

ΠΟΙΟΤΗΤΑ ΠΛΗΡΟΦΟΡΙΑΣ

ΠΕΡΙΕΧΟΜΕΝΑ

1.1	Διασφάλιση ποιότητας δομημένων δεδομένων	4
1.2	Γενικοί ορισμοί	4
1.3	Αρχιτεκτονική πληροφορίας – αξιολόγηση ποιότητας.....	5
1.4	Αρχές ποιότητας πληροφορίας	5
1.4.1	Η πληροφορία ως προϊόν	5
1.4.2	Στρατηγικός χαρακτήρας της πληροφορίας.....	6
1.4.3	Ορισμός ποιότητας πληροφορίας	6
1.5	Δράσεις διασφάλισης ποιότητας πληροφορίας.....	7
1.5.1	Βήματα διασφάλισης ποιότητας πληροφορίας	7
1.5.1.1	Αξιολόγηση ρίσκου χρήσης πληροφορίας κακής ποιότητας.....	7
1.5.1.2	Αξιολόγηση ποιότητας προδιαγραφών παραγωγής πληροφορίας : ορισμός δεδομένων και αρχιτεκτονική πληροφορίας.....	12
1.5.1.3	Αξιολόγηση ποιότητας πληροφορίας.....	18
1.5.1.4	Απόπειρα δόμησης και καθαρισμού δεδομένων.....	22
1.5.1.5	Βελτίωση διαδικασιών διασφάλισης ποιότητας πληροφορίας	29

ΛΙΣΤΑ ΕΙΚΟΝΩΝ

Εικόνα 1 – Κόστος χαμηλής ποιότητας πληροφορίας - παράδειγμα	11
Εικόνα 2 – Λεκτικά που δεν αποτυπώνονται σε σχεσιακή βάση δεδομένων	16
Εικόνα 3 – Προτυποποίηση πεδίου δεδομένων	24
Εικόνα 4 – Προτυποποίηση και ενοποίηση δεδομένων πελατών.....	27
Εικόνα 5 - Διάγραμμα τύπου fishbone, προσαρμοσμένο από [English].....	30

1.1 Διασφάλιση ποιότητας δομημένων δεδομένων

Η εγκαθίδρυση πολιτικών, προτύπων και διαδικασιών που θα καθορίζουν, επικοινωνούν, διασφαλίζουν, μετρούν και βελτιώνουν την διαχείριση πληροφορίας, αποτελεί επιτακτική ανάγκη στο σύγχρονο Οργανωτικό Περιβάλλον.

Σε πρόσφατες διεθνείς έρευνες αξιολογήθηκε το πλαίσιο ποιότητας πληροφορίας Οργανισμών και ο βαθμός εφαρμογής αρχών και μηχανισμών που συμβάλλουν στην βελτίωση (π.χ. έρευνα στα πλαίσια του συνεδρίου DAMA-I/Meta-data, 2001). Στις έρευνες αυτές αναδεικνύεται η περιορισμένη διείσδυση σχετικών αρχών και μηχανισμών, κάτι που δημιουργεί έντονο προβληματισμό.

Οι παράγοντες που αξιολογήθηκαν στις έρευνες αυτές αφορούσαν:

- Τήρηση προτύπων ορισμού δεδομένων (data standards) σε όλες τις δράσεις ανάπτυξης νέων συστημάτων
- Τήρηση σημαντικού ποσοστού των επιχειρησιακών όρων, σε υποδομή που δημιουργεί λεξικό (data dictionary)
- Πολιτικές διαχείρισης πληροφορίας έχουν εκδοθεί στο ίδιο διοικητικό επίπεδο με πολιτικές οικονομικής διαχείρισης και διαχείρισης προσωπικού, και περιλαμβάνουν αναλόγου βαρύτητας κυρώσεις με τις τελευταίες, όταν δεν εφαρμόζονται

Ανάλογος προβληματισμός υπάρχει και στο Ελληνικό Περιβάλλον. Πρόσφατα έχουν αρχίσει προσπάθειες βελτίωσης της ποιότητας δεδομένων σε μεγάλους Οργανισμούς (Τράπεζες, Κυβερνητικούς Φορείς, ΔΕΚΟ). Αρκετές όμως είναι αποσπασματικές, δεδομένου ότι εστιάζονται μόνο στον καθαρισμό των επιχειρησιακών δεδομένων, χωρίς να εξετάζουν το σύνολο της αξιακής αλυσίδας πληροφορίας του Οργανισμού και να θέτουν τις προϋποθέσεις για την άρση των αιτιών που προκαλούν προβλήματα στην ποιότητα πληροφορίας. Είναι ξεκάθαρη η ανάγκη για την εγκαθίδρυση μηχανισμών που διασφαλίζουν την ποιότητα της πληροφορίας, από την στιγμή που εισάγεται στον Οργανισμό μέχρι την στιγμή που αξιοποιείται παραγωγικά. Είναι επιπλέον σαφές ότι δεδομένα που δεν εισάγονται στον Οργανισμό σε σωστό πλαίσιο, δεν μπορούν να αποτελέσουν στην συνέχεια βάση για πληροφορία υψηλής ποιότητας.

1.2 Γενικοί ορισμοί

Στην παρούσα ενότητα περιγράφεται ο τρόπος ‘μετατροπής’ των δεδομένων σε αξιοποιήσιμη γνώση.

Δεδομένα είναι στοιχεία ή χαρακτηριστικά που περιγράφουν κάποια πραγματική οντότητα.

Η **πληροφορία** είναι δεδομένα σε συγκεκριμένο πλαίσιο (data in context). Αν τα δεδομένα είναι η πρώτη ύλη, η πληροφορία είναι το τελικό προϊόν.

Γνώση είναι πληροφορία, η σημασία της οποίας γίνεται κατανοητή και αξιοποιήσιμη από το συγκεκριμένο πρόσωπο που την έχει. Η έννοια της γνώσης είναι συνυφασμένη με την Ανθρώπινη διάσταση.

Η ικανότητα του Οργανισμού να αποκτά νέα γνώση και να την διαχέει αποτελεσματικά στο εσωτερικό του, θα παίζει μεγάλο ρόλο στην δυνατότητα επιβίωσης και επέκτασης του, στην εποχή της Πληροφορίας.

1.3 Αρχιτεκτονική πληροφορίας – αξιολόγηση ποιότητας

Επιχειρησιακή πληροφορία είναι η γλώσσα της επιχείρησης. Η αποτυχία συστηματικού ορισμού της γλώσσας, μπορεί να έχει αρνητικές συνέπειες στην επικοινωνιακή ικανότητα εσωτερικά και ‘προς τα έξω’.

Ο ορισμός των δεδομένων και της αρχιτεκτονικής της πληροφορίας (data definition & information architecture), είναι για την επιχειρησιακή πληροφορία, το ανάλογο των προδιαγραφών κατασκευής για το βιομηχανικό προϊόν. Η επιδίωξη συστηματικής παραγωγής ποιοτικής πληροφορίας, επιβάλλει την εφαρμογή ποιοτικής ‘προδιαγραφής κατασκευής’ για την παραγωγή του ‘πληροφοριακού προϊόντος’.

Κριτήριο ποιότητας της ‘προδιαγραφής’ των δεδομένων είναι το πόσο καλά αυτή διατυπώνει και επικοινωνεί την πραγματική σημασία των πραγματικών οντοτήτων που τα δεδομένα περιγράφουν.

Η ‘προδιαγραφή’ των δεδομένων περιλαμβάνει ονοματοδοσία, ορισμούς και περιγραφές σχέσεων μεταξύ οντοτήτων και χαρακτηριστικών αυτών, καθώς και επιχειρησιακούς κανόνες που διέπουν την επιχειρησιακή δραστηριότητα και την ακεραιότητα των δεδομένων.

Κριτήρια αξιολόγησης της ποιότητας της ‘προδιαγραφής’ όσον αφορά τον ορισμό των δεδομένων είναι:

- η ορθότητα (π.χ. προδιαγραφής σχέσεων οντοτήτων ή οντοτήτων με χαρακτηριστικά)
- η πληρότητα (π.χ. προδιαγραφή του συνόλου νομίμων τιμών για ένα χαρακτηριστικό (domain values)),
- η ακρίβεια και η σαφήνεια αυτής.

Κριτήρια αξιολόγησης της ποιότητας της ‘προδιαγραφής’ όσον αφορά τον ορισμό της αρχιτεκτονικής πληροφορίας είναι:

- η σταθερότητα του μοντέλου δεδομένων
- η ευελιξία του μοντέλου δεδομένων
- η ικανότητα του μοντέλου δεδομένων να καταγράφει τις πληροφοριακές ανάγκες του Οργανισμού.

1.4 Αρχές ποιότητας πληροφορίας

1.4.1 Η πληροφορία ως προϊόν

Έχει ιδιαίτερο ενδιαφέρον η ανάλυση της πληροφορίας ως προϊόντος που παρέχεται από ‘Παραγωγούς πληροφορίας’ και αξιοποιείται από ‘Πελάτες πληροφορίας’. Ξεκινώντας από τα δεδομένα που είναι η ‘πρώτη ύλη’ καταλήγουμε με κατάλληλη επεξεργασία στην πληροφορία που είναι το τελικό προϊόν.

Η πληροφορία έχει το ιδιαίτερο χαρακτηριστικό της δυνατότητας επαναχρησιμοποίησης, υπό την έννοια ότι αυτή δεν καταναλώνεται.

Πελάτες πληροφορίας είναι :

- Οι Εργαζόμενοι-γνώσης (knowledge workers) του Οργανισμού είτε είναι στο ίδιο τμήμα με τους παραγωγούς πληροφορίας είτε είναι σε προχωρημένο στάδιο της αξιακής αλυσίδας πληροφορίας
- Οι τελικοί Πελάτες του Οργανισμού

1.4.2 Στρατηγικός χαρακτήρας της πληροφορίας

Η πληροφορία πρέπει να αντιμετωπίζεται ως στρατηγικός πόρος της επιχείρησης που πρέπει να αξιοποιείται από τον Οργανισμό με τον καλύτερο τρόπο.

1.4.3 Ορισμός ποιότητας πληροφορίας

Μέσω της ανάλυσης της γενικότερης έννοιας της ποιότητας για τον Οργανισμό και θεωρώντας την πληροφορία ως ένα προϊόν, μπορεί να αναλυθεί σε αναλογία η έννοια της ποιότητας της πληροφορίας.

Ένας τρόπος γενικού ορισμού της ποιότητας είναι: ‘η συστηματική ικανοποίηση των προσδοκιών του Πελάτη’. Εξειδικεύοντας τον ορισμό για την ποιότητα πληροφορίας μπορεί να διατυπωθεί: συστηματική ικανοποίηση των προσδοκιών του ‘Πελάτη πληροφορίας’.

