

UNIVERZA V LJUBLJANI  
Fakulteta za elektrotehniko

Ivan Pavlović

ZBIRANJE PODATKOV KLINIČNEGA PRESKUŠANJA  
S POMOČJO SPLETNO ZASNOVANEGA SISTEMA  
Z REDUCIRANIM UPRAVLJANJEM PODATKOV

DOKTORSKA DISERTACIJA

LJUBLJANA, 2010



UNIVERZA V LJUBLJANI  
Fakulteta za elektrotehniko

Ivan Pavlović

ZBIRANJE PODATKOV KLINIČNEGA PRESKUŠANJA  
S POMOČJO SPLETNO ZASNOVANEGA SISTEMA  
Z REDUCIRANIM UPRAVLJANJEM PODATKOV

DOKTORSKA DISERTACIJA

Mentor: profesor dr. Damijan Miklavčič

LJUBLJANA, 2010



## Zahvala

*Iskreno se zahvaljujem svojemu mentorju, prof. dr. Damijanu Miklavčiču, za strokovno vodstvo, koristne nasvete in konstruktivno kritiko pri nastajanju doktorske disertacije, kot tudi za vso podporo v profesionalnem in privatnem življenju, brez katere bi me marsikateri izziv, ki ga novo okolje prinaša, lahko oddaljil od doktorata in tudi od življenja, ki mi ga je uspelo v zadnjih desetih letih zgraditi.*

*Zahvaljujem se Bredi Slekovec z Onkološkega inštituta v Ljubljani in Katarini Macan iz podjetja Clinres Farmacija iz Zagreba, ker sta mi posvetili svoj dragoceni čas in mi pomagali razumeti organizacijo in procese zbiranja podatkov kliničnih študij.*

*Posebej se zahvaljujem Ivanu Lazareviću in podjetju Omnicom iz Beograda, ker so načrtovani sistem spravili v življenje.*

*Svojim roditeljima se zahvaljujem što su mi bili podrška u prelomnim odlukama koje su me kroz život dovele do doktorata.*

*Hvala Tjaši, ker je bila in je še vedno zraven in ker mi je podarila Lolo.*



## Povzetek

Klinično preskušanje ali klinična študija je skrbno načrtovana raziskava učinkov zdravila, medicinske terapije, ali naprave na skupini subjektov. Praviloma klinično študijo zasnuje, financira in zanjo odgovarja sponzor, ki je lahko farmacevtsko podjetje, podjetje z medicinskimi pripomočki, ali neprofitna organizacija. Za registracijo zdravila slovenska in evropska zakonodaja zahteva obvezno klinično preskušanje, s katerim se lahko pridobi ocena njegove kakovosti, varnosti in učinkovitosti.

Razlikujemo štiri faze kliničnih preskušanj:

- faza I – začetna raziskava na majhnem številu oseb za ugotovitev varnosti in toksičnosti;
- faza II – ugotovitev učinka zdravila na različne vrste bolezni;
- faza III – dokazovanje varnosti in učinkovitosti zdravila za določeno bolezen;
- faza IV – študije po odobritvi prodaje zdravila (večinoma namenjena tržnim raziskavam).

Za razliko od faze I, ki večinoma poteka na zdravih prostovoljcih, so faze II do IV prava klinična preskušanja in se vedno izvajajo na bolnikih. Število bolnikov vključenih v preskušanje je lahko tudi več tisoč. Običajno takšna preskušanja potekajo v več raziskovalnih centrih (klinike, medicinski inštituti), pogosto v več državah, kar zagotavlja hitrejšo izvedbo študije. Takšnim preskušanjem rečemo multicentrične študije.

Pri multicentričnih študijah je spremljanje študije in zbiranje podatkov zelo zahtevno. Še vedno je najbolj razširjena oblika zbiranja podatkov na papirnih obrazcih. V tem primeru ima raziskovalec za vsakega bolnika poseben obrazec, po navadi z več deset stranmi, v katerega vnaša podatke o bolniku, terapiji in odzivu na terapijo. Takšnemu obrazcu rečemo CRF (Case Report Form) in je za vsako študijo posebej sestavljen in enak za vse bolnike. Med potekom kliničnega preskušanja monitorji (strokovne osebe zadolžene za spremljanje študije) občasno obiskujejo raziskovalne centre in preverjajo, ali študija poteka po načrtu in če so podatki v obrazcih pravilno vneseni. Na koncu študije raziskovalni centri pošljejo obrazce sponzorju na obdelavo zbranih podatkov.

Sodobnejši način zbiranja podatkov je elektronsko zbiranje podatkov (EDC - Electronic Data Collection). Pri EDC raziskovalci vnašajo podatke neposredno v računalnik, v tako imenovane elektronske CRF – eCRF. Na ta način se zmanjša število napačnih podatkov, ki so pogosto posledica prepisovanja podatkov iz papirnih CRF-ov v računalniške baze podatkov brez prisotnosti raziskovalcev. EDC prinaša številne druge prednosti, kot so preverjanje pravilnosti podatkov že pri vnosu, povezljivost z drugimi viri podatkov (npr.

laboratorijskimi izvidi, CT ali MR slikami ipd.), ali elektronsko spremljanje poteka študije z dinamično generiranimi poročili in vnaprej definiranimi delnimi obdelavami podatkov.

Čeprav EDC rešitve obstajajo že poldrugo desetletje, zbiranje podatkov na papirnih obrazcih še vedno prevladuje. Zadnjih nekaj let so s pospešenim razvojem Interneta spletno zasnovane EDC rešitve (iEDC) stopile v središče pozornosti. S spletnimi rešitvami so načeloma tehnološko zahtevne EDC rešitve postale bolj preproste, ker na strani uporabnika praviloma ne zahtevajo nobene programske opreme razen spletnega brskalnika.

Nova tehnološka rešitev pogosto sama po sebi ne prinaša veliko prednosti brez istočasne spremembe procesa zbiranja podatkov, ki mora ustrezati novi tehnološki rešitvi. Zato je pri načrtovanju nove rešitve za iEDC sistem zelo pomembno, da smo enako pozorni na uvajanje novega procesa kot na tehnološke posege. Nova rešitev mora enako ali manj obremenjevati osebe vključene v proces klinične študije, ali s prerazporeditvijo dela in nalog optimizirati porabo človeških in materialnih virov.

Za potrebe multicentričnih kliničnih študij smo razvili sistem za zbiranje podatkov klinične študije zasnovan na spletni tehnologiji (iEDC). Pri zasnovi sistema smo upoštevali obstoječe zakonske akte in priporočila na področju kliničnih preskušanj in elektronskega zbiranja podatkov.

iEDC sistem je sestavljen iz platforme in podatkovne baze in omogoča hitro postavitve nove študije. Platforma je programsko okolje, v katerem je mogoče postaviti katerokoli klinično študijo. Osnovna vloga platforme je, da zagotovi pretok podatkov med uporabniki in delegiranje nalog med študijo. Platforma omogoča tudi generiranje poročil o poteku študije in zgodovini vnosa in sprememb podatkov ter prikaz statistik, prikaz večpredstavnih vsebin in povezljivost z drugimi podatkovnimi viri. Platforma je razvita in deluje v Microsoft .NET okolju. Baza podatkov je sestavljena iz dveh segmentov: prvega, ki vsebuje objekte potrebne za delovanje platforme; in drugega, ki vsebuje objekte značilne za posamezno klinično študijo (tabele podatkov, procedure za generiranje poročil ipd.). Baza podatkov je lahko postavljena na Microsoft SQL Server strežniku.

Skupaj z novo tehnološko rešitvijo smo uvedli tudi nov proces za vnos in preverjanje podatkov zbranih med klinično študijo. Ta proces je fleksibilen glede števila udeležencev v procesu. Zaradi tega je sistem mogoče uporabljati tudi v primeru majhnega števila udeležencev procesa zbiranja podatkov. iEDC sistem skupaj z novim procesom vnosa in preverjanja podatkov prispeva k boljši kakovosti zbranih podatkov in krajšemu trajanju študije glede na klasično zbiranje podatkov.

Pri zasnovi novega procesa smo izhajali iz načela, da morata biti v proces vnosa in preverjanja vsakega podatka vključeni najmanj dve osebi. Na ta način se pomembno



zmanjša število napak med vnesenimi podatki. V prenovljenem procesu smo uporabnikom dodelili različne vloge:

- raziskovalec / raziskovalna sestra – vnaša podatke;
- monitor – preverja in potrjuje vnesene podatke v enem ali več raziskovalnih centrih;
- koordinator – kontrolira potek raziskave v vseh centrih vključenih v klinično raziskavo.

Omogočili smo podvajanje in zamenjavo vlog raziskovalec in monitor za enega uporabnika, tako da je lahko raziskovalec po potrebi tudi monitor za obrazce, ki jih je izpolnil drugi raziskovalec iz istega raziskovalnega centra. Na ta način smo omogočili izvajanje študije z manjšim številom udeležencev, kar je posebej primerno za t.i. nekomercialne študije, katere pogosto nimajo finančne možnosti za formalno zagotavljanje monitoringa.

V skladu z načeli načrtovanja uporabnosti (angl. »usability engineering«) pri razvoju interaktivnih sistemov smo do končne rešitve za EDC sistem prišli prek več korakov. Te korake smo imenovali pilotni sistemi. Prvi pilotni sistem EDCp1 smo razvili in uporabili za potrebe mednarodne onkološke multicentrične študije. Po dveh letih uporabe sistema smo prosili uporabnike sistema, da ocenijo sistem. Ponudili smo jim spletno anketo, v kateri so ocenjevali različne funkcionalnosti in aspekte uporabnosti EDCp1 (glej Dodatek: Spletna anketa uporabnosti pilotnega sistema EDCp1). Poleg tega smo sledili uporabi sistema in naša opazovanja izkoristili za predlog izboljšav prototipa. V skladu z rezultati spletne ankete in našimi ugotovitvami smo predlagali izboljšave prototipa in jih vgradili v naslednji prototip EDCp2. Prototip EDCp2 smo razvili in uporabili na enocentrični onkološki študiji ter s sledenjem uporabe prototipa preverili uspešnost izboljšav.

Z modeliranjem procesov je mogoče analizirati prednosti in pomanjkljivosti spremembe procesov v pogledu stroškov in trajanja procesov. Na osnovi naših modelov zbiranja podatkov na papirju in elektronskega zbiranja podatkov smo naredili izračune stroškov procesov in ocenili razliko v trajanju procesov za vzorčno klinično študijo in za variacije parametrov študije in procesov. Z enako metodo, ob upoštevanju dejanskih vrednosti parametrov posamezne študije in procesov, je mogoče ugotoviti prednosti oziroma pomanjkljivosti elektronskega zbiranja podatkov pred zbiranjem podatkov na papirju tako glede stroškov kot tudi trajanja zbiranja podatkov.

Naši rezultati kažejo, da do večine prihrankov ob uporabi EDC sistema prihaja zaradi zmanjšanja stroškov monitoringa in upravljanja s podatki. Točne vrednosti prihrankov so odvisne od parametrov, ki se v izračunih pojavljajo. Na primer, velikost klinične študije pomembno vpliva na prihranke, ki jih EDC sistem prinaša, medtem ko so stroški manj občutljivi na spremembe vrednosti povprečnih naporov za izvajanje posameznih nalog. Takšni rezultati niti niso presenetljivi, ampak naš pristop omogoča, da jih kvantificiramo. Z

natančnejšimi ocenami parametrov za posamezno študijo lahko pridemo do zanesljivih ocen prihrankov ob prehodu s PDC na EDC sistem, zaradi katerih se lažje odločamo za takšen prehod. Z enakim pristopom, kot smo ga predstavili v naši nalogi, je mogoče narediti tudi celotno oceno stroškov klinične študije z modeliranjem klinične študije kot enotnega poslovnega procesa.

## Abstract

A clinical trial or clinical study is a carefully designed study of effects of drug, or medical treatment, or a device on a group of entities. Usually, a clinical trial is initiated and financed by a sponsor, who may be a pharmaceutical company, a medical device company, or some non-profit organization. According to Slovenian and European legislation the process of drug registration requires an estimation of the quality, safety and efficacy through a clinical trial.

There are four types of clinical trials:

- Phase I - initial survey on a small number of persons to determine safety and toxicity;
- Phase II - findings on drug impact on different types of diseases;
- Phase III - proving safety and efficacy of a drug for a specific disease;
- Phase IV - studies after drug registration (mainly intended for marketing purposes).

Unlike Phase I, which is usually performed on healthy volunteers, Phase II to IV are real clinical trials and should always be carried out on patients. The number of patients enrolled in the trial may be several thousand. Typically, such clinical trials are simultaneously run in several research centers (clinics, medical institutes), often in several countries, providing faster execution of the study. Such clinical trials are known as multicenter clinical trials.

Study monitoring and data collection within multicenter clinical trials are very difficult. However, data collection on paper forms (paper data collection - PDC) is still the most widespread way of collecting multicenter clinical trial data. In PDC the investigator uses a special paper form for each patient, usually with dozens of pages, to collect the data on patient treatment and response to therapy. Such forms are called Case Report Forms (CRF). Each study has a different CRF, which is the same for all patients included in the study. During the course of a clinical trial monitors (professionals in charge of monitoring study) occasionally visit research centers in order to check whether the study is proceeding according to the protocol and if the information are correctly collected in CRF. At the end of study research centers send all CRFs to the sponsors, who further process and analyze all the collected data.

Another method of data collection is electronic data collection (EDC). In the case of EDC researchers enter data directly into computers through the so-called electronic CRF (eCRF). In this way the number of incorrect data, which often results from copying data collected on paper-based CRFs by researcher to a computer database by another person

(not researcher), may be reduced. EDC brings many other advantages, such as checking the correctness of the information at the entry point, the connectivity with other data sources (e.g. laboratory information system, CT or MR devices, etc.), or remote monitoring of the study by dynamically generated reports and predefined intermediate data processing.

Although EDC solutions exist for more than two decades, the collection of data on paper forms still prevails. Over the past few years with the accelerated development of Internet web-based EDC solutions (IEDC) became more popular. With the utilization of web-technology the EDC solutions became less complicated for setup and maintenance as the iEDC user usually does not require any software other than a web browser.

New technological solutions often do not bring many advantages without the simultaneous changes in the process (here: data collection process), which should exploit the functionalities of the new technological solution. Accordingly, when designing new solutions for the IEDC system, it is very important to be equally attentive to the introduction of a new process as well as to the technological interventions. The new solution must require equal or less effort by the actors involved in the data collection process, or it should optimize the use of human and material resources through the reallocation of work and tasks.

For the needs of multicenter clinical trials we have developed a system for collecting study data based on web technology (IEDC). When designing the system we took into account the existing legal acts and recommendations in the field of clinical trials and electronic data collection.

The IEDC system consists of a platform and database and allows rapid setup of a new study. The platform is a software environment, in which it is possible to set up any clinical trial (study). The basic role of the platform is to ensure the flow of information between users and the delegation of tasks during the study. The platform also allows generating reports on study execution and history of data entry and data changes as well as display of real-time statistics and multimedia content and even connectivity with other data sources. The platform is developed and operates in Microsoft .NET environment. The database consists of two segments: the first, which includes platform-related system objects, and the second, which contains objects specific to each clinical study (tables, procedures for generating reports, etc.). The database is designed for the Microsoft SQL Server.

Together with the new technological solution we have introduced a new data collection and revision process. This process is flexible regarding the number of participants in the process. Therefore, the system can be used in case of a small number of participants in the data collection process. The IEDC system together with the new data collection and

revision process contributes to a better quality of the collected data and consequently results in shorter duration of the study compared to paper data collection.

While redesigning the process we postulated that at least two different persons should be included in the entry and verification of any information. In this way the number of errors among collected data is significantly reduced. The new process includes the following roles:

- investigator / investigator assistant - data entry;
- monitor - checking and verifying data entered in one or more research centers;
- coordinator – keeping an overview of the course of study in all the centers involved in clinical trial.

We allowed duplication and replacement of the investigator and monitor roles for one user, and hence the investigator may, if necessary, monitor the forms completed by another investigator from the same research center. In this way we enabled data collection with fewer participants in the process, which is particularly appropriate for non-commercial studies, which often lack the financial support for inclusion of professional monitors.

According to the usability engineering principles for development of interactive information systems, we had to take several steps to reach the final solution for the EDC system. In our development process each of these steps resulted in a particular pilot system. The first pilot system (EDCp1) we have developed and used for the needs of an international oncology multicenter clinical trial. After two years of the system usage, we asked users of the system to evaluate the system. We offered them an online survey, which assessed different aspects of functionality and usability of EDCp1. In addition, we monitored the system usage and based on our observations we proposed improvements of the system. According to the online survey and our findings, we suggested improvements and included them in the next prototype (EDCp2). EDCp2 prototype has been developed and used for a single-center oncology clinical trial. The usage of the prototype was observed in order to verify the improvements of the system.

By modeling processes it is possible to analyze advantages and disadvantages of process changes in respect of costs and length of the process. Based on our models of paper data collection and electronic data collection processes, we have made calculations of the process costs and assessed the difference in the duration of processes for the sample clinical trial as well as for the variations of study and process parameters. Using the same method, and while taking into account the actual values of particular study parameters and processes, it is possible to identify the advantages and disadvantages of electronic

data collection compared to paper data collection in terms of both cost and duration of data collection.

Our results show that most benefit comes from reducing monitoring and data management costs. The exact value depends on the estimation of parameters, which affect the calculations. For example, the variations in clinical trial size significantly affect the savings that EDC brings, while the costs are not that sensitive to changes in average efforts for particular tasks. These results are not surprising, but with our approach we offer the way to quantify them. With more reliable estimation of average costs, error and query rates, and considering specific values, we can get more reliable results and use them to decide on switching from paper data collection to electronic data collection. However, we have to emphasize that our models do not include all the aspects of organization and implementation of clinical trials which are affected by the change in data collection approach. The same approach that we presented in our study can nevertheless be used in assessing the costs of a clinical trial by modeling the entire clinical trial as a business process.

# Kazalo

<b>I. Uvod</b> .....	<b>1</b>
I.1. Opis področja .....	1
I.1.1. Klinične študije.....	1
I.1.2. Klinično preskušanje .....	2
I.1.3. Zbiranje podatkov, zgodovina in obstoječe rešitve .....	4
I.1.4. Veljavna zakonodaja in standardi .....	8
I.1.5. Načrtovanje uporabnosti.....	12
I.2. Cilji in struktura naloge.....	14
<b>II. Metode</b> .....	<b>17</b>
II.1. Opisovanje procesov .....	17
II.1.1. eEPC tehnika modeliranja procesov .....	17
II.1.2. Analiza procesov .....	18
II.2. Določanje funkcionalnega okvirja .....	22
II.2.1. Predlog EDC procesa.....	22
II.2.2. Predlog funkcionalnosti .....	22
II.3. Določanje tehnološkega okvirja .....	23
II.4. Končna rešitev iEDC.....	23
II.5. Ocenjevanje.....	23
II.5.1. Stroški .....	24
II.5.2. Čas za končanje zbiranja podatkov.....	26
<b>III. Rezultati</b> .....	<b>27</b>
III.1. Opis PDC procesa .....	27
III.1.1. PDC Zajem podatkov.....	28
III.1.2. PDC Monitoring .....	30
III.1.3. PDC Upravljanje s podatki .....	32
III.2. Funkcionalni okvir rešitve.....	34
III.2.1. Predlog EDC procesa.....	34
III.2.2. Prototipa .....	38

III.2.3. Predlog ključnih funkcionalnosti .....	42
III.3. Tehnološki okvir rešitve .....	44
III.4. Pilotni sistem - iEDC .....	45
III.4.1. Model elektronskega CRF obrazca - eCRF .....	45
III.4.2. Primeri uporabe ( <i>use-cases</i> ).....	51
III.4.3. Model upravljanja z obrazci .....	55
III.4.4. Opis strukture sporočil v sistemu .....	60
III.4.5. Uporabniški vmesnik .....	61
III.4.6. Varnostni mehanizmi .....	68
III.4.7. Arhitektura in tehnološke zahteve sistema.....	69
III.5. Ocene .....	70
III.5.1. Stroški .....	70
III.5.2. Čas za končanje zbiranja podatkov .....	81
<b>IV. Razprava .....</b>	<b>83</b>
<b>V. Sklep .....</b>	<b>89</b>
<b>VI. Prispevki k znanosti.....</b>	<b>91</b>
<b>VII. Literatura .....</b>	<b>93</b>
<b>VIII. Seznam kratic.....</b>	<b>99</b>
<b>IX. Priloge.....</b>	<b>101</b>

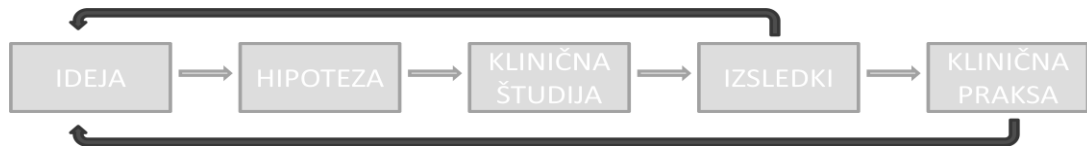


# I. Uvod

## I.1. Opis področja

### I.1.1. Klinične študije

Medicinska praksa se nahaja v obdobju »na izsledkih temelječe medicine« (angl. »evidence based medicine«), kar pomeni, da svojo prakso ne temelji le na intuiciji in logiki ampak tudi na znanstveno utemeljenih dognanjih (Slika 1). Kot osnovno orodje za ustvarjanje medicinskih znanstvenih dognanj, medicina uporablja klinične študije (angl. »clinical study«), v katerih preverja ali ustvarja medicinske hipoteze.



Slika 1. Prikaz principa na izsledkih temelječe medicine

Klinične študije delimo v dve skupini študij, in sicer »eksperimentalne« in »opazovalne« študije. Na Sliki 2 je prikazan način klasifikacije kliničnih študij.

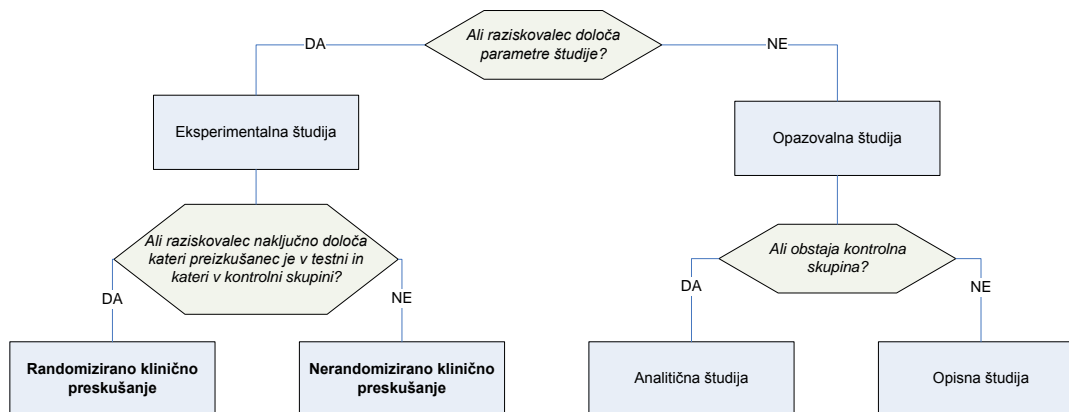
#### Opazovalne klinične študije

V primeru opazovalnih kliničnih študij raziskovalec nima nadzora nad izbiro parametrov študije, kot je na primer doziranje zdravila ali izbira terapije. Pri takšni študiji prihaja do »samo-izbire« parametrov, ali pa gre za tako imenovan »eksperiment narave«. Takšne študije so edine sprejemljive v primeru, ko je neetično izbirati parametre študije. Opazovalne študije ponujajo slabša empirična dokazila, ker brez randomizacije ni mogoče odpraviti sistematičnega odklona (angl. »bias«), ki se lahko pojavi pod vplivom nekontroliranih parametrov študije. Največja vrednost takšnih študij je to, da zagotovijo preliminarne dokaze, ki se lahko uporabijo kot osnova za hipoteze pri močnejših eksperimentalnih študijah, kot je randomizirano klinično preskušanje.

#### Eksperimentalne klinične študije

Značilnost eksperimentalnih študij je ta, da raziskovalec lahko določa parametre študije. Še več, pri randomiziranih kliničnih študijah raziskovalec s postopkom naključne izbire (randomizacija) lahko odstrani eventualni vpliv parametrov, katerih v študiji ne upošteva. Na primer, raziskovalec lahko določa različne količine apliciranega zdravila za različne preizkušance, pri čem preizkušance izbere naključno, in na ta način v dovolj veliki populaciji preizkušancev zagotovi, da razen kontroliranega parametra (količina zdravila)

drugi, nekontrolirani, parametri (na primer starost preizkušanca) ne vplivajo na povprečje rezultatov študije. Dobro načrtovane eksperimentalne klinične študije, t.i. »klinična preskušanja«, nudijo najbolj kakovostna empirična dokazila za medicinske predpostavke, ki jih preskušamo.



Slika 2. Klasifikacija kliničnih študij

### I.1.2. Klinično preskušanje

Kot pravi evropska zakonodaja (Evropski parlament in Svet, 2001), je klinično preskušanje »kaksna koli raziskava na ljudeh z namenom odkriti ali potrditi klinične, farmakološke in/ali druge farmakodinamske učinke enega ali več zdravil v preskušanju in/ali odkriti kakršne koli neželene učinke enega ali več zdravil v preskušanju in/ali preučiti absorpcijo, porazdelitev, presnovo in izločanje enega ali več zdravil v preskušanju, da bi ugotovili njegovo (njihovo) varnost in/ali učinkovitost«. Klinična preskušanja se lahko izvajajo v enem ali več krajih preskušanja v eni ali več državah. V primeru preskušanja pri enem samem raziskovalnem (preskuševalnem) centru govorimo o »enocentričnem kliničnem preskušanju«. »Multicentrično klinično preskušanje« se izvaja po enem samem protokolu, vendar v več kot enem centru.

Klinična preskušanja so pogoj za pridobitev dovoljenja za promet z zdravilom, brez katerega pridelovalci zdravil ne morejo tržiti svojih izdelkov (Vlada RS, 2006). Ker izvajanje kliničnega preskušanja predstavlja znaten strošek, je v interesu pridelovalcev zdravil (oziroma naročnikov preskušanj), da stroške preskušanj čim bolj zmanjšajo z bolj učinkovitim izvajanjem preskušanja. Namreč, stroški razvoja novega zdravila lahko znašajo tudi več kot milijardo evrov, od česar znašajo stroški kliničnih preskušanj skoraj 50 % (Pharmaceutical Research and Manufacturers of America, 2008).

Klinična preskušanja niso vedno namenjena registraciji zdravil, ampak tudi preverjanju učinkovitosti nove kombinacije zdravil oziroma novega terapijskega postopka. Pogosto so nosilci takšnih preskušanj same zdravstveno-raziskovalne organizacije, kot so denimo klinične ustanove ali medicinska združenja (npr. European Organization for Research and Treatment of Cancer - EORTC). Približno polovica vseh investicij v klinična preskušanja gre

za takšna »nekomercialna« preskušanja (Karlberg, 2008). Tudi v primeru nekomercialnih kliničnih preskušanj je cilj organizatorja preskušanja zmanjšati stroške kliničnega preskušanja.

Razlikujemo štiri tipe kliničnih preskušanj:

- faza I – začetna raziskava na majhnem številu oseb za ugotovitev varnosti in toksičnosti;
- faza II – ugotovitev učinka zdravila na različne vrste bolezni;
- faza III – dokazovanje varnosti in učinkovitosti zdravila za določeno bolezen;
- faza IV – študije po odobritvi prodaje zdravila (večinoma namenjena tržnim raziskavam).

Za razliko od kliničnih preskušanj faze I, ki večinoma potekajo na zdravih prostovoljcih, se klinična preskušanja faze II do IV vedno izvajajo na bolnikih. Število bolnikov vključenih v preskušanje faze III je lahko tudi več tisoč. Običajno takšna preskušanja potekajo v več raziskovalnih centrih (klinike, bolnišnice), pogosto v več državah, kar zagotavlja hitrejšo izvedbo preskušanja. Z organizacijo takšnih multicentričnih kliničnih preskušanj poskušajo naročniki študij skrajšati čas izvedbe študije, ker je na več lokacijah možno v krajšem času pridobiti dovolj veliko populacijo preizkušancev.

### *Vloge v procesu kliničnega preskušanja*

Za klinična preskušanja so v zakonodaji jasno določene ključne vloge oz. ključni akterji, ki v tem procesu nastopajo, in sicer:

- "naročnik": posameznik, podjetje, ustanova ali organizacija, ki prevzame odgovornost za organizacijo, vodenje in/ali financiranje kliničnega preskušanja;
- "raziskovalec": zdravnik ali oseba, ki opravlja poklic, ki je zaradi znanstvenih izkušenj in izkušenj z oskrbo bolnikov, ki jih ta poklic zahteva, v državi članici sprejet za raziskave. Raziskovalec je odgovoren za izvajanje kliničnega preskušanja na kraju preskušanja. Če preskušanje na kraju preskušanja izvaja skupina posameznikov, je raziskovalec vodja, ki je odgovoren za skupino, in se lahko imenuje glavni raziskovalec;
- "preizkušanec": posameznik, ki sodeluje v kliničnem preskušanju kot prejemnik zdravila v preskušanju ali kot član kontrolne skupine.

