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ  
2007-2013  
Πρόγραμμα για την ανάπτυξη  
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



<b>Δράση 1</b>	<b>Αναπαράσταση και ενσωμάτωση μη-παραδοσιακών δεδομένων</b>				
<b>Ομάδα</b>	Ερ. Ομάδα 1	<b>Έναρξη</b>	01/02/2012	<b>Λήξη</b>	30/11/2015
<b>Συντονιστής ΕΟ1</b>	Δ. Πλεξουσάκης (Παν. Κρήτης)				
<b>Υποδράση: ΥΔ 1.1</b>	Μοντελοποίηση μη-παραδοσιακών δεδομένων				
<b>Συμμετέχοντες</b>	<i>Μέλη ΚΕΟ</i>	Δ. Πλεξουσάκης (Παν. Κρήτης), Γ. Θεοδωρίδης (Παν. Πειραιώς), Σ. Σκιαδόπουλος (Παν. Πελοποννήσου)			
	<i>Μέλη ΟΕΣ</i>	Θ. Πάτκος (Παν. Κρήτης), Ν. Πελέκης (Παν. Πειραιώς), Ε. Ζερβάκης (Παν. Πελοποννήσου)			
<b>Σύντομη Περιγραφή</b>	Ο σκοπός της υποδράσης Δ1.1 είναι να διερευνήσει, στο γενικότερο πλαίσιο των οικοσυστημάτων δεδομένων, μεθόδους για την κοινή αναπαράσταση μη-παραδοσιακών δεδομένων και μετα-δεδομένων που προέρχονται από διαφορετικές πηγές (data streams, χωροχρονικά δεδομένα, δεδομένα αισθητήρων, επιστημονικά δεδομένα όπως ιατρικά και βιολογικά, δεδομένα ηλεκτρονικών αγορών, δεδομένα κοινωνικών δικτύων, δεδομένα από το Παγκόσμιο Ιστό) καθώς και τρόπους συσχέτισής τους.				
<b>Παραδοτέο</b>	<u>Π.1.1</u> Μοντελοποίηση μη-παραδοσιακών δεδομένων				
<b>Στόχος στο Τ.Δ.</b>	Τεχνική αναφορά που θα περιλαμβάνει τουλάχιστον 1 δημοσίευση.				
<b>Επίτευξη στόχου</b>	100%				

## Περιεχόμενα

1	Εισαγωγή.....	5
1.1	Πλαίσιο έρευνας και στόχοι.....	6
2	Μοντελοποίηση μη παραδοσιακών δεδομένων διαφόρων τύπων με στόχο την ιδιωτικότητα και την ανωνυμία.....	8
2.1	Δεδομένα συνόλων και συναλλαγών.....	8
2.2	Δεδομένα τροχιών.....	10
3	Χωρο-χρονικά δεδομένα μελετημένα υπό το πρίσμα των αναγκών διαφορετικών εφαρμογών.....	11

## 1 Εισαγωγή

Ο βασικός στόχος του έργου ΕΙΚΟΣ είναι να προσφέρει τη μεθοδολογία, τη θεωρητική θεμελίωση, τις αλγοριθμικές τεχνικές και την αρχιτεκτονική του λογισμικού που απαιτείται ώστε τα πληροφοριακά συστήματα να μπορούν να προσφέρουν στους χρήστες αφενός την δυνατότητα εξατομίκευσης της παρεχόμενης πληροφορίας και αφετέρου τη δυνατότητα χρήσης ενσωματωμένων ετερογενών δεδομένων, ενδεχομένως διαφορετικής προέλευσης, με διαφανή τρόπο.