Οι διαστάσεις που καθορίζουν την ποιότητα της πληροφορίας είναι οι ακόλουθες:

- Ο ορισμός της σημασίας των δεδομένων (ή το λεξικό των δεδομένων). Όπως κάθε καταγεγραμμένη γλώσσα έχει λεξικό που περιγράφει την σημασία των λέξεων, έτσι και τα δεδομένα πρέπει να συνοδεύονται από τον ορισμό της σημασίας τους. Ποιότητα σε αυτήν την διάσταση, είναι ο βαθμός που ο ορισμός περιγράφει με ακρίβεια την σημασία του στοιχείου της πραγματικής οντότητας.
- Η ακρίβεια των δεδομένων. Η ακρίβεια με την οποία τα δεδομένα εκφράζουν τα χαρακτηριστικά της πραγματικής οντότητας (π.χ. η διεύθυνση ενός Πελάτη έχει καταγραφεί με ακρίβεια ώστε να μπορεί να χρησιμοποιηθεί επιχειρησιακά χωρίς περαιτέρω ελέγχους ή επιβεβαιώσεις).
- Το πλαίσιο παρουσίασης των δεδομένων. Ποιότητα σε αυτήν την διάσταση είναι ο βαθμός που:
 - Τα δεδομένα είναι προσβάσιμα και διαθέσιμα έγκαιρα
 - Τα δεδομένα παρουσιάζονται με τρόπο κατανοητό, δεν υπάρχουν ασάφειες και συνδυάζονται σωστά μεταξύ τους
- Η αρχιτεκτονική της πληροφορίας. Ποιότητα σε αυτήν την διάσταση είναι ο βαθμός που η δομή των δεδομένων :
 - Υλοποιεί σχέσεις μεταξύ δεδομένων που αντικατοπτρίζουν τις σχέσεις των στοιχείων των πραγματικών οντοτήτων
 - Είναι ευέλικτη ώστε να μπορεί να καλύψει μελλοντικές αλλαγές και επεκτάσεις στις επιχειρησιακές διαδικασίες που υποστηρίζει
 - Είναι σταθερή ώστε να μπορεί να εξυπηρετήσει νέες εφαρμογές χωρίς σημαντικές αλλαγές στην δομή.

1.5 Δράσεις διασφάλισης ποιότητας πληροφορίας

Η διασφάλιση ποιότητας της επιχειρησιακής πληροφορίας απαιτεί παρεμβάσεις στις βασικές διαστάσεις του Οργανισμού δηλαδή:

- ο στην Οργανωτική δομή
- ο στις επιχειρησιακές διαδικασίες
- ο στα συστήματα υποστήριξης της διαχείρισης πληροφορίας

Δανείζεται και εφαρμόζει αρχές από την γενικότερη θεωρία συστημάτων ποιότητας και συνεχούς βελτίωσης, όπως:

- ο την εστίαση στον Πελάτη, εσωτερικό ή εξωτερικό
- ο την θέσπιση διαδικασιών και Οργανωτικών ρόλων που θα παρακολουθούν την ποιότητα
- ο την ενεργοποίηση διαδικασιών συνεχούς βελτίωσης

Η διασφάλιση της ποιότητας πληροφορίας βρίσκει εφαρμογή σε δυο βασικές δραστηριότητες :

- ο Την ανάπτυξη πληροφορικών συστημάτων κατά την οποία καθορίζονται πληροφοριακές ανάγκες και επιχειρησιακές διαδικασίες που θα υποστηρίζονται και σχεδιάζεται η αρχιτεκτονική πληροφορίας και η βάση (ή βάσεις) δεδομένων.
- ο Την καταγραφή, επικαιροποίηση και γενικότερη διαχείριση δεδομένων σε κατάλληλο νοηματικό πλαίσιο για την παραγωγή και παρουσίαση της πληροφορίας στους 'Πελάτες'.

1.5.1 Βήματα διασφάλισης ποιότητας πληροφορίας

Παρακάτω παρατίθενται τα βασικά βήματα διασφάλισης ποιότητας πληροφορίας. Η σειρά με την οποία οι σχετικές δράσεις εφαρμόζονται στον Οργανισμό εξαρτάται από το στάδιο ωρίμανσης του Οργανισμού στην χρήση πληροφορικής αλλά και την αξιολόγηση των ρίσκων χρήσης πληροφορίας κακής ποιότητας.

- ο Αξιολόγηση ρίσκου χρήσης πληροφορίας κακής ποιότητας
- ο Αξιολόγηση ποιότητας προδιαγραφών παραγωγής πληροφορίας : ορισμός δεδομένων και αρχιτεκτονική πληροφορίας
- ο Αξιολόγηση ποιότητας πληροφορίας
- ο Απόπειρα αναδόμησης και καθαρισμού δεδομένων
- ο Βελτίωση διαδικασιών διασφάλισης ποιότητας πληροφορίας

Παρακάτω αναλύονται τα προαναφερθέντα βήμα διασφάλισης ποιότητας.

1.5.1.1 Αξιολόγηση ρίσκου χρήσης πληροφορίας κακής ποιότητας

Η χρήση πληροφορίας κακής ποιότητας μπορεί να έχει αρνητικά αποτελέσματα για τον Οργανισμό, ανάλογα με την δραστηριότητα που η πληροφορία υποστηρίζει:

- ο Προβλήματα στην παροχή υπηρεσιών στον Πελάτη
- ο Κόστη παραγωγής που υπό άλλες συνθήκες θα μπορούσαν να αποφευχθούν

- Χαμένες ευκαιρίες πωλήσεων και διαφυγόντα έσοδα
- Δυσκολία στην βελτίωση προϊόντων / υπηρεσιών βάσει αναγκών Πελάτη ή δημιουργία νέων προϊόντων
- Λήψη αποφάσεων σε όλα τα επίπεδα, βασισμένη σε ανακριβή ή λάθος στοιχεία

1.5.1.1.1 Τύποι κινδύνου χρήσης πληροφορίας κακής ποιότητας

Παρακάτω κατηγοριοποιούνται και περιγράφονται σχετικοί κίνδυνοι.

1.5.1.1.1.1 Κακολειτουργία επιχειρησιακής διαδικασίας

Η κακολειτουργία επιχειρησιακής διαδικασίας που προκαλείται από πληροφορία κακής ποιότητας είναι συχνό φαινόμενο. Η κακή ποιότητα μπορεί να συνίσταται στην ανακρίβεια ή μη πληρότητα των στοιχείων, παράγοντες που δημιουργούν:

- πρόσθετα κόστη λειτουργίας εφόσον το πρόβλημα εντοπίζεται και διορθώνεται
- λειτουργικά κόστη που δεν παράγουν αξία για τον Οργανισμό (π.χ. αποστολή γραμμάτων σε λάθος ταχυδρομικές διευθύνσεις)
- προβληματική εξυπηρέτηση Πελάτη που μπορεί να έχει ποικίλες συνέπειες ανάλογα με την σοβαρότητα του θέματος όπως το αξιολογεί ο Πελάτης, από την απλή δυσαρέσκεια, την απώλεια Πελάτη έως την δικαστική δίωξη του Οργανισμού (π.χ. λανθασμένες ή διογκωμένες τιμολογήσεις λόγω κακής διαχείρισης της πληροφορίας)

1.5.1.1.1.2 Εκ νέου παραγωγή της πληροφορίας

Προβλήματα στην ποιότητα σημαντικής ή κρίσιμης επιχειρησιακής πληροφορίας πρέπει να αντιμετωπίζονται από τον Οργανισμό. Οι τρόποι αντιμετώπισης περιλαμβάνουν:

- Επανεκτέλεση επιχειρησιακών διαδικασιών π.χ. επανεπικοινωνία με τον Πελάτη για την άντληση ή επιβεβαίωση στοιχείων που δεν καταγράφηκαν ορθά
- Αναζήτηση πληροφορίας σε προχωρημένο στάδιο ολοκλήρωσης διαδικασίας με συνέπεια την αναστολή ολοκλήρωσης. Εφόσον η διαδικασία αυτή τροφοδοτεί ευρύτερες διαδικασίες ή επιτρέπει την έναρξη τους μπορεί να δημιουργήσει σοβαρό πρόβλημα (π.χ. παραγωγή στοιχείων μισθοδοσίας Προσωπικού)
- Επαναλαμβανόμενη αξιολόγηση της ποιότητας της πληροφορίας κατά μήκος της αξιακής αλυσίδας, όταν οι ρόλοι που την αξιοποιούν, δεν εκτιμούν ότι είναι αξιόπιστη.

1.5.1.1.3 Κόστος ευκαιρίας

Η κακή ποιότητα πληροφορίας και κατά συνέπεια η μη δυνατότητα αποτελεσματικής αξιοποίησης αυτής, δημιουργεί χαμένες ευκαιρίες για τον Οργανισμό σε διάφορα επίπεδα:

- αξιοποίηση συνολικού ιστορικού Πελατών (λόγω προβλημάτων αρχιτεκτονικής πληροφορίας)
- αύξηση εσόδων μέσω Cross-selling και up-selling με την αξιοποίηση πληροφορίας του προφίλ πελατών
- ανάπτυξη ‘πιστών’ Πελατών (customer loyalty) και αξιοποίηση της δια βίου αξίας Πελάτη (Customer lifetime value) μέσω της διαφοροποιημένης παροχής αξίας βασισμένης στην αξιοποίηση πληροφορίας (π.χ. Amazon book preview)
- δημιουργία νέων Πελατών και αγορών βάσει ανάλυσης στοιχείων (π.χ. στοιχεία προτιμήσεων Πελατών, χαρακτηριστικά προϊόντων και υπηρεσιών που έχουν σημαντική αξία για τον Πελάτη) και ανάπτυξης ελκυστικής πρότασης παροχής αξίας (value proposition)
- διαφοροποίηση μέσω του περιορισμού του κόστους παραγωγής με την αποτελεσματική καταγραφή και αξιοποίηση πληροφορίας απόδοσης παραγωγικών διαδικασιών
- αποτελεσματική υποστήριξη στην λήψη αποφάσεων βάσει αναλυτικών στοιχείων που αντλούνται από παραγωγικές διαδικασίες και αξιοποιούνται αποτελεσματικά

1.5.1.1.4 Εκ των υστέρων ‘καθαρισμός’ των δεδομένων

Η προσπάθεια εκ των υστέρων ‘καθαρισμού’ των δεδομένων έχει συχνά ‘πυροσβεστικό’ χαρακτήρα, ειδικά όταν αυτό είναι δύσκολο και έχει περιορισμένο αποτέλεσμα (χαμηλό ποσοστό των δεδομένων καθαρίζεται) με υψηλό κόστος. Είναι συχνό φαινόμενο ειδικά σε παλαιότερα πληροφορικά συστήματα όπου δεν υπήρχε ενιαία και πελατοκεντρική αρχιτεκτονική πληροφορίας ή σε συστήματα που συγχωνεύονται λόγω συγχώνευσης των Οργανισμών.

Η ποιότητα της πληροφορίας (και των δεδομένων που την παράγουν) πρέπει να διασφαλίζεται σε αρχικό στάδιο της αξιακής αλυσίδας της πληροφορίας. Η διαδικασία ‘καθαρίσματος’ των δεδομένων αποτελεί κοστοβόρα και συχνά αναποτελεσματική λύση.

Στόχο πρέπει να αποτελεί η εξ αρχής άρτια δόμηση των δεδομένων και η εγκαθίδρυση μηχανισμών πρόληψης των ελαττωμάτων πληροφορίας (defect prevention).

1.5.1.1.5 Ανορθόδοξη ανάπτυξη αρχιτεκτονικής πληροφορίας

Σταδιακή ανάπτυξη νησίδων πληροφορίας που σχεδιάζονται ανεξάρτητα, έχουν διαφορετικό ορισμό δεδομένων, διαφορετικό μοντέλο δεδομένων, και δεν διασυνδέονται με υφιστάμενο κεντρικό πληροφορικό σύστημα

Δημιουργία συμπληρωματικών βάσεων δεδομένων με βελτιωμένη πληροφορία που όμως είναι προσβάσιμη από περιορισμένο τμήμα του Οργανισμού (ατομική ή τμηματική πρόσβαση).