Naročnik preskušanja je odgovoren za izvajanje preskušanja. V njegovem interesu je, da je preskušanje izvedeno s čim manjšimi stroški in da so zbrani podatki čim bolj kakovostni (natančni). Kot vodilo za dobro organizacijo kliničnega preskušanja je ICH (International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use) določila nabor navodil (angl. »guidelines«), imenovan

»Dobra klinična praksa« (angl. »Good Clinical Practice« - GCP). Ker ICH sestavljajo predstavniki evropskih, ameriških in japonskih oblasti ter strokovnjaki farmacevtske industrije, je sledenje tem navodilom (GCP) zahtevano pri večini kliničnih preskušanj.

Ena od ključnih idej GCP je ta, da mora biti preskušanje dokumentirano na takšen način, da je tudi po končanem preskušanju mogoče preveriti vse korake preskušanja in vse podatke zbrane med preskušanjem. Naročniki zaradi tega veliko naporov posvetijo zagotavljanju visoke kakovosti zbranih podatkov. Dva osnovna mehanizma za zagotavljanje visoke kakovosti zbranih podatkov, ki ju naročniki uporabljajo, sta monitoring in centralizirano upravljanje s podatki. Monitoring je mehanizem, pri katerem t.i. »monitorji« (strokovni sodelavci, ki jih angažira naročnik) občasno obiskujejo raziskovalne centre in ugotavljajo morebitna odstopanja od protokola preskušanja ter primerjajo zbrane podatke z izvirnimi podatki v originalni zdravstveni dokumentaciji preizkušanca. Za namen centraliziranega zbiranja podatkov naročnik praviloma vzpostavi podatkovni center, v katerem t.i. »koordinator« preskušanja skupaj z upravljavci podatkov skrbi, da so podatki posredovani iz vseh raziskovalnih centrov pravočasno zbrani in preverjeni ter shranjeni v digitalni obliki za potrebe analize rezultatov preskušanja. Torej, naročnik kliničnega preskušanja za potrebe zbiranja podatkov zagotavlja naslednje ključne vloge:

- »monitor«: preverja podatke v raziskovalnem centru;
- »koordinator«: koordinira delo raziskovalnih centrov in njihovo prispevanje podatkov;
- »upravljevec podatkov«: skrbi za centralizirano digitalizacijo podatkov in preverjanje kakovosti podatkov.

V tem poglavju omenjeni akterji se praviloma pojavljajo pri vseh kliničnih preskušanjih in smo jih zato tudi izpostavili. V nadaljevanju so bolj natančno opisani različni pristopi k elektronskemu zbiranju podatkov kliničnih preskušanj.

### **I.1.3. Zbiranje podatkov, zgodovina in obstoječe rešitve**

V začetku 80-tih let so se pojavili osebni računalniki in kmalu potem se je začela njihova uporaba za poslovne in osebne namene. Sredi 80-tih let se je začela uporaba osebnih računalnikov za zbiranje podatkov kliničnih študij. To je prineslo veliko spremembo v načinu zbiranja kliničnih podatkov. Namreč, pred uporabo osebnih računalnikov za zbiranje podatkov, so raziskovalci podatke zbirali izključno na papirnate CRF (angl. »Case Report Forms«) obrazce. Izpolnjene obrazce so raziskovalci posredovali podatkovnemu centru naročnika študije, kjer so bili podatki potem digitalizirani za potrebe nadaljnje obdelave podatkov, analiz in statističnih izračunov. Takšnemu »centraliziranemu« načinu zbiranja podatkov, kjer so podatki digitalizirani na enem mestu s strani strokovnjakov za digitalizacijo podatkov, rečemo tudi papirno zbiranje podatkov (angl. »Paper Data

Collection« - PDC). Glavni nalogi raziskovalcev sta bili, da izpolnijo papirnate CRF obrazce ter da po pregledu digitaliziranih podatkov naročniku pojasnijo morebitne nejasnosti med podatki. Računalniške rešitve za te potrebe se uporabljajo že več kot dve desetletji. Po navadi v tem primeru raziskovalci podatkovnemu centru pošljejo izpolnjene CRF obrazce po pošti ali po faksu. Osebe v podatkovnem centru potem ročno vnaša podatke v računalniško rešitev. Glede na rezultate naše ankete (Dodatek X) kot tudi na podatke iz številnih virov (Atoji, 2007; Mastro, 2007; Welker, 2007; The eClinical Forum & PhRMA EDC/eSource Taskforce, 2006; Kush, 2006; Mary Ann Beyster, David H. Hardison, & Gary M. Lubin, 2005; Paul, Seib, & Prescott, 2005; Bleicher, 2003) so takšne rešitve še vedno največ v uporabi, čeprav je potek postopka zbiranja, digitalizacije in validacije podatkov dolg, ker se pojavljajo problemi pri branju zapisov na papirnatih CRF obrazcih, problemi pri tolmačenju nepravilnih vnosov ter problemi z manjkajočimi podatki.

Uporaba osebnih računalnikov je omogočila »decentralizacijo« procesa zbiranja podatkov. V 90-tih letih so se začele pojavljati računalniške rešitve, ki so omogočale oddaljen vnos podatkov (angl. »Remote Data Capture« – RDC ali »Remote Data Entry« - RDE) (Christiansen, Hosking, Dannenberg, & Williams, 1990; Gerritsen, Sartorius, vd Veen, & Meester, 1993; Krischer et al., 1991). V primeru RDC lahko raziskovalci sami vnašajo podatke v osebni računalnik. Tako zbrani podatki se potem prenesejo v centralno bazo podatkov v podatkovnem centru. V preteklosti se je za prenos podatkov do centralne baze najpogosteje uporabljala telefonska povezava med računalniki. Za te prvotne rešitve je bil značilen t.i. tekstualni uporabniški vmesnik (angl. »character-based user interface« - CUI), ki je, kar se tiče uporabnosti, precej omejen.

Naslednja generacija RDC rešitev je že uporabljala t.i. grafični uporabniški vmesnik (angl. »graphical user interface« - GUI), ki je omogočal bolj uporabne rešitve s stališča vnosnih mask oziroma izdelave elektronskih obrazcev, ki so lahko spominjali na papirnate obrazce. Nekatero med temi rešitvami so izkoriščale le možnosti HTML tehnologije (T. Kiuchi & Kaihara, 1997; Takahiro Kiuchi et al., 1996; Weber, Kassai, Csontos, Czuczor, & Horváth, 1999), druge pa so uporabljale še Java applete za bolj zahtevne obrazce (Eich & Ohmann, 1999; Kelly & Oldham, 1997).

Nekatere rešitve so bile namenjene avtomatični digitalizaciji podatkov iz papirnatih obrazcev (Jørgensen & Karlsmose, 1998). Te rešitve bolj ali manj uspešno uporabljajo različne algoritme za branje podatkov s skeniranih papirnatih obrazcev.

Z razvojem svetovnega spleta so RDE rešitve še bolj napredovale. V začetkih so obstajale znatne omejitve za uporabo spletnih rešitev za zbiranje podatkov kliničnih študij. Predvsem je problem predstavljala slaba razširjenost Interneta. V primeru, da kateri od raziskovalnih centrov vključenih v študijo ni imel dostopa do Interneta, je bilo treba podatke iz tega centra posebej zbrati na papirnatih obrazcih in v centru za upravljanje s

podatki (angl. »Data Management Center») podatke ročno vnesti v centralno bazo podatkov. Poleg omenjenega tudi sama spletna tehnologija ni omogočala kompleksnih uporabniških vmesnikov, oziroma je bila njihova izdelava precej zahtevna. Rešitve so pogosto zahtevale precejšnjo moč uporabniških računalnikov. Uporabniki pogosto niso bili pripravljeni sprejeti novega načina zbiranja podatkov, če je bil ta novi način manj prijazen in je zahteval več napora s strani uporabnikov. Še en problem je predstavljala zaskrbljenost glede varnosti komunikacij in podatkov ob uporabi Interneta.

Konec 90-tih so se pojavile prve celovite spletno zasnovane rešitve za zbiranje podatkov kliničnih študij. Te rešitve so zasnovane tako, da zagotavljajo zbiranje podatkov in imajo vse zakonsko zahtevane varnostne mehanizme ter da omogočajo vpogled v potek študije in zbrane podatke. Že prve objavljene rešitve so izpostavljale potrebo po prilagodljivosti rešitve študijam oziroma hitremu in enostavnemu nastavljanju elektronskih obrazcev študije in funkcij za avtomatično preverjanje podatkov (Hollingsworth, Hay, & Richards, 1999; Wübbelt, Fernandez, & Heymer, 2000; Brandt et al., 2000; Pepine et al., 1998). S spletnimi rešitvami je zaživel tudi naziv »elektronsko zbiranje podatkov« (angl. »Electronic Data Collection« - EDC), ki je danes skoraj edini, ki se uporablja kot alternativa papirnemu zbiranju podatkov.

Razvoj ustrezne generične rešitve, ki bi lahko bila uporabljena za zbiranje podatkov katerekoli klinične študije, je zahtevna naloga, ki zahteva predvsem rešitve semantičnih izzivov. To vključuje razvoj enotnih standardov, ontologij in slovarjev za področje kliničnih študij in medicine nasploh. Določen uspeh je že dosežen na področju poročanja rezultatov klinične študije (angl. »trial summary«) (I Sim, Owens, Lavori, & Rennels, 2000; Strang, Cucherat, Yzebe, & Boissel, 2000).

EDC rešitve, ki se uporabljajo danes, lahko glede na topologijo sistema razdelimo na štiri tipe:

- »nepovezane« (angl. »off-line«);
- »težki odjemalec« (angl. »thick client«);
- »lahki odjemalec« (angl. »thin client«);
- »hibridne«.

### *»Nepovezane« rešitve*

Pri »nepovezanih« (angl. »off-line«) rešitvah, znanih tudi kot »samostojne« (angl. »stand alone«) rešitve, se programska oprema in podatki nahajajo na prenosnem ali osebнем računalniku na lokaciji raziskovalca. Takšen pristop je zelo zaželen na področjih, kjer je telefonsko in/ali internetno povezavo težko vzpostaviti ali vzdrževati, oziroma je povezava nezanesljiva. Strojna in programska oprema mora biti validirana in določeni morajo biti mehanizmi za zagotovitev validnosti opreme tekom celotne študije. Možnosti za prenos podatkov iz raziskovalnega centra v centralni podatkovni center so različne – od pošiljanja

zgoščenk s podatki do občasnih prenosov podatkov po telefonskih ali internetnih povezah. Značilnost nepovezanih rešitev je ta, da klinični podatki fizično obstajajo na najmanj dveh lokacijah (lokalnem računalniku in centralnem računalniku).

### *»Težki odjemalec« rešitve*

»Težki odjemalec« (angl. »thick client«) rešitve so nepovezanim rešitvam podobne v tem, da je programska oprema naložena na lokalni računalnik in da morajo biti vse mere validacije podobno zagotovljene. Razlika je le v tem, da so takšne rešitve zasnovane tako, da lahko delujejo na povezan ali nepovezan način.

### *»Lahki odjemalec« rešitve*

»Lahki odjemalec« (angl. »thin client«) rešitve so zasnovane na uporabi interneta in spletnega brskalnika. Uporabniki uporabljajo spletni brskalnik za varen dostop do vseh vsebin, pri čemer so celotna EDC programska oprema, obrazci in podatki dejansko na centralnem strežniku (v podatkovnem centru). Takšen pristop je zaželen v primerih, ko je enostavno in zanesljivo vzpostaviti spletno povezavo do vseh sodelujočih raziskovalnih centrov. Podatke uporabniki prek spletnega brskalnika vnašajo neposredno v centralno bazo, tako da imajo vpogled v podatke in možnost ukrepanja v realnem času. Potrebe po validaciji lokalnih računalnikov so minimalne, ker se lokalni računalniki uporabljajo le kot terminali, ki ne shranjujejo ali procesirajo podatkov. Praviloma mora biti EDC sistem sposoben le tega, da ugotovi, ali je ustrezen operacijski sistem in spletni brskalnik nameščen na lokalnem računalniku. Za nemoteno delovanje sistema je treba zagotoviti dovolj zmogljivo internetno povezavo in ustrezno prilagoditi nastavitve morebitnih »požarnih zidov« na strani uporabnikov (raziskovalni centri). Kljub fleksibilnosti pri zahtevah po strojni in programski opremi na strani uporabnikov, se nekateri naročniki študij včasih odločijo, da za svoje EDC študije raziskovalce oskrbijo z validiranimi prenosnimi računalniki s ciljem, da na ta način zmanjšajo možnosti za nedelovanje povezave med odjemalcem in strežnikom.

### *»Hibridne« rešitve*

»Hibridna« rešitev je lahko katera koli rešitev, ki kombinira zgornji dve rešitvi. Na primer, pri hibridni rešitvi je lahko EDC programska oprema nameščena na prenosnikih, podatki pa se shranjujejo v centralnem strežniku. Drugi primer bi bila rešitev, pri kateri lokalni odjemalec deluje v »nepovezanem« režimu in se le občasno (predvidoma, ko je internetna povezava dovolj hitra) poveže na centralni strežnik in posreduje zbrane podatke. Nekatere hibridne rešitve omogočajo odjemalcu, da deluje tako v »nepovezanem« kakor tudi v »povezanem« režimu, tako da bi uporabnik lahko izkoristil prednosti spletnih rešitev in bi obenem lahko po potrebi delal tudi takrat, ko ne more vzpostaviti povezave na splet. Pri takšnih rešitvah je ključnega pomena sinhronizacija podatkov med odjemalcem in

strežnikom. Praviloma se preverjanje podatkov izvaja šele po prenosu podatkov na strežnik.

### I.1.4. Veljavna zakonodaja in standardi

Področje kliničnih preskušanj je zakonsko urejeno. To vpliva tudi na EDC rešitve, ki morajo zagotoviti predvsem varnost podatkov in sledljivost spremembam podatkov. V nadaljevanju je podan kratek pregled zakonskih predpisov in priporočil ter standardov, ki zadevajo področje informatizacije kliničnih študij.

#### *Zakonodaja*

Direktiva Sveta EU 65/65/EGS z dne 26. januarja 1965 o približevanju določb zakonov ali drugih predpisov, ki se nanašajo na zdravila (Svet Evrope, 1965), zahteva, da se vlogam za pridobitev dovoljenja za promet z zdravilom priloži dokumentacija s podrobnostmi in dokumenti o rezultatih preskusov in kliničnih preskušanj, ki so bili z zdravilom izvedeni.

Sprejeta podlaga za izvajanje kliničnega preskušanja na ljudeh temelji na varovanju človekovih pravic in človekovega dostojanstva glede uporabe biologije in medicine, kot ga na primer odraža besedilo Helsinške deklaracije iz leta 1964. Zaščito preizkušancev kliničnega preskušanja zagotavljajo ocena tveganja, ki temelji na rezultatih toksikoloških poskusov pred kakršnikoli kliničnim preskušanjem, pregled odborov za etiko in pristojnih organov držav članic ter pravila o varstvu osebnih podatkov.

Mednarodna konferenca o usklajevanju (ICH) je leta 1995 dosegla soglasje o usklajenem pristopu k dobri klinični praksi (angl. »Good Clinical Practice« - GCP). Dobra klinična praksa je skupek mednarodno priznanih etičnih in znanstvenih zahtev glede kakovosti, ki se morajo spoštovati pri oblikovanju, izvajanju in evidentiranju kliničnih preskušanj ter poročanju o kliničnih preskušanjih, v katerih sodelujejo ljudje. Dokument o soglasju bi bilo treba upoštevati tako, kot je bil dogovorjen v okviru Odbora za zdravila za ljudi (CHMP) Evropske agencije za zdravila.

Evropska Komisija je leta 2001 določila direktivo o približevanju zakonov in drugih predpisov držav članic v zvezi z izvajanjem dobre klinične prakse pri kliničnem preskušanju zdravil za ljudi, ki državam članicam EU nalaga, na kakšen način mora zakonodaja urejati klinična preskušanja (Evropski parlament in Svet, 2001). Ta direktiva določa posebne določbe o izvajanju intervencijskih kliničnih preskušanj na ljudeh, vključno z multicentričnim kliničnim preskušanjem, ki vključujejo zdravila, zlasti glede izvajanja dobre klinične prakse. Skladnost z dobro klinično prakso zagotavlja, da so pravice, varnost in dobrobit preizkušancev zavarovane in da so rezultati kliničnih preskušanj verodostojni. Načela in podrobne smernice je Evropska Komisija podala leta 2005 z Direktivo 2005/28/ES (Evropska komisija, 2005).



Ameriška agencija za hrano in zdravila (Food and Drug Administration - FDA) regulira vse postopke registracije zdravil na ameriškem trgu, ki med drugim vključujejo tudi registracijo kliničnih preskušanj in prijavo rezultatov preskušanj za potrebe postopka registracije zdravil. FDA je leta 2007 objavila priporočila glede uporabe računalniških sistemov v kliničnih raziskavah (FDA, 2007). Namen priporočil je, da pomagajo zagotoviti zanesljivost, kakovost in celovitost elektronskega vira podatkov in dokumentacije (npr. elektronskih zapisov), ki nastanejo tekom kliničnih raziskav in so pogoj za registracijo zdravil v ZDA (FDA, 1997). Ta priporočila se nanašajo tudi na klinična preskušanja, ki so ena od vrst kliničnih raziskav.

Kot posebej občutljivo področje v zakonodaji (tako slovenski in evropski kot tudi ameriški) je izpostavljeno varovanje osebnih podatkov. To se nanaša tudi na podatke zbrane med klinično študijo. Ta zakonodaja je večinoma povsod podobna in predvsem usmerjena v zaščito posameznikov pred zlorabo osebnih podatkov. Namreč, Zakon o varstvu osebnih podatkov (ZVOP (DZ RS, 2004)) v 13. členu pravi, da je uporaba in obdelava občutljivih osebnih podatkov (med katere sodijo zdravstveni podatki) dovoljena le v primeru, če posameznik za to poda izrecno osebno privolitev. Še več, ZVOP v 16. členu določa, da se podatki lahko zbirajo le za določene in zakonite namene ter se ne smejo nadalje obdelovati tako, da bi bila njihova obdelava v neskladju s temi nameni. Zaradi tega mora vsak preizkušane ob vključitvi v študijo (preskušanje) naročniku podati pisno potrdilo o sodelovanju v študiji (angl. »written consent«), v katerem naročniku študije dovoljuje uporabo in obdelavo podatkov za namen študije.

V 17. členu ZVOP se dovoljuje nadaljnja uporaba osebnih podatkov za znanstveno–raziskovalne namene pod pogojem, da so podatki anonimizirani, kar pomeni, da iz podatkov ni mogoče ugotoviti, za katero osebo gre. Na ta način je omogočeno izkoriščanje ogromne količine kliničnih podatkov zbranih v kliničnih študijah s tem, da morajo lastniki podatkov (naročniki študij) zagotoviti anonimizacijo podatkov, preden jih posredujejo uporabnikom za znanstveno–raziskovalne namene. To je dvorezen meč, ker zadostni mehanizmi anonimizacije niso zakonsko določeni, kar pomeni, da v primeru identifikacije osebe z združevanjem podatkov iz različnih podatkovnih virov lahko pride do posredne identifikacije osebe. Znan je takšen primer iz Islandije, kjer so morali ustaviti projekt združevanja treh baz anonimiziranih zdravstvenih in genealoških podatkov, ko je zaradi majhnosti populacije na Islandiji iz bogastva tako združenih podatkov postalo precej bolj verjetno, da bo kakšna oseba identificirana (Kalra, Gertz, Singleton, & Inskip, 2006).

Glede zavarovanja osebnih podatkov ZVOP določa, da morajo biti zagotovljeni tehnično–logistični postopki, ki morajo poleg zaščite programske in strojne opreme in preprečevanja nepooblaščenega dostopa do podatkov omogočiti tudi poznejše ugotavljanje tega, kdaj so bili posamezni osebni podatki vneseni v zbirko osebnih podatkov, uporabljeni ali drugače obdelani in kdo je to storil. Pri tem morajo biti postopki in ukrepi za zavarovanje osebnih

podatkov ustrezni glede na tveganje, ki ga predstavlja obdelava in narava določenih osebnih podatkov, ki se obdelujejo.

Zakonske določbe in njihova implementacija pri kliničnih študijah in EDC rešitvah so pogosto predmet polemik strokovne javnosti (Al-Shahi & Warlow, 2000; Strobl, Cave, & Walley, 2000), ker v precejšnji meri otežujejo uporabo kliničnih podatkov za nadaljnje znanstveno-raziskovalne namene.

### *Standardi*

Na področju informatike v medicini so aktivne številne mednarodne standardizacijske organizacije in iniciative. Ključne med njimi so:

- ISO (International Organisation for Standardisation) - največja svetovna organizacija za mednarodne standarde;
- CEN (European Committee for Standardisation) - ključna Evropska standardizacijska organizacija;
- IHTSDO (International Health Terminology) - snovalec široko sprejetega terminološkega standarda SNOMED-CT;
- HL7 (Health Level 7) - skrbnik široko sprejetih standardov za izmenjavo elektronskih sporočil v zdravstvenih sistemih;
- DICOM (Digital Imaging and Communications in Medicine) - de-facto standard za shranjevanje in izmenjavo digitalnih slik;
- OpenEHR - kot obetajoča evropsko-avstralska odprtokodna iniciativa za standardizacijo elektronskega zdravstvenega zapisa;
- IHE (Integrating the Healthcare Enterprise) - kot ključna iniciativa za interoperabilnost e-zdravstvenih sistemov;
- CDISC (Clinical Data Interchange Standards Consortium), ključna iniciativa za standardizacijo izmenjave podatkov kliničnih študij.

Za naše področje uporabe bi kot najbolj pomembno iniciativo izpostavili CDISC. CDISC je neprofitna organizacija sestavljena iz prostovoljcev strokovnjakov tako iz akademske sfere kakor tudi iz farmacevtske industrije, ki zaradi trenutne nekompatibilnosti EDC sistemov različnih proizvajalcev vidi velik interes v standardizaciji tega področja informatizacije. Naloga CDISC-a je razvoj in podpora globalnih in od platforme neodvisnih podatkovnih standardov, ki omogočajo interoperabilnost informacijskih sistemov s ciljem izboljšave medicinskih raziskav in njim bližnjih področij zdravstva. CDISC podatkovni standardi so zasnovani kot nabori modelov, ki jih lahko izrazimo v elektronskem formatu. Zaželen elektronski format je XML (eXtended Meta Language) z uporabo ODM (Operational Data Model) modela kot osnovne XML sheme. Standardi oziroma modeli, ki jih CDISC razvija so naslednji:

- SDTM (Study Data Tabulation Model), ki je od leta 2004 priporočen standard za podatkovni model za pošiljanje rezultatov kliničnih preskušanj ameriškim državnim organom (FDA);
- ADaM (Analysis Data Model), katerega namen je dopolnitev SDTM z detajli o statistični analizi narejeni na rezultatih kliničnega preskušanja;
- SEND (Standard for Exchange of Non-clinical Data), ki je ekvivalent SDTM namenjen študijam na živalih;
- ODM (Operational Data Model), katerega ključne značilnosti so: uporaba XML, zapis oziroma format razumljiv tako računalniku kakor tudi človeku, vključitev informacij pomembnih za nadzor študije (zapisi o vseh spremembah podatkov ipd.), informacije so neodvisne od podatkovne baze ter shranjevanje ODM je neodvisno od strojne in programske opreme;
- LAB (Laboratory Data Model), ki je namenjen izmenjavi laboratorijskih podatkov;
- CDASH (Clinical Data Acquisition Standards Harmonization), ki definira minimalen nabor podatkov za šestnajst SDTM domen posvečenih varnosti skupaj z imeni elementov, definicijami in metapodatki z namenom vzpostavitve standardnega nabora podatkov, ki bi bil enoten pri vseh kliničnih preskušanjih;
- CDISC Terminology, ki definira terminologijo za SDTM in CDASH s ciljem harmonizacije zbranih oziroma posredovanih podatkov iz različnih kliničnih preskušanj.

ODM standard je bistveni standard za EDC sisteme. Namreč, implementacija ODM standarda v posameznih EDC rešitvah omogoča izmenjavo informacij (podatkov) med temi EDC rešitvami brez potrebe po prilagajanju samih rešitev. ODM format zapisa vključuje tako podatke zbrane v študiji kakor tudi metapodatke (podatke o podatkih), iz katerih je razvidno, na kaj se podatki nanašajo.

Čeprav so CDISC standardi usmerjeni predvsem na pogoje, ki jih postavlja FDA, in zaradi tega mogoče nekoliko manj globalni, Evropska agencija za medicino (European Medicines Agency - EMEA) intenzivno sodeluje s CDISC-om pri standardizaciji naslednje generacije evropske baze kliničnih preskušanj (EudraCT), nastale kot rezultat Direktive 2001/20/EC (Evropski parlament in Svet, 2001).

CDISC sodeluje tudi s HL7 v okviru katerega je razvil BRIDG (Biomedical Research Integrated Domain Group Model) HL7 RIM model za predstavitev kliničnega preskušanja. Poleg tega CDISC v okviru HL7 dela na razvoju RLUS (Retrieve, Locate and Update Service) standarda za vmesnike, prek katerih bodo sistemi na standardiziran način dostopali in upravljali s podatki.

### I.1.5. Načrtovanje uporabnosti

Z naraščanjem uporabe računalnikov v vsakdanjih procesih raste tudi pomembnost načrtovanja uporabnosti (angl. »usability engineering«). Namreč, uporabniki se v svojem delu vse bolj pogosto srečujejo ali poslužujejo številnih računalniških aplikacij, katerih namen je olajšati in/ali izboljšati delo uporabnikov. Iz tega razloga se pri zasnovi računalniških rešitev ustrezna pozornost posveča tudi načrtovanju uporabnosti. Načrtovanje uporabnosti je, kot znanstvena in strokovna veja, nastala v 80-tih letih prejšnjega stoletja z namenom raziskovanja interakcije človek-računalnik (angl. »human-computer interaction«), predvsem pa s ciljem razvoja uporabniku čim bolj prijaznih (angl. »user friendly«) vmesnikov. Kot eno od temeljnih del za afirmacijo načrtovanja uporabnosti velja knjiga »Usability Engineering« znanega strokovnjaka s tega področja Jakoba Nielsena (Nielsen, 1994). Načrtovanje uporabnosti v kontekstu razvoja programske opreme predstavlja posebne postopke med razvojem programske opreme, katerih namen je doseganje želene uporabnosti programske opreme.

Za kakovostno ocenjevanje uporabnosti je treba biti pozoren na več meril uporabnosti, ki jih je strokovna javnost sprejela kot ključne. Izhajajoč iz ISO 9241 standarda, ki je osredotočen na doseganje učinkovitosti in zadovoljstva, kot merila uporabnosti, so identificirani:

- čas učenja;
- hitrost uporabljanja;
- pogostost napak uporabnika;
- zapomljivost;
- zadovoljstvo uporabnika.

Načrtovanje uporabnosti lahko znatno podraži ceno razvoja produkta, kar ni vedno primerno. Pristopi k načrtovanju uporabnosti so številni in so pogosto predmet raziskav (Kushniruk & V. L. Patel, 2004; C. M. Johnson, T. R. Johnson, & Zhang, 2005; Schriver, 1989; Beuscart-Zéphir et al., 2005; Lecerof & Paterno, 1998; Kushniruk, C. Patel, V. L. Patel, & Cimino, 2001; van Waes, 2000). Pri izboru primerne metodologije načrtovanja uporabnosti je treba upoštevati ključne elemente uporabnosti, ki so za posamezno aplikacijo oziroma uporabnika značilne. Po Shneidermanu (Shneiderman & Plaisant, 2004) lahko identificiramo pet skupin računalniških aplikacij/sistemov, od katerih ima vsaka značilne motive uporabnosti, in sicer:

- življenjsko-kritični sistemi (npr. kontrola zračnega prometa, nuklearnega reaktorja, zdravstvenih naprav itn.) – daljši čas učenja je sprejemljiv za doseganje zanesljivega upravljanja sistema;
- sistemi za industrijsko in komercialno uporabo (bančni sistemi, zavarovanje, naročila, upravljanje z zalogami itn.) – izobraževanje uporabnikov je drago,

zato mora biti čim hitrejša, zapomljivost zagotavlja dolgoročna uporaba sistema, pogostost napak uporabnikov mora biti cenovno upravičena na dolgi rok;

- aplikacije za pisarno, dom in zabavo (e-mail, bančni terminali, igrice, izobraževalni programi, iskalniki, mobilne naprave itn.) – morajo biti enostavne za uporabo, s kratkim časom učenja in visokim zadovoljstvom uporabnikov;
- vmesniki za raziskovanje/brskanje, ustvarjanje in sodelovanje (npr. spletni brskalniki, znanstveni simulatorji, orodja za poslovno odločanje, sistemi za dizajniranje, sistemi za on-line sestanke itn.) – morajo popolnoma zadovoljiti strokovne zahteve in omogočiti čim bolj kompleksne operacije z možnostjo ustvarjanja lastnih bližnjic, in na splošno čim manj obremeniti uporabnika z uporabniškim vmesnikom, tako da lahko ostane osredotočen na delo;
- socio-tehnični sistemi (kompleksni sistemi, ki vključujejo veliko število uporabnikov v dolgem obdobju kot npr. volilni sistem, zdravstveni sistem, kriminalistični sistem itn.) – morajo omogočati uporabniku dostop do zanesljivih informacij, se odzivati na aktivnosti uporabnikov in ponujati informacije o statusu aktivnosti.