Στο πλαίσιο του έργου, η Δράση 1 «Αναπαράσταση και ανάκτηση μη παραδοσιακών δεδομένων» αποσκοπεί στον ορισμό μοντέλων αναπαράστασης για μη παραδοσιακούς τύπους δεδομένων που είναι πλούσια σε μεταδεδομένα, καθώς και στο σχεδιασμό μηχανισμών για την αυτόματη εξαγωγή της εν λόγω μετα-πληροφορίας για νέα δεδομένα. Τέλος αποσκοπεί στο σχεδιασμό νέων τεχνικών ευρετηρίασης των δεδομένων, που θα επιτρέψουν την αποδοτική ανάκτηση των δεδομένων με βάση την προαναφερθείσα μετα-πληροφορία. Η Δράση οργανώνεται σε τρεις θεμελιώδεις δράσεις, εκ των οποίων η πρώτη αφορά στη μοντελοποίηση μη-παραδοσιακών δεδομένων, η δεύτερη σε τεχνικές ανάκτησης πληροφορίας από πηγές μη-παραδοσιακών δεδομένων και η τρίτη σε μηχανισμούς δεικτοδότησης μη-παραδοσιακών δεδομένων.

Το παρόν Παραδοτέο Π.1.1 περιλαμβάνει τα αποτελέσματα της υποδράσης ΥΔ1.1. Στην ενότητα 1 παρουσιάζουμε το γενικότερο πλαίσιο του προβλήματος και την επισκόπηση των περιοχών που μελετήθηκαν. Στην ενότητα 2 αναλύονται οι ανάγκες μοντελοποίησης διαφορετικών πεδίων με έμφαση σε ευαίσθητα δεδομένα και τις μεθόδους ανωνυμίας που αναπτύχθηκαν. Στην ενότητα 3 η έμφαση δίνεται σε χωρο χρονικά δεδομένα τόσο με σκοπό τη διαχείριση ασάφειας όσο και με στόχο την παροχή διασύνδεσης διαφορετικών πηγών και σημασιολογικού εμπλουτισμού αυτών.

## 1.1 Πλαίσιο έρευνας και στόχοι

Οι νέες προκλήσεις που ανακύπτουν μέσα στο γενικότερο πλαίσιο των οικοσυστημάτων δεδομένων περιλαμβάνουν την ικανότητα αναπαράστασης ετερογενών δεδομένων τα οποία προέρχονται από διαφορετικού τύπου πηγές (π.χ., δίκτυα αισθητήρων, εφαρμογές ηλεκτρονικών αγορών) , βρίσκονται σε διαφορετικά επίπεδα αφαίρεσης (π.χ., raw sensor data, βιβλιογραφικά δεδομένα ή δεδομένα από κοινωνικά δίκτυα), παράγονται με ασύγχρονο τρόπο και σε μεταβαλλόμενους όγκους και ρυθμούς (π.χ., remote sensing applications, high-throughput scientific experiment data), είναι δομημένα, ημιδομημένα ή δεν ακολουθούν συγκεκριμένη δομή (π.χ., δεδομένα βάσεων δεδομένων, δεδομένα προερχόμενα από πηγές Web 2.0, δεδομένα από microarray experiments στη βιολογία), είναι στατικά ή δυναμικά (π.χ., χάρτες, δεδομένα αισθητήρων) και εξαρτώμενα από το χρόνο ή το χώρο (π.χ. ροές δεδομένων, γεωγραφικά δεδομένα) .

Επιπλέον, οι μέθοδοι αναπαράστασης πρέπει να συνοδεύονται από κατάλληλους μηχανισμούς που να υποστηρίζουν την εξέλιξη των δεδομένων και μεταδεδομένων καθώς και των ίδιων των σχημάτων αναπαράστασης. Η εξαγωγή αξίας από τα δεδομένα, προϋποθέτει επίσης σημασιολογικά εμπλουτισμένες περιγραφές των μεταδεδομένων ώστε να καθίσταται δυνατή η ολοκλήρωσή και η διασύνδεσή τους.