1.5.1.1.2 Αξιολόγηση κινδύνων

Όσον αφορά τον κίνδυνο ‘κακολειτουργίας επιχειρησιακών διαδικασιών’ μπορεί να γίνει αξιολόγηση της σοβαρότητας για κάθε διαδικασία ως εξής:

- i. Σε συνεργασία με τους ειδικούς κάθε θεματικής περιοχής αναπτύσσεται πίνακας διαδικασιών – οντοτήτων που καταγράφει τις οντότητες δεδομένων που απαιτούνται για την εύρυθμη λειτουργία κάθε επιχειρησιακής διαδικασίας
- ii. Εκτιμάται το αποτέλεσμα της έλλειψης, μη πληρότητας ή ανακρίβειας δεδομένων για την κάθε διαδικασία (business impact analysis).
- iii. Επιλέγονται οι διαδικασίες με υψηλό αντίκτυπο για να αναλυθούν περαιτέρω
- iv. Κρίσιμοι ρόλοι σε κάθε διαδικασία αναλαμβάνουν να περιγράψουν τις ενέργειες που γίνονται όταν η πληροφορία είναι ανεπαρκής. Αναλύεται η κατανάλωση πόρων για την διεκπεραίωση των ενεργειών αυτών από τις ακόλουθες κατηγορίες:
 - ο Χρόνος Προσωπικού
 - ο Οικονομικοί πόροι
 - ο Πρώτες ύλες, ημιτελή και έτοιμα προϊόντα
 - ο Υποδομές και εξοπλισμός
 - ο Υποδομές πληροφορικής,

και αξιολογείται το πρότυπο μοναδιαίο κόστος χρήσης για κάθε κατηγορία και επιμέρους πόρο καθώς και η απαιτούμενη ποσότητα αυτού, όπου αυτό είναι εφικτό. Στοιχεία πρότυπου κόστους παρακολουθούνται από τους Οργανισμούς για λόγους κοστολόγησης.

Περαιτέρω ποσοτικοποίηση του κινδύνου απαιτεί την εκτίμηση της συχνότητας εμφάνισης κάθε προβλήματος που σχετίζεται με ανεπαρκή πληροφορία.

Έστω και αν η μέθοδος δεν καταλήξει σε αναλυτική ποσοτικοποίηση, έχει σημαντική αξία γιατί εντοπίζει τις επιχειρησιακές διαδικασίες και τα σημεία λειτουργίας που πρέπει να αντιμετωπιστούν με προτεραιότητα.

Όσον αφορά το κόστος ευκαιρίας, η αξιολόγηση του δεν είναι εύκολη δεδομένου ότι πιθανή απόπειρα θα βασιζόταν σε σειρά εκτιμήσεων (π.χ. της δια βίου αξίας Πελάτη, της αυξημένης ζήτησης διαφοροποιημένου προϊόντος, του όγκου πωλήσεων που θα προέκυπταν λόγω ‘πληροφοριακά ενισχυμένου’ (information-enabled) cross selling.

Ειδικότερα δε, το κόστος ευκαιρίας που αφορά την υποβοήθηση λήψης αποφάσεων δεν υπολογίζεται. Συχνά όμως η ποσοτικοποίηση σχετικών μεγεθών, στην προσπάθεια δικαιολόγησης μιας επένδυσης σε συστήματα υποβοήθησης λήψης αποφάσεων (DSS) δεν έχει νόημα, γιατί η ανάπτυξη τέτοιων συστημάτων απλά αποτελεί στρατηγική επιλογή του Οργανισμού.

Στην παρακάτω εικόνα 26 παρατίθεται παράδειγμα ποσοτικής εκτίμησης του κόστους ευκαιρίας λόγω κακής ποιότητας πληροφορίας. Εταιρεία προσπαθεί να προσελκύσει νέους πελάτες με την αποστολή καταλόγων. Στα στοιχεία επιλεγμένων υποψηφίων πελατών (prospects) συνόλου 500.000, η εταιρεία είχε 10% διπλές καταχωρήσεις και 15% λάθος ταχυδρομικές διευθύνσεις. Θεωρώντας κόστος καταλόγου 2 ευρώ και ‘δια βίου αξία Πελάτη’ που έχει εκτιμηθεί σε 1000

ευρώ, προκύπτει κόστος σπατάλης 250.000 ευρώ και χαμένη αξία Πελάτη 750.000 ευρώ.

Κόστος ευκαιρίας σε διαδικασία marketing	
Κατάλογοι που εστάλησαν	500.000
Διπλές διευθύνσεις (10%)	50.000
Λάθος διευθύνσεις (15%)	75.000
Σύνολο καταλόγων που σπαταλήθηκαν	125.000
Κόστος καταλόγου (ευρώ)	2
Κόστος σπατάλης (ευρώ)	250.000
Ποσοστό ανταπόκρισης Πελατών	0,01
Δια βίου αξία Πελάτη (ευρώ)	1000
Απωλεσθείσα αξία Πελάτη (ευρώ)	750.000

Εικόνα 1 – Κόστος χαμηλής ποιότητας πληροφορίας - παράδειγμα

Η προσέγγιση που αναλύθηκε μπορεί να αναδείξει το μέγεθος των προβλημάτων που δημιουργούν στον Οργανισμό τα προβλήματα στην ποιότητα πληροφορίας, και σε κάποιες περιπτώσεις να οδηγήσει σε ιδιαίτερα πειστικές εκτιμήσεις αριθμητικών μεγεθών.

Μελέτες παρέχουν ενδείξεις ότι πληροφορία κακής ποιότητας καταναλώνει 15-25% των εσόδων ή λειτουργικών εξόδων του Οργανισμού, διεθνώς.

1.5.1.2 Αξιολόγηση ποιότητας προδιαγραφών παραγωγής πληροφορίας : ορισμός δεδομένων και αρχιτεκτονική πληροφορίας

Ο όρος ‘ορισμός δεδομένων’ αναφέρεται στην πληροφορία που περιγράφει την ακριβή σημασία των τύπων πραγματικών οντοτήτων, για τις οποίες ο Οργανισμός πρέπει να έχει στοιχεία για να εκπληρώσει την αποστολή του.

Ο βαθμός ακριβούς αποτύπωσης και επικοινωνίας της σημασίας των πραγματικών οντοτήτων, είναι βασικό κριτήριο ποιότητας του ‘ορισμού δεδομένων’.

Θεωρώντας την πληροφορία ως τελικό προϊόν, ο ‘ορισμός δεδομένων’ είναι για την πληροφορία, ότι είναι οι προδιαγραφές κατασκευής για ένα βιομηχανικό προϊόν.

Στην τεχνική ορολογία χρησιμοποιείται συχνά ο όρος μετα-δεδομένα (metadata), υπό την έννοια των δεδομένων που περιγράφουν τα ίδια τα επιχειρησιακά δεδομένα. Επίσης χρησιμοποιείται ο όρος λεξικό δεδομένων (data dictionary).

Στα πλαίσια ανάπτυξης άρτιας αρχιτεκτονικής πληροφορίας η ποιότητα αξιολογείται σε 3 επίπεδα ποιότητας:

1.5.1.2.1 Πρότυπα που διέπουν την μέθοδο ‘ορισμού δεδομένων’.

Τα πρότυπα διασφαλίζουν τον τυποποιημένο και συστηματικό ορισμό δεδομένων, ώστε να υπάρχει δυνατότητα ενοποιημένης αξιοποίησης πληροφορίας από διαφορετικές πηγές.

Θέματα που θεωρούνται είναι:

- Τα πρότυπα είναι σαφή, πλήρη (αφορούν το σύνολο της σημαντικής πληροφορίας) και εφαρμόζονται ‘οριζόντια’ στον Οργανισμό (για το σύνολο των νέων εφαρμογών που αναπτύσσονται)
- Τα πρότυπα ανανεώνονται μέσω σχετικής διαδικασίας αξιολόγησης των αποτελεσμάτων στην επικοινωνία και την παραγωγικότητα

1.5.1.2.2 Ποιότητα ‘ορισμού δεδομένων’.

Παρακάτω αναλύονται οι παράμετροι ορισμού δεδομένων και ο τρόπος αξιολόγησης της ποιότητας.

Επιχειρησιακή ορολογία

Λεξικογραφική περιγραφή της επιχειρησιακής ορολογίας. Η περιγραφή πρέπει να αντανακλά με σαφήνεια την σημασία των όρων που χρησιμοποιούνται. Η ύπαρξη ενιαίου γλωσσαρίου επιχειρησιακών όρων (Corporate dictionary) συμβάλλει στην αποφυγή συγχύσεων και την βελτίωση της ενδοοργανισμιακής επικοινωνίας. Η ακριβής περιγραφή των επιχειρησιακών όρων βοηθάει στον ‘ορισμό δεδομένων’.

Τύπος οντότητας (entity type)

Οι τύποι οντοτήτων αφορούν οτιδήποτε έχει επιχειρησιακό ενδιαφέρον: Πελάτες (Φυσικά και Νομικά Πρόσωπα), Προϊόντα, οικονομικά γεγονότα κ.α..

Ονοματοδοσία τύπων οντοτήτων (entity type naming). Η ονομασία των τύπων οντοτήτων πρέπει να ανταποκρίνεται στην κατηγορία οντοτήτων που αντιπροσωπεύει. Θέματα προς αξιολόγηση είναι:

- ο τύπος οντότητας αντανakλά πραγματική οντότητα με ένα μοναδικό κοινό πλαίσιο χαρακτηριστικών
- το όνομα περιλαμβάνει επιχειρησιακούς όρους που έχουν περιγραφεί (π.χ. Πελάτης)
- το όνομα είναι κατανοητό από τους ‘Πελάτες Πληροφορίας’

Περιγραφή τύπου οντότητας

Η περιγραφή τύπου οντότητας πρέπει να περιλαμβάνει ένα πλήρες παράδειγμα και να είναι κατανοητή από τους ‘Πελάτες Πληροφορίας’.

Παράδειγμα επαρκούς ορισμού οντότητας:

‘Πελάτης: Πρόσωπο που αγόρασε υπηρεσίες ή προϊόντα από τον Οργανισμό. Περιλαμβάνει ενεργούς ή ανενεργούς πελάτες καθώς και πρόσωπα που εκδήλωσαν ενδιαφέρον για υπηρεσίες ή προϊόντα’

Χαρακτηριστικά τύπου οντότητας (entity type attributes)

Χαρακτηριστικό τύπου οντότητας είναι ένα στοιχείο που συμβάλλει στην περιγραφή της πραγματικής οντότητας που ενδιαφέρει τον Οργανισμό. Λαμβάνοντας μια τιμή για ένα χαρακτηριστικό, ένας Εργαζόμενος με σχετική με τον τύπο οντότητας αντίληψη, μπορεί να συσχετίσει την τιμή αυτή με κάποια πραγματική οντότητα του τύπου αυτού και να αρχίσει να διαμορφώνει εικόνα της οντότητας.

Ονοματοδοσία χαρακτηριστικών τύπου οντότητας.

Θέματα ονοματοδοσίας χαρακτηριστικών τύπου οντότητας προς αξιολόγηση είναι:

- το όνομα περιλαμβάνει επιχειρησιακούς όρους που έχουν περιγραφεί (π.χ. Ταχυδρομική_Διεύθυνση) και είναι κατανοητό από τους Εργαζόμενους γνώσης
- εφόσον χρησιμοποιείται συντομογραφία, η τελευταία συνδέεται με τους επιχειρησιακούς όρους που είναι κατανοητοί
- ονόματα χαρακτηριστικών πρέπει να χρησιμοποιούνται κατά ενιαίο τρόπο στα διάφορα στάδια παραγωγής και παρουσίασης πληροφορίας.

Περιγραφή χαρακτηριστικών τύπου οντότητας

Η περιγραφή χαρακτηριστικών τύπου οντότητας πρέπει να βασίζεται σε προκαθορισμένους επιχειρησιακούς όρους ή κοινή γλώσσα και να περιλαμβάνει πλήρες παράδειγμα. Το όνομα πρέπει να είναι νοηματικά συναφές με την περιγραφή.