Sisteme za elektronsko zbiranje podatkov kliničnih študij, zaradi njihove uporabe za posamezne projekte (študije) farmacevtske industrije, razumemo predvsem kot sisteme za industrijsko in komercialno uporabo, čeprav imajo tudi določene značilnosti vmesnikov za raziskovanje in sodelovanje ter socio-tehnoloških sistemov malih dimenzij. Namreč, vsaka klinična študija ima značilnosti, ki so lahko zelo specifične, in posledično je tudi EDC rešitev v uporabi specifična za vsako študijo. Izobraževanje uporabnikov mora biti čim hitrejša. Uporabniški vmesnik ne sme obremeniti uporabnika, ki mora vso pozornost usmeriti v zajem podatkov (včasih tudi med samim zdravstvenim posegom). Ker gre za občutljive osebne podatke mora sistem delovati zanesljivo ter uporabnika obveščati o statusu aktivnosti, ki jo izvaja.

Za doseganje želene uporabnosti sistema načrtovalci uporabnosti izvajajo ocenjevanje uporabnosti obstoječe ali predlagane rešitve že pred ali med samim razvojem. Na osnovi rezultatov ocenjevanja načrtovalci sistema izvajajo potrebne izboljšave sistema oziroma nadaljujejo razvoj. Najbolj razširjene metode ocenjevanja uporabnosti so:

- testiranje uporabnosti;
- intervjuji;
- fokusne skupine;
- vprašalniki;
- kognitivni sprehodi (angl. »cognitive walkthroughs«);
- hevristično ocenjevanje;

- kognitivna analiza nalog (angl. »cognitive task analysis«);
- kontekstualna preiskava (angl. »contextual inquiry«);
- protokol glasnega razmišljanja (angl. »think aloud protocol«).

Na izbor primernih metod za ocenjevanje uporabnosti med drugim vplivajo tudi planirani stroški razvoja. V primeru komercialnih in industrijskih sistemov morajo biti stroški načrtovanja uporabnosti uravnoteženi s pridobitvami, ki jih načrtovanje uporabnosti lahko prinese. Zaradi tega se takšni sistemi pogosto zanašajo na cenejše metode ocenjevanja uporabnosti kot so intervjuji, vprašalniki in podobno. Pri tem nekateri avtorji (Nielsen, 2000; Virzi, 1992) zagovarjajo stališče, da je že štiri do pet uporabnikov dovolj za ugotavljanje 80 % težav z uporabnostjo sistema, kar je v večini primerov sprejemljiv strošek načrtovanja uporabnosti.

## I.2. Cilji in struktura naloge

Osnovni cilj naloge je predlog računalniške rešitve, ki bo izboljšala postopek zbiranja podatkov klinične študije v primerjavi z zbiranjem podatkov na papirnatih obrazcih, pri čemer mislimo na izboljšave v pogledu stroškov zbiranja podatkov, časa potrebnega za zaključek zbiranja podatkov ter prijaznosti samega procesa za udeležence v procesu. Rešitev naj bi bila primerna za multicentrične klinične študije in sicer tudi takšne, ki nimajo na razpolago velika denarna sredstva za kontrolo kakovosti podatkov (monitoring). Poleg tega, predlagana rešitev naj bi upoštevala veljavno zakonodajo ter bila zasnovana v skladu z načeli načrtovanja uporabnosti.

Predlagana rešitev mora izhajati iz optimiziranega procesa zbiranja podatkov, ki bo zagotovil najmanj enako kakovost zbranih podatkov kot papirni proces. Zato je smo se v našem delu lotili popisa papirnega procesa zbiranja podatkov ter ugotovili možne izboljšave ob uporabi računalniške tehnologije. Ker mora rešitev temeljiti na primerni računalniški tehnologiji, smo naredili pregled možnih tehnoloških rešitev ter predlagali najbolj primerno rešitev za podporo optimalnega procesa. Pri razvoju pilotne rešitve smo uporabili metode načrtovanja uporabnosti z namenom zagotavljanja uporabniku čim bolj prijazne rešitve.

Na koncu smo naredili analizo in primerjavo PDC in EDC pristopa z aspekta stroškov in trajanja zbiranja podatkov. V diskusiji smo primerjali različne scenarije zbiranja podatkov v izvedbi papirnega ali elektronskega procesa s predlagano računalniško rešitvijo. Iz rezultatov primerjav smo prišli do ugotovitev o tem, kateri so ključni dejavniki, ki vplivajo na stroške zbiranja podatkov ter koliko EDC pristop pripomore k zniževanju stroškov klinične študije.

V skladu z zgoraj omenjenimi cilji in doseženimi rezultati je disertacija v nadaljevanju organizirana na naslednji način:

Poglavje Metode vsebuje opis metod, ki smo jih v našem delu uporabili. Predstavili smo metode uporabljene pri opisovanju procesov, določanju funkcionalnega in tehnološkega okvirja EDC rešitve, zasnove končne EDC rešitve ter ocenjevanja predlagane rešitve.

V poglavju Rezultati smo predstavili rezultate našega dela, in sicer: opis PDC procesa, funkcionalni in tehnološki okvir predlagane rešitve, samo EDC rešitev (pilotni sistem) ter primerjavo ocen stroškov in časa za PDC in EDC proces.

V Razpravi smo podrobneje analizirali prednosti in pomanjkljivosti elektronskega zbiranja podatkov v primerjavi z zbiranjem podatkov na papirju. Diskutirali smo naš predlog reducirane upravljanja podatkov z vpeljavo internega monitoringa, kot tudi strukturo elementov, možnosti usklajevanja EDC rešitve s pomembnimi standardi ter primernost spletno zasnovane rešitve. Večji del razprave smo posvetili ocenam stroškov in vplivu sprememb parametrov študije in procesa zbiranja podatkov na izračune. Ob tem smo izpostavili tudi prednosti in pomanjkljivosti našega pristopa ter podali svoje predloge, kako naš pristop prilagoditi značilnostim posamezne klinične študije.

Sklepno poglavje vsebuje povzetek našega dela in naših zaključkov.





## II. Metode

### II.1. Opisovanje procesov

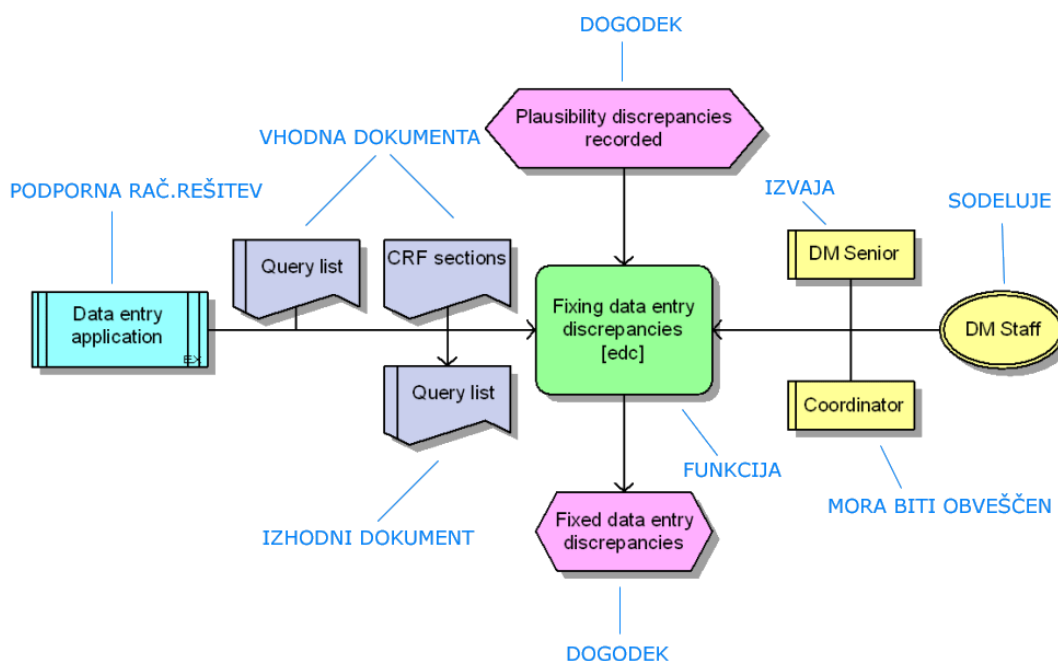
Za opisovanje procesov zbiranja podatkov smo izbrali primerno tehniko opisovanja (modeliranja) procesov in v skladu z izbrano tehniko modeliranja naredili analizo procesov. V nadaljevanju je predstavljena izbrana tehnika modeliranja procesov (eEPC) ter opisano, kako smo se lotili analize procesov.

#### II.1.1. eEPC tehnika modeliranja procesov

Obstajajo številne tehnike za opisovanje procesov, med katerimi kot najbolj razširjene lahko omenimo eEPC (Extended Event-driven Process Chains), BPEL (Business Process Execution Language) in UML (Unified Modeling Language). Mi smo se odločili za eEPC tehniko, ker je usmerjena na uporabnika in nam omogoča, da modeliramo tako funkcije podprte z rešitvami informacijske tehnologije kot tudi funkcije, ki se izvajajo »ročno«.

eEPC model je zaporeden graf dogodkov in funkcij. Vsebuje različne tipe povezav, ki omogočajo zaporedno in paralelno izvajanje procesov. Poleg tega omogoča uporabo logičnih operatorjev »ali« (OR), »ekskluzivno ali« (XOR) ter »in« (AND). eEPC velja za najbolj enostavno tehniko modeliranja z zelo preprosto in razumljivo notacijo. Zato je tudi ena od najbolj sprejetih tehnik za opisovanje poslovnih procesov. Za eEPC tehniko modeliranja je razvito programsko orodje (ARIS, IDS Scheer AG, Nemčija), ki se uporablja in razvija že več kot desetletje (vir: Gartner Group, 2007).

Slika 3 prikazuje primer uporabe osnovnih elementov eEPC grafa, ki smo jih uporabljali za modeliranje EDC in PDC procesov. Iz primera je razvidno, da je vsaka funkcija sprožena in zaključena z dogodkom (lahko jih je tudi več). Različni viri (človeški in materialni) so lahko na različne načine vključeni v izvajanje funkcije. Kot nosilci informacij (vhodnih in izhodnih) se lahko pojavljajo dokumenti ali na primer računalniške rešitve. Poleg elementov prikazanih na Sliki 3, smo pogosto uporabljali tudi logične elemente "∧" (AND), "∨" (OR) in "×" (XOR).



Slika 3. Primer uporabe elementov eEPC grafa

Podroben opis eEPC modelov je predstavljen v dokumentaciji ARIS-a (Scheer, 2000).

### II.1.2. Analiza procesov

Proces zbiranja podatkov klinične študije ni nikjer strogo določen. Različni naročniki se zbiranja podatkov lotevajo na različne načine. Na primer, v primeru zbiranja podatkov v papirnate CRF-je lahko raziskovalec pošlje izpolnjene CRF-je naročniku potem, ko jih monitor verificira, ali jih monitor pošlje potem, ko jih pregleda, ali raziskovalec pošlje izpolnjene CRF-je po faksu že preden jih monitor pogleda. Ne glede na to nedoločenost procesa pa je praksa pokazala, da so nekateri postopki v tem procesu skupni večini pristopov. Te skupne postopke smo poiskali in povzeli v analizi procesov. Pri tem smo uporabili več metod, in sicer: pregled literature; zbiranje informacij prek spletne ankete; in zbiranje informacij prek intervjujev z osebami z izkušnjami izvajanj kliničnih študij.

#### *Pregled literature*

Pregled literature je vključeval tri sklope virov: strokovne knjige; znanstvene članke; in strokovne novice.

#### Strokovne knjige

Strokovnih knjig na temo zbiranja podatkov kliničnih študij tako rekoč ni. Edina, ki že več let objavlja navodila na tem področju, je ameriška avtorica Susan Prokscha. Njeno knjigo »Practical Guide to Clinical Data Management« (Prokscha, 1999) smo uporabili kot glavni vir informacij o procesu kliničnih študij. Poleg te knjige smo analizirali še dobro splošno preglednico upravljanja z medicinskimi podatki evropskih avtorjev - »Medical Data Management: A Practical Guide« (Leiner, Gaus, Haux, Knaup-Gregori, & Wagner, 2003).

## II. Metode

Na ta način smo želeli dobiti vpogled tudi v morebitne razlike v praksah v dveh sredinah z največjo klinično-raziskovalno aktivnostjo.

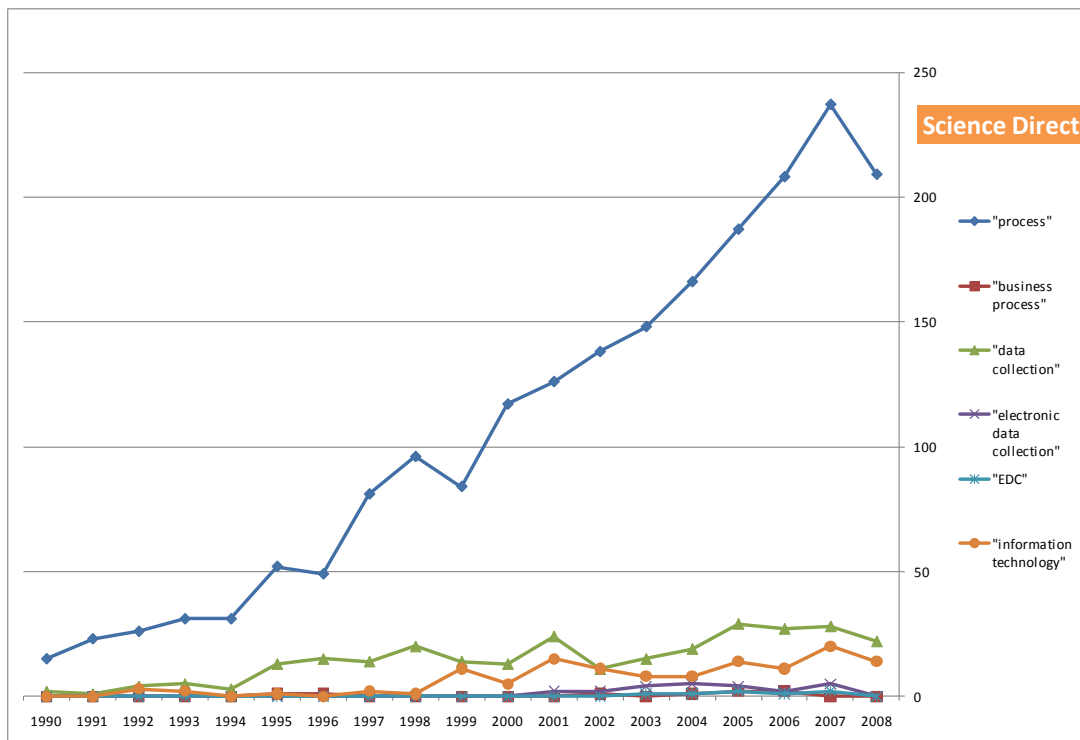
### Znanstveni članki

Čeprav je znanstvenih člankov na področju medicine in informacijske tehnologije ogromno, se jih le majhen delež ukvarja z organizacijo kliničnih študij. Z uporabo spletnih servisov »Science Direct« (<http://www.sciencedirect.com/>) in »PubMed« (<http://www.ncbi.nlm.nih.gov/pubmed/>) smo poiskali vse članke v katerih je v naslovu, povzetku, ali med ključnimi besedami omenjen termin »clinical trial« v kombinaciji s termini »process«, »business process«, »data collection«, »electronic data collection«, »EDC«, ali »information technology«, ter jih pregledali. Rezultati poizvedb so prikazani na:

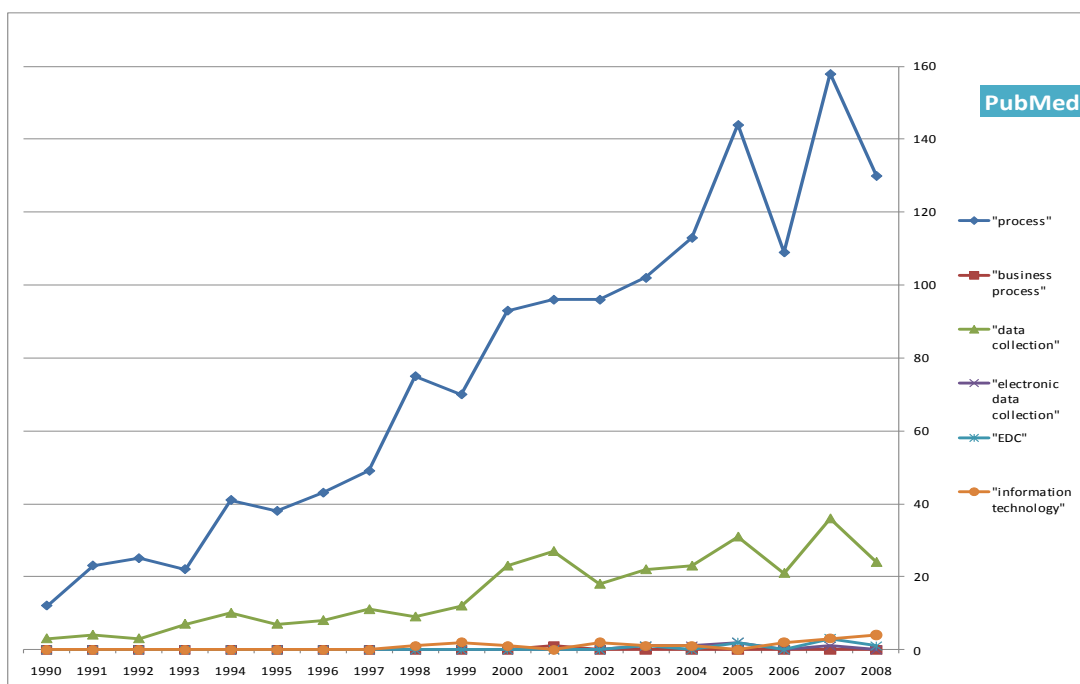
Tabela 1. Rezultati iskanja člankov objavljenih na Science Direct in PubMed spletnih bazah člankov od leta 1990 do 2008. Tabela prikazuje število letno objavljenih člankov, ki vsebujejo iskalni termin »clinical trial« v kombinaciji z vsakim od terminov: »process«, »business process«, »data collection«, »electronic data collection«, »EDC« ter »information technology«.

Science Direct	PubMed	Iskalni niz = "clinical trial" + ...											
		"process"		"business process"		"data collection"		"electronic data collection"		"EDC"		"information technology"	
št.zadetkov po letih	2008	209	130	0	0	22	24	0	0	0	1	14	4
	2007	237	158	0	0	28	36	5	1	2	3	20	3
	2006	208	109	2	0	27	21	2	0	1	0	11	2
	2005	187	144	2	0	29	31	4	2	2	2	14	0
	2004	166	113	1	0	19	23	5	1	1	0	8	1
	2003	148	102	0	0	15	22	4	1	1	1	8	1
	2002	138	96	1	0	11	18	2	0	0	0	11	2
	2001	126	96	0	1	24	27	2	0	0	0	15	0
	2000	117	93	0	0	13	23	0	0	0	0	5	1
	1999	84	70	0	0	14	12	0	0	0	0	11	2
	1998	96	75	0	0	20	9	0	0	0	0	1	1
	1997	81	49	0	0	14	11	0	0	0	0	2	0
	1996	49	43	1	0	15	8	0	0	0	0	0	0
	1995	52	38	1	0	13	7	0	0	0	0	1	0
	1994	31	41	0	0	3	10	0	0	0	0	0	0
	1993	31	22	0	0	5	7	0	0	0	0	2	0
	1992	26	25	0	0	4	3	0	0	0	0	3	0
	1991	23	23	0	0	1	4	0	0	0	0	0	0
	1990	15	12	0	0	2	3	0	0	0	0	0	0

## II. Metode



Slika 4. Rezultati iskanja člankov objavljenih na Science Direct spletni bazi člankov od leta 1990 do 2008. Diagram prikazuje število letno objavljenih člankov, ki vsebujejo iskalni termin »clinical trial« v kombinaciji z vsakim od terminov: »process«, »business process«, »data collection«, »electronic data collection«, »EDC« ter »information technology«.



Slika 5. Rezultati iskanja člankov objavljenih na PubMed spletni bazi člankov od leta 1990 do 2008. Diagram prikazuje število letno objavljenih člankov, ki vsebujejo iskalni termin »clinical trial« v kombinaciji z vsakim od terminov: »process«, »business process«, »data collection«, »electronic data collection«, »EDC« ter »information technology«.

### Strokovne novice

Področje optimizacije procesov kliničnih študij je v večji meri predmet interesa farmacevtske industrije in manj akademskih krogov. Zaradi tega je mogoče zaslediti številne objave na temo zbiranja podatkov kliničnih študij tudi med strokovnimi novicami ali drugimi podobnimi objavami z majhno akademsko težo. Vseeno nam je pregled tovrstnih objav dal vpogled v informacije relevantne za farmacevtsko industrijo in njene pristope k optimizaciji procesov, ki posledično optimizira tudi stroške. Kot vir novic najbolj osredotočen na IT rešitve za podporo kliničnih študij, smo pregledali objave spletne revije »Bio IT World« na temo »eClinical Trials« (<http://www.bio-itworld.com/BioIT/eClinicalTrials.aspx>). Poleg tega smo prebrskali splet s pomočjo spletnega brskalnika »Google« (<http://www.google.com>) ter za enake iskalne termine kot pri iskanju znanstvenih člankov odkrili številne »white paper« prispevke industrije, povezane s kliničnimi študijami, in strokovnih združenj.

Tabela 2. Rezultati iskanja strokovnih objav s pomočjo spletnega iskalnika Google. V tabeli je podano približno število zadetkov vsebin objavljenih do konca leta 2008.

Google	Iskalni niz = "white paper" + "clinical trial" + ...					
	"process"	"business process"	"data collection"	"electronic data collection"	"EDC"	"information technology"
približno št.zadetkov	25000	3000	7500	100	5000	13000

### Anketa

Anketo smo zasnovali kot spletno aplikacijo z omejenim dostopom. Ločili smo tri skupine anketirancev, in sicer: farmacevtska podjetja, CRO podjetja in klinične ustanove. Vsaka od skupin je imela sebi primeren nabor vprašanj (glej Priloga A, B, C).

Seznam farmacevtskih podjetij smo naredili na osnovi podatkov o odobrenih kliničnih študijah s strani Agencije za zdravila RS v zadnjih petih letih. Pri podjetjih z večjo aktivnostjo v zadnjih petih letih smo poiskali ustrezen kontaktni naslov, na katerega smo naslovili prošnjo za sodelovanje v anketi. Na osnovi podatkov ARRS smo prav tako naredili seznam slovenskih ustanov, na katerih potekajo klinične študije. Seznam smo razširili z nekaj tujimi kliničnimi ustanovami. CRO podjetja smo poiskali prek interneta. Torej, vsaki od skupin smo poslali vabilo, da se udeleži spletne ankete. Spletna anketa je vsebovala stran za dostop z uporabniškim imenom in geslom ter več strani vprašalnikov. Kot motivacijo za izpolnjevanje vprašalnikov smo anketirancem, potem ko izpolnijo anketo, ponudili možnost vpogleda v rezultate ankete oziroma anonimizirane odgovore drugih anketirancev. Kontakte na seznamu anketirancev smo v dveh mesecih dvakrat opomnili, da izpolnijo anketo, in potem zaključili z zbiranjem odgovorov. Rezultate ankete smo med drugim uporabili tudi za modeliranje procesov.

Tabela 3. Odziv anketirancev na povabilo, da izpolnijo anketo. (CRO – Clinical Research Organization)

	Število oseb		% odziva
	povabljeno	izpolnilo	
CRO podjetja	116	17	15 %
farmacevtska podjetja	33	7	21 %
klinične ustanove	65	10	15 %

### *Intervjuji*

En intervju smo izvedli z osebo z izkušnjami v vlogi raziskovalne sestre. Opravili smo dva neposredna razgovora v njenem delovnem okolju, kjer nam je lahko demonstrirala različne korake v procesu zbiranja podatkov za različne naročnike. Sproti smo zahtevali, da nam komentira osnutke procesov, ki smo jih sestavili na osnovi pregleda literature in ankete. Poleg tega smo dodatna vprašanja posredovali po elektronski pošti in si tako pojasnjevali morebitne nejasnosti.

Drugi intervju smo izvedli z osebo z izkušnjami v vlogi monitorja, zaposleno pri tujem CRO podjetju. Enako smo jo v neposrednem razgovoru prosili za komentarje naših osnutkov procesov ter primere funkcij, ki jih izvaja pri svojem delu za različne naročnike. Tudi to osebo smo po razgovoru prosili za dodatna pojasnila po elektronski pošti.

Rezultate intervjujev smo uporabili med drugim tudi za modeliranje procesov.

## II.2. Določanje funkcionalnega okvirja

### II.2.1. Predlog EDC procesa

V prvi fazi smo modelirali PDC proces. Ugotovili smo kje v procesu lahko računalniška rešitev nadomesti ročne funkcije in predlagali spremembo procesa tako, da določenih korakov ni treba izvajati (predvsem tistih povezanih s komunikacijo med udeleženci v procesu). Model EDC procesa smo tako kot model PDC procesa modelirali v eEPC tehniki.

### II.2.2. Predlog funkcionalnosti

Pri določanju funkcionalnosti računalniške rešitve smo izhajali iz predloga EDC procesa. Pri tem smo se usmerjali glede na tuje izkušnje, ki smo jih poiskali v literaturi na enak način kot pri analizi procesov. Potrebe uporabnikov smo ugotovili s pomočjo spletne ankete, katere del je bil neposredno usmerjen na ocene pomembnosti predlaganih

funkcionalnosti. Poleg tega smo prek intervjujev opisanih v Metodah v poglavju »II.1.2 Analiza procesov« prišli do dodatnih pojasnil glede potreb uporabnikov.

V skladu z načeli načrtovanja uporabnosti (angl. »usability engineering«) pri razvoju interaktivnih sistemov smo do končne rešitve za EDC sistem prišli prek več korakov. Te korake smo imenovali pilotni sistemi. Prvi pilotni sistem EDCp1 smo razvili in uporabili za potrebe mednarodne onkološke multicentrične študije (Pavlovic & Miklavcic, 2007). Po dveh letih uporabe sistema smo prosili uporabnike sistema, da ocenijo sistem. Ponudili smo jim spletno anketo, v kateri so ocenjevali različne funkcionalnosti in aspekte uporabnosti EDCp1 (glej Dodatek: Spletna anketa uporabnosti pilotnega sistema EDCp1). Poleg tega smo sledili uporabi sistema in naša opazovanja izkoristili za predlog izboljšav prototipa. V skladu z rezultati spletne ankete in našimi ugotovitvami smo predlagali izboljšave prototipa in jih vgradili v naslednji prototip EDCp2. Prototip EDCp2 smo razvili in uporabili na enocentrični onkološki študiji ter skozi sledenje uporabe prototipa preverili uspešnost izboljšav.

### II.3. Določanje tehnološkega okvirja

Pri določanju tehnološkega okvirja smo ugotovili možne različice izpeljav računalniške rešitve za zbiranje podatkov kliničnih študij, jih primerjali v pogledu prednosti, ki jih prinašajo, in se odločili za optimalno tehnološko rešitev za razvoj naše EDC rešitve.

### II.4. Končna rešitev iEDC

Končno rešitev, imenovano iEDC, smo zasnovali na predlaganem funkcionalnem in tehnološkem okviru. Za opisovanje posameznih vidikov sistema smo uporabili UML (Unified Modeling Language) tehniko.

Naredili smo model elektronskih CRF obrazcev (eCRF), ki omogoča implementacijo vseh zelenih funkcionalnosti sistema. Določili smo vloge v sistemu in njihove primere uporabe (angl. »use cases«) tako, upoštevajoč potrebo po reduciranjem upravljanju podatkov. Na osnovi primerov uporabe in zahtevanih funkcionalnosti sistema smo zasnovali uporabniški vmesnik za vsakega od tipov uporabnikov. Vpeljali smo varnostne mehanizme, ki zagotavljajo varno in zanesljivo delovanje in uporabo sistema. Za konec smo opisali arhitekturo in tehnološko implementacijo sistema ter tehnične zahteve za njegovo postavitev.

### II.5. Ocenjevanje

Primerjavo EDC procesa s podporo iEDC rešitve in PDC procesa smo izvedli s pomočjo simulacijske študije, v kateri smo primerjali več različnih scenarijev zbiranja podatkov

klinične študije. Kot izhodišče za izračune smo vzeli zamišljeno vzorčno študijo (klinično preskušanje) običajne velikosti. Značilnosti naše vzorčne študije so naslednje:

- 10 raziskovalnih centrov sodeluje v študiji;
- 100 preizkušancev prispeva vsak od centrov (skupaj 1.000 preizkušancev);
- CRF je razdeljen na 10 približno enakih sekcij (map);
- en CRF vsebuje približno 1.000 podatkov;
- preskušanje traja 24 mesecev;
- monitor obiskuje raziskovalni center enkrat na mesec.