Στόχος της έρευνας στο πλαίσιο της δράσης Δ1.1 και του παρόντος παραδοτέου ήταν να καλυφθεί ένα μεγάλο εύρος μη παραδοσιακών τύπων δεδομένων, να παρουσιαστούν δυνατότητες μοντελοποίησής τους, καθώς και να αναλυθούν οι απαιτήσεις που εισάγονται κατά την χρήση τους σε διαφορετικά πεδία. Η κατανόηση των χαρακτηριστικών και των απαιτήσεων αυτών τροφοδοτούν την έρευνα τόσο στις δύο υποδράσεις ΥΔ1.2 «Τεχνικές ανάκτησης πληροφορίας από πηγές μη-παραδοσιακών δεδομένων» και ΥΔ1.3 «Μηχανισμοί δεικτοδότησης μη-παραδοσιακών δεδομένων» που συμπληρώνουν τη Δράση 1 «Αναπαράσταση και ενσωμάτωση μη-παραδοσιακών δεδομένων» όσο και στο γενικότερο ερευνητικό αντικείμενο το έργου ΕΙΚΟΣ.

Συγκεκριμένα, η ανάλυση των χαρακτηριστικών και απαιτήσεων με σκοπό τη μοντελοποίηση μη παραδοσιακών δεδομένων δομήθηκε σε τρεις κεντρικούς άξονες: τους *τύπους* των δεδομένων αυτών, τα *πεδία εφαρμογής* τους και τους

στόχους που χρειάζεται να εξυπηρετηθούν από τη μοντελοποίησή τους. Εξ αρχής απώτερος στόχος ήταν η όσο το δυνατόν ευρύτερη κάλυψη καθενός εκ των αξόνων αυτών, για τη καλύτερη κατανόηση των απαιτήσεων που υπεισέρχονται. Ο Πίνακας 1 συνοψίζει τις περιοχές που μελετήθηκαν στο πλαίσιο της ΥΔ1.1. περιγράφονται αναλυτικότερα στη συνέχεια του παρόντος παραδοτέου.

**Πίνακας 1 Συνοπτική Περιγραφή Περιοχών Μελέτης της ΥΔ1.1**

<b>Είδος Δεδομένων</b>	<b>Πεδίο Εφαρμογής</b>	<b>Στόχος</b>
Προσωπικά και ευαίσθητα δεδομένα συνόλων [TLMS12]	Ιατρικά δεδομένα, δεδομένα ιστορικών, ημερολογίων πιστωτικών καρτών, επίσκεψης ιστοτόπων	Ανωνυμία
Χωρο-χρονικά δεδομένα [PDP14]	Αρχαιολογία, γεωλογία, στρωματογραφία, γεωγραφία	Χρονική ασάφεια, συσχέτιση χωρο-χρονικών όγκων
Δεδομένα συναλλαγών [PGLST14, PLGS13]	Υγειονομική περίθαλψη, μάρκετινγκ κ.α.	Ανωνυμία
Γεωγραφικά δεδομένα, συνδυασμένα με εικόνες κ άλλη πληροφορία οικοσυστημάτων [TABM15]	Θαλάσσια βιολογία, ιχθυολογικά δεδομένα	Διασύνδεση και σημασιολογικός εμπλουτισμός
Δεδομένα τροχιών [PSLG14, PSLG13a, PSLG13b]	Τεχνολογίες εντοπισμού θέσης	Ιδιωτικότητα και ανωνυμία

## 2 Μοντελοποίηση ευαίσθητων και προσωπικών δεδομένων διαφόρων τύπων με στόχο την ιδιωτικότητα και την ανωνυμία

Η μοντελοποίηση και διαχείριση ευαίσθητων δεδομένων αποτελούν κομβικές δραστηριότητες σε πλήθος εφαρμογών και αφορούν μη παραδοσιακά δεδομένα διαφόρων τύπων. Αρκετές από τις εργασίες που εκπονήθηκαν στο πλαίσιο της συγκεκριμένης δράσης επικεντρώθηκαν σε αυτό τον τύπο, προσεγγίζοντάς τον από ποικίλες σκοπιές.