Ορισμός πεδίου αποδεκτών τιμών (domain type definition) χαρακτηριστικού τύπου οντότητας. Το πεδίο αποδεκτών τιμών ανταποκρίνεται στο σύνολο των τιμών που αναμένεται για το χαρακτηριστικό αυτό. Παραδείγματος χάριν το χαρακτηριστικό 'Ημερομηνία_πώλησης' αντλεί τιμές από το πλαίσιο τύπου 'Ημερομηνία', και ο αποδέκτης της πληροφορίας αναμένει να δει στο αντίστοιχο πεδίο νόμιμες 'τιμές' ημερομηνίας.

Προβλήματα καταγραφής πληροφορίας μπορεί να προκληθούν όταν:

- ο ένα χαρακτηριστικό τύπου οντότητας χρησιμοποιείται για να καταγράψει 2 διαφορετικούς τύπους στοιχείων για την πραγματική οντότητα
- οι αποδεκτές τιμές έχουν επικάλυψη μεταξύ τους.

Χρήση συντομογραφίας ή ακρώνυμου

Επιχειρησιακοί όροι που περιγράφονται με μακροσκελή διατύπωση απαιτούν συχνά τον ορισμό αντίστοιχης συντομογραφίας.

Θέματα χρήσης συντομογραφίας προς αξιολόγηση είναι:

- ο Το σύνολο των συστημάτων του Οργανισμού χρησιμοποιεί την ίδια συντομογραφία για ίδιο επιχειρησιακό όρο. Αν αυτό δεν συμβαίνει, πρέπει εναλλακτικές συντομογραφίες να χαρακτηριστούν ως συνώνυμα
- ο Η χρήση γενικά αποδεκτών συντομογραφίσεων, όπου τέτοιες υφίστανται, είναι προτιμότερη σε σχέση με την δημιουργική ανάπτυξη νέων
- ο Όταν όμως η πληροφορία φτάνει στο στάδιο αξιοποίησης από συστήματα λήψης αποφάσεων, η χρήση συντομογραφιών πρέπει να αντικαθίσταται από επιχειρησιακούς όρους περιφραστικά διατυπωμένους ώστε να διευκολύνεται η κατανόηση από τον επιχειρησιακό Αναλυτή.
- ο Η χρήση ενιαίας προσέγγισης συντομογραφίσεων. Παραδείγματος χάριν η προσέγγιση χρήσης των πρώτων γραμμάτων του όρου (π.χ. 'Ημερομ' για τον όρο 'Ημερομηνία) ή άλλη προσέγγιση (π.χ. 'Ημ/νία') πρέπει να τηρείται σε όλες τις εφαρμογές.

Περιγραφές επιχειρησιακών κανόνων

Οι επιχειρησιακοί κανόνες διέπουν την λειτουργία του Οργανισμού. Συχνά οδηγούν σε καθορισμό σχέσεων μεταξύ οντοτήτων, σε περιορισμούς του πεδίου τιμών χαρακτηριστικών, σε κανονισμούς ακεραιότητας δεδομένων (integrity constraints).

Οι επιχειρησιακοί κανόνες πρέπει να είναι διατυπωμένοι με σαφήνεια, χρησιμοποιώντας περιγεγραμμένη και οικεία επιχειρησιακή ορολογία. Περιγράφουν τι πρέπει να γίνει και σε ποιες περιπτώσεις, καθώς επίσης και εξαιρέσεις στην εφαρμογή κανόνων.

Οι επιχειρησιακοί κανόνες δεν πρέπει να επηρεάζονται από περιορισμούς υφιστάμενων συστημάτων.

1.5.1.2.3 Ποιότητα αρχιτεκτονικής πληροφορίας.

Η αξιολόγηση της ποιότητας της αρχιτεκτονικής πληροφορίας βασίζεται στα προαναφερθέντα κριτήρια (ενότητα Ποιότητα ‘ορισμού δεδομένων’), και επιπρόσθετα εξετάζει τα ακόλουθα:

- Ορθότητα σχέσεων δεδομένων. Οι σχέσεις μεταξύ πραγματικών οντοτήτων και μεταξύ χαρακτηριστικών αυτών πρέπει να αναγνωρίζονται με ακρίβεια
- Σχέσεις μεταξύ τύπων οντοτήτων (entity type to entity type relationship). Κριτήρια αξιολόγησης της παραμέτρου είναι:
 - Και οι δύο τύποι οντοτήτων περιγράφονται με σαφήνεια και αντιπροσωπεύουν πραγματικούς τύπους οντοτήτων που ενδιαφέρουν τον Οργανισμό
 - Η σχέση μεταξύ των δύο τύπων οντοτήτων περιγράφεται με σαφήνεια και πληρότητα και αντισταθμίζει την σχέση των πραγματικών τύπων οντοτήτων (π.χ. εάν η σχέση τύπου οντότητας Πελάτη – τύπου οντότητας τραπεζικού λογαριασμού έχει υλοποιηθεί σαν ένα-προς-ένα, τότε δεν αποτυπώνει ορθά Πελάτες που στην πραγματικότητα έχουν 2 λογαριασμούς)
- Σχέσεις μεταξύ χαρακτηριστικών και τύπου οντότητας
 - Όλες οι περιπτώσεις (instances) του τύπου οντότητας περιλαμβάνουν τιμή του εν λόγω χαρακτηριστικού.
 - Το χαρακτηριστικό του τύπου οντότητας που αποτυπώνεται στην αρχιτεκτονική πληροφορία αποτελεί αναπόσπαστο στοιχείο του πραγματικού τύπου οντότητας που ενδιαφέρει
 - Το χαρακτηριστικό του τύπου οντότητας που αποτυπώνεται στην αρχιτεκτονική πληροφορία αφορά μόνο ένα αναπόσπαστο στοιχείο της πραγματικής τύπου οντότητας που ενδιαφέρει και όχι συνδυασμό στοιχείων

1.5.1.2.3.1 Μοντέλο επιχειρησιακής πληροφορίας

Αποτελεί απεικόνιση υψηλού επιπέδου (conceptual) των πληροφοριακών απαιτήσεων του Οργανισμού.

Μεθοδολογίες μοντελοποίησης της Αρχιτεκτονικής του Οργανισμού (Enterprise Architecture) ενσωματώνουν απεικόνιση τέτοιου τύπου (π.χ. πλαίσιο Zachman, 2ο επίπεδο) με την περιγραφή των βασικών επιχειρησιακών διαδικασιών (business processes) και οντοτήτων που εμπλέκονται.

Στην μοντελοποίηση αυτή, που κάθε Οργανισμός πρέπει να επιδιώκει πριν την ανάπτυξη μοντέλων δεδομένων, αξιολογείται η τήρηση των ακόλουθων:

- Η χρήση επιχειρησιακής ορολογίας και η εφαρμογή των προδιαγραφών και κριτηρίων της ενότητας Ποιότητα ‘ορισμού δεδομένων’
- Το μοντέλο είναι κατανοητό από την διοίκηση του Οργανισμού
- Το μοντέλο ορίζει τους βασικούς τύπους οντοτήτων (master entity types)
- Το μοντέλο δομείται κατά θεματική περιοχή (subject area) και όχι κατά Τμήμα του Οργανισμού
- Οι θεματικές περιοχές δεν επικαλύπτονται. Τύποι οντοτήτων πρέπει να κατηγοριοποιούνται σε μόνο μία θεματική περιοχή.

1.5.1.2.3.2 Λογικό Μοντέλο δεδομένων

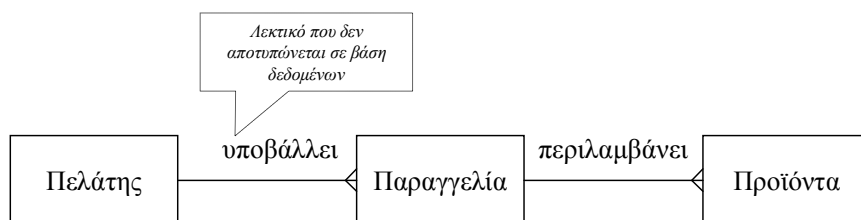
Το επόμενο επίπεδο εξειδίκευσης της Αρχιτεκτονικής Οργανισμού (Enterprise Architecture) αφορά τον σχεδιασμό μοντέλων δεδομένων που εξειδικεύουν το Μοντέλο επιχειρησιακής πληροφορίας σε λογικό επίπεδο (logical) (συνήθως με χρήση σχεσιακού μοντέλου δεδομένων).

Στην μοντελοποίηση αυτή, αξιολογείται η τήρηση των ακόλουθων:

- ο η εφαρμογή των προδιαγραφών και κριτηρίων της ενότητας ‘Ποιότητα ορισμού δεδομένων’
- ο σαφήνεια και δυνατότητα κατανόησης οντοτήτων και σχέσεων
- ο σταθερότητα μοντέλου: επεκτάσεις στο μοντέλο πρέπει να γίνονται χωρίς αλλαγή της βασικής δομής αυτού
- ο ευελιξία μοντέλου: επιχειρησιακές αλλαγές προκαλούν περιορισμένες αλλαγές δομής στο μοντέλο
- ο δυνατότητα επαναχρησιμοποίησης του μοντέλου κατά την ανάπτυξη νέων συστημάτων και εφαρμογών

1.5.1.2.3.3 Φυσική υλοποίηση μοντέλων δεδομένων

Φυσική υλοποίηση (physical design) είναι η διαδικασία εφαρμογής του λογικού σχεδιασμού (λογικό μοντέλο δεδομένων) σε ηλεκτρονικό σύστημα διαχείρισης βάσης δεδομένων (DBMS). Κατά την υλοποίηση βάσεων δεδομένων επιδιώκεται η ακριβής αναπαραγωγή όλων των ιδιοτήτων του λογικού σχεδιασμού, κάτι που συχνά δεν είναι εφικτό (π.χ. οι σχέσεις μεταξύ οντοτήτων που φαίνονται με λεκτικό στο λογικό μοντέλο δεν είναι δυνατόν να αποτυπωθούν σε σχεσιακή βάση δεδομένων. Επίσης οι σχέσεις τύπου πολλοί –προς – πολλούς δεν αποτυπώνονται απευθείας).



Εικόνα 2 – Λεκτικά που δεν αποτυπώνονται σε σχεσιακή βάση δεδομένων

Όσον αφορά την φυσική υλοποίηση μοντέλων δεδομένων, αξιολογείται η τήρηση των ακόλουθων:

- ο Υφίσταται λογικό μοντέλο δεδομένων (logical data model) και αναπτύχθηκε προ της υλοποίησης της ηλεκτρονικής βάσης δεδομένων
- ο Εφαρμόζονται όλοι οι περιορισμοί ακεραιότητας (integrity constraints) που είναι εφικτή η εφαρμογή τους
- ο σταθερότητα μοντέλου. Επεκτάσεις στην βάση, στους πίνακες και στα πεδία αυτών, γίνονται χωρίς αλλαγή της βασικής δομής αυτού
- ο ευελιξία. Επιχειρησιακές αλλαγές προκαλούν περιορισμένες ανάγκες αναδόμησης της βάσης δεδομένων
- ο δυνατότητα χρησιμοποίησης της βάσης δεδομένων από καινούργιες εφαρμογές

1.5.1.2.3.4 Κατανεμημένες βάσεις δεδομένων

Βασική αρχή αποτελεί η υλοποίηση κοινής ‘βάσης δεδομένων αναφοράς’ (record of reference data base), από την οποία αντλούν στοιχεία όλοι οι Χρήστες.

Εφόσον αυτό δεν είναι εφικτό, πρέπει να συντηρούνται ενεργά αντίγραφα της βάσης αναφοράς μέσω μεθόδου database replication, που θα διασφαλίζουν την αντιστοιχία (consistency) των δεδομένων.