Iz naštetih značilnosti vzorčne študije smo izpeljali naslednje vrednosti parametrov, ki smo jih kasneje uporabili v naših izračunih:

- skupno število CRF obrazcev (enako številu preizkušancev) -  $N_{CRF} = 1.000$ ;
- skupno število sekcij (map) CRF obrazcev -  $N_{sec} = 10.000$ ;
- skupno število zbranih podatkov -  $N_D = 1.000.000$ ;
- skupno število obiskov monitorjev -  $N_M = 240$ .

### II.5.1. Stroški

Ocena stroškov poslovnega procesa je včasih zelo zahtevna, ker na njih vplivajo številni dejavniki, kot so denimo vloženi napor, cena dela, cena in amortizacija drugih uporabljenih virov (npr. oprema, materiali, itn.) kot tudi drugi ekonomski parametri. Stroške smo razdelili na tri skupine: stroške priprave na zbiranje podatkov, stroške izvajanja procesa zbiranja podatkov in materialne stroške procesa zbiranja podatkov. Primerjali smo ocene teh treh skupin stroškov za PDC in EDC proces. Pri vseh ocenah smo izhajali iz podatkov pridobljenih prek intervjujev, literature ali naših prejšnjih izkušenj.

Za izračun stroškov izvajanja procesa zbiranja podatkov so poleg parametrov vzorčne študije pomembni tudi napor udeležencev v procesu, cene njihovih ur dela ter predvideni parametri kakovosti podatkov. V nadaljevanju smo te tri dejavnike podrobneje predstavili.

#### *Napori*

Napor za izvajanje ene naloge (funkcije na eEPC diagramu) je določen s časom, ki ga posamezni udeleženec procesa uporabi za dokončanje te naloge. Ta čas vključuje: čas za pripravo; čas za izvajanje; in čas za zaključek. V naših izračunih smo vse tri čase obravnavali skupaj kot čas za opravljanje naloge in ga v enačbah predstavili kot funkcijo  $F$ .

Pri izračunu vsake funkcije  $F$  smo upoštevali parametre študije, ki lahko vplivajo na trajanje opravljanja naloge (npr. število podatkov v CRF obrazcu ali število CRF dokumentov itn.). Napore smo potem računali na ta način, da smo ocenili povprečne napore za vsako nalogo normirane na relevantne parametre naloge (npr. čas potreben za zapisovanje enega podatka v CRF obrazec) in skupen napor izračunali z linearnimi



enačbami, ki upoštevajo parametre. V naših enačbah smo povprečne napore predstavljali s simbolom  $\mu$ .

### *Cene dela*

Stroški procesa so odvisni od cen dela posameznih udeležencev v procesu. V naših izračunih smo na osnovi informacij pridobljenih prek intervjujev uporabili naslednje ocene cen dela:

- raziskovalec:  $\varphi_I = 20 \frac{\text{€}}{h}$ ;
- raziskovalna sestra:  $\varphi_{IA} = 10 \frac{\text{€}}{h}$ ;
- monitor (angl. »Clinical Research Associate« - CRA):  $\varphi_{CRA} = 15 \frac{\text{€}}{h}$ ;
- vnašalec podatkov v bazo:  $\varphi_{DE} = 6 \frac{\text{€}}{h}$ ;
- upravljavec podatkov (angl. »Data Manager«):  $\varphi_{DM} = 20 \frac{\text{€}}{h}$ ;
- koordinator študije:  $\varphi_C = 25 \frac{\text{€}}{h}$ .

Seveda, cene dela so lahko različne za različne raziskovalne centre, za različne države in različne naročnike. Kakorkoli, za konkretne izračune za konkretno študijo je treba uporabiti le konkretne cene dela udeležencev v procesu.

### *Kakovost podatkov*

Za večino kliničnih študij je določena kakovost podatkov postavljena kot pogoj. To pomeni, da so podatki zbrani med študijo zadostno kakovostni le v primeru, da je delež napak med podatki pod določeno mejo. S primerno oblikovanim procesom zbiranja podatkov (predvsem zajema, preverjanja in odpravljanja napak) je mogoče zagotoviti ustrezno kakovost podatkov. V naših modelih na kakovost podatkov vpliva več parametrov, za katere smo v izračunih uporabili naslednje ocene.

- Delež podatkov pri katerih monitor ugotovi napake: po podatkih pridobljenih prek intervjujev za PDC je ta delež  $\varepsilon_{mon}^P = 10 \%$ ; po podatkih objavljenih v (Spink, 2002) za EDC je ta delež  $\varepsilon_{mon}^E$  lahko ocenjen kot  $20\% \varepsilon_{mon}^P$ , kar pomeni  $\varepsilon_{mon}^E = 2 \%$ .
- Delež napak med podatki, ki jih prispevajo vnašalci podatkov (angl. »entry error rate« -  $\varepsilon_{ent}^P$ ): po podatkih v literaturi to vrednost lahko ocenimo kot 1 % (Day, Fayers, & Harvey, 1998; Kawado et al., 2003; King & Lashley, 2000; Kleinman, 2001; Wahi, Parks, Skeate, & Goldin, 2008).
- Delež podatkov za katere se pojavijo poizvedbe ( $\varepsilon_{qry}$ ): glede na podatke pridobljene prek intervjujev za PDC je ta delež  $\varepsilon_{qry}^P = 2 \%$ ; za EDC različni viri navedeni v (Bart, 2003) ocenjujejo zmanjšanje poizvedb v primerjavi s PDC procesom za 80 % do 95 %, tako da lahko uporabimo oceno  $\varepsilon_{qry}^E = 0,2 \%$ .

- Delež neverodostojnih vnosov podatkov (angl. »plausibility discrepancies« - podatki zunaj predvidenih meja, manjkajoči podatki, nepravilne kombinacije podatkov itn.): po informacijah pridobljenih prek intervjujev in po podatkih objavljenih v (Spink, 2002) za PDC je njihov delež 80 % vseh poizvedb ( $\varepsilon_{pla}^P = 80 \% \varepsilon_{qry}^P$ ).
- Delež napak ugotovljenih prek analize podatkov (angl. »analysis discrepancies« - npr. kršitve protokola, odstopanja od običajnih vrednosti itn.): to so vse preostale poizvedbe -  $\varepsilon_{ana}^P = \varepsilon_{qry}^P - \varepsilon_{pla}^P = 20 \% \varepsilon_{qry}^P$ .
- Delež podatkov, ki jih je treba popraviti po odgovoru na poizvedbo: glede na informacije pridobljene prek intervjujev, ki smo jih opravili, ta delež lahko ocenimo kot polovico vseh poizvedb -  $\varepsilon_{cor}^P = 50 \% \varepsilon_{qry}^P$ .

Ocene, ki smo jih v tem poglavju predstavili, so uporabljene za naše izračune in komentirane v Razpravi.

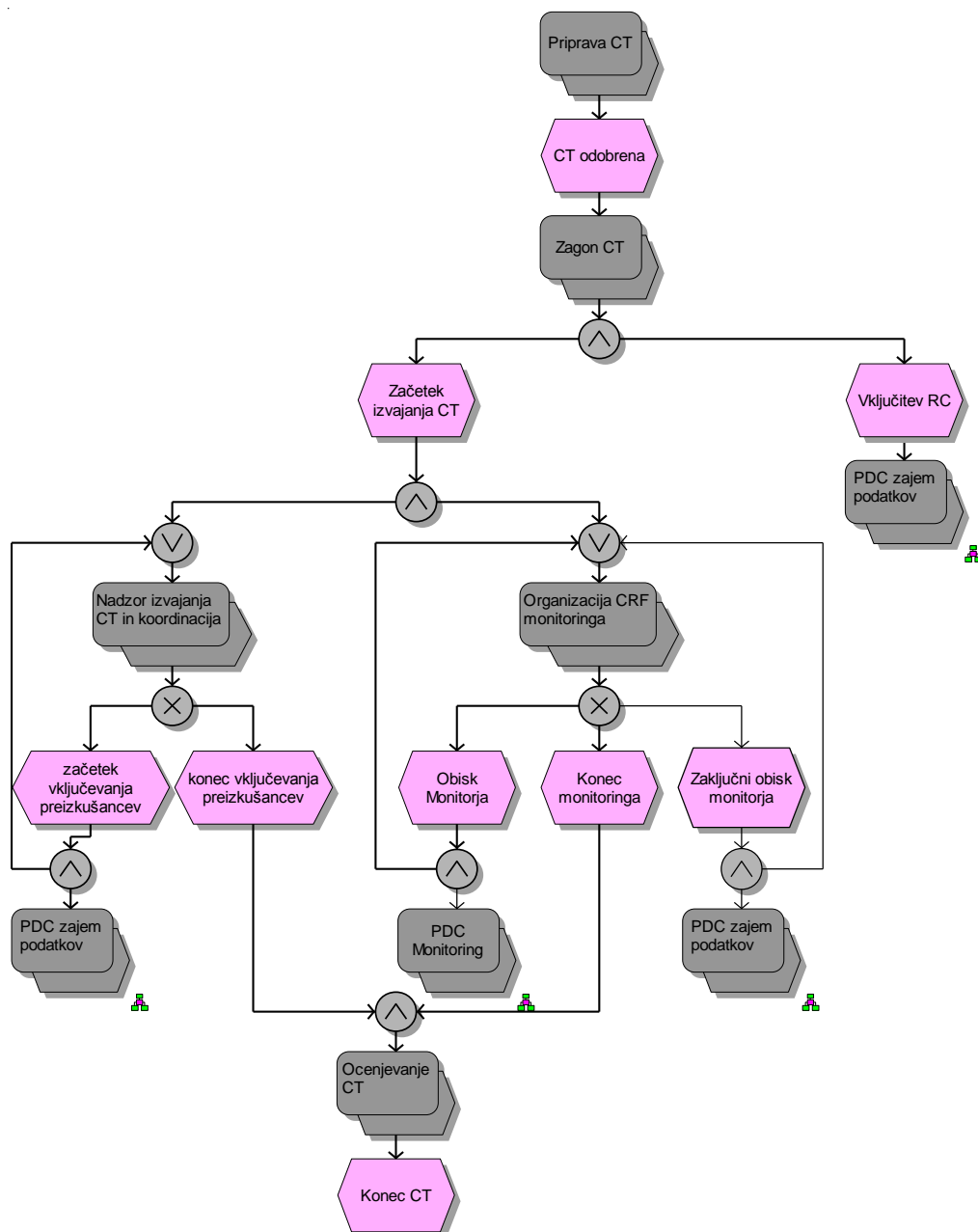
### II.5.2. Čas za končanje zbiranja podatkov

Čas za končanje procesa zbiranja podatkov je odvisen od samega trajanja klinične študije (dela z bolniki) in časa, ki ga po končani študiji podatkovni center (v katerem se zbirajo podatki iz vseh raziskovalnih centrov) potrebuje, da zbere in zaklene vse podatke v bazi. Glede na dejstvo, da sam postopek zbiranja podatkov (PDC ali EDC) ne vpliva na trajanje klinične študije, je razlika v času potrebnem za končanje zbiranja podatkov dejansko le posledica razlike v času od konca zajema podatkov zadnjega bolnika do zaklepanja baze podatkov. Če upoštevamo, da so obiski monitorja v obeh primerih (PDC in EDC) v enakih časovnih intervalih, potem dejansko lahko primerjamo le čas od zadnjega obiska monitorja do zaklepanja baze podatkov oziroma čas, potreben za zadnji cikel upravljanja s podatki.

## III. Rezultati

### III.1. Opis PDC procesa

PDC proces je modeliran s pomočjo eEPC (Extended Event-driven Process Chains) tehnike predstavljene v Metodah (poglavje II.1.1 eEPC tehnika modeliranja procesov). PDC modeli procesov so razdeljeni na več diagramov. Izhodiščni diagram, ki opisuje podproces upravljanja klinične študije je prikazan na Sliki 6.

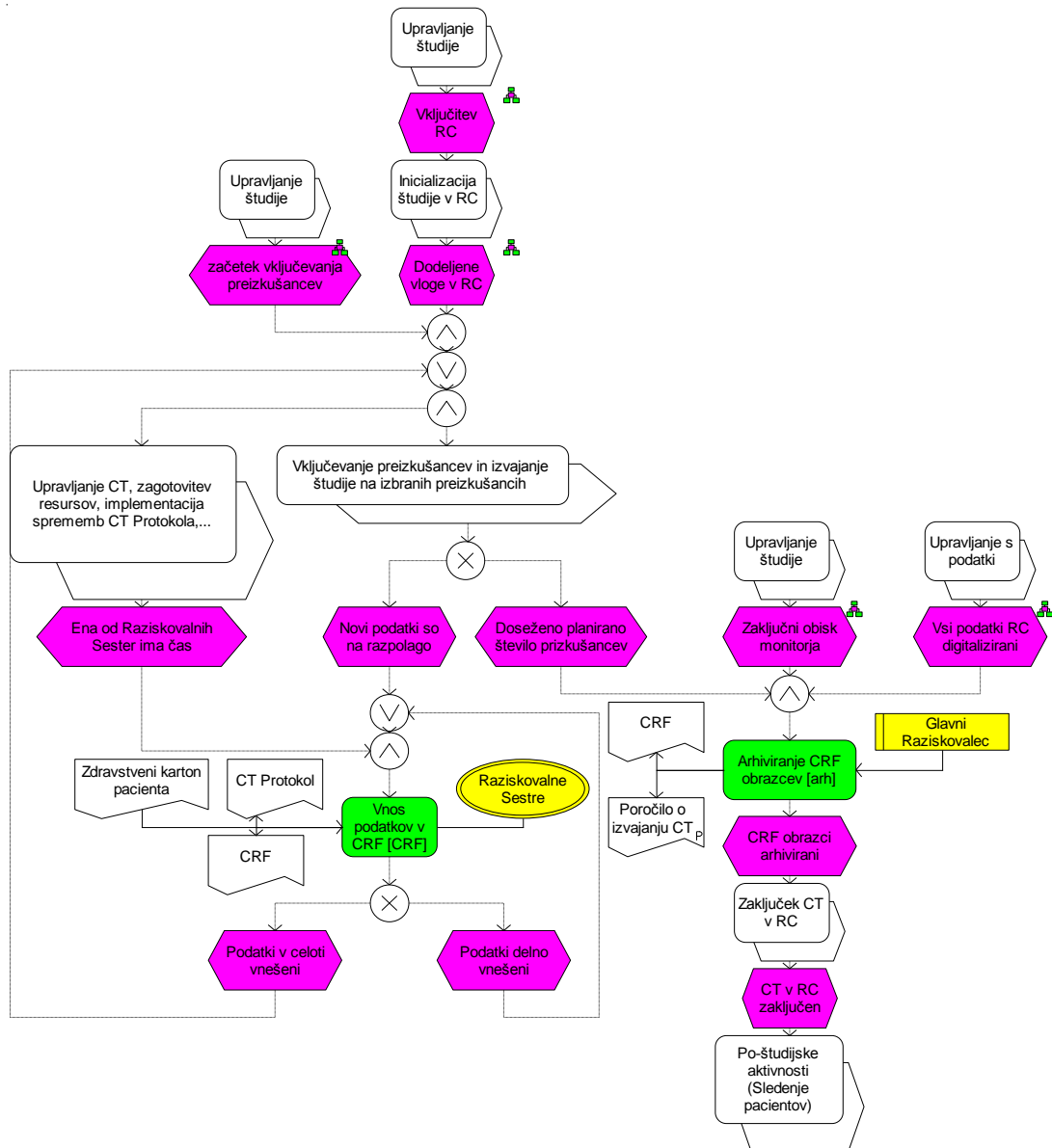


Slika 6. eEPC model upravljanja klinične študije (CT – klinična študija; RC - raziskovalni center; CRF – obrazec za zbiranje podatkov preizkušancev)

Upravljanje klinične študije zahteva številne aktivnosti, med katerimi so le nekatere povezane z zbiranjem podatkov klinične študije. Na diagramu prikazanem na Sliki 6 lahko opazimo podprocese »PDC Zajem podatkov«, »PDC Monitoring« in »PDC Upravljanje s podatki«. V nadaljevanju so ti podprocesi bolj podrobno opisani.

#### III.1.1. PDC Zajem podatkov

Ena izmed osnovnih nalog raziskovalnega centra je, da zagotovi vse pomembne informacije, ki so določene v Protokolu klinične študije in spremljajoči dokumentaciji. Ta proces smo imenovali »zajem podatkov«, da bi ga ločili od celotnega procesa »zbiranja podatkov« (angl. »data collection«), ki se nanaša na celotno študijo in vključuje tudi aktivnosti na strani naročnika (termin »zbiranje podatkov« uporabljamo v besednih zvezah »elektronsko zbiranje podatkov« - EDC - in »papirno zbiranje podatkov« - PDC). Naloga zajema podatkov je, da so podatki popolni in natančni ter pravočasno posredovani naročniku. Na Sliki 7 je prikazan diagram procesa zajema podatkov.



Slika 7. eEPC model papirnega zajema podatkov (CT – klinična študija; RC – raziskovalni center; CRF – obrazec za zbiranje podatkov preizkušanca)

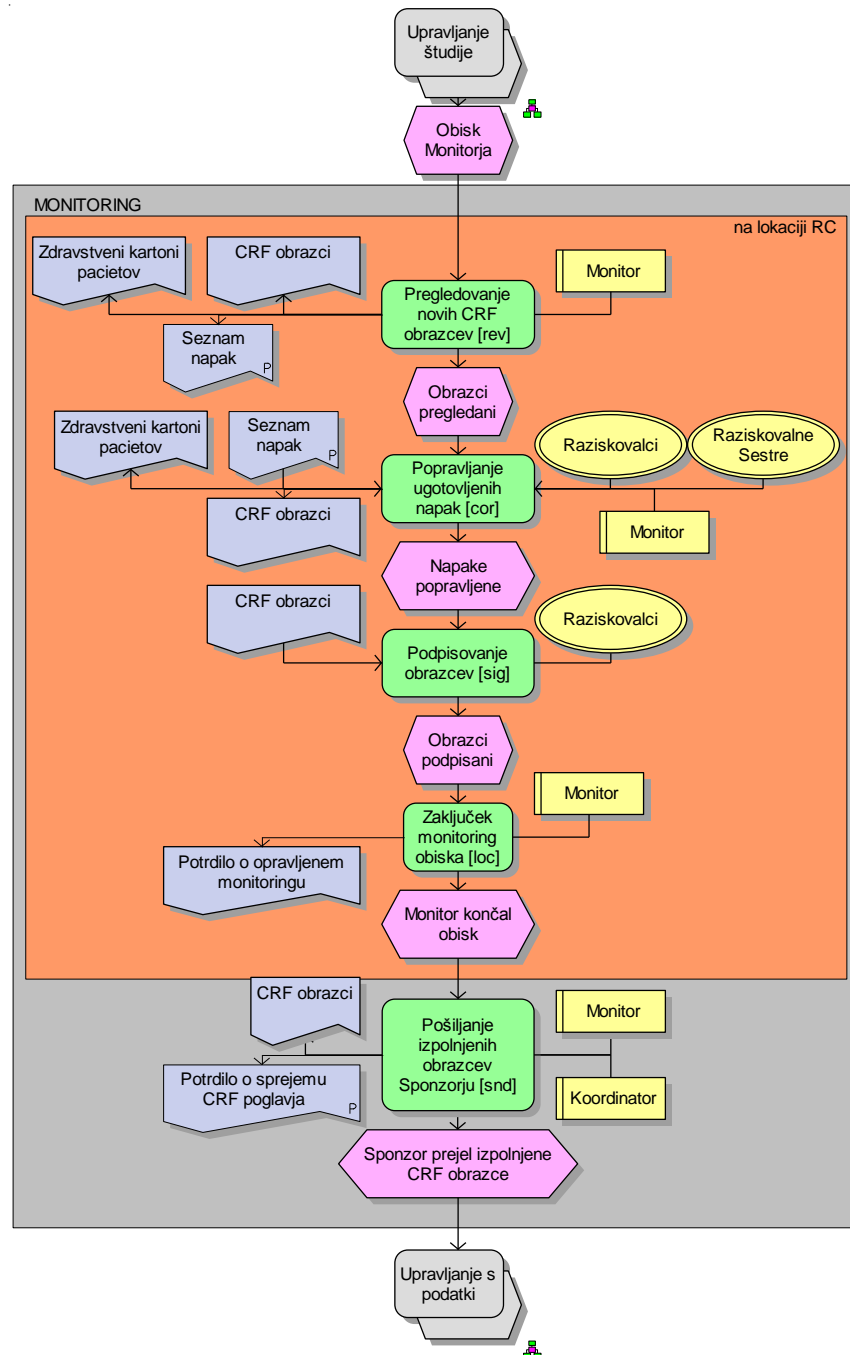
Čeprav vodenje študije ter vključevanje preizkušancev in izvajanje kliničnih posegov ne sodita v naloge zajema podatkov, dogodki, ki jih ta dva podprocesa ustvarjata, vplivajo na potek zajema podatkov. In sicer, sam potek študije, vključevanje preizkušancev in posegi na preizkušancih vplivajo na dotok novih podatkov. Po drugi strani pa upravljanje z viri, kot so raziskovalne sestre ali dostopnost zdravstvene dokumentacije preizkušancev, vpliva na čas vnosa novih podatkov v CRF obrazce.

Dve ključni nalogi raziskovalnega centra pri zajemu podatkov sta vnos podatkov v CRF obrazce in arhiviranje obrazcev ob koncu študije. Vnos podatkov v CRF obrazce praviloma izvajajo raziskovalne sestre. Kot vir podatkov uporabljajo originalne zdravstvene zapise preizkušanca, kot vodilo pri izpolnjevanju CRF obrazcev pa Protokol klinične študije. Arhiviranje CRF obrazcev je poseg, ki ga v raziskovalnem centru izvedejo ob koncu študije.

Ko je opravljen zadnji predviden poseg na zadnjem preizkušancu, ko si vsi podatki pregledani s strani monitorja in potrjeni ter »zaklenjeni« s strani podatkovnega centra, takrat glavni raziskovalec v raziskovalnem centru obrazce arhivira in po potrebi pošlje poročilo o zaključku študije. Vse nadaljnje aktivnosti niso povezane z zbiranjem podatkov klinične študije in jih zaradi tega v naših modelih niti ne obravnavamo.

#### III.1.2. PDC Monitoring

Proces monitoringa (angleško besedo »monitoring« uporabljamo, ker je slovenska stroka to besedo sprejela kot strokovni termin) je neizogiben podproces klinične študije, katerega cilj je zagotoviti kakovost (popolnost in točnost) zbranih podatkov. Naročnik dodeli eni ali več osebam (monitorjem) naloge, da občasno obiskujejo raziskovalne centre in preverjajo, kako poteka postopek izvajanja študije in zajema podatkov. Monitorji ob vsakem obisku opravijo številne aktivnosti, med katerimi so le nekatere neposredno vezane na proces zbiranja podatkov. Na Sliki 8 je prikazan podproces PDC Monitoring, ki vključuje le aktivnosti vezane na zbiranje podatkov.

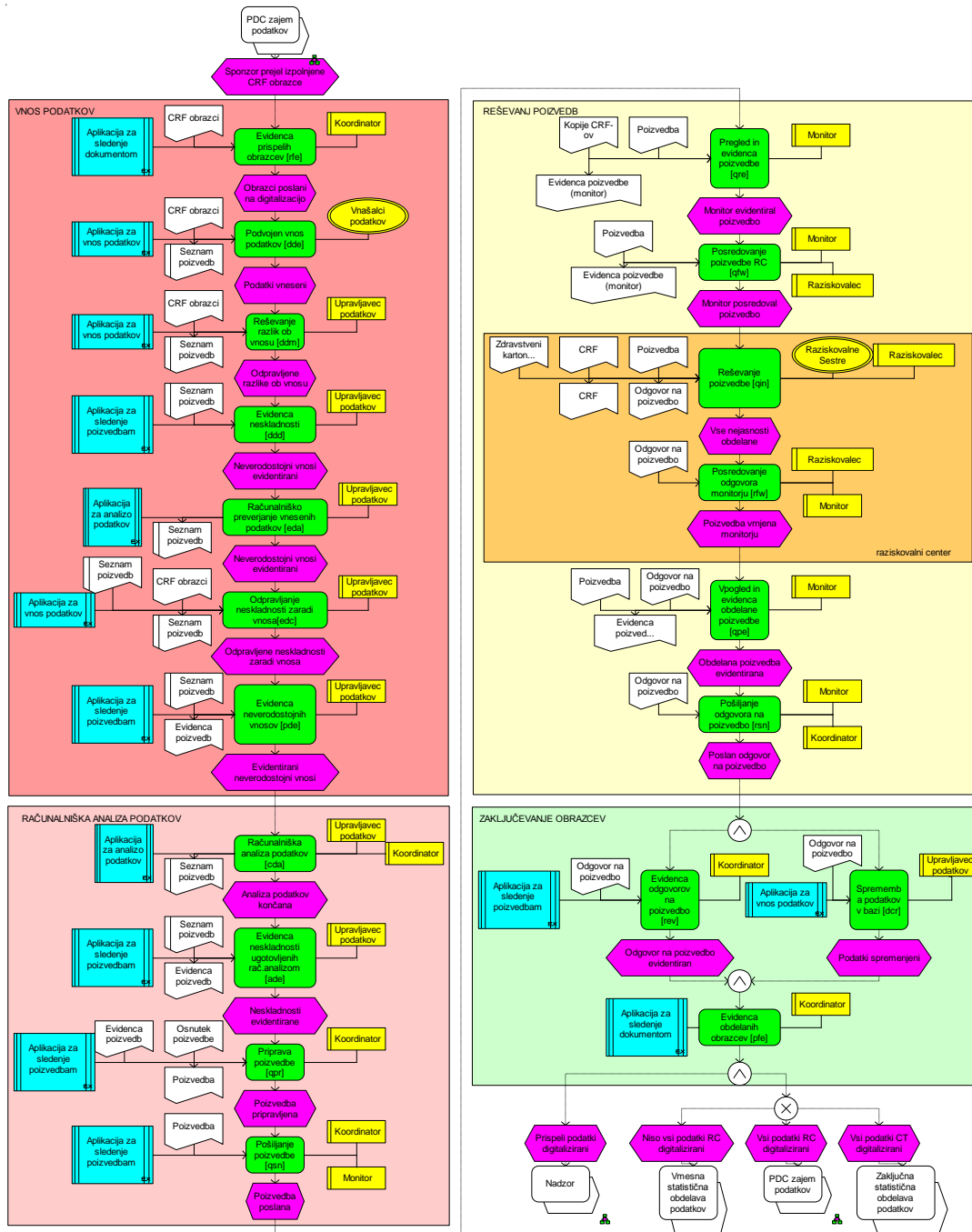


Slika 8. eEPC model podprocesa PDC Monitoring (CRF – obrazec za zbiranje podatkov preizkušanca)

V sklopu procesa zbiranja podatkov mora monitor pregledati vsebine CRF obrazcev na ta način, da primerja vnesene podatke s podatki v originalni zdravstveni dokumentaciji preizkušanca. Pri tem si zabeleži morebitne nepravilne vnose in potem od raziskovalcev in raziskovalnih sester zahteva, da te podatke preverijo in jih po potrebi popravijo. Raziskovalci tako preverjene in popravljene obrazce potem podpišejo. Monitor zaključi obisk in kopije podpisanih CRF obrazcev pošlje naročniku študije (koordinatorju) v nadaljnjo obdelavo.

### III.1.3. PDC Upravljanje s podatki

Podatke, ki jih prispevajo raziskovalni centri potem, ko so pregledani in potrjeni s strani monitorjev, digitalizirajo in temeljito preverijo v podatkovnem centru. Podatkovni center je zadožen za nadzor nad procesom zbiranja podatkov in kakovost podatkov. Proces upravljanja s podatki je prikazan na Sliki 9.



Slika 9. eEPC model pod-procesa PDC Upravljanje s podatki (CT – klinična študija; RC – raziskovalni center; CRF – obrazec za zbiranje podatkov preizkušanca)



Kot je iz slike razvidno, je proces upravljanja s podatki najbolj zapleten od vseh podprocesov zbiranja podatkov. Proces lahko razdelimo v štiri logične sklope, in sicer: vnos podatkov; računalniška analiza podatkov; reševanje poizvedb; ter zaključevanje obrazcev.

#### Vnos podatkov

Ob sprejemu novih podpisanih CRF obrazcev koordinator evidentira prispele obrazce v aplikacijo za sledenje dokumentom (ali pa kje drugje). Vnašalci podatkov (osebje zadolženo za pretipkavanje podatkov iz CRF obrazcev v bazo podatkov) podatke vnesejo v bazo s pomočjo aplikacije za vnos podatkov. Pri tem morebitne nejasnosti med podatki evidentirajo na poseben seznam poizvedb. Zaradi želje po zmanjševanju napak ob vnosu naročnik praviloma zahteva dvojni vnos podatkov (angl. »double data entry«), pri katerem dva vnašalca vnašata iste podatke, ki jih potem upravljavec podatkov primerja. Ugotovljene razlike pri dvojnem vnosu podatkov upravljavec podatkov potem odpravi. Preostale neskladnosti upravljavec podatkov evidentira v aplikaciji za sledenje poizvedbam. Z računalniško analizo vnesenih podatkov upravljavec podatkov ugotovi morebitne druge neverodostojnosti med podatki ter odpravi tiste, ki so nastale ob vnosu podatkov. Vse preostale neverodostojnosti med podatki upravljavec podatkov evidentira v aplikaciji za sledenje poizvedbam.