### 2.1 Δεδομένα συνόλων και συναλλαγών

Μη-παραδοσιακά δεδομένα συνόλων, όπως ιατρικά δεδομένα, ιστορικά, ημερολογία πιστωτικών καρτών και επίσκεψης ιστοτόπων, απαρτίζουν σημαντικό ποσοστό στο οικοσύστημα δεδομένων. Σε πολλές περιπτώσεις, τα διαθέσιμα δεδομένα περιέχουν ευαίσθητα προσωπικά δεδομένα (π.χ., ασθένειες, θεραπείες και αγορές χρηστών). Κατά συνέπεια, είναι πολύ σημαντική η δημοσίευση των δεδομένων αυτών με τρόπο που διασφαλίζει την ανωνυμία των εμπλεκόμενων. Στην εργασία [TLMS12] παρουσιάστηκε μια πρωτοπόρα μέθοδος δημοσιοποίησης των παραπάνω δεδομένων συνόλου που αφενός εξασφαλίζει την ανωνυμία των χρηστών και αφετέρου διατηρεί την ποιότητα της ανάλυσης των δημοσιευμένων δεδομένων.

Η συγκεκριμένη εργασία παρείχε μεθόδους προστασίας της ιδιωτικότητας κατά την δημοσίευση πολυδιάστατων δεδομένων. Οι υπάρχουσες μέθοδοι για την διασφάλιση της ανωνυμίας (α) προστατεύουν την ιδιωτικότητα γενικεύοντας ή διαγράφοντας δεδομένα ή προσθέτοντας θόρυβο και (β) θεωρούν την ξεκάθαρη διάκριση μεταξύ ευαίσθητων και μη-ευαίσθητων δεδομένων. Σε πολλές όμως εφαρμογές οι παραπάνω μέθοδοι δεν μπορούν να εφαρμοστούν. Οι μέθοδοι γενίκευσης και διαγραφής αφαιρούν σημαντική πληροφορία από τα δεδομένα. Επιπρόσθετα, τα σημερινά δεδομένα δεν μπορούν ξεκάθαρα να διαχωριστούν σε ευαίσθητα και μη, κάποια δεδομένα μπορεί να είναι ευαίσθητα για ένα χρήστη και μη-ευαίσθητα για κάποιον άλλο. Με βάση την παρατήρηση αυτή, παρουσιάστηκε μια μέθοδος ανωνυμίας που την ονομάζουμε αποσυσχέτιση η



οποία διατηρεί τα αρχικά δεδομένα ανέπαφα αλλά αποκρύπτει την κοινή τους εμφάνιση. Η ανωνυμία εξασφαλίζεται αποσυσχετίζοντας εγγραφές που συμμετέχουν σε συνδυασμούς που μπορούν να αποκαλύψουν την ταυτότητα των χρηστών. Η μέθοδος που παρουσιάζουμε είναι η πρώτη που χρησιμοποιεί αποσυσχέτιση για την εξασφάλιση της ανωνυμίας. Η εργασία παρείχε τόσο την υλοποίηση την μεθόδου όσο και την πειραματική σύγκριση με αντίστοιχες μεθόδους γενίκευσης, διαγραφής και εισαγωγής θορύβου.