Επιπλέον πρέπει να αποφεύγεται η δημιουργία επικαλυπτόμενων (redundant) βάσεων αναφοράς που οδηγούν σε αναντιστοιχίες (inconsistencies) καθώς και η χρήση πολλών διεπαφών μετατροπής δεδομένων (data transformation interfaces). Σε περίπτωση που υπάρχουν επικαλυπτόμενες βάσεις δεδομένων ή/και διεπαφές μετατροπής δεδομένων, πρέπει να υπάρχει μηχανισμός ελέγχου και διασφάλισης της ποιότητας των επικαλυπτόμενων δεδομένων ή/και των δεδομένων που μετατρέπονται και φορτώνονται σε άλλη βάση δεδομένων.

Κατά την ανάπτυξη συστημάτων διαχείρισης πληροφορίας, είναι επιθυμητό να σχεδιαστεί κατάλληλη αρχιτεκτονική πληροφορίας (information architecture) που θα επιδιώκει:

- Την χρήση κοινών βάσεων δεδομένων που οι χρήστες μοιράζονται διατηρηματικά για την άντληση στοιχείων. Αυτό βέβαια με την χρήση έτοιμων λύσεων (τύπου ERP) επιτυγχάνεται εφόσον το ERP είναι πλήρως ολοκληρωμένο και αντικαθιστά προηγούμενες ενδομηματικές-μεμονωμένες εφαρμογές
- Την αποφυγή ανάπτυξης πολλαπλών πηγών για τις ίδιες οντότητες δεδομένων (αποφυγή data redundancy) και κατά συνέπεια την μέγιστη δυνατή αποφυγή χρήσης διεπαφών μετατροπής δεδομένων (transformation interfaces)
- Την μοναδική και ηλεκτρονική σύλληψη (single electronic data capture) κάθε επιχειρησιακού δεδομένου,

καθώς και να εφαρμοστεί συστηματικό πλαίσιο τεκμηρίωσης των δεδομένων που θα περιλαμβάνει:

Τον ενιαίο ορισμό των δεδομένων (τήρηση ενιαίων κωδικοποιήσεων κατηγοριών επιχειρησιακής πληροφορίας στα διάφορα τμήματα – π.χ. κωδικοποίηση υλικών παραγωγής)

Η επιχειρησιακή πληροφορία είναι η γλώσσα της επιχείρησης. Ο χαμηλής ποιότητας ‘λεξικογραφικός’ ορισμός της γλώσσας αυτής, δημιουργεί προβλήματα στην επικοινωνία του Οργανισμού και κατά συνέπεια στην επιχειρησιακή λειτουργία.

1.5.1.3 Αξιολόγηση ποιότητας πληροφορίας

Αξιολόγηση του βαθμού συστηματικής κάλυψης των πληροφοριακών αναγκών των 'Πελατών πληροφορίας'.

Η μέτρηση της ποιότητας στοχεύει στα ακόλουθα:

- ο Διασφάλιση της επαρκούς πληροφοριακής υποστήριξης των εξαρτώμενων από την πληροφορία και της καλής λειτουργίας των επιχειρησιακών διαδικασιών
- ο Εντοπισμός και προσπάθεια επίλυσης προβλημάτων ποιότητας πληροφορίας κατά μήκος της αξιακής αλυσίδας παραγωγής πληροφορίας

Η μέτρηση της ποιότητας πληροφορίας, πρέπει να γίνεται βάσει κριτηρίων-χαρακτηριστικών που ορίζουν οι 'Πελάτες πληροφορίας' και υποστηρίζουν στοχεύσεις για βελτίωση της ικανοποίησης του 'Πελάτη' (εσωτερικού και εξωτερικού ως προς τον Οργανισμό).

1.5.1.3.1 Εσωτερικά χαρακτηριστικά ποιότητας δεδομένων

Συμμόρφωση με τον ορισμό.

Αντιστοιχία νοήματος τιμών δεδομένων με τον ορισμό των δεδομένων. Για παράδειγμα το χαρακτηριστικό 'ώρα εξέτασης' μπορεί να ορίζεται ως η ώρα που Ασθενής εξετάστηκε. Αν όμως το σύστημα εισάγει αυτόματα την ώρα και η εισαγωγή στοιχείων εξέτασης στο σύστημα γίνεται ετεροχρονισμένα, τότε η ώρα θα είναι λάθος. 'Συμμόρφωση με τον ορισμό', είναι ο βαθμός συμφωνίας μεταξύ του νοήματος που οι χρήστες αποδίδουν σε δεδομένα και του επίσημου ορισμού αυτών.

Πληρότητα δεδομένων

Ύπαρξη τιμών στα προβλεπόμενα πεδία. Μετριέται ως το ποσοστό εγγραφών που έχουν τιμή σε κάποιο συγκεκριμένο πεδίο. Για το συγκεκριμένο χαρακτηριστικό χρησιμοποιείται και ο όρος: ποσοστό κάλυψης (coverage).

Συχνά η συμπλήρωση κάποιων πεδίων που αφορούν πληροφορία μη απαραίτητη για την ολοκλήρωση μιας διαδικασίας είναι προαιρετική. Η πληροφορία αυτή όμως μπορεί να συμβάλλει στην παραγωγή χρήσιμων συμπερασμάτων.

Συμμόρφωση με επιχειρησιακούς κανόνες

Αφορά την διαμόρφωση τιμών δεδομένων εντός πεδίου αποδεκτών τιμών και την συμβατότητα αυτών με προκαθορισμένους επιχειρησιακούς κανόνες. Παραδείγματος χάριν, η τιμή κωδικού υλικού πρέπει να αντιστοιχεί σε μια από τις αποδεκτές τιμές κωδικών βάσει κωδικοποίησης. Το χαρακτηριστικό μετριέται ως το ποσοστό εγγραφών που έχουν τιμή συμμορφούμενη με τους κανόνες.

Αξιοσημείωτο είναι ότι η συμμόρφωση με επιχειρησιακούς κανόνες δεν διασφαλίζει την ακρίβεια των τιμών. Συνεπώς αυτοματοποιημένες εφαρμογές ελέγχου συμμόρφωσης **με επιχειρησιακούς κανόνες** δεν ελέγχουν παράλληλα την ακρίβεια των τιμών. (π.χ. αλγόριθμος ελέγχου εγκυρότητας αριθμού ΑΦΜ, δεν ελέγχει την ύπαρξη του αριθμού αυτού).

Ακρίβεια

Ακρίβεια των δεδομένων είναι ο βαθμός στον οποίο αυτά αντανακλούν με ακρίβεια τα στοιχεία που χαρακτηρίζουν την πραγματική οντότητα.

Συχνά ο έλεγχος της ακρίβειας δεδομένων σε σχέση με την πραγματικότητα (π.χ. ένα οικονομικό γεγονός) δεν είναι εφικτός ή έχει απαγορευτικό κόστος. Στην περίπτωση αυτή, πιθανόν να είναι εφικτός ο έλεγχος της ακρίβειας δεδομένων σε σχέση με 'υποκατάστατη πηγή' (surrogate source) που θεωρείται αξιόπιστη (έντυπο που κατέγραψε το οικονομικό γεγονός, εξωτερικά ηλεκτρονικά ή έντυπα αρχεία).

Έλεγχος της ακρίβειας σε σχέση με την πραγματικότητα δικαιολογείται όταν το αντίκτυπο της χαμηλής ποιότητας δεδομένων, είναι σημαντικό για τον Οργανισμό. Το κόστος ελέγχου της ακρίβειας σε σχέση με την πραγματικότητα μπορεί να περιοριστεί με κατάλληλη μέθοδο δειγματοληψίας. Παράδειγμα ελέγχου ακρίβειας σε σχέση με την πραγματικότητα είναι η περιοδική φυσική απογραφή αποθήκης και σύγκριση με τηρούμενα δεδομένα.

Η μέτρηση της ακρίβειας σε σχέση με υποκατάστατη πηγή δεν είναι πάντα η καλύτερη προσέγγιση και πρέπει να είναι σαφής η διαφορά μεταξύ της ακρίβειας σε σχέση με την πραγματικότητα και της ακρίβειας σε σχέση με πηγή που θεωρείται αξιόπιστη.

Στον ακόλουθο πίνακα δίνονται παραδείγματα υποκατάστατων πηγών ελέγχου ακρίβειας.

<i>Τύπος οντότητας</i>	<i>Πραγματική πηγή ελέγχου</i>	<i>Υποκατάστατη πηγή ελέγχου</i>
Πελάτης	Τηλεφωνική κλήση ή αποστολή ερωτήσεων προς απάντηση	Έντυπα παραγγελιών, Βοηθητικά αρχεία
Χαρακτηριστικά Προϊόντος	Δείγμα προϊόντος	Έντυπα με χαρακτηριστικά του προϊόντος, έντυπα υποστήριξης παραγωγής του γεγονότος
Διαθεσιμότητα προϊόντος	Έλεγχος αποθηκευτικού χώρου	Έλεγχος στο πληροφορικό σύστημα υποστήριξης αποθήκης
Οικονομικό γεγονός	Αυτόματη καταγραφή σε πραγματικό χρόνο (π.χ. scan προϊόντων σε ταμείο supermarket)	Καταγραφή σε αρχείο

Τιμές δεδομένων σε κατάλληλο επίπεδο λεπτομέρειας

Αξιολογείται αν οι τιμές δεδομένων τηρούνται σε κατάλληλο επίπεδο λεπτομέρειας ανάλογα με τις ανάγκες της επιχειρησιακής διαδικασίας που εξυπηρετείται. Αν τα δεδομένα αξιοποιούνται από διάφορες διαδικασίες τότε πρέπει να εφαρμόζεται η αυστηρότερη απαίτηση.

Μη ύπαρξη πολλαπλών καταχωρήσεων

Αφορά την τήρηση σχέσεων ένα-προς-ένα μεταξύ εγγραφών και πραγματικών οντοτήτων. Μετράται ως το ποσοστό των καταχωρήσεων που δεν τηρούν την άνω σχέση ένα προς ένα.

Η τήρηση σχέσεων ένα-προς-ένα μεταξύ εγγραφών και πραγματικών οντοτήτων είναι πρόκληση σε μεγάλες Επιχειρήσεις που έχουν μεγάλη πελατειακή βάση καθώς και δεδομένα τηρούμενα σε διάφορες ασύνδετες βάσεις δεδομένων.

Η αδυναμία τήρησης της άνω σχέσης ένα προς ένα, δημιουργεί προβλήματα στον Οργανισμό, δεδομένου ότι :

- Οδηγεί σε αλλοίωση της πραγματικότητας (το σημαντικότερο πρόβλημα)
- Οδηγεί σε υποβέλτιστη αξιοποίηση των δεδομένων, όταν σε κάθε εγγραφή δεν καταγράφεται το σύνολο των διαθέσιμων στοιχείων
- Απαιτεί χρόνο καθαρισμού των δεδομένων

Στις σύγχρονες βάσεις δεδομένων είναι εύκολο να ανιχνευθούν πανομοιότυπες εγγραφές σε πίνακες. Όμως όταν η δόμηση δεν είναι πελατοκεντρική αλλά βασίζεται π.χ. στο λογαριασμό (account centric), τότε είναι εύκολο να αναπτυχθούν πολλαπλές εγγραφές για τον ίδιο πελάτη που έχουν διαφορετικό αριθμό λογαριασμού.

Αντιστοιχία επικαλυπτόμενων ή καταναμημένων δεδομένων

Χρονική αντιστοιχία

Μετράται το χρονικό διάστημα που απαιτείται από την ενημέρωση μιας βάσης δεδομένων με νέα εγγραφή μέχρι την ενημέρωση μιας επικαλυπτόμενης ή καταναμημένης βάσης δεδομένων με την ίδια εγγραφή (γνωστό και ως electronic information float). Το θέμα αυτό μπορεί να δημιουργήσει προβλήματα στην λειτουργία μιας επιχείρησης (π.χ. λειτουργία βάσει παλιών δεδομένων, χαμένες ευκαιρίες).