#### Računalniška analiza podatkov

Potem, ko so podatki čim bolj kakovostno vneseni v bazo podatkov, koordinator skupaj z upravljavcem podatkom izvede računalniško analizo novih podatkov, pri čemer v analizo vključi tudi vse do tedaj zbrane podatke. Pri tem skupaj ugotavljajo morebitne kršitve protokola, nenavadna odstopanja od običajnih vrednosti, manjkajoče podatke, ali druge skrite napake med podatki, ki so jih posredovali raziskovalci. Vse ugotovljene neskladnosti upravljavec podatkov evidentira v aplikaciji za sledenje poizvedbam. Koordinator potem pripravi dopise s poizvedbami in jih pošlje ustreznemu monitorju.

#### Reševanje poizvedb

Reševanje poizvedb je ključni segment procesa upravljanja s podatki. Različni naročniki kliničnih študij imajo različne pristope k temu procesu. Lahko se odločijo za neposredno komunikacijo koordinatorja z raziskovalnim centrom, ali pa vključijo monitorja v ta proces. V našem modelu smo prikazali pristop, ko koordinator poizvedbe pošilja monitorju, ki je na ta način vključen v proces reševanja poizvedb in seznanjen z morebitnimi spremembami vrednosti v CRF obrazcih. Monitor pregleda in evidentira prispele poizvedbe ter jih posreduje v ustrezen raziskovalni center (razen, če so poizvedbe takšne, da lahko sam na njih odgovori). V raziskovalnem centru raziskovalno osebje pristopi k reševanju poizvedb ter svoje odgovore in morebitne popravke CRF obrazcev posreduje nazaj monitorju, ki jih evidentira, po potrebi dopolni s svojimi pojasnili in vse skupaj pošlje nazaj koordinatorju.

#### Zaključevanje obrazcev

Po sprejemu odgovorov na poizvedbe lahko koordinator pristopi k zaključevanju obrazcev, oziroma dela na upravljanju s podatki. Koordinator evidentira odgovore na poizvedbe (in jih sprejeme, ali zavrne). Upravljavec podatkov v bazi izvede morebitne spremembe podatkov, ki izhajajo iz odgovorov na poizvedbe. Dokončno obdelane obrazce koordinator evidentira in s tem zaključi delo na teh obrazcih. Nadaljnje aktivnosti so povezane z nadzorom študije, statistično obdelavo podatkov in novim zajemom podatkov v raziskovalnih centrih.

## III.2. Funkcionalni okvir rešitve

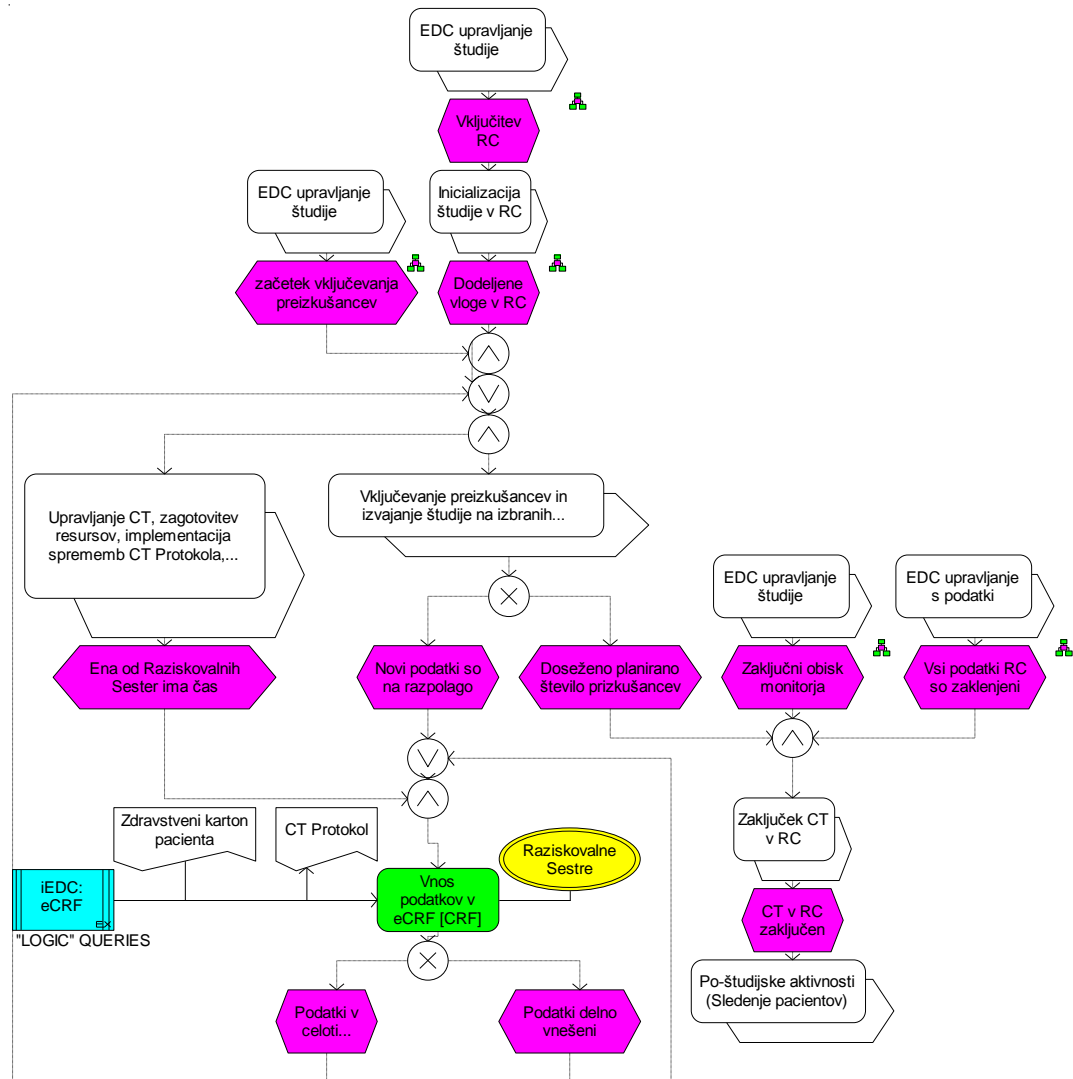
### III.2.1. Predlog EDC procesa

Kot osnovo za razvoj EDC procesa smo vzeli PDC proces. Za vsakega od identificiranih podprocesov (Zajem podatkov, Monitoring in Upravljanje s podatki) smo poiskali primerno zamenjavo z uporabo računalniške tehnologije. Pri tem je bilo pomembno ohraniti vhodne in izhodne informacije. Za same nosilce informacij (v primeru PDC je to papir) smo predlagali elektronske nosilce informacij (npr. elektronske obrazce CRF). V nadaljevanju so predstavljeni vsi trije podprocesi EDC procesa.

#### *EDC Zajem podatkov*

Kot osnovno izboljšavo procesa Zajema podatkov predlagamo spremembo orodja za zbiranje podatkov, in sicer namesto papirnatih obrazcev predlagamo uporabo računalniške rešitve, v katero raziskovalci in njihovi sodelavci (raziskovalne sestre) vnašajo podatke. Avtoriziran uporabnik (npr. raziskovalna sestra) lahko vnaša podatke v eCRF obrazce in spreminja že vnesene podatke. Ob tem se poleg samih podatkov lahko shranjujejo tudi informacije o tem, kdo in kdaj je podatek vnesel ali spremenil.

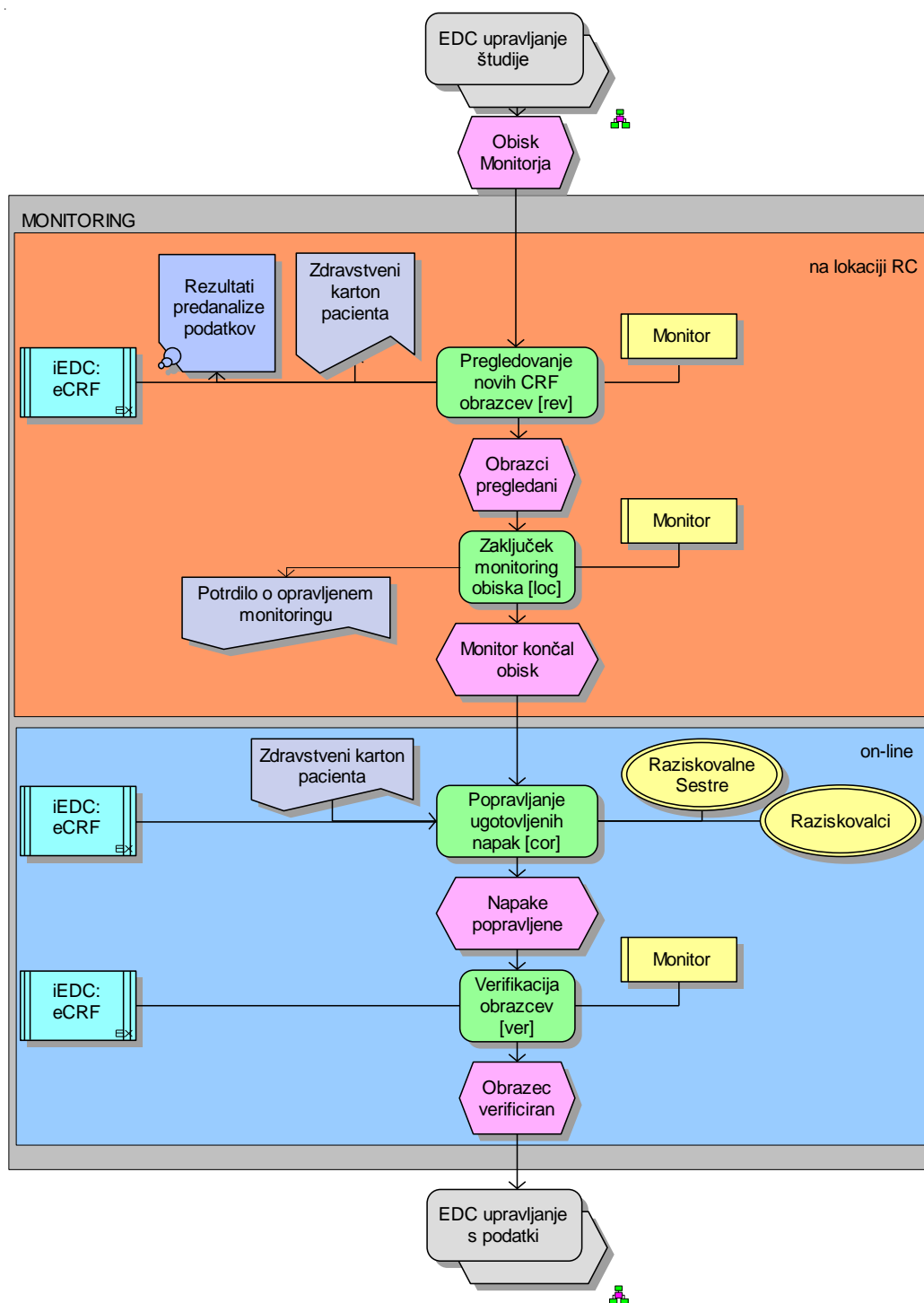
Sprememba orodja za zbiranje podatkov omogoča še optimizacijo postopka arhiviranja CRF obrazcev. Vsak dokončan eCRF obrazec je avtomatično shranjen v sistemu in po potrebi dostopen pooblaščenim osebam na vpogled tudi na daljavo. Zaradi tega je, kot je razvidno s Slike 10, zajem podatkov nekoliko bolj enostaven.



Slika 10. eEPC diagram podprocesa EDC Zajem podatkov (CT – klinična študija; RC – raziskovalni center; eCRF – elektronski obrazec za zbiranje podatkov preizkušanca)

### EDC Monitoring

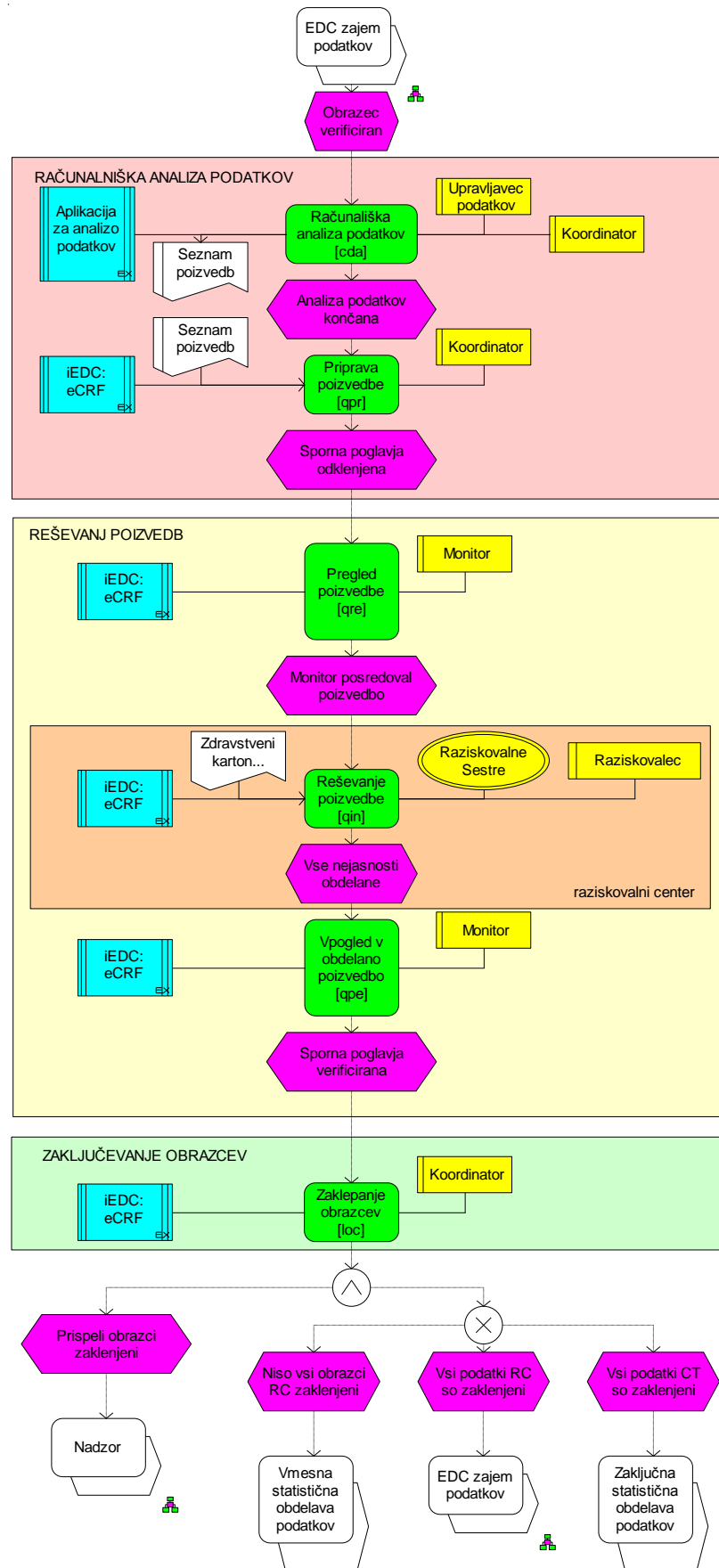
Uporaba računalniške rešitve omogoča monitorju, da podatke pregleduje in poizvedbe o podatkih beleži neposredno v računalniško aplikacijo, ki skrbi, da so ustrezni uporabniki o tem obveščeni. Poleg tega računalniška rešitev z oddaljenim dostopom omogoča monitorju, da sledi reševanju poizvedb, brez da bi bil prisoten. Odgovore na poizvedbe lahko pregleda in odobri ali vrne nazaj v reševanje. Za celotno komunikacijo pri tem skrbi računalniška rešitev za zbiranje podatkov. Računalniška rešitev ob tem hrani informacije o spremembah podatkov in verifikaciji podatkov ter avtomatično pošilja ustrezna obvestila akterjem v sistemu. Iz naštetih razlogov je diagram EDC podprocesa EDC Monitoring (Slika 11) bolj enostaven, kot je primer pri PDC procesu.



Slika 11. eEPC model podprocesa EDC Monitoring (eCRF – elektronski obrazec za zbiranje podatkov preizkušanca; iEDC – spletna EDC aplikacija)

### EDC Upravljanje s podatki

Elektronsko zbiranje podatkov prinaša največ sprememb podprocesu Upravljanja s podatki. Ker so podatki že na viru podatkov (raziskovalni center) digitalizirani, se lahko podatkovni center posveti le analizi zbranih podatkov in t.i. »čiščenju podatkov« (napak). Podproces EDC Upravljanje s podatki je prikazan na Sliki 12.



Slika 12. ePC model podprocesa EDC Upravljanje s podatki (CT – klinična študija; RC – raziskovalni center; eCRF – elektronski obrazec za zbiranje podatkov preizkušanca; iEDC – spletna EDC aplikacija)

Spremembe procesa so zaradi elektronskega zbiranja podatkov številne. Kot prvo, postopek digitalizacije podatkov ni več potreben (sklop »VNOS PODATKOV« na Sliki 9), kar posledično zagotavlja manj napak med podatki in zmanjšuje napor za njihovo odpravljanje. Računalniški sistem bistveno prispeva k lažji komunikaciji in poenostavitvi procesa zbiranja podatkov pri reševanju poizvedb (sklop »REŠEVANJE POIZVEDB«), ker skrbi za sledenje spremembam podatkov in obveščanje udeležencev v procesu. Zaključevanje obrazcev je tudi poenostavljeno (sklop »ZAKLJUČEVANJE OBRAZCEV«). Med vsemi omenjenimi spremembami je treba poudariti, da je bistveno zmanjšano tudi število nosilcev informacij (dokumentov), ki so prisotni v procesu. Večina informacij (vse razen vira kliničnih podatkov preizkušanca – zdravstvene dokumentacije) je lahko shranjena znotraj računalniške rešitve za zbiranje podatkov.

#### III.2.2. Prototipa

Izhajajoč iz opisa PDC procesa in želje po nadomestilu tega procesa z EDC procesom smo izdelali dva prototipa EDC rešitve. Oba prototipa sta uporabljena v realnem okolju (za podpora kliničnim študijam).

##### *Prototip 1 - EDCp1*

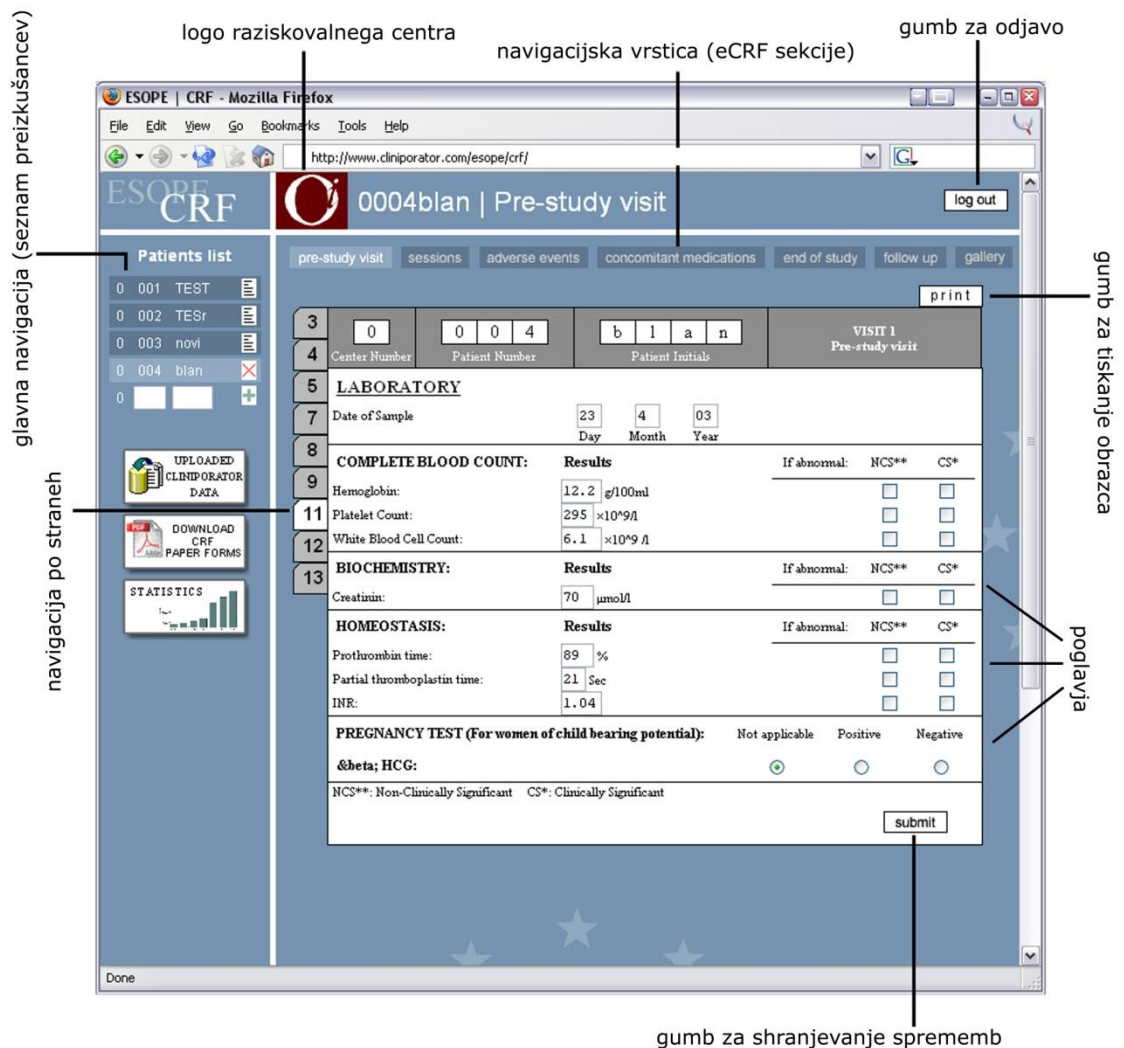
Prototip 1 (EDCp1) smo zasnovali in razvili za potrebe mednarodne multicentrične onkološke klinične študije. Pri razvoju EDCp1 smo osnovni funkcionalni okvir nadgradili z naslednjimi funkcionalnimi zahtevami:

- možnost shranjevanja digitalnih slik (tumorjev);
- interaktivna mapa preizkušanca za določanje lokacij tumorjev;
- možnost prenosa podatkov iz zunanjega vira podatkov (medicinske naprave);
- on-line prikaz osnovnih statistik;
- interaktivni graf odziva na zdravljenja tumorja.

Tehnološke značilnosti prototipa EDCp1 so naslednje:

- podatkovni in spletni strežnik na istem računalniku;
- podatkovni strežnik: Microsoft SQL Server 2000;
- spletni strežnik: Microsoft Internet Information Server 5;
- spletna aplikacija: Active Server Pages (ASP);
- spletni brskalnik: Internet Explorer 5 ali več in njemu kompatibilni z vgrajenim plug-in-om za Flash™;
- varna povezava: Secure Socket Layer (SSL) protokol;
- avtentikacija in avtorizacija uporabnikov: le z uporabniškim imenom in geslom.

Uporabniški vmesnik smo zasnovali, kot je prikazano na Sliki 13.



Slika 13. Uporabniški vmesnik EDCp1 prototipa

Značilno za elektronske obrazce eCRF prvega prototipa je to, da so oblikovani enako kot papirnati CRF obrazci. Polja so grupirana na strani v tematske sklope (poglavja) ter razporejena po celotni strani obrazca. Posamezne strani so grupirane v sekcije (npr. »Pre-Study Visit«) in oštevilčene enako kot v papirnatem CRF obrazcu. Uporabnik ima tako tri nivoje navigacije: po preizkušancih; po sekcijah posameznega eCRF obrazca; ter po posameznih straneh znotraj sekcij.

Druga značilnost EDCp1 prototipa sta dva osnovna tipa uporabnikov in dva dodatna tipa uporabnikov. Dva osnovna tipa uporabnikov sta: uporabnik (vnaša in spreminja podatke) in »read-only« uporabnik (ima le vpogled v podatke in nadzor nad potekom študije). Dva dodatna tipa uporabnikov sta: obiskovalec (nima dostopa do podatkov raziskovalnih centrov); nadzornik medicinske naprave (ima vpogled le v podatke v centralni bazi, prenesene iz medicinskih naprav).

EDCp1 prototip je detajlno predstavljen v (Pavlović, 2004). V članku (Pavlovic & Miklavcic, 2007) so povzeti tudi rezultati spletne ankete, katere namen je bil oceniti uporabnost sistema na osnovi menja uporabnikov. Rezultati ankete so pokazali, da so uporabniki predvsem zadovoljni s funkcionalno zasnovo rešitve. Po drugi strani pa so omenili določene težave, ki izhajajo iz tehnološke zasnove prototipa. Namreč, aplikacija je delovala nekoliko bolj počasi, kot bi si uporabniki želeli. Kot vir teh težav smo prepoznali postavitev obeh strežnikov (spletnega in podatkovnega) na enem računalniku in uporabo skriptnega programskega jezika (ASP).

Naloge podatkovnega centra smo izvajali sami. EDCp1 ni imel vgrajene funkcije za upravljanje s poizvedbami. To pomeni, da so bili enkrat vneseni in digitalno podpisani podatki zaklenjeni in jih uporabnik ni mogel več spreminjati. Ob pregledu podatkov s strani monitorja, ali ponovnem pregledu podatkov s strani raziskovalca, so bile pogosto ugotovljene napake ali manjkajoči podatki. V tem primeru je raziskovalec ali monitor moral zaprositi koordinatorja študije, da dovoli spremembo podatkov. Po dovoljenju koordinatorja je skrbnik baze odklenil podatke z uklanjanjem digitalnega podpisa neposredno v bazi podatkov in o tem obvestil raziskovalca, monitorja in koordinatorja. Takšna rešitev se je pokazala kot zamudna, ker je vsiljevala odvečno komunikacijo na relaciji raziskovalec-monitor-koordinator in zahtevala ukrepe skrbnika baze neposredno na bazi podatkov.

V treh letih delovanja prototipa smo imeli več kot 150 zahtev za odklepanje ali spremembe podatkov v bazi z več kot 1000 posameznimi spremembami podatkov. Skoraj ena petina se jih je nanašala na izpuščene podatke. Poleg tega so pogosto že izpolnjeni obrazci veliko časa ostali nepodpisani, kar je koordinatorju in monitorju dajalo napačno sliko o poteku študije. Namreč, šele ob podpisu obrazca sta koordinator in monitor dobila obvestilo o izpolnjenosti obrazca.

Med uporabo EDCp1 je večkrat prišlo do manjših sprememb CRF obrazcev. Vse spremembe so se nanašale na dodajanje ali odstranitev določenih polj v obrazcih. Zaradi tega je bilo treba nadgraditi tudi elektronske obrazce eCRF z novimi vnosnimi polji, ali odstraniti odvečna polja.

Čeprav so uporabniki na splošno ocenili sistem kot odličen in izrazili željo po njegovi uporabi namesto papirnatih obrazcev, so rezultati ankete ter potrebe po naših intervencijah na sistemu med njegovo uporabo nakazali, v katero smer bi morali izboljšati sistem:

- izboljšave zmožnosti sistema;
- poenostavitev postopka manjših sprememb obrazcev med študijo;
- izboljšave vmesnika na način, da bo bolj enostavno sledenje izpolnjenosti obrazcev;



### III. Rezultati

- bolj natančna razdelitev vlog med uporabniki (več tipov uporabnikov);
- neposredno generiranje poizvedb in odklepanje podatkov znotraj sistema.

V smeri teh izboljšav smo zasnovali naš naslednji prototip EDCp2.

#### *Prototip 2 - EDCp2*

Drugi prototip rešitve za zbiranje podatkov kliničnih študij EDCp2 je uporabljen za enocentrično onkološko študijo. Zaradi enostavnosti študije smo EDCp2 zasnovali tako, da zagotovi le prve tri točke izboljšav nakazane s prototipom EDCp1. Uporabniški vmesnik je bil oblikovan, kot je prikazano na Sliki 14.

The screenshot displays the GEMCIS web application interface. The browser address bar shows the URL: <https://www.edc-trial.com/gemcis/form2.asp?patientid=0003TR>. The page title is "0003TRET | Medical History". The interface features a navigation menu on the left with a "LIST OF PATIENTS" section containing entries for patients 001 PRVI, 002 DRUG, 003 TRET (highlighted), 004 CETR, 005 PETI, 006 SEST, 007 seda, 008 abcd, 009 novi, and a new patient button. The main content area has a header with "GENERAL", "MEDICAL HISTORY" (active), "BASELINE", "CYCLES", "FOLLOW UP", "ASSESSM RESPON", "SUMMARY", and "GALLERY" tabs. Below the tabs, there is a "print" button and a form with the following sections:

- SMOKING:** A dropdown menu for "Smoking: -- select --".
- PREVIOUS DISEASES:** A section with radio buttons for "Previous Cancer: Yes" (selected) and "No", and a text input field for "If yes, specify:". There are checkboxes for "N/A" and "Yes".
- Diabetes:** Radio buttons for "Yes" and "No", with a checkbox for "N/A".
- Hypertension:** Radio buttons for "Yes" and "No", with a checkbox for "N/A".
- Cardiac:** Radio buttons for "Yes" and "No", with a checkbox for "N/A".
- Vascular:** Radio buttons for "Yes" and "No", with a checkbox for "N/A".
- Respiratory:** Radio buttons for "Yes" and "No", with a checkbox for "N/A".
- Neurologic:** Radio buttons for "Yes" (selected) and "No", with a checkbox for "N/A".
- OtherSpec:** A text input field for "OtherSpec:" with a checkbox for "N/A".