Από διαφορετική σκοπιά, οι εργασίες [PGLST14, PLGS13] μελέτησαν τη δημοσίευση συνόλων δεδομένων με προσωπικά δεδομένα, τα οποία περιέχουν τόσο σχεσιακά όσο και δεδομένα συναλλαγών, και τα οποία είναι σημαντικά σε μια ευρεία γκάμα εφαρμογών, που εκτείνονται από την υγειονομική περίθαλψη ως και το μάρκετινγκ. Η διατήρηση την ιδιωτικότητας και της χρηστικότητας αυτών των συνόλων δεδομένων αποτελεί πρόκληση, καθώς απαιτεί (i) προστασία από κακόβουλους χρήστες, των οποίων η γνώση καλύπτει και τους δύο τύπους χαρακτηριστικών, και (ii) την ελαχιστοποίηση της συνολικής απώλειας πληροφορίας. Οι υπάρχουσες τεχνικές ανωνυμίας, δεν μπορούν να εφαρμοστούν σε τέτοια σύνολα δεδομένων, και το πρόβλημα δεν μπορεί να αντιμετωπιστεί με βάση τις δημοφιλείς, στρατηγικές βελτιστοποίησης πολλαπλών στόχων. Οι εργασίες που εκπονήθηκαν στο πλαίσιο της Δ1.1 κατάφεραν να προτείνουν την πρώτη προσέγγιση για την αντιμετώπιση του προβλήματος αυτού και να αναπτύξουν δύο πλαίσια ανωνυμοποίησης, με προκαθορισμένη απώλεια πληροφοριών στον ένα τύπο δεδομένων και ελάχιστη απώλεια πληροφορίας στον άλλο. Για την εφαρμογή κάθε πλαισίου, προτάθηκαν αλγόριθμοι ανωνυμοποίησης οι οποίοι διατηρούν αποτελεσματικά τη χρηστικότητα των δεδομένων, όπως επαληθεύεται και από εκτεταμένα πειράματα.

Μάλιστα, αναπτύχθηκε το σύστημα SECRETΑ, ένα σύστημα για την ανάλυση της αποτελεσματικότητας και της αποδοτικότητας των διαφόρων αλγορίθμων ανωνυμίας. Το σύστημα αυτό επιτρέπει σε όσους επιθυμούν να διαμοιραστούν δεδομένα, να αξιολογήσουν έναν συγκεκριμένο αλγόριθμο, να συγκρίνουν πολλαπλούς αλγόριθμους, και να συνδυάσουν αλγόριθμους για την ανωνυμοποίηση συνόλων δεδομένων τα οποία περιέχουν σχεσιακά δεδομένα

και δεδομένα συναλλαγών. Η ανάλυση των αλγορίθμων γίνεται, με ένα διαδραστικό και προοδευτικό τρόπο, και τα αποτελέσματα, συμπεριλαμβανομένων των στατιστικών και διαφόρων δεικτών χρησιμότητας των δεδομένων, μπορούν να συνοψιστούν και να παρουσιαστούν γραφικά.

## 2.2 Δεδομένα τροχιών

Η διάδοση των συσκευών με GPS (π.χ., smartphones και tablets) και κοινωνικών δικτύων βασισμένων στον διαμοιρασμό της θέσης των χρηστών έχει οδηγήσει σε σημαντική αύξηση των δεδομένων τροχιών. Η δημοσίευση των δεδομένων αυτών ανοίγει νέες κατευθύνσεις στην ανάλυση, μελέτη και κατανόηση της ανθρώπινης συμπεριφοράς. Ωστόσο, θα πρέπει να πραγματοποιείται με έναν τρόπο, ο οποίος να διαφυλάττει την ιδιωτικότητα, επειδή οι ταυτότητες των ατόμων, των οποίων η κίνηση καταγράφεται στις τροχιές, μπορεί να αποκαλυφθεί ακόμη και μετά την αφαίρεση προσδιοριστικών πληροφοριών.

Οι υπάρχουσες προσεγγίσεις ανωνυμοποίησης των δεδομένων τροχιών προσφέρουν προστασία της ιδιωτικότητας, αλλά με υψηλό κόστος στην χρησιμότητα των δεδομένων, δεδομένου ότι είτε δεν παράγουν αληθή δεδομένα (μια σημαντική απαίτηση πολλών εφαρμογών), ή περιορίζονται από τους αλγόριθμους ανωνυμίας που χρησιμοποιούν. Ως επί το πλείστον, οι υπάρχουσες προσεγγίσεις, είτε περιορίζονται από το μοντέλο ανωνυμοποίησης το οποίο χρησιμοποιούν, ή επιφέρουν σημαντική, και συχνά περιττή, αλλοίωση των δεδομένων.