Σημασιολογική αντιστοιχία

Εφόσον επικαλυπτόμενες βάσεις δεδομένων έχουν διαφορές στον ορισμό των δεδομένων, αξιολογείται η σημασιολογική αντιστοιχία μεταξύ δεδομένων διαφορετικών βάσεων, που αντιστοιχούν στην ίδια πραγματική οντότητα. Παραδείγματος χάριν κωδικός 'Ε' για χαρακτηρισμό ημέρας την ημερομηνία 05-04-2005 και κωδικός '1' για τύπο ημερολογιακής ημέρας την ημερομηνία 05-04-2005, είναι σημασιολογικά ισοδύναμα και σημαίνουν ότι η ημέρα είναι εργάσιμη. Η μη ύπαρξη σημασιολογικής αντιστοιχίας μπορεί να έχει επιχειρησιακό αντίκτυπο, δεδομένου ότι ο Οργανισμός μπορεί να μην είναι σε θέση να αναγνωρίσει την ορθή εκδοχή και Τμήματα αυτού να λειτουργούν βάσει λανθασμένων δεδομένων (π.χ. για στοιχεία Πελατών).

1.5.1.3.2 Χαρακτηριστικά επιχειρησιακής αξιοποίησης των δεδομένων

Προσβασιμότητα στα δεδομένα

- Διαθεσιμότητα στον κατάλληλο χρόνο. Ελαχιστοποίηση του χρόνου που μεσολαβεί από την στιγμή που ο Οργανισμός έχει στην διάθεση του κρίσιμα δεδομένα μέχρι την στιγμή που είναι διαθέσιμα για ηλεκτρονική ανάκτηση (information float).
- Δυνατότητα πρόσβασης σε δεδομένα, διαθέσιμα εντός του Οργανισμού
- Ευκολία πρόσβασης σε διαθέσιμα δεδομένα
- Παρουσίαση των δεδομένων με τρόπο πλήρως κατανοητό και χρηστικό (στον ακόλουθο πίνακα αποτυπώνεται παράδειγμα χαμηλής ποιότητας παρουσίασης απόδοσης ομολόγων)
- Ορθότητα διαδικασίας επεξεργασίας πρωτογενών δεδομένων για την παραγωγή επεξεργασμένης πληροφορίας

Ομόλογο	30/08/05	6/9/05	13/09/05	20/09/05	27/09/05	04/10/05	11/10/05
OM _A	89,52	92,34	93,09	92,67	91,52	90,95	90,77
OM _B	110,12	112,88	113,03	112,90	111,87	111,38	110,90
OM _Γ	57,43	58,65	59,20	59,05	58,87	58,90	58,30
OM _Δ	95,32	96,56	97,12	97,10	96,90	96,95	98,50

1.5.1.3.3 Προσέγγιση αξιολόγησης της ποιότητας δεδομένων

Η διαμόρφωση εικόνας για τα πιο σημαντικά χαρακτηριστικά ποιότητας δεδομένων σε ένα Οργανισμό πρέπει να γίνεται αντλώντας την γνώμη Εργαζόμενων από διάφορα τμήματα. Έτσι διαμορφώνονται οι πλέον κατάλληλες μετρήσεις ποιότητας για τον συγκεκριμένο Οργανισμό.

Μια αποτελεσματική προσέγγιση αξιολόγησης της ποιότητας δεδομένων περιλαμβάνει τα ακόλουθα:

- Ορθή επιλογή κατηγοριών δεδομένων προς αξιολόγηση – μέτρηση (αναγνώριση καταλλήλων υποσυνόλων πληροφορίας με υψηλό επιχειρησιακό αντίκτυπο)
- Ορθή επιλογή των χαρακτηριστικών ποιότητας που θα αποτελέσουν κριτήρια αξιολόγησης και των τεχνικών μέτρησης που κρίνονται κατάλληλες
- Κατάλληλη αξιολόγηση των αποτελεσμάτων και πρόταση τρόπων βελτίωσης

1.5.1.4 Απόπειρα δόμησης και καθαρισμού δεδομένων

Επιδιώκεται η βελτίωση της ποιότητας των δεδομένων, ώστε να παράγεται πληροφορία που ικανοποιεί καλύτερα της ανάγκες των εργαζομένων.

Η δόμηση αφορά την τοποθέτηση αδόμητων δεδομένων σε κατάλληλη αρχιτεκτονική πληροφορίας.

Ο καθαρισμός δεδομένων περιλαμβάνει δράσεις όπως:

- Προτυποποίηση δεδομένων (data standardizing)
- Συμπλήρωση δεδομένων που λείπουν
- Διόρθωση λάθος τιμών
- Αφαίρεση διπλών εγγραφών

Μια ολοκληρωμένη παρέμβαση περιλαμβάνει τα ακόλουθα βήματα:

1.5.1.4.1 Αναγνώριση πηγών δεδομένων

Αφορά τον εντοπισμό των αρχείων ή βάσεων αναφοράς για κάθε τύπο οντότητας ή θεματική περιοχή. Περιλαμβάνει τις ακόλουθες δραστηριότητες:

- εντοπισμό όλων των παραγωγικών βάσεων δεδομένων
- εντοπισμό των διαδικασιών που χρησιμοποιούν τα δεδομένα

Τρόποι επιλογής των πλέον αξιόπιστων πηγών:

- η διαδικασία που εξαρτάται περισσότερο από την ακρίβεια ενός στοιχείου είναι η πλέον αξιόπιστη (π.χ. η διεύθυνση Πελάτη είναι ιδιαίτερα κρίσιμη για την διαδικασία τιμολόγηση του, συνεπώς το λογισμικό που υποστηρίζει τιμολογήσεις πρέπει να θεωρείται κατάλληλη πηγή δεδομένων διευθύνσεων Πελατών)
- δεδομένα που ανανεώνονται συχνά
- δεδομένα που επικαιροποιούνται από κάποιον που εξαρτάται από αυτά για την διεκπεραίωση των εργασιών του
- το παλαιότερο αρχείο μεταξύ των διαθέσιμων σχετικών αρχείων, αποτελεί την πλέον αξιόπιστη πηγή για ιστορικά δεδομένα

Επιλογή των πλέον αξιόπιστων πηγών μπορεί να γίνει και με την διενέργεια αξιολόγησης ποιότητας περιορισμένης έκτασης, στις διάφορες υποψήφιες πηγές.

1.5.1.4.2 Ανάκτηση και ανάλυση πηγών δεδομένων

Περιλαμβάνει τις ακόλουθες εργασίες:

- Ανάκτηση επαρκούς μεγέθους δείγματος
- Ανάλυση δεδομένων με τους ειδικούς θεματικών περιοχών για την κατανόηση της πραγματικής σημασίας πεδίων
- Τεκμηρίωση του ορισμού πεδίων αποδεκτών τιμών και επιχειρησιακών κανόνων για κάθε χαρακτηριστικό τύπου οντότητας

1.5.1.4.3 Προτυποποίηση δεδομένων (data standardization)

Αφορά την προτυποποίηση δεδομένων σε ένα ενιαίο για τον Οργανισμό, ενοποιήσιμο, σετ τύπων οντοτήτων και χαρακτηριστικών αυτών. Ορίζονται ενιαίες δομές και αποδεκτές τιμές δεδομένων. Η προτυποποίηση δεδομένων συμβάλλει στην καλύτερη επικοινωνία στοιχείων και τον συστηματικότερο καθαρισμό των δεδομένων.

Παρακάτω αναφέρονται κατηγορίες περιπτώσεων που απαιτούν προτυποποίηση δεδομένων:

Επικάλυψη αποδεκτών τιμών δεδομένων (domain value redundancy)

Μη προτυποποιημένα δεδομένα μπορεί να τηρούν διαφορετικές τιμές για την ίδια πραγματική οντότητα. Παραδείγματος χάριν η τιμή ημερομηνίας '10-02-2005' είναι διαφορετική από την τιμή '02-10-2005', αλλά μπορεί να αντιστοιχεί στην ίδια ημερομηνία. Στην περίπτωση αυτή πρέπει να καθορίζεται και να δρομολογείται η εφαρμογή ενιαίας για τον Οργανισμό κωδικοποίησης ημερομηνίας π.χ. Ημέρα-Μήνας-Έτος.

Αναντιστοιχίες στην δόμηση πεδίων (field format inconsistencies)

Δομημένα πεδία δεδομένων μπορεί να τηρούνται με αναντιστοιχίες στην δόμηση. Παραδείγματος χάριν η δόμηση αριθμού τηλεφώνου μπορεί να είναι '2106804000' ή '210-6804000' ή '30-210-6804000'. Στην περίπτωση αυτή πρέπει να καθορίζεται και να δρομολογείται η εφαρμογή ενιαίου προτύπου δόμησης του πεδίου.

Διαιρούμενες τιμές δεδομένων (nonatomic values)

Χαρακτηριστικά οντοτήτων πρέπει να ορίζονται κατά τρόπο ώστε τα πεδία να περιέχουν το πλέον αδιαίρετο επίπεδο (π.χ. η τιμή του πεδίου Όνομα 'Γιάννης Παπαδόπουλος' δεν είναι στο πλέον αδιαίρετο επίπεδο και μπορεί να χωριστεί σε δύο πεδία).

Αν και η δημιουργία πεδίων σε 'αδιαίρετο επίπεδο' συχνά δεν αποτελεί απαίτηση, διευκολύνει την εκτέλεση ερωτημάτων. Στο επίπεδο καθαρισμού δεδομένων, διευκολύνει την απαλοιφή διπλοεγγραφών και την ενοποίηση δεδομένων. Επιπλέον διευκολύνει την εφαρμογή επιχειρησιακών κανόνων.

Ενσωμάτωση πολλών στοιχείων οντότητας σε μοναδικό πεδίο

Συμβαίνει συχνά σε δομημένες κωδικοποιήσεις επιχειρησιακής πληροφορίας. Παραδείγματος χάριν κωδικός προϊόντος τύπου EAN-13 σε προϊόν περιλαμβάνει:

- Τμήμα κωδικού για την Χώρα που τηρεί τον κωδικό
- Τμήμα κωδικού για τον Προμηθευτή
- Τμήμα κωδικού για το προϊόν

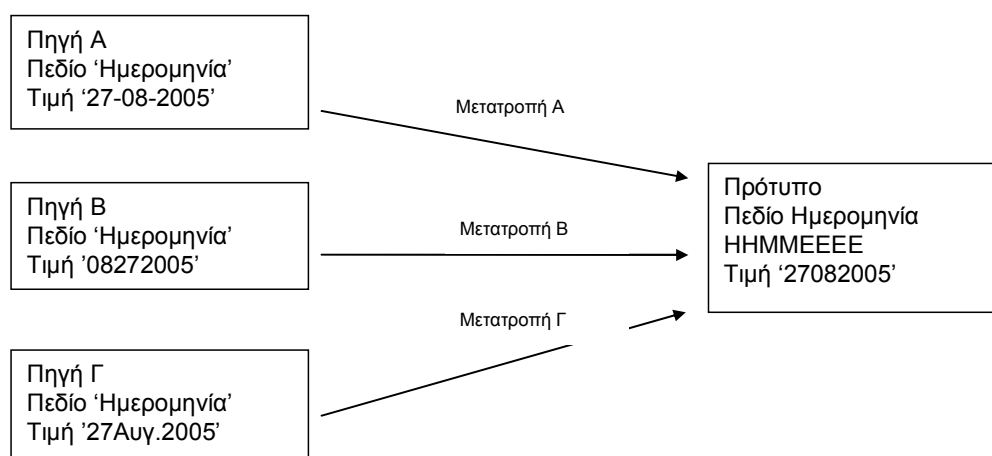
Η αξιοποίηση της πληροφορίας που ενσωματώνεται σε δομημένες ή 'πολυαξονικές' κωδικοποιήσεις, απαιτεί συχνά την εκτέλεση ερωτημάτων σε τμήμα του πεδίου (δηλαδή του κωδικού).