At the bottom, there is a section titled "MEDICAL HISTORY OF THIS CANCER" with fields for "Date of initial diagnosis:" and "Initial Stage: -- select --".

Slika 14. Uporabniški vmesnik prototipa EDCp2

Osnovna funkcionalna izboljšava v primerjavi z EDCp1 je prikaz informacije o izpoljenosti eCRF obrazca oziroma njegovih sekcij. Uporabnik lahko v navigaciji zasledi zeleni kvadratek (npr. na gumbu »General« in gumbu »Gallery« na Sliki 14). Na ta način lahko uporabnik

ugotovi, katerih obrazcev še ni izpolnil, oziroma katere je izpolnil in jih lahko podpiše. Še več, vnosna polja v obrazcu so postavljena v zaporedne vrstice eno pod drugim. Takšna razporeditev polj omogoča bolj pregleden vpogled v izpolnjenost posamezne strani. Ozadje vrstice s poljem se spremeni v primeru, ko je polje izpolnjeno, ali ko je označeno, da podatka ni (opcija »N/A«). Uporabnik lahko hitro ugotovi, da določen podatek manjka in ga dopolni.

Omenjene funkcionalne spremembe so zahtevale bolj kompleksno programsko kodo, ki omogoča hitro dodajanje novih vrstic in sestavljanje novih obrazcev, oziroma prenos informacij o izpolnjenosti prek objektov eCRF (polje, stran, sekcija, eCRF). Ob vsaki spremembi podatkov v obrazcu se stanje obrazca znova izračuna in statusi (izpolnjen/neizpolnjen) posodobijo. Prikaz novega stanja obrazcev je takojšnji.

Z namenom izboljšave zmožnosti sistema smo ločili podatkovni od spletnega strežnika. Nova aplikacija je zaradi bolj kompleksne programske kode bolj zahtevna za strojno opremo, čemur je tudi zadoščeno s postavitvijo spletnega strežnika na zasebni računalnik.

V treh letih delovanja sistema so bile aktivnosti podatkovnega centra izjemno majhne. Dve zahtevnejši spremembi obrazcev sta bili implementirani v manj kot eni uri. Zahtev za popraviljanje podatkov v že zaklenjenih obrazcih več kot 200 preizkušancev tako rekoč ni bilo. Študija še vedno nemoteno poteka in vsi trije uporabniki na delovanje EDCp2 za zdaj nimajo nobenih pripomb. Takšni rezultati sledenja uporabe sistema so nam potrdili, da so vpeljane spremembe izboljšale uporabnost sistema.

#### **III.2.3. Predlog ključnih funkcionalnosti**

Kot je razvidno iz predloga EDC procesa, mora računalniška rešitev za zbiranje podatkov kliničnih študij (imenovana iEDC) zagotoviti podporo izvajanju številnih funkcij EDC procesa. Še več, z uporabniškega vidika mora biti podpora tem funkcijam čim bolj kakovostna. V nadaljevanju so našteje funkcionalnosti iEDC rešitve, ki v celoti zagotavljajo podporo predlaganemu EDC procesu, upoštevajoč ugotovljene potrebe uporabnikov ter rezultate uporabe prototipov.

##### *Delo s podatki (vnos, spreminjanje)*

Osnovna funkcionalnost, ki jo mora zagotoviti EDC rešitev, je delo s podatki, ki vključuje vnos podatkov in možnost spremembe podatkov. Uporabnik zadolžen za vnos podatkov mora imeti možnost vnosa različnih tipov podatkov glede na potrebe klinične študije. Polja za vnos podatkov morajo biti prilagojena za vnos različnih tipov podatkov, in sicer: numeričnih podatkov, tekstualnih podatkov, kodiranih podatkov (izbor ene ali več vrednosti s seznama) ter datumov (v poljudni ali strogo definirani obliki). EDC rešitev mora omogočati pošiljanje poizvedb v zvezi z posameznimi podatki ali celotnim obrazcem. Omogočeno mora biti tudi spreminjanje podatkov v skladu z varnostno politiko študije

(npr. podatke lahko spreminja le oseba, ki je podatke vnesla, ali tudi monitor ipd.). Spremembe podatkov so lahko omogočene le dokler podatki niso končno potrjeni in zaklenjeni.

#### Upravljanje z obrazci

EDC rešitev poleg samih podatkov v eCRF obrazcu lahko vsebuje tudi informacije o izpoljenosti obrazca oziroma različnih statusih obrazca v procesu elektronskega zbiranja podatkov. EDC rešitev mora različnim tipom uporabnikov omogočati spremembe statusov obrazcev v skladu z njihovimi vlogami v procesu. Na primer: raziskovalcu mora biti omogočeno, da posamezni obrazec označi kot izpolnjen; monitorju mora biti omogočeno, da pregledni obrazec označi kot verificiran ali zavržen; koordinatorju mora biti omogočeno, da obrazec ali skupino obrazcev označi kot zaklenjene in podobno.

#### Prikaz statusov obrazcev

V postopku zbiranja podatkov eCRF obrazci prehajajo od enega do drugega uporabnika (udeleženca v procesu). Ta različna stanja obrazca so lahko predstavljena kot statusi obrazca (npr. izpolnjen, pregledan, z napako, podpisan ipd.). Informacija o statusu obrazca mora biti razvidna uporabnikom EDC rešitve v skladu z njihovimi pravicami vpogleda v postopek zbiranja podatkov (npr. raziskovalec lahko samo pri obrazcu svojega preizkušanca vidi status, da je obrazec z napako – obrazce »tujih« preizkušancev lahko vidi samo kot dokončane ali še odprte).

#### Zgodovina sprememb vsakega podatka

Za vsak podatek shranjen v EDC rešitvi morajo biti shranjene tudi informacije o vnosu in morebitnih spremembah podatka. Informacije, ki jih EDC rešitev mora shranjevati so: kdo in kdaj je podatek vnesel ter kdo, kdaj in zakaj je podatek spremenil.

#### Avtomatično sporočanje

EDC rešitev mora uporabnike obveščati o pomembnih dogodkih v procesu zbiranja podatkov in o čakajočih nalogah. Na primer, za koordinatorja je pomemben dogodek, da je določen obrazec izpolnjen. Ta dogodek je pomemben tudi za monitorja in ob tem dogodku je treba monitorju poslati obvestilo o čakajoči nalogi (pregled izpolnjenega obrazca).

#### Reducirano upravljanje podatkov

Pri uporabi sistema za tako imenovane »nekomercialne« klinične študije, katere financirajo in izvajajo same raziskovalne ustanove, lahko interni monitoring (oz. nadzor nad potekom zbiranja podatkov) izvajajo sami raziskovalci oziroma raziskovalne sestre namesto posebej določenega monitorja. V tem primeru obrazce, ki jih izpolni en uporabnik (npr. raziskovalna sestra), drugi uporabnik (npr. druga raziskovalna sestra) lahko preverja na enak način, kot bi to počel monitor. EDC rešitev zato mora omogočati,

da koordinator študije dodeli uporabnikom iz raziskovalnega centra naloge monitoringa, zato da bi zagotovil višjo raven kakovosti zbranih podatkov.

## III.3. Tehnološki okvir rešitve

Obstaja več možnih pristopov tehnološke implementacije EDC rešitev. Vsak od teh pristopov ima svoje prednosti v določenih primerih uporabe.

Pri **»off-line«** pristopu (v industriji znan še kot **»stand-alone«**) je EDC program nameščen na vsakem računalniku za vnos podatkov. Vsak raziskovalec na svojem računalniku zbira podatke. Podatki se potem periodično pošiljajo v podatkovni center: ali fizično (recimo na zgoščenci), ali z vzpostavitvijo internetne ali telefonske povezave.

**»Thick client«** pristop je podoben **»off-line«** pristopu. EDC program je nameščen na vsakem računalniku posebej, s tem da lahko deluje v off-line ali v on-line načinu – torej, lahko je povezan na podatkovni center ali pa ne.

**»Thin client«** pristop ali spletna rešitev predvideva, da je vsa EDC programska oprema nameščena na centralnem strežniku v podatkovnem centru. Vsi uporabniki do EDC rešitve dostopajo prek Interneta s pomočjo spletnega brskalnika.

**»Hibridni«** pristop je lahko katerakoli kombinacija **»thick client«** in **»thin client«** pristopa. Na primer, hibridni pristop je lahko rešitev, pri kateri je EDC program naložen na prenosnik, podatki pa se shranjujejo v centralno bazo podatkov.

Kot najprimernejši pristop za implementacijo računalniške rešitve za podporo procesa zbiranja podatkov klinične študije, kot je predstavljen v Metodah v poglavju **»II.2 Določanje funkcionalnega okvirja«**, smo izbrali **»thin client«** pristop oziroma spletno rešitev. Tehnološke značilnosti predlagane rešitve so naslednje:

- EDC programska oprema na spletnem strežniku;
- centralna baza podatkov na podatkovnem strežniku;
- na strani uporabnika navaden spletni brskalnik;
- avtentikacija in avtorizacija uporabnikov z uporabniškim imenom in geslom;
- varna povezava med brskalnikom in spletnim strežnikom na osnovi Secure Socket Layer (SSL) protokola.

Ključni razlogi za izbor spletne rešitve so enostavnost in posledično nizka cena vzdrževanja takšnega sistema, centraliziranost podatkov, hitra penetracija interneta na vseh koncih sveta ter navajenost uporabnikov na uporabo spletnega brskalnika, kar olajšuje sprejem uporabe nove tehnologije.

Slabost, katere smo se zavedali ob izboru spletne rešitve, je pred vsem dejstvo, da spletni vmesnik omogoča manj fleksibilnosti v pogledu zasnove uporabniškega vmesnika kot

navadna aplikacija. Zmogljivosti spletnih aplikacij so tudi slabše od navadnih aplikacij zaradi omejitev, ki jih postavlja spletni brskalnik. Kakorkoli, nobena od teh slabosti nam ni bila ključnega pomena, saj bistveno ne vpliva na kvaliteto končne rešitve, kar se je izkazalo tudi prek uporabe prototipov.

## III.4. Pilotni sistem - iEDC

### III.4.1. Model elektronskega CRF obrazca - eCRF

#### *Opis strukturnih elementov*

Elektronski CRF obrazci (eCRF) so zasnovani tako, da je mogoče sledenje statusom posameznih segmentov eCRF. Celoten eCRF vsake študije je sestavljen iz hierarhičnih strukturnih elementov tipa mapa (*folder*), stran (*page*) in podatek (*ITEM*).

**ITEM** je logična skupina polj za vnos in prikaz podatkov (npr. podatek o višini, ki je le eno polje, ali pa datum obiska, ki je sestavljen iz treh vnosnih polj – dan, mesec in leto). Vsak **ITEM** ima lahko status, ki je odvisen od samodejnega stanja polja (npr. izpolnjeno/neizpolnjeno), ali od strani uporabnika določenega stanja polja (npr. napačno/pravilno izpolnjeno). Poleg tega se za vsak **ITEM** shranjuje zgodovina sprememb. Shranjujejo se podatki: kdo, kdaj in zakaj je spremenil podatek, ter stara in nova vrednost podatka.

Stran (*page*) je strukturni element, ki vsebuje **ITEM**-e. Stran deduje statuse **ITEM**-ov, ki jih vsebuje. Spremembe podatkov se shranjujejo na nivoju strani, kar pomeni, da se v bazo podatkov vedno shranjuje cela stran podatkov. Obstajata dva tipa strani: **PAGE** in **MULTIPAGE**.

**PAGE** je strukturni element, ki vsebuje enega ali več **ITEM**-ov, in predstavlja eno stran obrazca CRF. Vsak **PAGE** ima točno določeno strukturo.

**MULTIPAGE** je strukturni element podoben elementu **PAGE**. Razlika je le v tem, da je **MULTIPAGE** lahko večkrat ponovljen. Število ponovitev je lahko vnaprej določeno ali odvisno od vrednosti nekaterih podatkov v eCRF. En primer uporabe **MULTIPAGE**-a bi lahko bila po ena enaka stran podatkov za vsakega od petih bolnikovih tumorjev (lezij).

Višji strukturni element od strani (*page*) je mapa (*folder*). Obstajajo štiri tipi map:

- **CONTAINER FOLDER;**
- **MULTIFOLDER;**
- **DYNAMIC FOLDER;**
- **STATIC FOLDER.**

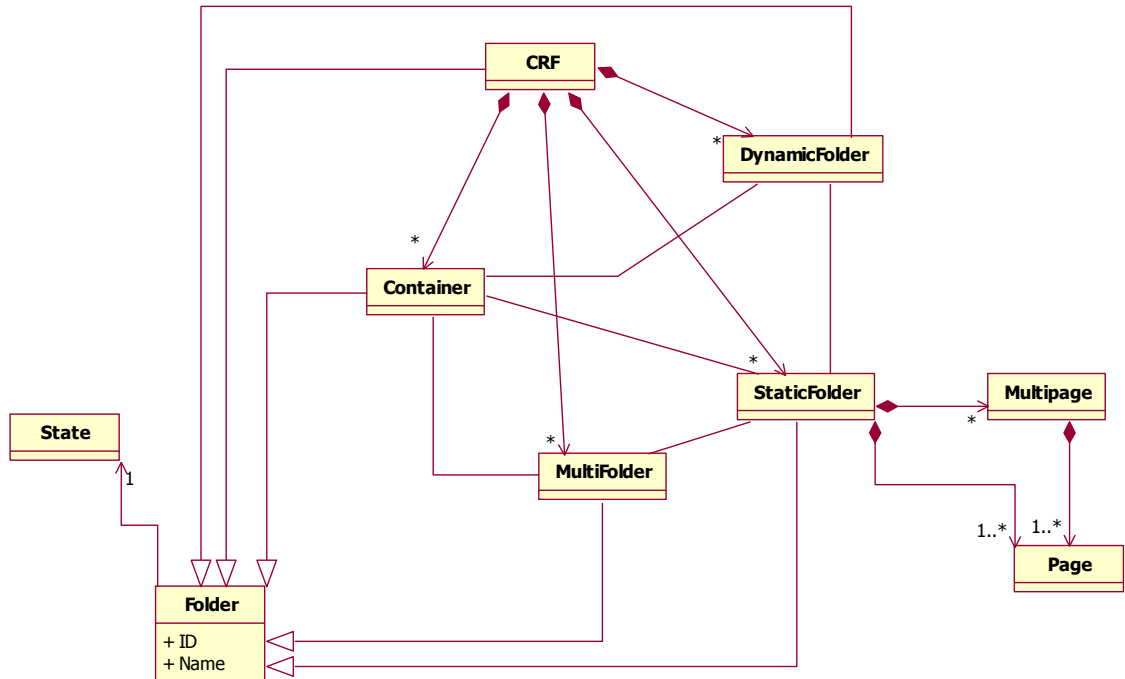
**CONTAINER FOLDER** je mapa, ki lahko vsebuje le mape (na pa tudi strani) kateregakoli tipa. En primer take mape je sam eCRF. Statusa *CONTAINER FOLDER*-ja ni mogoče neposredno spremeniti. *CONTAINER FOLDER* deduje statuse map, ki jih vsebuje. Na primer, v primeru, da so vse mape znotraj enega *CONTAINER FOLDER*-ja verificirane (pregledane in potrjene s strani monitorja), ima tudi sam status verificiranega.

**MULTIFOLDER** je mapa, ki vsebuje **določeno** število enakih, vnaprej določenih map, kateregakoli tipa. Število map znotraj *MULTIFOLDER*-ja je lahko vnaprej določeno ali odvisno od vrednosti določenega podatka v eCRF. Tako kot *CONTAINER FOLDER*, tudi *MULTIFOLDER* deduje statuse map, ki jih vsebuje.

**DYNAMIC FOLDER** je mapa, ki vsebuje **nedoločeno** število enakih, vnaprej določenih map, kateregakoli tipa. Uporabnik, ki vnaša podatke, lahko dodaja nove instance map znotraj *DYNAMIC FOLDER*-ja. Tako kot *CONTAINER FOLDER* tudi *DYNAMIC FOLDER* deduje statuse map, ki jih vsebuje. En primer take mape bi lahko bila mapa »OBISKI«, v kateri uporabnik ob vsakem novem obisku bolnika dodaja novo mapo »OBISK št. X« z enako strukturo.

**STATIC FOLDER** vsebuje vnaprej določeno množico strani, v katere uporabniki vnašajo podatke. Strani so lahko neponavljajoče (*PAGE*) ali ponavljajoče (*MULTIPAGE*). Status *STATIC FOLDER*-ja lahko postavi uporabnik (npr. »verificiran«), lahko pa je odvisen od statusov vseh strani, ki jih vsebuje (npr.: če imajo vse strani znotraj mape tipa *STATIC FOLDER* status »izpolnjena«, potem ima tudi mapa status »izpolnjena«).

Zgoraj opisana hierarhija strukturnih elementov eCRF je prikazana na UML razrednem diagramu na Sliki 15.



Slika 15. Hierarhija strukturnih elementov

Za vsako študijo je treba določiti predlogo (angl. »template«) eCRF obrazca, ki je sestavljena iz predlog map in strani organiziranih v določeno strukturo. Za vsakega obravnavanega bolnika je treba na osnovi ustrezne predloge eCRF kreirati njegov lasten eCRF obrazec.

### *Specifične značilnosti strukturnih elementov*

Značilnosti *ITEM*-a so njegov enotni ID, ime, ID strani, kateri pripada, ter vrstni red prikaza na strani. *ITEM* vsebuje še informacijo o tem, ali podatek iz kateregakoli razloga ni na razpolago (označena opcija »N/A«), ter morebitni komentar raziskovalca, v katerem pojasni posredovan podatek. *ITEM* vsebuje še informacijo o tem, ali ga je monitor označil za pravilnega ali nepravilnega. V primeru, da je podatek označen kot nepravilen, potem mora *ITEM* obvezno vsebovati še sporočilo monitorja (ALERT). Enako v primeru, da koordinator dvomi v vrednost podatka, potem *ITEM* vsebuje še poizvedbo (WARNING/QUERY).

Poleg tega vsak *ITEM* v enem tekstualnem polju hrani tudi zgodovino sprememb. Ta zgodovina sprememb je shranjena v XML obliki, kot je opisana v APPENDIX INFO XML. Z vsako spremembo kateregakoli od parametrov *ITEM*-a (označitev podatka kot nepravilnega, sprememba vrednosti podatka itn.), se v INFO XML dopiše nov element (angl. »node«), ki vsebuje informacije o trenutnem stanju *ITEM*-a.

Značilnosti *PAGE*-a so njegov enoten ID, ime in ID predloge, na osnovi katere je kreiran, pot na vir te strani ter vrstni red strani v prikazu v mapi. Poleg tega *PAGE* vsebuje še informacijo o tem, v katerem stanju se nahaja (status) ter koliko *ITEM*-ov vsebuje.

Za vse tipe map (*CONTAINER FOLDER*, *MULTIFOLDER*, *DYNAMIC FOLDER*, *STATIC FOLDER* in *MULTIPAGE*) kot tudi sam eCRF lahko opazimo številne skupne značilnosti, ki nam dovoljujejo, da vse te elemente skupaj obravnavamo kot generalni element *FOLDER*. Značilnosti tega elementa so: njegov ID, ime, tip, tip in ID elementa, znotraj katerega se nahaja (nadrejenega elementa – angl. »parent«), vrsti red prikaza znotraj nadrejenega elementa, realno stanje (na osnovi katerega se uporabniku prikazuje ustrezen status elementa) ter ID predloge *FOLDER*-ja, na osnovi katere je primerek (angl. »instance«) *FOLDER*-ja kreiran.

Poleg omenjenih skupnih značilnosti vseh map lahko za posamezne tipe map identificiramo določene specifične značilnosti, in sicer:

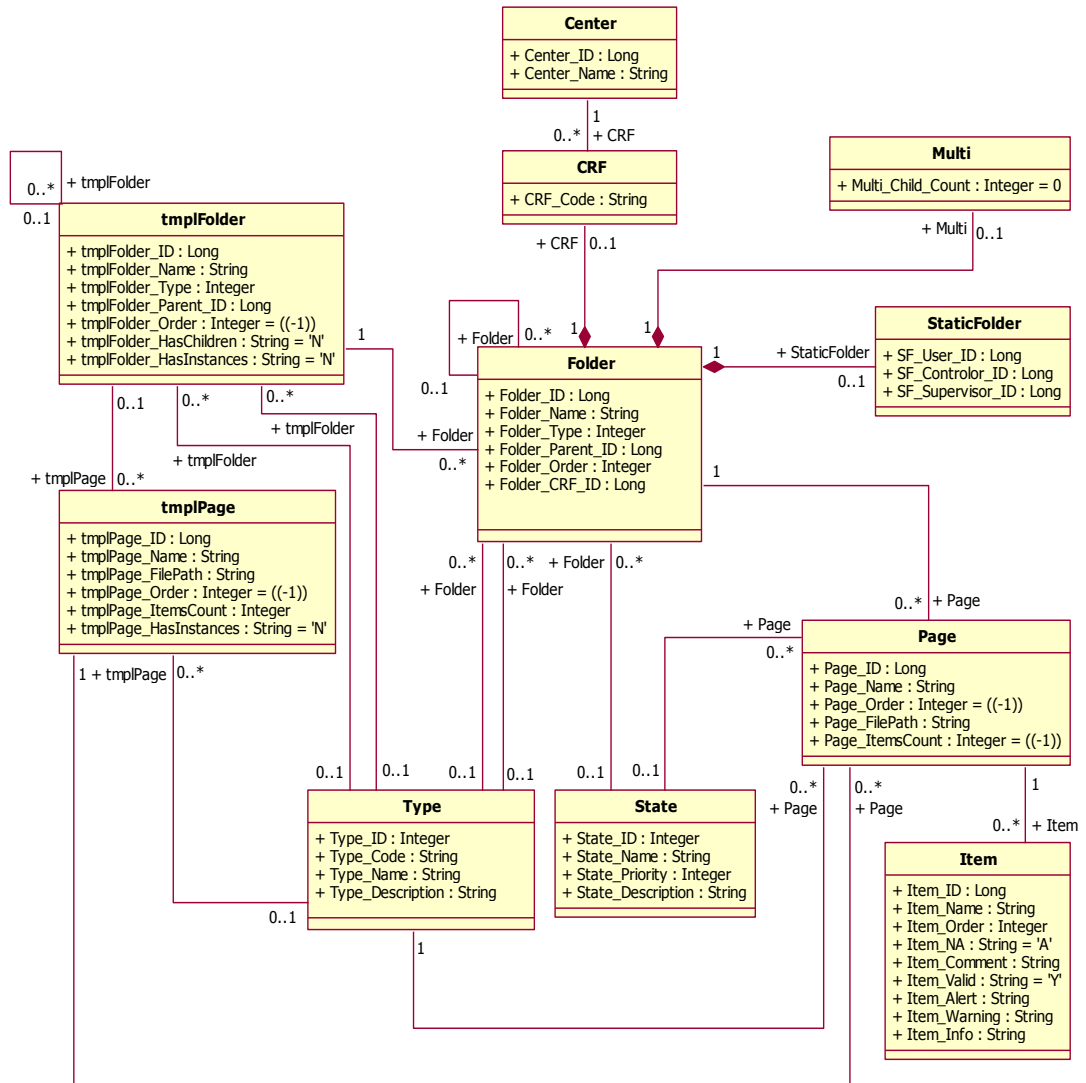
Za **eCRF** (celoten obrazec) sta značilna še ID raziskovalnega centra, kateremu pripada, in *CRF\_Code*, ki je specifičen za vsakega bolnika;

Za **MULTIFOLDER** in **MULTIPAGE** je značilno še število ponovitev ustreznega elementa (mape ali strani).

Za **STATIC FOLDER** je značilna še informacija o tem, kdo je dokončno izpolnil mapo. Poleg tega *STATIC FOLDER* vsebuje tudi informacijo o tem, kdo preverja (verificira) podatke potem, ko je mapa izpolnjena.

Izhajajoč iz opisov strukturnih elementov in opaženih skupnih lastnosti elementov z generalizacijo, smo prišli do modela podatkov, kot je prikazan na Sliki 16.





Slika 16. Analitični razredni diagram elektronskega obrazca eCRF

### Postavitev strukture eCRF za posamezno študijo

Za postavitev nove študije na iEDC sistemu moramo določiti strukturo eCRF. Za določanje strukture eCRF uporabljamo administracijski modul (Slika 17), ki administratorju študije (koordinatorju ali komu drugemu) omogoča enostavno gradnjo predloge eCRF.

Slika 17. Primer uporabe administracijskega modula za postavitve nove študije

Administrator predlogo eCRF gradi z dodajanjem gradnikov (strukturnih elementov) in določanjem njihovih parametrov. Struktura se shrani v bazi podatkov. Administracijski modul omogoča dodajanje gradnikov do nivoja strani in števila *ITEM*-ov na njej. Posamezne *ITEM*-e je potem potrebno ročno dodati v predloge posameznih strani. Poleg prikaza *ITEM*-ov vsebujejo te predloge tudi funkcije kontrole vrednosti posameznih polj kot tudi povezave med vnosnimi polji in bazo študijskih podatkov (kamor se shrani podatki). Za vsako novo študijo je treba postaviti tudi primerno bazo študijskih podatkov.

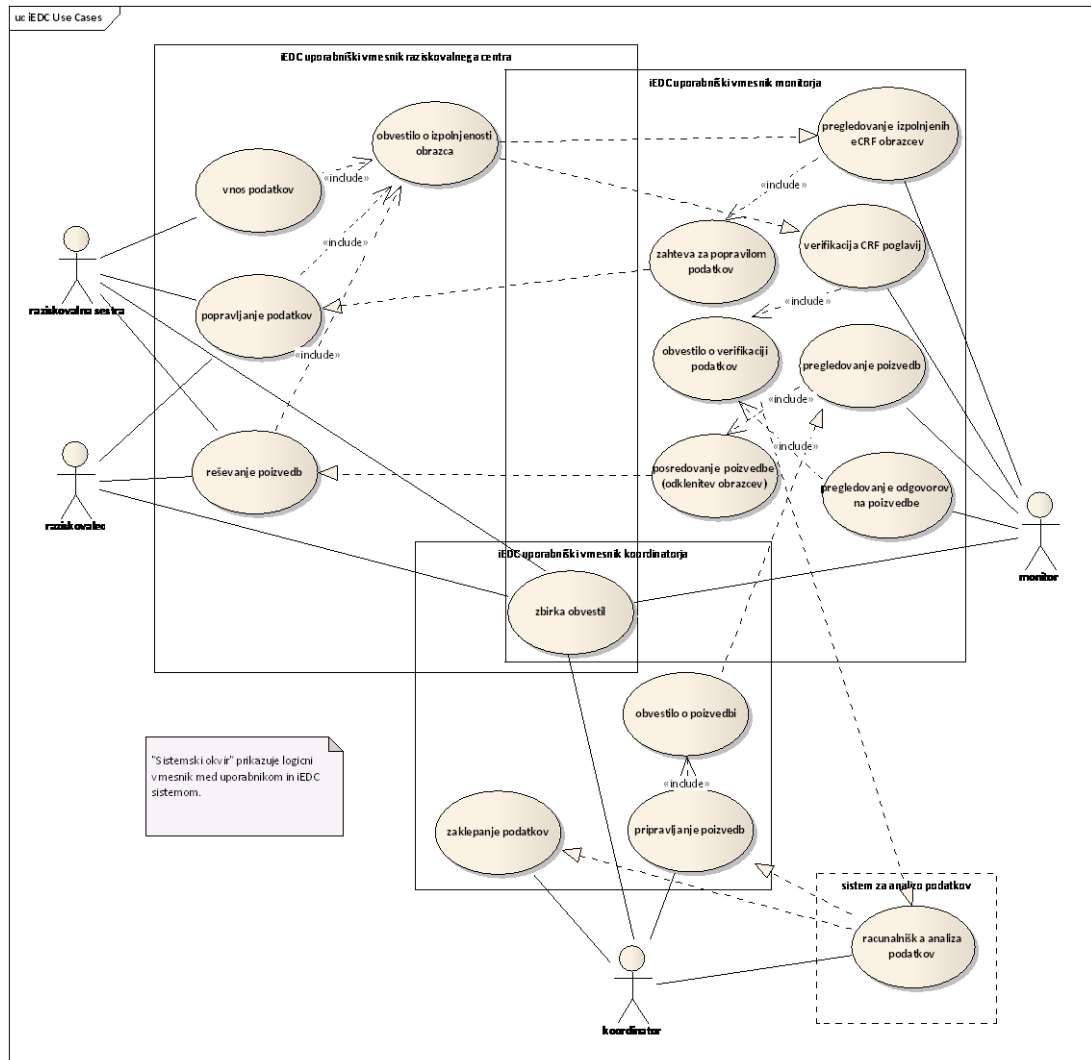
#### ***Kreiranje novega primerka eCRF***

Ko uporabnik v sistem želi dodati eCRF za novega bolnika na osnovi vnaprej določene predloge eCRF za konkretno študijo, se kreira nov primerek (angl. »instance«) eCRF. Na osnovi predlog posameznih map eCRF se kreirajo ustrezne mape za konkretnega preizkušanca in se jim ob tem dodelijo vrednosti vseh parametrov (enotni ID, ime, ID nadrejenega elementa, vrstni red prikaza itn.) razen tistih, ki se med študijo spreminjajo (npr. stanje, zgodovina sprememb itn.).