Στις δημοσιεύσεις [PSLG13b, PSLG14], προτάθηκε μια νέα προσέγγιση που αντιμετωπίζει τις ανεπάρκειες αυτές, χρησιμοποιώντας το μοντέλο της  $k^m$ -ανωνυμίας σε δεδομένα τροχιών. Για να υλοποιήσουμε την προσέγγισή μας, αναπτύξαμε τρεις αποδοτικούς και αποτελεσματικούς αλγόριθμους ανωνυμίας, οι οποίοι βασίζονται στην αρχή *Argiori*. Αυτοί οι αλγόριθμοι στοχεύουν στη διατήρηση διαφορετικών χαρακτηριστικών των δεδομένων, όπως η απόσταση μεταξύ σημείων, η σημασιολογική ομοιότητα, καθώς επίσης και στη διατήρηση διαφόρων κριτηρίων χρησιμότητας που ορίζονται από το χρήστη, τα οποία

πρέπει να πληρούνται για να διασφαλιστεί ότι τα δημοσιευμένα στοιχεία μπορούν να αναλυθούν ουσιαστικά. Τα εκτεταμένα πειράματα μας, χρησιμοποιώντας συνθετικά και πραγματικά δεδομένα, επαλήθευσαν ότι οι προτεινόμενοι αλγόριθμοι είναι αποδοτικοί και αποτελεσματικοί στη διατήρηση της χρηστικότητας των δεδομένων.

Επιπρόσθετα, στην εργασία [PSLG13a] προτάθηκε ένα νέο πλαίσιο για την ανωνυμοποίηση δεδομένων τροχιών, το οποίο αποτρέπει την αποκάλυψη πληροφοριών τόσο για την ταυτότητα όσο και για τις ευαίσθητες περιοχές που επισκέφτηκε ο χρήστης, διατηρώντας παράλληλα την χρηστικότητα των δεδομένων. Το πλαίσιο μας περιλαμβάνει: (i) την επιλογή παρόμοιων τροχιών, με τη χρησιμοποίηση είτε ενός αλγορίθμου Z-ordering ή χρησιμοποιώντας προβολές δεδομένων στις συχνά εμφανιζόμενες υποτροχιές, (ii) την ομαδοποίηση των επιλεγμένων τροχιών σε προσεκτικά κατασκευασμένες ομάδες, και (iii) την ανωνυμοποίηση κάθε τέτοιας ομάδας ξεχωριστά. Αναπτύξαμε αλγορίθμους για την υλοποίηση του πλαισίου μας, οι οποίοι είναι αποτελεσματικοί και αποδοτικοί, όπως επαληθεύεται και από εκτεταμένα πειράματα.

### **3 Χωρο-χρονικά δεδομένα μελετημένα υπό το πρίσμα των αναγκών διαφορετικών εφαρμογών**

Τα χωρο-χρονικά δεδομένα αποτελούν σημαντικό όγκο πληροφορίας που απασχολεί μεγάλη ποικιλία εφαρμογών. Τα προβλήματα που εισάγονται αφορούν τόσο την ασάφεια της χωρο-χρονικής πληροφορίας όσο και την ανάγκη διασύνδεσης της από διαφορετικές πηγές και εμπλουτισμού της με σημασιολογικά μετα-δεδομένα.