Η δράση προτυποποίησης δεδομένων περιλαμβάνει τα ακόλουθα βήματα:

- Εντοπισμό κοινών χαρακτηριστικών τύπων οντοτήτων σε όλα τα αρχεία ή βάσεις δεδομένων που έχουν επιλεχθεί κατά το στάδιο 'Αναγνώριση πηγών δεδομένων'
- Εντοπισμό ρόλων που αξιοποιούν τα δεδομένα από τα αρχεία αυτά

- Σύνταξη πρότασης για τον ορισμό δεδομένων, επιχειρησιακών κανόνων και την προτυποποίηση αποδεκτών τιμών δεδομένων. Αποστολή της πρότασης στους εμπλεκόμενους στην διαχείριση πληροφορίας ρόλους, για να γίνουν παρατηρήσεις.
- Τεκμηρίωση των συμφωνημένων: ορισμού δεδομένων, επιχειρησιακών κανόνων και προτυποποίησης αποδεκτών τιμών δεδομένων. Ορισμός ρόλου υπευθύνου για την συντήρηση της τεκμηρίωσης (information steward).
- Αξιολόγηση ποιότητας ορισμών και αρχιτεκτονικής πληροφορίας βάσει των κριτηρίων της ενότητας ‘Αξιολόγηση ποιότητας πληροφορίας’.
- Αντιστοίχιση δεδομένων επιλεγμένων αρχείων (‘πηγές δεδομένων’) με προτυποποιημένα δεδομένα (mapping source data to standardized data)
- Τεκμηρίωση κανόνων μετατροπής δεδομένων επιλεγμένων αρχείων σε προτυποποιημένους τύπους οντοτήτων και χαρακτηριστικών αυτών

Στην ακόλουθη εικόνα 28 περιγράφεται παράδειγμα προτυποποίησης δεδομένων: έχουν επιλεγθεί 3 πηγές δεδομένων που τηρούν στοιχεία ημερομηνίας συναλλαγών. Η δόμηση του πεδίου ‘Ημερομηνία’ είναι διαφορετική, όπως φαίνεται στο αριστερό μέρος. Στο δεξί μέρος φαίνεται η προτυποποιημένη δόμηση που προβλέπει τα ψηφία 1-2 για την ημέρα (HH), ψηφία 3-4 για τον μήνα (MM), ψηφία 5-8 για το έτος (EEEE). Η μετατροπή σε κάθε μια περίπτωση είναι διαφορετική. Στην πηγή Α η δόμηση αναγνωρίζεται από επαρκές δείγμα ότι είναι HH-MM-EEEE, οπότε η μετατροπή Α, αφορά την αφαίρεση του 3^{ου} και του 6^{ου} ψηφίου.



Εικόνα 3 – Προτυποποίηση πεδίου δεδομένων

1.5.1.4.4 Διόρθωση, προτυποποίηση και συμπλήρωση δεδομένων

Η διαδικασία διόρθωσης, προτυποποίησης και συμπλήρωσης, συμβάλλει στην βελτίωση της ποιότητας δεδομένων του Οργανισμού. Περιλαμβάνει τα ακόλουθα:

- I. Εντοπισμός δεδομένων προς συμπλήρωση. Εντοπίζονται πεδία όπου δεν υπάρχουν τιμές ή υπάρχουν ‘κενές τιμές’ (null values). Επιπλέον με την εφαρμογή κανόνων ακεραιότητας εντοπίζονται ανύπαρκτες εγγραφές. Π.χ. στην περίπτωση διασύνδεσης οντότητας τραπεζικού λογαριασμού με οντότητα του

- Πελάτη- ιδιοκτήτη μέσω κλειδιού Πελάτη, για κάθε τραπεζικό λογαριασμό πρέπει να υπάρχει διασύνδεση σε εγγραφή Πελάτη. Με κατάλληλο ερώτημα εντοπίζονται κλειδιά σε εγγραφές που πιθανόν λείπουν.
- II. Εντοπισμός εμφανώς λανθασμένων ή υπόπτων δεδομένων:
- Δεδομένα εκτός πεδίου αποδεκτών τιμών (domain value set)
 - Διπλές εγγραφές σε πεδία που πρέπει να παίρνουν μοναδικές τιμές (π.χ. αποτελούν πρωταρχικό κλειδί (primary key) εύρεσης στοιχείων πραγματικής οντότητας)
- III. Καθορισμός προσέγγισης για τον καθαρισμό δεδομένων:
- Εφαρμογή κανόνων αυτοματοποιημένης προτυποποίησης-διόρθωσης-συμπλήρωσης δεδομένων. Σε κάποιες περιπτώσεις είναι εφικτή η εφαρμογή λογισμικού που υλοποιεί κανόνες για τον καθαρισμό. Η ύπαρξη βάσεων δεδομένων αναφοράς, έναντι των οποίων μπορεί να γίνει έλεγχος και διόρθωση ή συμπλήρωση δεδομένων, διευκολύνει την αυτοματοποίηση.
 - Καθαρισμός δεδομένων με χειρονακτικό έλεγχο έναντι της πραγματικής οντότητας και μη αυτοματοποιημένες παρεμβάσεις. Π.χ. η κλήση προς Πρόσωπο με σκοπό την επιβεβαίωση ή/και συμπλήρωση των δεδομένων που υπάρχουν για αυτό.
 - Καθαρισμός δεδομένων με χειρονακτικό έλεγχο έναντι αξιόπιστης υποκατάστατης πηγής. Π.χ. έλεγχος στοιχείων έναντι εντύπων που καταγράφουν μια συναλλαγή.
- IV. Προτεραιοποίηση διαδικασιών καθαρισμού, λαμβάνοντας υπόψη την αξία του καθαρισμού σε σχέση με το εκτιμώμενο κόστος. Επιλέγονται οι κατηγορίες δεδομένων που έχουν σημαντική αξία για τον Οργανισμό και αποδεκτό κόστος καθαρισμού. Για κάθε κατηγορία που επιλέγεται καταγράφεται ο τρόπος καθαρισμού.
- V. Εκτέλεση του καθαρισμού στα επιλεγμένα δεδομένα και τεκμηρίωση των εργασιών και των αποτελεσμάτων – ευρημάτων:
- καταγραφή των εγγραφών που διορθώθηκαν και του τρόπου διόρθωσης (π.χ. χρήση εγχειριδίων κωδικοποιήσεων που εφαρμόζει ο Οργανισμός για την διόρθωση κωδικών, εφαρμογή επιχειρησιακών κανόνων),
 - καταγραφή των εγγραφών – πεδίων που δεν ήταν εφικτός ο καθαρισμός τους,
 - καταγραφή του κόστους καθαρισμού: χρόνος που αναλώθηκε, κόστος ανάπτυξης λογισμικού, κόστος πρόσβασης σε πηγές προς σύγκριση στοιχείων
- VI. Καθορισμός του τρόπου χειρισμού κατηγοριών και περιπτώσεων δεδομένων που δεν επιλέχθηκαν προς καθαρισμό. Επιλογές είναι:
- η απόρριψη των δεδομένων και η καταγραφή τους ως ανεπαρκούς ποιότητας,
 - η αποδοχή των δεδομένων (θεωρώντας ότι το κόστος χαμηλής ποιότητας είναι αποδεκτό), με καταγραφή των προβληματικών σημείων,
 - αποδοχή των δεδομένων και συμπλήρωση κάποιων πεδίων με προσεγγιστικές τιμές βάσει κάποιων εκτιμήσεων, επιλογή που ενέχει κινδύνους σημαντικής απώλειας της ακρίβειας
- Οι ακόλουθες κατηγορίες δεδομένων απαιτούν ειδικές τεχνικές ‘καθαρισμού’ δεδομένων:
- Οντότητες με σχετικά σταθερά χαρακτηριστικά που όμως μπορεί να αλλάζουν στην διάρκεια του χρόνου (στην ορολογία μοντέλων διαστάσεων τα χαρακτηριστικά αυτά λέγονται ‘αργά μεταβαλλόμενες διαστάσεις’ (slowly changing dimensions)). Στην περίπτωση αυτή η διόρθωση βάσει ιστορικών τιμών δεν είναι ασφαλής πρακτική.

- Οντότητες που περιγράφουν γεγονότα (πώληση, κατάθεση). Μόλις ολοκληρωθεί ο κύκλος ζωής μιας συναλλαγής, δεν επιτρέπονται αλλαγές στις τιμές.
- Ονόματα και διευθύνσεις. Χρήση τηλεφωνικών καταλόγων για την επιβεβαίωση διευθύνσεων και τηλεφώνων βάσει Ονόματος (συνήθως με χειρονακτική εξέταση). Χρήση καταλόγων αντιστοίχισης συνωνύμων ονομάτων (π.χ. Γιάννης, Ιωάννης).

1.5.1.4.5 Ενοποίηση δεδομένων από διαφορετικές πηγές (data consolidation)

Επιδιώκεται η ενοποίηση δεδομένων που περιγράφουν τον ίδιο τύπο οντότητας αλλά τηρούνται σε διαφορετικά αρχεία ή βάσεις δεδομένων (π.χ. διάφορα αρχεία που τηρούν στοιχεία Πελατών).

Δεδομένα διαφόρων πηγών στα οποία έχουν εφαρμοστεί διαδικασίες προτυποποίησης-διόρθωσης-συμπλήρωσης, μπορούν να ενοποιηθούν σε μοναδική βάση αναφοράς που περιέχει μοναδική εγγραφή για κάθε πραγματική οντότητα (π.χ. υλικό παραγωγής που χρησιμοποιείται).

Παρατήρηση: Αρκετοί Οργανισμοί διαπιστώθηκε στο παρελθόν ότι διαθέτουν πολλαπλά και ασύνδετα αρχεία για τον ίδιο τύπο δεδομένων (π.χ. αρχεία Πελατών).

Σε διεθνές επίπεδο παρατηρείται δυσκολία των Επιχειρήσεων να ενοποιήσουν πληροφορία Πελατών. Στο πλαίσιο αυτό έχει αναπτυχθεί αγορά ενοποίησης πληροφορίας Πελατών (Customer Data Integration – CDI) για να βοηθήσει τις Επιχειρήσεις στην στρατηγική επιλογή για την καλύτερη αξιοποίηση της πληροφορίας Πελατών.