Za vsak kreiran primerki strani se na enak način nastavijo vrednosti parametrov. Poleg tega se avtomatično kreirajo še primerki vseh *ITEM*-ov, ki so navedeni v predlogi strani. Primerki *ITEM*-ov vsebujejo le enotni ID, vrstni red prikaza na strani, ID strani, kateri pripadajo, ter ime *ITEM*-a. Parametri, ki se nanašajo na pravilnost podatka, komentar raziskovalca itn., se med samim zbiranjem podatkov sproti nastavljajo in spreminjajo glede na aktivnosti uporabnikov in njihove vloge v sistemu.

### III.4.2. Primeri uporabe (use-cases)

Pri zasnovi določanja vlog v pilotnem sistemu iEDC (v nadaljevanju »sistem«) smo izhajali iz eEPC modelov EDC procesa. Na osnovi omenjenih modelov smo naredili diagram primerov uporabe sistema (angl. »use case diagram«; Slika 18).



Slika 18. Diagram primerov uporabe sistema

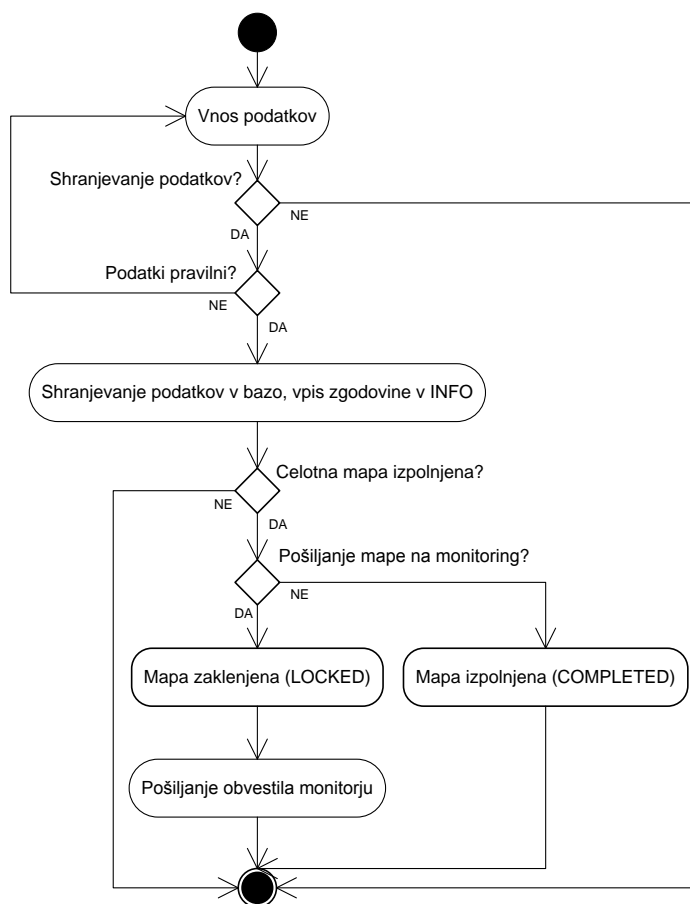
Sistem vsebuje tri različne uporabniške vmesnike; za vsakega od treh skupin uporabnikov svojega:

- uporabniški vmesnik raziskovalnega centra;
- uporabniški vmesnik monitorja;
- uporabniški vmesnik koordinatorja.

Uporabnik dostopa do sistema z uporabniškim imenom in geslom. Glede na tip uporabnika sistem prikaže ustrezen uporabniški vmesnik.

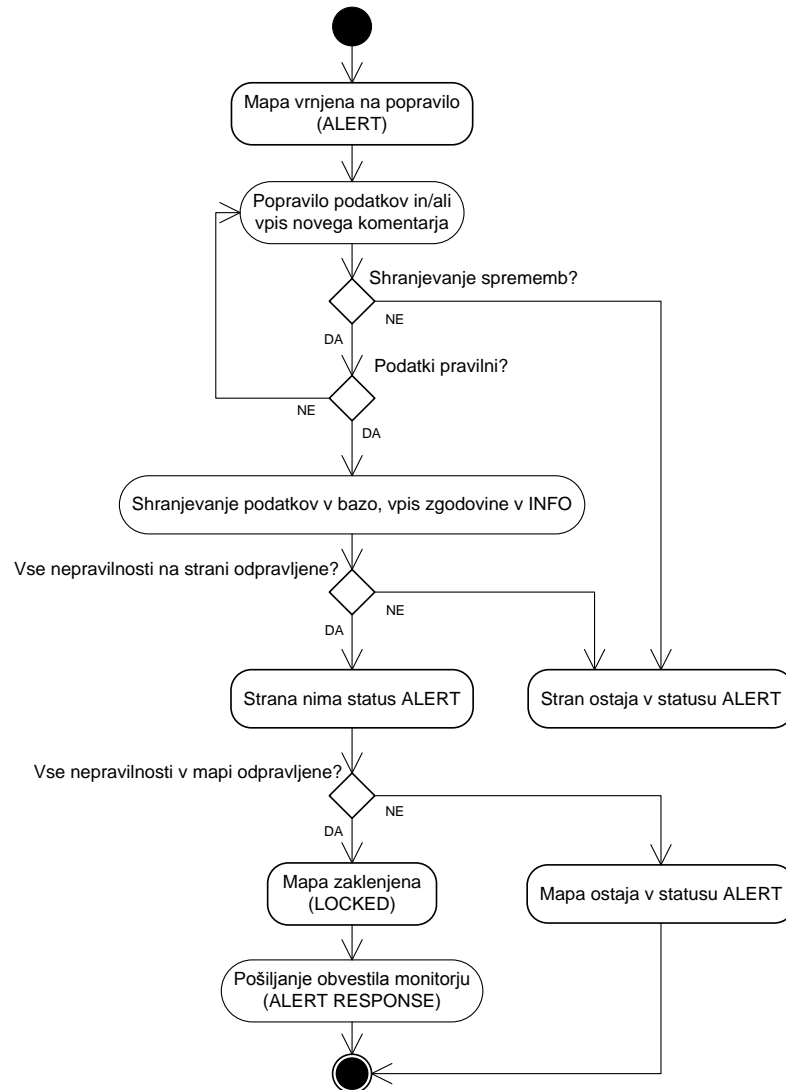
*Raziskovalni center*

Iz eEPC modelov EDC procesa je razvidno, da na strani raziskovalnega centra obstajata dve vlogi: raziskovalec in raziskovalna sestra. V procesu zbiranja podatkov ju loči le to, da raziskovalec podatkov ne vnaša v elektronske obrazce, ampak to namesto njega počne raziskovalna sestra. Pogosto sta zaradi pomanjkanja raziskovalnega osebja ti dve vlogi združeni v eno, ki jo v celoti izvaja raziskovalec. Zato smo se odločili za enoten uporabniški vmesnik za obe vlogi: raziskovalca in raziskovalno sestro. Takšen uporabniški vmesnik omogoča izvajanje vseh EDC funkcij raziskovanega centra, in sicer: vnos podatkov, popravljanje podatkov in odgovarjanje na poizvedbe koordinatorja (reševanje poizvedb). Diagram aktivnosti ob vnosu podatkov v iEDC je prikazan na Sliki 19.



Slika 19. Diagram aktivnosti ob vnosu podatkov v iEDC

Ko uporabnik v celoti izpolni obrazec in ga označi kot izpolnjen, sistem samodejno pošlje obvestilo monitorju o izpolnjenosti obrazca. V primeru, da monitor po pregledu podatkov vrne obrazec raziskovalcu/raziskovalni sestri, mora ta označene podatke popraviti in spet označiti obrazec kot izpolnjen (Slika 20). Enako v primeru, da monitor posreduje poizvedbe o podatkih: raziskovalec/raziskovalna sestra mora na te poizvedbe odgovoriti in spet označiti obrazec kot izpolnjen. Poleg omenjenih nalog ima raziskovalec/raziskovalna sestra možnost upravljanja svoje zbirke obvestil.

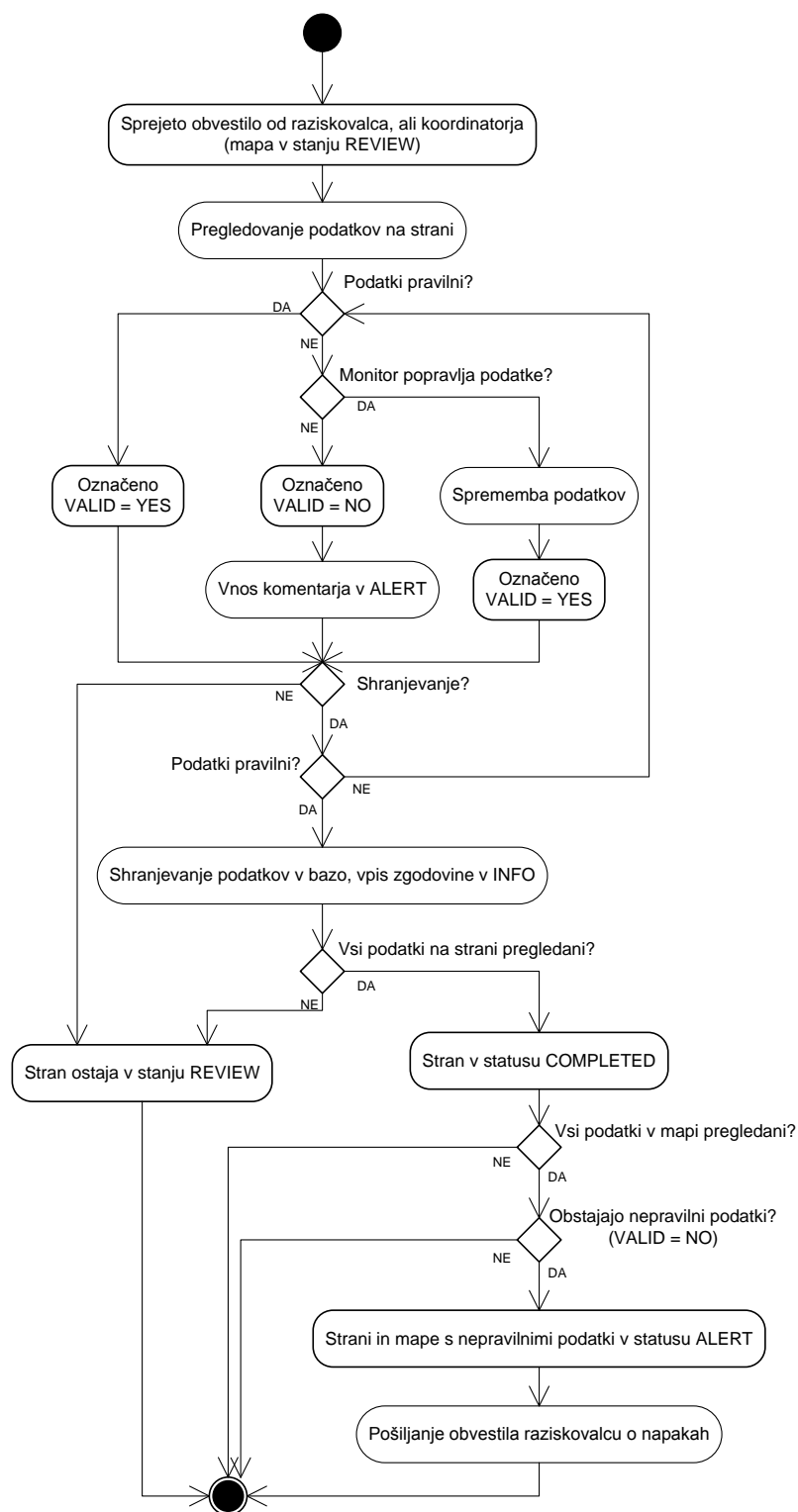


Slika 20. Diagram aktivnosti ob popravljanju podatkov v iEDC

### Monitor

Osnovna funkcionalnost uporabniškega vmesnika monitorja je pregledovanje izpolnjenih eCRF obrazcev. Vmesnik omogoča, da monitor pregleduje izpolnjene eCRF obrazce in jih po potrebi skupaj s svojimi opombami vrne raziskovalcu/raziskovalni sestri v popravilo podatkov. V primeru, da je obrazec pravilno izpolnjen, ga monitor lahko verificira in o tem obvesti koordinatorja. Slika 21 prikazuje diagram aktivnosti monitorja ob pregledovanju podatkov v iEDC.

Poleg tega monitor od koordinatorja prejema poizvedbe za določene obrazce (podatke) in jih posreduje odgovornim raziskovalcem v reševanje. Obrazce z odgovori na poizvedbe dobi nazaj od raziskovalcev in jih spet pregleda ter verificira. Na ta način so vsi udeleženci v procesu obveščeni o morebitnih spremembah pri podatkih. Poleg omenjenih nalog ima monitor možnost upravljanja svoje zbirke obvestil.

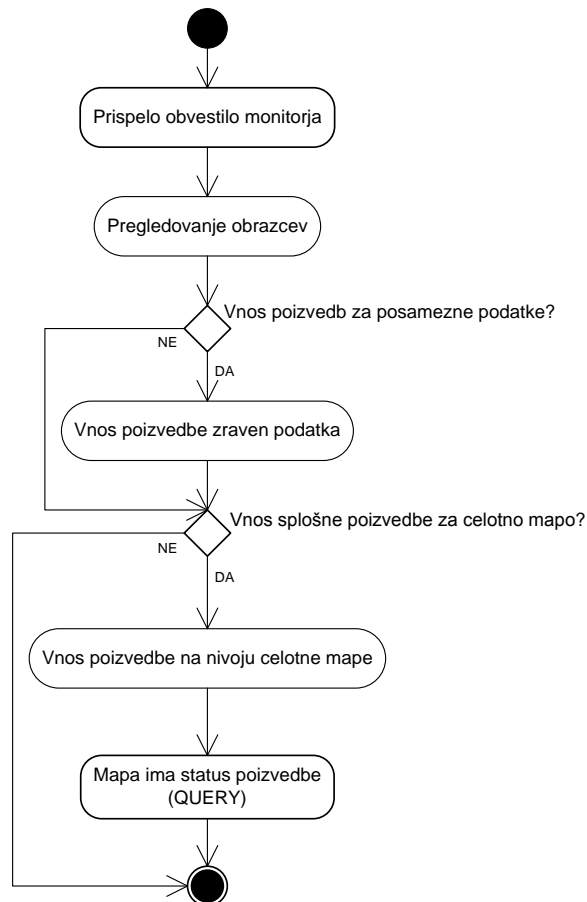


Slika 21. Diagram aktivnosti ob monitoringu podatkov v iEDC

V primeru, da študija nima posebej določenih monitorjev (npr. »nekomercialna« klinična študija), lahko koordinator znotraj skupine uporabnikov raziskovalnega centra določi, kateri uporabnik katerega nadzoruje (t.i. »internega monitorja«). V tem primeru ima lahko en ali več uporabnikov podvojeno vlogo: vnaša podatke in tudi preverja obrazce, ki jih izpolni drugi uporabnik.

### *Koordinator*

Koordinatorjev uporabniški vmesnik omogoča izvajanje dveh osnovnih nalog koordinatorja, in sicer: pošiljanje poizvedb in zaklepanje podatkov. Koordinator lahko na že izpolnjenem in verificiranem eCRF obrazcu dopiše poizvedbe o podatkih ter o tem obvesti monitorja (Slika 22). V primeru, da po analizi podatkov ni ugotovljenih nobenih napak med podatki, lahko koordinator obrazce oziroma podatke v njih dokončno zaklene. Seveda lahko tudi koordinator upravlja svojo zbirko obvestil.



Slika 22. Diagram aktivnosti ob postavljanju poizvedb v iEDC

V primeru študije brez monitorja lahko koordinator za vsakega uporabnika v raziskovalnem centru določi enega od ostalih uporabnikov za »internega monitorja«. To lahko koordinator izvede na začetku študije za vse uporabnike ali med samim potekom študije, tako da ob vsakem prispelem obvestilu o izpolnjenem obrazcu dodeli »internega monitorja« za ta obrazec, upoštevajoč obremenjenost posameznih uporabnikov v raziskovalnem centru.

### III.4.3. Model upravljanja z obrazci

Uporabniki vsak na svoj opisan način v primerih uporabe sodelujejo pri zbiranju podatkov. Ob tem elektronski obrazci (eCRF) za vsakega preizkušanca oziroma njihovi strukturni elementi (mape, strani, podatki) spreminjajo stanja in statuse glede na aktivnosti

uporabnikov. Stanje elementa vpliva na to, kaj lahko uporabnik počne s tem elementom. Obstajajo tri mogoča stanja elementa, in sicer: *urejanje* (angl. »edit« - E), *zaklenjen* (angl. »locked« - L), ali *pregledovanje* (angl. »review« - R). Status elementa uporabniku pove, v kateri fazi postopka zbiranja podatkov je posamezni element. Možni statusi so: *izpolnjen* (angl. »completed« - C), *opozorilo* (angl. »alert« - A), *poizvedba* (angl. »query« - Q), ali *zaključen* (angl. »terminated« - T). Obnašanje posameznih strukturnih elementov je opisano v nadaljevanju.

*ITEM* je izpolnjen takrat, ko je podatek v njem izpolnjen, ali ko je označeno, da je podatek manjkajoč (izbrana opcija »N/A«) ter dopisan komentar (pojasnilo), zakaj podatka ni. Ko monitor pregleda podatek, označi pravilnost podatka (pravilen/nepravilen). V primeru, da koordinator dvomi v pravilnost že verificiranega podatka, mu lahko pripiše poizvedbo (angl. »query«).

Element *PAGE* deduje status *ITEM*-ov, ki jih vsebuje, na naslednji način: stran je izpolnjena, če so vsi *ITEM*-i izpolnjeni; stran je nepravilna, če je katerikoli od njenih *ITEM*-ov nepravilen; stran je s poizvedbo, če na kateremkoli *ITEM*-u obstaja poizvedba (angl. »query«). Možna kombinacija stanj in statusov elementa *PAGE* je prikazana v Tabeli 4.



### III. Rezultati

Tabela 4. Možna stanja in statusi elementa PAGE (stanja: E - urejanje, L - zaklenjen, R - pregledovanje; statusi: C - izpolnjen, A - opozorilo, Q - poizvedba, T - zaključen)

ID	opis faze zbiranja podatkov	raziskovalec		podgovorni raziskovalec		monitor		koordinator	
		stanje	stat	stanje	status	stanje	status	stanje	status
P0	stran je v fazi izpolnjevanja	E	L	E	L	E	L	E	L
P1	stran je izpolnjena	E	C						
P2	stran pripada mapi, ki čaka na koordinirano določitev monitorja	L	C	L					
P3	stran posredovana monitorju ni pregledana in nima napak	L		L		R			
P4	stran posredovana monitorju ni pregledana in ima napake	L		L		R	a		
P5	stran posredovana monitorju je pregledana in nima napak	L		L		R	C	a	
P6	stran posredovana monitorju je pregledana in ima napake	L		L		R	C	a	
P7	stran vnijena raziskovalcu ne vsebuje napak	L		L	C				
P8	stran vnijena raziskovalcu vsebuje napake, ki niso vse popravljene	L	E	E	A			a	
P9	stran vnijena raziskovalcu vsebuje napake, ki so vse popravljene	L	E	E	a				
P10	stran je verificirana in nima poizvedb	L		L		L		R	
P11	stran je verificirana in nima poizvedb	L		L		L		R	Q
P12	stran je vnijena monitorju in nima poizvedb	L		L		L	C		
P13	stran je vnijena monitorju, ima poizvedbe, nije pregledana in nima napak	L		L		R		Q	L
P14	stran je vnijena monitorju, ima poizvedbe, nije pregledana in ima napake	L		L		R		a	Q
P15	stran je vnijena monitorju, ima poizvedbe, je pregledana in nima napak	L		L		R	C	q	L
P16	stran je vnijena monitorju, ima poizvedbe, je pregledana in ima napake	L		L		R	C	a	q
P17	stran vnijena monitorju je vnijena raziskovalcu, ima poizvedbe in nima napak	L		L	C			q	L
P18	stran vnijena monitorju je vnijena raziskovalcu, ima poizvedbe in ima napake, ki niso vse popravljene	L	E	E	A			a	q
P19	stran vnijena monitorju je vnijena raziskovalcu, ima poizvedbe in ima napake, ki so vse popravljene	L	E	E	a				q
P20	stran, ki je imela poizvedbe je verificirana in nima novih poizvedb	L		L		L		R	
P21	stran, ki je imela poizvedbe je verificirana in ima nove poizvedbe	L		L		L		R	q
P22	stran je zaključena	L		L		L			Q

Mape (element *FOLDER*) dedujejo stanja in statuse podrejenih (angl. »child«) elementov.

*STATIC FOLDER* je izpolnjen takrat, ko so vse strani, ki jih vsebuje, izpolnjene. Ko sistem zazna, da so vse strani znotraj enega *STATIC FOLDER*-ja (mape) izpolnjene, uporabniku (raziskovalec/raziskovalna sestra) ponudi možnost, da mapo označi kot »izpolnjeno«, in s tem avtomatsko obvesti monitorja, da je mapa pripravljena za verifikacijo podatkov. V tem primeru se mapa zaklene za nadaljnje spremembe podatkov in med parametre mape se vpiše podatek o tem, kateri uporabnik je mapo poslal na verifikacijo in kdo bo mapo pregledal. Informacija o tem, kdo bo mapo pregledal je vezana na posameznega uporabnika, ki podatke vnaša (vsak raziskovalec/raziskovalna sestra ima vnaprej določenega monitorja). Vsaka nadaljnja sprememba podatkov v obrazcu je v primeru vrnitve mape na popravilo podatkov dovoljena le tistemu uporabniku, ki je podatke poslal na verifikacijo. V Tabeli 5 so prikazana možna stanja in statusi map (element *FOLDER*).

### III. Rezultati

Tabela 5. Možna stanja in statusi elementa *FOLDER* (stanja: E - urejanje, L - zaklenjen, R - pregledovanje; statusi: C - izpolnjen, A - opozorilo, Q - poizvedba, T - zaključen)

ID	faza zbiranja podatkov	raziskovalec			odgovorni raziskovalec			monitor			koordinator		
		stanje	status	stanje	status	stanje	status	stanje	status	stanje	status		
F0	mapa je v fazi izpolnjevanja	E	L	C	E	L	C	A	T	E	L	Q	T
F1	mapa je izpolnjena	E		C									
F2	mapa čaka na koordinatorjevo določitev monitorja	L			L								
F3	mapa posredovana monitorju ni pregledana in nima napak	L			L				R				
F4	mapa posredovana monitorju ni pregledana in ima napake	L			L				R	a			
F5	mapa posredovana monitorju je pregledana in nima napak	L			L				R	C			
F6	mapa posredovana monitorju je pregledana in ima napake	L			L				R	C	a		
F7	mapa vrnjena raziskovalcu ni popravljena	L			E		A		L	a			
F8	mapa vrnjena raziskovalcu je popravljena	L			E		C	a	L	a			
F9	mapa je verificirana in nima poizvedb	L			T		T		L		T	R	T
F10	mapa je verificirana in ima poizvedbe	L			L		T		L		T	R	Q
F11	mapa vrnjena monitorju ni pregledana in nima napak	L			L				R		Q		q
F12	mapa vrnjena monitorju ni pregledana in ima napake	L			L				R	a	Q		q
F13	mapa vrnjena monitorju je pregledana in nima napak	L			L				R	C	q		q
F14	mapa vrnjena monitorju je pregledana in ima napake	L			L				R	C	a	q	q
F15	mapa vrnjena monitorju je vrnjena operaterju in ni popravljena	L			E		A		L	a	q		q
F16	mapa vrnjena monitorju je vrnjena operaterju in je popravljena	L			E		C	a	L	a	q		q
F17	mapa, ki je imela poizvedbe je verificirana in nima novih poizvedb	L			T		T		L		T	R	q
F18	mapa, ki je imela poizvedbe je verificirana in ima nove poizvedbe	L			T		T		L		T	R	Q
F19	mapa je zaključena	L			L				L			L	T

Mapa deduje stanja podrejenih (angl. »child«) elementov na naslednji način (E - *urejanje*; R - *pregledovanje*; L - *zaklenjen*): če je katerikoli od podrejenih elementov v stanju R, potem je R; v nasprotnem primeru, če je katerikoli od podrejenih elementov v stanju E, potem je E, drugače je L. V nadaljevanju so prikazana pravila za dedovanje statusov map.

Tabela 6. Pravila po katerih mape dedujejo statuse podrejenih (angl. »child«) elementov (C – *izpolnjen*; A – *opozorilo*; Q – *poizvedba*; T - *zaključen*)

ID	Pravilo
a.	<i>STATIC FOLDER</i> ima status C, če so vsi otroci C
b.	vsii ostali <i>FOLDER</i> -ji ( <i>CONTAINER FOLDER</i> , <i>MULTIFOLDER</i> , <i>DYNAMIC FOLDER</i> ) imajo status C, če je katerikoli od otrokov C
c.	<i>FOLDER</i> ima status A, če je katerikoli od otrokov A
d.	<i>FOLDER</i> ima status Q, če je katerikoli od otrokov Q
e.	<i>FOLDER</i> ima status T, če so vsi otroci T
f.	če ima katerikoli od otrokov v mapi izpostavljen status (velika črka v tabeli statusov), potem tudi mapa deduje <i>izpostavljen</i> status; drugače je <i>neizpostavljen</i>

#### III.4.4. Opis strukture sporočil v sistemu

V sistemu obstaja šest tipov sporočil, ki jih sistem avtomatično pošilja ustreznim uporabnikom ob določenih aktivnostih drugih uporabnikov. Vsako sporočilo ima enako strukturo, ki vsebuje ime uporabnika, ki sporočilo pošilja, ime uporabnika, ki sporočilo sprejema, tip sporočila, datum in čas, ko je sporočilo poslano, ter povezavo na mapo ali stran eCRF, na katero se sporočilo nanaša.

Tabela 7. Opis tipov sporočil, ki jih sistem avtomatski pošilja uporabnikom (<sup>1</sup> V primeru, ko *koordinator* določa *monitorja* (po sprejemu sporočila *MAPA IZPOLNJENA* od *raziskovalca*), takrat je *koordinator* za *monitorja* pošiljatelj; <sup>2</sup> Sporočilo, da je mapa izpolnjena, se pošilja *koordinatorju* v primeru, ko *koordinator* določa *monitorja* za vsako mapo posebej ob izpolnitvi mape.)

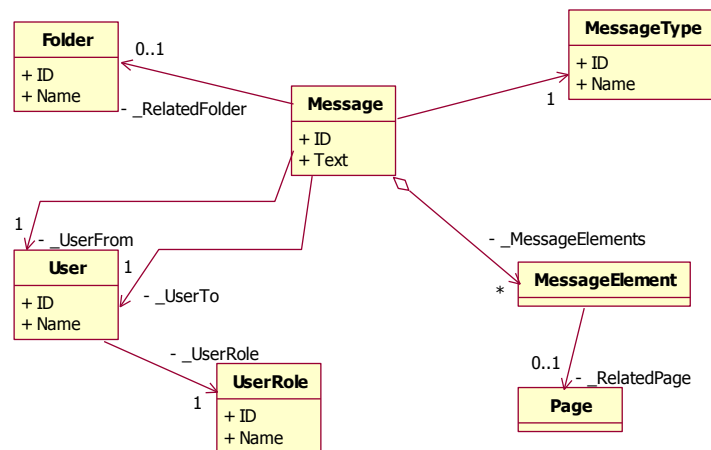
Tip sporočila	Pošiljatelj	Sprejemnik	Vsebina sporočila	Datum in čas
<b>MAPA IZPOLNJENA</b>	raziskovalec ali monitor <sup>1</sup>	kontrolor ali monitor <sup>2</sup>	povezava na mapo	kdaj je poslano
<b>OPOZORILO</b>	monitor	uporabnik	povezave na strani z napakami	kdaj je poslano
<b>ODGOVOR NA OPOZORILO</b>	raziskovalec	monitor	povezava na mapo	kdaj je poslano
<b>MAPA VERIFICIRANA</b>	monitor	koordinator	povezava na mapo	kdaj je poslano
<b>POIZVEDBA</b>	koordinator	monitor	generalna pripomba s povezavo na mapo ali povezave na strani	kdaj je poslano
<b>ODGOVOR NA POIZVEDBO</b>	monitor	koordinator	povezave na strani z odgovori na poizvedbe	kdaj je poslano

Sporočila so pri vsakem uporabniku razvrščena v dve skupini: aktivna sporočila in arhiv sporočil. Uporabnik lahko sporočilo sam prestavi iz aktivnih med arhivska. V določenih primerih sistem avtomatično prestavi aktivno sporočilo med arhivska, kot je prikazano v Tabeli 8.