Σε αυτό το πλαίσιο, στο άρθρο [PDP14] μελετήθηκε και προτάθηκε μια μεθοδολογία για την μοντελοποίηση της έκτασης τετραδιάστατων οντοτήτων όπως οι χωρο-χρονικοί όγκοι χρησιμοποιώντας εμπειρικές αποδείξεις. Επίσης εμβαθύνει στην μοντελοποίηση της χρονικής ασάφειας και στην χρονική συσχέτιση χωρο-χρονικών όγκων. Πιο συγκεκριμένα, τα ασαφή χρονικά διαστήματα που περιγράφουν ένα γεγονός περιέχουν εσωτερικά και εξωτερικά

όρια. Τα επίπεδα που σχηματίζονται αντιπροσωπεύουν την γνώση μας για την κατάσταση του γεγονότος. Για παράδειγμα τα χρονικά σημεία ανάμεσα στα εσωτερικά όρια δηλώνουν την βεβαιότητα ότι το γεγονός είναι ενεργό, ενώ οι περιοχές ανάμεσα στα εξωτερικά και εσωτερικά όρια υποδηλώνουν την άγνοια και κατ'επέκταση την ασάφεια. Η χρονική τοπολογία πάνω σε χωροχρονικούς όγκους προκύπτει μέσω των επιμέρους σχέσεων που σχηματίζονται στα κοινά χωρικά σημεία δύο όγκων. Η επικρατούσα σχέση από ένα σύνολο πιθανών επιμέρους, εξάγεται εφαρμόζοντας την ακόλουθη ιεραρχία: overlaps  $\leq$  equals  $\leq$  {starts, finishes}  $\leq$  meets  $\leq$  {before, during} (Allen's operators, η φορά του βέλους υποδεικνύει την πιο ισχυρή σχέση).

Η συγκεκριμένη εργασία πρότεινε μία μεθοδολογία για την μοντελοποίηση της χρονικής ασάφειας και της χρονικής συσχέτισης χωρο-χρονικών όγκων δεδομένων (τετραδιάστατοι όγκοι). Εξέτασε τη χρονική τοπολογία των χωροχρονικών όγκων όπως αυτή σχηματίζεται από τις σχέσεις μεταξύ των κοινών χωρικών σημείων και προτείνει μια άλγεβρα για το χειρισμό των σχέσεων αυτών. Τέτοια δεδομένα προκύπτουν σε πολλούς επιστημονικούς τομείς (αρχαιολογία, γεωλογία, στρωματογραφία, γεωγραφία) και ο χειρισμός και η ολοκλήρωσή τους με άλλους τύπους δεδομένων δεν έχει μελετηθεί επαρκώς στη διεθνή βιβλιογραφία. Το αποτέλεσμα παρείχε μια τυπική θεμελίωση για την απάντηση ερωτημάτων που δεν μπορούν να απαντηθούν με παραδοσιακές τεχνικές.

Σε ό,τι αφορά τις ανάγκες διασύνδεσης ετερογενών πηγών μη παραδοσιακών δεδομένων, μελετήθηκαν οι απαιτήσεις στον τομέα της θαλάσσιας βιολογίας, όπου τα προβλήματα έχουν εντοπιστεί είναι σαφή και καταγραφεί εδώ και καιρό. Δεδομένα που αφορούν τη θαλάσσια βιολογία σχετικά με ιχθυολογικά είδη (marine species data) βρίσκονται διασκορπισμένα σε πλήθος ετερογενών πηγών, βάσεων δεδομένων και θησαυρών όρων. Συχνά, η αναζήτηση όρων σχετικά με αυτά τα δεδομένα απαιτεί τη χρήση εντελώς διαφορετικών μεταξύ τους πρωτοκόλλων και μορφών, καθώς το εκάστοτε ινστιτούτο διατηρεί τη δική του βάση γνώσης με δεδομένα σε μη-επεξεργασμένη μορφή, ενώ η χρήση οντολογιών για την περιγραφή τους είναι αρκετά περιορισμένη.