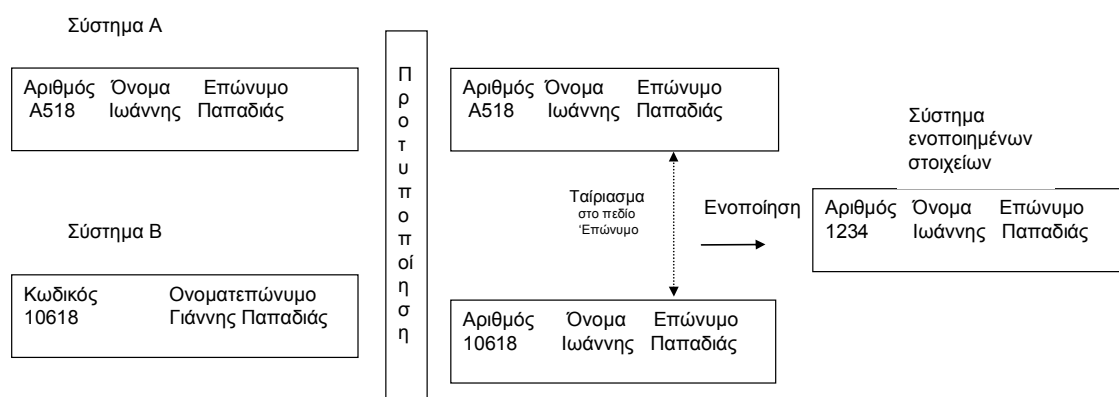
Η ενοποίηση των δεδομένων περιλαμβάνει τα ακόλουθα βήματα:

- I. Καθορισμός κριτηρίων ενοποίησης εγγραφών. Επιλέγονται πεδία ή συνδυασμοί πεδίων που θα μπορούσαν να αποτελέσουν βάση για την διασύνδεση εγγραφών διαφορετικών πηγών, που θεωρείται ότι αντανakλούν την ίδια πραγματική οντότητα και είναι υποψήφιες προς ενοποίηση. Τέτοια πεδία μπορεί να είναι αριθμοί ταυτοποίησης, ονόματα, διευθύνσεις που δεν αλλάζουν ή δεν αλλάζουν συχνά. Αξιολογούνται τα πεδία ή οι συνδυασμοί αυτών, που αποτελούν την πλέον αξιόπιστη βάση διασύνδεσης εγγραφών. Π.χ. κοινή τιμή στο πεδίο αριθμού ταυτοποίησης και παρόμοια τιμή στο όνομα, είναι ισχυρότερο κριτήριο από κοινή τιμή στο πεδίο ονόματος και παρόμοια τιμή αριθμού ταυτοποίησης.
- II. Απόδοση συντελεστή βαρύτητας σε κάθε κριτήριο διασύνδεσης εγγραφών.
- III. Αξιολόγηση του αρνητικού επιχειρησιακού αντίκτυπου στην ενοποίηση στοιχείων που χαρακτηρίζουν διαφορετικές πραγματικές οντότητες (π.χ. ενοποίηση στοιχείων 2 διαφορετικών Πελατών σε μια εγγραφή).
- IV. Καθορισμός των τεχνικών που θα χρησιμοποιηθούν για την διασύνδεση εγγραφών: ακριβής ταυτοποίηση αλφαριθμητικών χαρακτήρων στο πεδίο, κοινό τμήμα πεδίου με χρήση wild card για το υπόλοιπο, παρόμοιες τιμές σε πεδίο με διαφορές σε 1-2 χαρακτήρες.
- V. Χρήση κριτηρίων διασύνδεσης για τον εντοπισμό εγγραφών που πιθανόν αντιστοιχούν στην ίδια πραγματική οντότητα.
- VI. Αξιολόγηση ευρημάτων και ενοποίηση στοιχείων σε μοναδική εγγραφή που κρατά τα πιο αξιόπιστα στοιχεία από τις εγγραφές που συγχωνεύονται

- VII. Τεκμηρίωση των κανόνων διασύνδεσης και ενοποίησης στοιχείων από πηγές δεδομένων στην νέα δομή
- VIII. Καταγραφή εγγραφών που διασυνδέονται (περιγράφουν την ίδια πραγματική οντότητα), μέσω της καταγραφής πεδίων ταυτοποίησης
- IX. Τήρηση στοιχείων που επιτρέπουν την διόρθωση λανθασμένης ενοποίησης

Έχοντας εκτελέσει τις εργασίες που περιγράφονται στην ενότητα ‘Διόρθωση, προτυποποίηση και συμπλήρωση δεδομένων’ και ‘Ενοποίηση δεδομένων από διαφορετικές πηγές’, καταγράφεται και αναλύεται ο τρόπος δημιουργίας ελαττωματικών δεδομένων με στόχο την βελτίωση διαδικασιών και την επίλυση προβλημάτων που προκύπτουν από την ανεπαρκή ποιότητα δεδομένων. Αναλύονται οι τύποι ελαττωμάτων, η συχνότητα, ο επιχειρησιακός αντίκτυπος των λαθών. Συγκεκριμένοι τύποι ελαττωμάτων έχουν συστηματικό τρόπο δημιουργίας αλλά και διόρθωσης.

Η συνολική διαδικασία προτυποποίησης και ενοποίησης δεδομένων Πελατών, περιγράφεται στην ακόλουθη εικόνα 29:



Εικόνα 4 – Προτυποποίηση και ενοποίηση δεδομένων πελατών

1.5.1.4.6 Μετατροπή δεδομένων

Αφορά την μετατροπή των καθαρισμένων δεδομένων για να υλοποιηθεί η φόρτωση στην νέα αρχιτεκτονική δεδομένων.

Στον παρακάτω πίνακα παρατίθεται παράδειγμα μετατροπής πεδίων αποδεκτών τιμών (domain value conversion) συστημάτων, ώστε να προκύψει μοναδική κωδικοποίηση.

Στο σύστημα Α η τιμή 4 αποτυπώνει το ‘άριστα’ ενώ στο σύστημα Β η τιμή 1 αποτυπώνει το ‘άριστα’. Στην προτυποποιημένη μορφή ακολουθείται η κωδικοποίηση του συστήματος Β.

Διαφορά υφιστάμενα πεδία αποδεκτών τιμών (domain values) που καταγράφουν το ίδιο χαρακτηριστικό.		Προτυποποιημένη χρήση τιμών
Σύστημα Α: 4, 3, 2, 1	Μετατροπή	1, 2, 3, 4
Σύστημα Β: 1, 2, 3, 4	-	1, 2, 3, 4
Σύστημα Γ: άριστα, πολύ καλά, καλά, μέτρια	Μετατροπή	1, 2, 3, 4

1.5.1.4.7 Έλεγχος διαδικασίας αναδόμησης δεδομένων

Ο έλεγχος επιδιώκει να διασφαλίσει την ποιότητα σε όλα τα στάδια της διαδικασίας αναδόμησης, βάσει των ακόλουθων στοιχείων:

- Δείγματα δεδομένων από όλες τις φάσεις αναδόμησης
- Ορισμούς δεδομένων και επιχειρησιακούς κανόνες ‘πηγής’ και τελικής δόμησης
- Κανόνες ενοποίησης και αναδόμησης δεδομένων

Ελέγχεται:

- η άντληση των δεδομένων από τις προβλεπόμενες πηγές με πληρότητα
- ο αριθμός εγγραφών που περνά από το ένα στάδιο στο άλλο είναι σύμφωνος με το αποτέλεσμα της επεξεργασίας σε κάθε στάδιο (π.χ. ο αριθμός άνευ αξίας εγγραφών που δεν προχωρούν στο επόμενο στάδιο)
- η ενοποίηση δεδομένων με επιτυχία
- η μετατροπή πεδίου αποδεκτών τιμών

1.5.1.5 Βελτίωση διαδικασιών διασφάλισης ποιότητας πληροφορίας

Μέχρι το σημείο αυτό περιγράφηκαν εργασίες βελτίωσης της ποιότητας των υφιστάμενων βάσεων δεδομένων του Οργανισμού. Παρατηρήθηκε ότι σε πολλές περιπτώσεις η βελτίωση της ποιότητας των δεδομένων είναι, είτε ιδιαίτερα κοστοβόρα εργασία, είτε ανέφικτη σε προχωρημένο στάδιο της αξιακής αλυσίδας.

Η προληπτική διαχείριση ποιότητας της διαδικασίας παραγωγής πληροφορίας, σε αντίθεση με την εκ των υστέρων ‘πυροσβεστική’ απόπειρα καθαρισμού των δεδομένων, έχει σημαντικά πλεονεκτήματα. Το κόστος της εκ των υστέρων βελτίωσης δεδομένων είναι συχνά μεγαλύτερο από το κόστος διαχείρισης ποιοτικών δεδομένων από την στιγμή αρχικής καταγραφής τους (αρχικό στάδιο αξιακής αλυσίδας πληροφορίας). Η αρχική πλήρης καταγραφή δεδομένων, έχει σημαντικά μικρότερο κόστος, σε σχέση με την προσπάθεια εκ των υστέρων συμπλήρωσης των δεδομένων.

Η προσπάθεια βελτίωσης διαδικασιών παραγωγής πληροφορίας, εντάσσεται σε φιλοσοφία συνεχούς βελτίωσης διαδικασιών (continuous process improvement).

Αποτελεσματικοί Οργανισμοί έντασης πληροφορίας, επιδιώκουν την βελτίωση των επιχειρησιακών διαδικασιών που δημιουργούν και διαχειρίζονται τους πληροφοριακούς τους πόρους, μεγιστοποιώντας την παραγόμενη αξία. Διαδικασίες που δημιουργούν προβλήματα στην ποιότητα πληροφορίας, τροποποιούνται και βελτιώνονται συστηματικά. Παραδείγματα τέτοιων περιπτώσεων είναι :

- η καταγραφή του ίδιου δεδομένου ή δεδομένων περισσότερες της μιας φορές, αντικαθίσταται από μοναδική καταγραφή δεδομένων (single data capture)
- η ύπαρξη περισσότερων του ενός σταδίων μέχρι την τελική εισαγωγή των δεδομένων σε ηλεκτρονική βάση δεδομένων αναφοράς (καταγραφή σε έντυπη μορφή και μετέπειτα εισαγωγή ή καταγραφή σε ηλεκτρονική μορφή αλλά εσωτερική διακίνηση σε έντυπη μορφή και επανεισαγωγή), αποτελεί κακή πρακτική που μπορεί να δημιουργήσει προβλήματα στην ποιότητα των δεδομένων. Τα ενδιάμεσα αυτά στάδια πρέπει να απαλείφονται.
- η πληροφορία είναι διαθέσιμη σε κάποιο Τμήμα, αλλά δεν γίνεται άμεσα διαθέσιμη και αξιοποιήσιμη στον ευρύτερο Οργανισμό. Στόχος είναι να περιορίζεται ο χρόνος μη διαθεσιμότητας της πληροφορίας (information float), με την άμεση εισαγωγή αυτής σε βάση αναφοράς για τον Οργανισμό.

Η ύπαρξη δεδομένων χαμηλής ποιότητας, αποτελεί σοβαρή ένδειξη για την ύπαρξη προβληματικής ή ατελούς διαδικασίας (π.χ. διαδικασίας που δεν υφίσταται ποιοτικό έλεγχο). Ενώ ο καθαρισμός δεδομένων επιλύει το πρόβλημα πρόσκαιρα ή βελτίωση της διαδικασίας επιλύει το πρόβλημα μακροπρόθεσμα, εμποδίζοντας την επανεμφάνιση του.

Στην παρακάτω εικόνα 30 απεικονίζεται μέθοδος ανάλυσης του αιτίου (fishbone analysis), που προκαλεί ελαττωματικά δεδομένα (στην συγκεκριμένη περίπτωση διπλές εγγραφές Πελάτη).



Εικόνα 5 - Διάγραμμα τύπου fishbone, προσαρμοσμένο από [English]

Συχνά η προσπάθεια αύξησης της παραγωγικότητας, οδηγεί στην πλημμελή συμπλήρωση πληροφορίας κατά την εισαγωγή αυτής σε σύστημα.

ΕΥΡΕΤΗΡΙΟ

- DSS, 9
information float, 28
ακρίβεια, 5
Γνώση, 3
Δεδομένα, 3, 24, 25
ενιαίου γλωσσάριου επιχειρησιακών
όρων (Corporate dictionary), 11
Εργαζόμενοι-γνώσης, 4
κανονισμούς ακεραιότητας δεδομένων
(integrity constraints), 13
Κριτήρια αξιολόγησης, 4
μοναδική καταγραφή δεδομένων, 28
πλαίσιο παρουσίασης, 5
πληροφορία, 3, 4, 6, 8, 9, 10, 11, 13,
16, 17, 21, 23, 25, 28
ρίσκου, 6
υποκατάστατη πηγή, 18