Η εργασία [TABM15] περιγράφει την οντολογία MarineTLO, μια οντολογία υψηλού επιπέδου (top-level ontology) που ενοποιεί πληροφορία από πλήθος πηγών, όπως οι FLOD, ECOSCOPE, WoRMS, FishBase και DBpedia, με σκοπό τη δημιουργία ενός ενιαίου, συνεπούς και αξιόπιστου επιπέδου γνώσης που θα παρέχει σημασιολογική διαλειτουργικότητα πάνω σε δεδομένα θαλάσσιων ειδών. Περιγράφονται μοντέλα και μεθοδολογίες δημοσίευσης, πλοήγησης και επερωτήσεων πάνω σε αυτά τα δεδομένα μέσω ενός εκφραστικά πλούσιου εννοιολογικού μοντέλου, αποκρύπτοντας τις διαφορές, διατηρώντας ωστόσο τις απαραίτητες συνδέσεις με τις υποκείμενες πηγές γνώσεις.

Στο πλαίσιο αυτό, η συγκεκριμένη εργασία πρότεινε μια πρωτότυπη οντολογία υψηλού επιπέδου, η οποία επιτρέπει τη διασύνδεση και τον εμπλουτισμό ιχθυολογικών δεδομένων που περιλαμβάνουν γεωγραφικά δεδομένα, εικόνες, πληροφορία σχετικά με οικοσυστήματα, κοινά και επιστημονικά ονόματα ειδών σε διάφορες γλώσσες κ.α., τα οποία βρίσκονται καταναμημένα σε βάσεις τόσο σε δομημένη όσο και σε αδόμητη μορφή. Παρουσιάστηκε επίσης ένας τρόπος αποθήκευσης, προσπέλασης και μετασχηματισμού της πληροφορίας που αξιοποιεί την οντολογία, λαμβάνοντας υπόψη μεταδεδομένα προέλευσης (provenance data).

## Δημοσιεύσεις

- [TLMS12] M. Terrovitis, J. Liagouris, N. Mamoulis, and S. Skiadopoulos. Privacy preservation by disassociation. PVLDB, 5(10):944–955, 2012.
- [PSLG14] Giorgos Poulis, Spiros Skiadopoulos, Grigorios Loukides, Aris Gkoulalas-Divanis: Apriori-based algorithms for km-anonymizing trajectory data. Transactions on Data Privacy 7(2): 165-194 (2014)
- [PGLST14] Giorgos Poulis, Aris Gkoulalas-Divanis, Grigorios Loukides, Spiros Skiadopoulos, Christos Tryfonopoulos: SECRET: A System for Evaluating and Comparing RELational and Transaction Anonymization algorithms. EDBT 2014: 620-623
- [PSLG13a] Giorgos Poulis, Spiros Skiadopoulos, Grigorios Loukides, Aris Gkoulalas-Divanis: Select-Organize-Anonymize: A Framework for Trajectory Data Anonymization. ICDM Workshops 2013: 867-874 (PADM)
- [PSLG13b] Giorgos Poulis, Spiros Skiadopoulos, Grigorios Loukides, Aris Gkoulalas-Divanis: Distance-Based  $k^m$ -Anonymization of Trajectory Data. MDM (2) 2013: 57-62 (PRISMO)
- [PLGS13] Giorgos Poulis, Grigorios Loukides, Aris Gkoulalas-Divanis, Spiros Skiadopoulos: Anonymizing Data with Relational and Transaction Attributes. ECML/PKDD (3) 2013: 353-369
- [TABM15] Yannis Tzitzikas, Carlo Allocca, Chryssoula Bekiari, Yannis Marketakis, Pavlos Fafalios, Martin Doerr, Nikos Minadakis, Theodore Patkos and Leonardo Candella, “Unifying Heterogeneous and Distributed Information about Marine Species through the Top Level Ontology MarineTLO”, accepted for publication in the International Peer-Reviewed Journal Program: Electronic library and Information Systems, Special Issue on Application of Metadata and Semantics in Science, Emerald (2015)
- [PDP14] M. Papadakis, M. Doerr and D. Plexousakis, “Fuzzy Times on Space-Time Volumes”, in Proceedings for e-2014 (24th International Conference on E-Challenges), Belfast, N. Ireland, October 2014



## Παράρτημα