Tabela 8. Pravila za avtomatsko prestavljanje sporočil iz aktivnih v arhivska

Tip sporočila	Sprejemnik	Aktivnost, ki sproži avtomatično arhiviranje sporočila
<b>MAPA IZPOLNJENA</b>	monitor	po verifikaciji mape, ali po vrnitvi mape na popravilo
<b>MAPA IZPOLNJENA</b>	koordinator	po izbiri monitorja za mapo
<b>OPOZORILO</b>	raziskovalec	po odpravljanju napak
<b>OPOZORILO</b>	monitor	po verifikaciji mape
<b>MAPA VERIFICIRANA</b>	koordinator	po zaključku mape (podatki zaklenjeni), ali po pošiljanju poizvedbe
<b>POIZVEDBA</b>	monitor	po odpravljanju napak
<b>ODGOVOR NA POIZVEDBO</b>	koordinator	po zaključku mape (podatki zaklenjeni)

Strukturni model sporočil v sistemu je prikazan na UML diagramu na Sliki 23.



Slika 23. Strukturni model sporočil

### III.4.5. Uporabniški vmesnik

#### *Organizacija uporabniškega vmesnika*

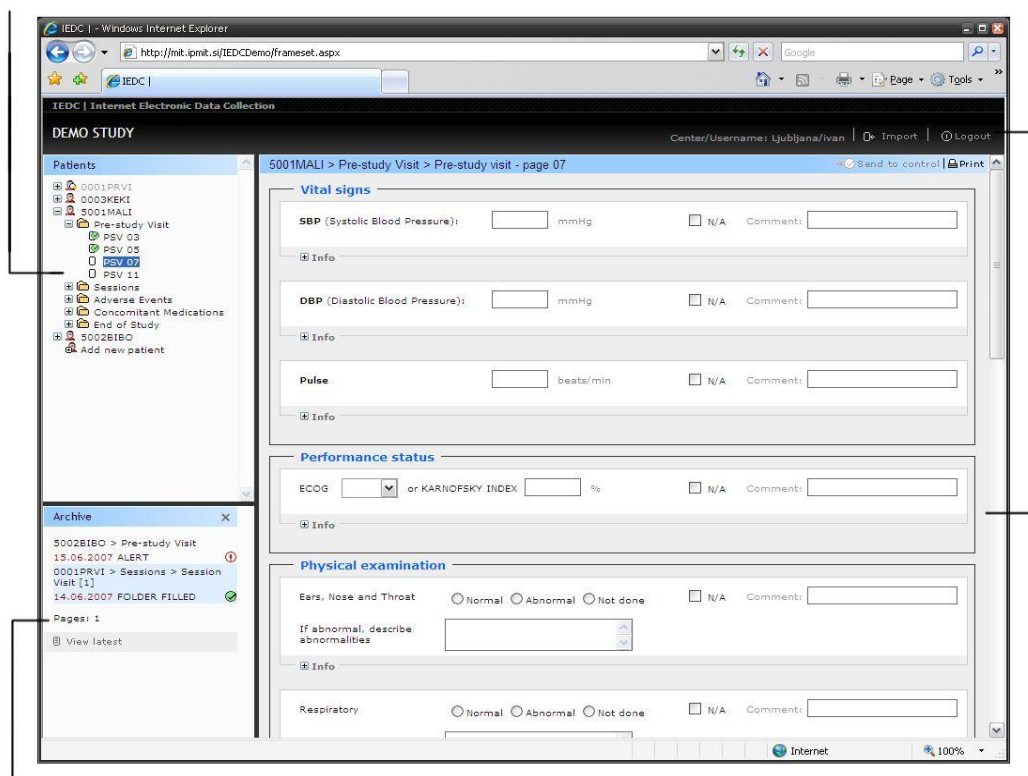
Pri zasnovi organizacije uporabniškega vmesnika smo upoštevali pravila izdelave uporabniku prijaznih spletnih vmesnikov povzeta v (Koyani, Bailey, & Nall, 2004). Stran smo razdelili v štiri funkcionalne cone, in sicer: vrstico z orodji, navigacijski panel eCRF

### III. Rezultati

obrazcev, prikaz eCRF obrazca in okno s sporočili (Slika 24). Zasnova in delovanje posameznih delov uporabniškega vmesnika so opisani v nadaljevanju.

navigacijski panel eCRF obrazcev

vrstica z orodji



okno s sporočili

prikaz eCRF obrazca

Slika 24. Prikaz uporabniškega vmesnika pilotne rešitve iEDC

#### Vrstica z orodji

Vrh uporabniškega vmesnika je rezerviran za vrstico, ki vsebuje naziv študije in podatke o uporabniku in raziskovalnem centru ter povezave na funkcionalnosti, ki so specifične za posamezno študijo (Slika 25).



Slika 25. Primer prikaza vrstice z orodji uporabniškega vmesnika iEDC

Na primer, v tej vrstici so lahko povezave na funkcionalnosti kot so: izvoz podatkov; uvoz podatkov iz zunanjega sistema ali datoteke; prikaz statistik itn. V isti vrstici je tudi gumb za odjavo uporabnika iz sistema.

#### Navigacijski panel eCRF obrazcev

Navigacijski panel eCRF obrazcev je izdelan v Macromedia Flash™ tehnologiji in vsebuje interaktivni seznam preizkušancev. Ob izbiri preizkušanca se prikaže drevesna struktura njegovega eCRF obrazca (Slika 26). Uporabnik lahko odpira in zapira posamezne elemente

drevesne strukture do zelene globine, podobno kot to lahko dela pri uporabi brskalnika Windows Explorer™ za navigacijo po drevesni strukturi map na disku.



Sistem drevesno strukturo navigacijskega panela gradi tako, da za vsakega preizkušanca (eCRF) iz baze prebere vse mape in strani, ki mu pripadajo, in jih prikaže v vrstnem redu določenim v bazi. V skladu z vlogo, ki jo ima uporabnik v sistemu, in na osnovi statusa strukturnih elementov eCRF (map in strani), se zraven elementov prikažejo ustrezne ikone, ki prikazujejo status elementa (npr. zeleni krogec ob strani »PSV 03« kaže na to, da je stran izpolnjena - Slika 26). Stanje mape določa barvo, s katero je izpisan tekst elementa v navigaciji. Element, v katerega uporabnik lahko vnaša podatke, je črn. Element, čigar podatke uporabnik lahko le bere, je siv (Slika 26). V primeru študije z »internim monitoringom« uporabnik (npr. raziskovalna sestra) tiste elemente (preizkušance in posamezne mape), katere je izpolnil drugi uporabnik in katere mora pregledati, vidi modro pobarvane.

Slika 26. Navigacijski panel iEDC

V navigacijskem panelu je mogoče dodajati nove elemente drevesne strukture v skladu s strukturo eCRF. Kot je s slike razvidno (Slika 26), lahko uporabnik za preizkušanca »5001MALI« poleg dveh že obstoječih obiskov (»Session Visit [1]« in »Session Visit [2]«) z izborom elementa »Add new« dodaja nove obiske. Prav tako lahko uporabnik doda novega preizkušanca v bazo, in sicer z izborom opcije v dnu drevesne strukture »Add new patient«.

Enkrat naložena drevesna struktura ne potrebuje novega branja podatkov iz baze vse dokler uporabnik samo brska po drevesni strukturi. Ob izbiri elementa strani (klik na ustrezno ikono v drevesni strukturi) v oknu za prikaz eCRF obrazca, se naloži izbrana stran, navigacijsko drevo pa se ponovno naloži. Med odpiranjem strani v navigacijskem panelu se predvaja animacija, ki uporabniku sporoča, da treba počakati na nalaganje novih vsebin.

#### *Prikaz eCRF obrazca*

Sistem avtomatično generira izpis izbrane strani v navigacijskem panelu. Na vrhu strani je prikazana celotna navigacijska pot (iz drevesne strukture) do strani (Slika 27). Poleg navigacijske poti so v tej vrstici prisotne še dodatne možnosti dela z obrazcem (npr. možnost tiskanja strani »Print«).

5001MALI > Sessions > Session Visit [1] > Session Visit - Vital Signs and Physical Examination - [Send to control](#) [Print](#)  
page 15

Date of session:     
Day Month Year  N/A Comment:

Info

Total number of nodules identified:   N/A Comment:

Info

Number of treated/evaluated nodules:   N/A Comment:

Info

Time of start of session:  Hr  Min  N/A Comment:

Info

Time of end of session:  Hr  Min  N/A Comment:

Info

**Vital signs - (before treatment)**

ECOG  or KARNOFSKY INDEX  %  N/A Comment:

Info

SBP  mmHg  N/A Comment:

Slika 27. Primer prikaz strani eCRF za raziskovalca

Stran obrazca je razdeljena v zaporedne vrstice, ki vsebujejo posamezne *ITEM*-e (v primeru prikazanem na Sliki 27: najprej »Date of session«, potem »Total number of nodules identified« itn.).

V primeru prikaza uporabniškega vmesnika raziskovalca je treba vsak *ITEM* »izpolniti« bodisi tako, da se v vnosno polje (eno ali več) vnese podatek bodisi tako, da uporabnik izbere opcijo »N/A« ter v polje »Comment« vnese razlago, zakaj podatek manjka. Po ponovnem dostopu do strani sistem s spremembo barve ozadja raziskovalcu pomaga ugotoviti, kateri *ITEM*-i še niso izpolnjeni. Namreč, neizpolnjeni *ITEM*-i imajo belo ozadje in so hitro opazni med izpolnjenimi *ITEM*-i s sivim ozadjem.

V primeru, da monitor vrne obrazec raziskovalcu na popravilo, raziskovalec vidi vse problematične *ITEM*-e (tiste, za katere monitor ve ali domneva, da ne vsebujejo pravih podatkov) uokvirjene z rdečim in izpisanim opozorilom monitorja ter ikono »klikaj« (Slika 28). Ostali *ITEM*-i so zaklenjeni in jih ni mogoče spreminjati. Polje z opozorilom je znova prikazano kot »neizpolnjeno«, vse dokler raziskovalec podatka ne popravi, ali v polje »Comment« ne vnese svojega komentarja/odgovora na monitorjevo opozorilo. Na enak način se raziskovalcu prikazujejo tudi poizvedbe koordinatorja, ki jih raziskovalcu posreduje monitor.



### III. Rezultati

5001MALI > Pre-study Visit > Pre-study Visit - Page 03 Send to control | Print

Visit Date:     
Day Month Year  N/A Comment:

**Alert:** before date of birth?!

**Info**

Date of written informed consent:     
Day Month Year  N/A Comment:

**Info**

**Demography**

Date of Birth:     
Day Month Year  N/A Comment:

Slika 28. Primer prikaza opozorila raziskovalcu na problematičen podatek (po presoji monitorja)

Monitor ima zelo podoben prikaz strani eCRF obrazca (Slika 29). Edina bistvena razlika je v tem, da se monitorju za vsak *ITEM* prikaže še polje za verifikacijo podatka (svetlo modra vrstica v polju *ITEM*« na Sliki 29). Monitor mora podatek označiti kot veljaven ali neveljaven. V primeru, da podatek označi kot neveljaven (pri polju »Valid« izbere opcijo »No«), je monitor obvezan, da vnese še tekst opozorila raziskovalcu (glej *ITEM* »Sex« na Sliki 29). Ob tem se monitorju na začetku polja za verifikacijo prikaže še diskretna ikona »klicaj«.

5001MALI > Pre-study Visit > Pre-study Visit - Page 03 Return to Op | Sign | Print

Visit Date:     
Day Month Year  N/A Comment:

Valid:  Yes  No

**Info**

Date of written informed consent:     
Day Month Year  N/A Comment:

Valid:  Yes  No

**Info**

**Demography**

Date of Birth:     
Day Month Year  N/A Comment:

Valid:  Yes  No

**Info**

Sex:  Male  Female  N/A Comment:

**Info**

Valid:  Yes  No

Height:  cm  N/A Comment:

Valid:  Yes  No

**Info**

Slika 29. Primer prikaza eCRF strani monitorju

### III. Rezultati

Monitor je obvezan verificirati vse podatke v obrazcu. Zato se tudi njemu *ITEM*-i prikazujejo z različnimi ozadji glede na to, ali je že označil veljavnost podatka ali ne.

Če koordinator pošlje poizvedbo, monitor *ITEM* s poizvedbo vidi z rumenim okvirjem (podobno kot raziskovalec vidi problematične *ITEM*-e z rdečim okvirjem). Vsi ostali *ITEM*-i so zaklenjeni in jih ni mogoče spreminjati.

Koordinator ima podoben prikaz strani kot monitor, s tem da ima namesto polja za verifikacijo *ITEM*-a polje za poizvedbe, v katerega lahko vnese tekst poizvedbe. Koordinatorju ni treba pregledati vseh polj v obrazcu, ampak po potrebi (recimo po ugotovitvah računalniške analize podatkov) pri določenih poljih vnese tekst poizvedbe. V primeru, da nima nobenih dvomov glede pravilnosti podatkov v obrazcih, koordinator obrazec (oziroma celotno mapo) dokončno zaklene ter s tem onemogoči nadaljnje spremembe podatkov v obrazcu.

Še ena značilnost prikaza eCRF obrazca je prikaz zgodovine sprememb *ITEM*-a. Vsak uporabnik ima pod vsakim *ITEM*-om opcijo »Info«. Ob kliku na to opcijo se pod *ITEM*-om odpre celotna zgodovina sprememb, ki vsebuje: ime uporabnika, ki je spremenil *ITEM*; datum in čas spremembe; ime polja; novo vrednost; ter druge pomembne podatke, kot je razvidno iz Slike 30.

The screenshot shows a data entry interface. At the top, there is a 'Visit Date' field with input boxes for Day (12), Month (2), and Year (2007). To the right, there is an 'N/A' checkbox and a 'Comment' field. Below this, there is a 'Valid' section with 'Yes' and 'No' checkboxes. An 'Info' button is visible. Below the form is a table with the following data:

Name	Date and Time	Not Available	Valid	FIELDS	Comment	Alert	Warning
Ivan Pavlovic	15/12/2008 01:12	<input type="checkbox"/>	/	Visit date	12.2.2007		
Janez Novak	15/12/2008 01:01	<input type="checkbox"/>	No	Visit date	12.2.1965		before date of birth?!
Janez Novak	15/12/2008 00:59	<input type="checkbox"/>	Yes	Visit date	12.2.1965		
Ivan Pavlovic	15/06/2007 11:31	<input type="checkbox"/>	/	Visit date	12.2.1965		

Slika 30. Prikaz zgodovine sprememb podatkov

#### Okno s sporočili

Uporabniki sprejemajo sporočila sistema, ki jih obveščajo o aktivnostih drugih uporabnikov sistema, ki vplivajo na njihov proces dela. Tipi sporočil in aktivnosti, ki jih sprožajo so razložena v poglavju »Opis strukture sporočil v sistemu«.

V osnovi so sporočila v oknu s sporočili prikazana v obliki seznama (Slika 31). Vsaka vrstica seznama vsebuje informacijo o tem, na katero mapo se sporočilo nanaša, ki je istočasno tudi povezava do te mape, potem tip sporočila (npr. »ALERT Removed«) skupaj z ikono, ki ustreza tipu sporočila (npr. »klikaj«) ter datum prejema sporočila. Datum in tip sporočila sta istočasno tudi povezavi na podrobnosti sporočila, ki ima poseben prikaz. Nova

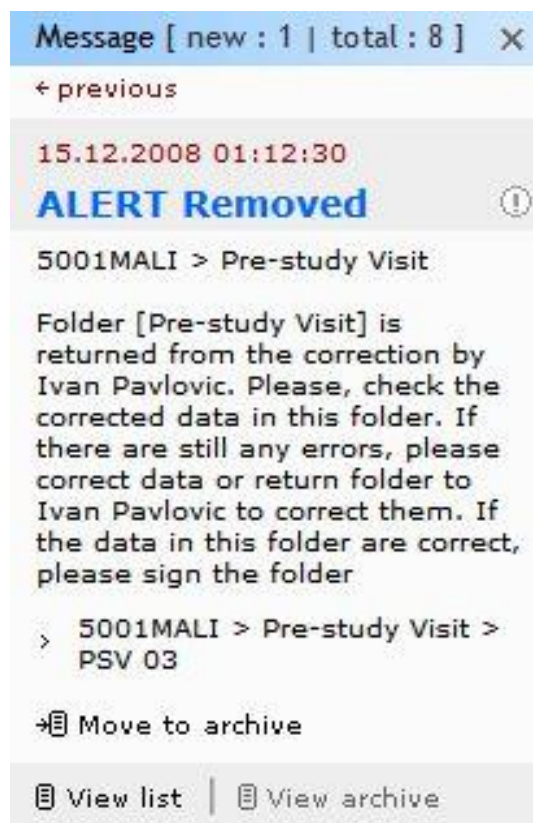
sporočila (tista pri katerih uporabnik še ni kliknil na povezavo) imajo status novih sporočil, kar se v prikazu seznama sporočil opazi kot opozorilo »new« (novo) zraven sporočila.



Slika 31. Primer prikaza okna s sporočili za monitorja

Na dnu seznama je še povezava na arhiv sporočil, v katerem se nahajajo vsa uporabnikova sporočila, ki jih je preselil med arhivska, ali jih je sistem avtomatično preselil tja, kot je razloženo v poglavju »Opis strukture sporočil v sistemu«. Prikaz arhiva sporočil je enak osnovnemu prikazu seznama sporočil.

Podrobnosti sporočila se prikažejo ob izbiri povezave na informaciji o datumu in tipu sporočila na seznamu sporočil. Prikaz podrobnosti sporočila je predstavljen na Sliki 32.



Slika 32. Primer prikaza podrobnosti sporočila

#### III.4.6. Varnostni mehanizmi

Ko govorimo o varnosti sistema lahko upoštevamo več različnih aspektov, in sicer: varnost sistema pred vdori, varnost osebnih podatkov pred zlorabo uporabnikov znotraj sistema ter varnost kakovosti podatkov.

iEDC sistem vsebuje osnovne varnostne mehanizme za zaščito spletnih aplikacij, in sicer:

- avtentikacijo uporabnikov z uporabniškim imenom in geslom;
- SSL protokol za zaščito komunikacije;
- uporabniška gesla kodirana in shranjena v bazi podatkov;
- zaščito pred zlonamernimi SQL ukazi;
- standardno zaščitno politiko za IIS strežnik.

Z namenom varovanja osebnih podatkov imajo uporabniki znotraj sistema dostop le do preizkušancev iz lastnega raziskovalnega centra. Še več, ob zasnovi eCRF se lahko predlaga zajem anonimiziranih podatkov na ta način, da se izključijo podatki, ki omogočajo identifikacijo osebe (kot so recimo ime, priimek, EMŠO, točen datum in mesto rojstva ipd.). Ključ do identitete osebe je lahko v kakšni kodirani obliki, kot je recimo kombinacija števila in nekaj začetnih črk imena in priimka, ali kakšen drug običajen način kodiranja identitete preizkušancev. Še posebej pomembno je shranjevanje informacij o dostopih do

sistema ter shranjevanje zgodovine sprememb podatkov, iz katere je razvidno kdo, kdaj in zakaj je spremenil vrednost posameznega podatka.

Z namenom zagotavljanja kakovosti podatkov sistem vsebuje tudi mehanizme, ki pomagajo uporabnikom pri pravilnem izpolnjevanju obrazcev, in sicer:

- opozorilo uporabniku ob kliku na novo stran, da so podatki na strani spremenjeni, brez da bi bili shranjeni;
- opozorilo uporabniku ob shranjevanju podatkov, da je vrednost podatka zunaj določenih meja;
- opozorilo uporabniku, da določena informacija manjka (npr. monitorju, da je podatek označil kot neveljaven, brez da bi dopisal pojasnilo).

Ob opozorilu se uporabniku ponudi možnost, da opozorilo ignorira, ali da odpravi domnevno nepravilnost in šele potem nadaljuje z delom.

#### III.4.7. Arhitektura in tehnološke zahteve sistema

Sistem je dizajniran kot trislojna aplikacija. Osnovni sloj je relacijska baza, v kateri so shranjeni vsi podatki o študiji in uporabnikih kot tudi podatki zbrani med študijo.

Srednji sloj predstavlja XML spletni servis. Srednji sloj vsebuje poslovno logiko aplikacije, skrbi za konsistentnost podatkov ter upravlja morebitne napake pri delovanju sistema in shranjuje informacije o delovanju sistema v »log« datoteke.

Najvišji sloj aplikacije je uporabniški spletno zasnovani »thin-client«. Ta sloj skrbi za delovanje spletnih obrazcev in tako imenovanih spletnih kontrol ter za avtentikacijo obrazcev.

Sistem je v celoti razvit na Microsoft™ programski tehnologiji, ki vključuje .NET tehnologijo in MS SQL tehnologijo. Tehnološke zahteve programske opreme na strani strežnika so naslednje:

- Microsoft Internet Information Server 6.0 (ali novejši);
- Microsoft .NET Framework 1.1 (ali novejši);
- Microsoft SQL Server 2000 (ali novejši).

Na strani uporabnika za delovanje »thin-client« aplikacije je potrebno zagotoviti naslednjo programsko opremo:

- Internet Explorer™ brskalnik, ali njemu kompatibilnega (npr. Mozilla Firefox™);
- Macromedia Flash™ dodatek za brskalnik.

Poleg omenjene programske opreme je potrebno zagotoviti tudi internetno povezavo med strežnikom in uporabniki z možnostjo delovanja po SSL protokolu (predvidoma port 443) za potrebe vzpostavitve varne povezave.

## III.5. Ocene

### III.5.1. Stroški

Stroške elektronskega zbiranja podatkov in zbiranja podatkov na papirju lahko razdelimo v tri sklope: stroške postavitve sistema, stroške uporabe sistema in stroške vzdrževanja sistema.

#### *Stroški vzpostavitve sistema*

Ko govorimo o stroških vzpostavitve sistema za zbiranje podatkov na papirju, mislimo na stroške izdelave papirne dokumentacije in dostavo raziskovalnim centrom. Ocenjeni stroški vzpostavitve PDC sistema za vzorčno študijo znašajo 30.000 € (Tabela 9).

Tabela 9. Stroški vzpostavitve PDC sistema (skupno število CRF obrazcev -  $N_{CRF} = 1.000$ ; število strani v CRF obrazcu -  $n_{p/CRF} = 50$  strani; število kopij CRF obrazca -  $k_{CRF} = 3$ ; cena tiskanja ene strani -  $\varphi_p = 0,10$  €/stran; cena pošiljanja enega CRF-ja -  $\varphi_s = 15$  €/CRF;)

Strošek	Izračun	Vrednost
Tiskanje papirnih CRF obrazcev	$N_{CRF} \times n_{p/CRF} \times k_{CRF} \times \varphi_p$	15.000 €
Pošiljanje obrazcev od naročnika študije do raziskovalnih centrov	$N_{CRF} \times \varphi_s$	15.000 €
<b>SKUPAJ</b>		<b>30.000 €</b>

Pri elektronskem zbiranju podatkov mora naročnik študije zagotoviti vso potrebno tehnično opremo za delovanje iEDC sistema, prilagoditi iEDC aplikacijo študiji ter izvesti izobraževanje uporabnikov za uporabo iEDC aplikacije. Za vzorčno študijo ocenjeni stroški vzpostavitve sistema za elektronsko zbiranje podatkov znašajo približno 30.000 € (Tabela 10).

Tabela 10. Stroški vzpostavitve EDC sistema (cena ure dela na postavitvi iEDC -  $\varphi_i = 160$  €/h; število raziskovalnih centrov -  $N_{RC} = 10$ ; cena ure izobraževanja uporabnikov -  $\varphi_t = 100$  €/h)

Strošek	Izračun	Vrednost
Strežniška strojna oprema za EDC sistem	Podatkovni strežnik - 2.000 € Spletni strežnik - 2.000 €	4.000 €
Strežniška programska oprema	Licence za 2 x MS Server in 1 x MS SQL Server	5.000 €
Postavitev iEDC sistema in nastavitve študije	- Analiza uporabniških zahtev in pomoč pri planiranju študije in EDC procesa - $5h \times \varphi_i$ ; - Izdelava specifikacije študije (uporabniki, eCRF) - $7h \times \varphi_i$ - Nastavitve uporabnikov - $2h \times \varphi_i$	19.600 €

Strošek	Izračun	Vrednost
	- Postavitev strukture eCRF dokumenta in dizajn - $8h \times \varphi_i$ - Postavitev baze podatkov in povezovanje na eCRF dokument - $30h \times \varphi_i$ - Postavitev iEDC študije in testiranje - $8h \times \varphi_i$ - Licenca za iEDC - 10.000 €	
Izobraževanje uporabnikov	Dodatna ura izobraževanja ob prvem obisku monitorja pri vsakem raziskovalnem centru: $N_{RC} \times \varphi_t$	1.000 €
	<b>SKUPAJ</b>	<b>29.600 €</b>

Pri izračunih smo predpostavljali, da raziskovalni centri že imajo lastne računalnike s spletnim brskalnikom, ki jih lahko uporabljajo za dostop do iEDC sistema.

Iz zgornjih izračunov stroškov smo ugotovili, da so stroški vzpostavitve PDC in EDC sistema za vzorčno študijo podobni.

### *Stroški izvajanja procesa zbiranja podatkov*

Primerjavo smo izvedli za primer vzorčne študije, kot je predstavljeno v Metodah v poglavju »II.5 Ocenjevanje«. Za vsakega od podprocesov zbiranja podatkov smo na osnovi eEPC modela postavili sistem enačb in upoštevajoč ocene vrednosti parametrov izračunali predvidene stroške procesa.

### *Zajem podatkov*

Glede na Sliko 7 in Sliko 10 izhaja razlika med stroški EDC in PDC procesa zajema podatkov iz razlik stroškov izvajanja naloge zajema podatkov (PDC: *Vnos podatkov v CRF*; EDC: *Vnos podatkov v eCRF*) in zaradi potrebe po fizičnem arhiviranju obrazcev pri PDC procesu (PDC: *Arhiviranje CRF obrazcev*). Če si dovolimo generalizacijo, da se te naloge enako izvajajo pri vseh raziskovalnih centrih, se lahko razlika med stroški PDC in EDC zajema podatkov prikaže kot:

$$\Delta\phi_{DG} = \phi_{DG^P} - \phi_{DG^E}$$

$$\phi_{DG^P} = F_{DG_{CRF}^P} \times \varphi_{IA} + F_{DG_{arh}^P} \times \varphi_I$$

$$\phi_{DG^E} = F_{DG_{CRF}^E} \times \varphi_{IA}$$

Skupen napor za zapisovanje podatkov na papirnate obrazce je izražen z  $F_{DG_{CRF}^P} = N_D \times \mu_{DG_{CRF}^P}$ , pri čem je  $\mu_{DG_{CRF}^P}$  povprečen napor zapisovanja enega podatka na papir. Skupen napor arhiviranja vseh CRF obrazcev je izražen z  $F_{DG_{arh}^P} = N_{CRF} \times \mu_{DG_{arh}^P}$ , pri čem  $\mu_{DG_{arh}^P}$  predstavlja povprečen napor za arhiviranje enega CRF obrazca.

Po drugi strani je skupen napor za vnos podatkov v eCRF izražen z  $F_{DG_{CRF}^E} = N_D \times \mu_{DG_{CRF}^E}$ , pri čem  $\mu_{DG_{CRF}^E}$  predstavlja povprečen napor za vnos enega podatka.

Na osnovi podatkov pridobljenih v intervjujih smo ocenili povprečno hitrost zapisovanja podatkov v CRF kot 5 podatkov v minuti za PDC in 10 podatkov v minuti za EDC. Hitrejši vnos podatkov v EDC je posledica tega, da številne podatke uporabnik vnaša tako, da izbere eno (ali več) od ponujenih možnosti (iz slovarja, kodne tabele ipd.) in možnosti avtomatičnega izračuna nekaterih vrednosti v obrazcu (npr. indeks telesne teže (BMI) na osnovi višine in teže). Poleg tega drži, da povprečen uporabnik računalnikov na splošno hitreje tipka tekst, kot ga piše na roko.

Torej, v naših izračunih smo uporabili naslednje ocene povprečnih naporov:

- PDC: zapisovanje podatka v CRF -  $\mu_{DG_{CRF}^P} = 12 \frac{sec}{data}$ ;
- EDC: vnos podatka v eCRF -  $\mu_{DG_{CRF}^E} = 6 \frac{sec}{data}$ ;
- PDC: arhiviranje papirnatega CRF -  $\mu_{DG_{arh}^P} = 2 \frac{min}{doc}$ .

Upoštevajoč naše ocene za vzorčno klinično študijo, pridemo do naslednjih izračunov stroškov zajema podatkov:

$$\phi_{DG^P} \approx 33,500 \text{ €}$$

$$\phi_{DG^E} \approx 16,500 \text{ €}$$

$$\Delta\phi_{DG} \approx 17,000 \text{ €}$$

#### Monitoring

Sprememba procesa zbiranja podatkov s papirnega na elektronskega vnaša spremembe tudi v proces monitoringa (Slika 8 in Slika 11). PDC Monitoring zahteva več naporov za organizacijo dela, ker mora biti celoten monitoring narejen med obiskom monitorja na lokaciji raziskovalnega centra. Vse napake med zbranimi podatki morajo biti identificirane med obiskom in tudi odpravljene. To zahteva istočasno prisotnost vseh udeležencev v procesu (monitorja, raziskovalcev in raziskovalnih sester) kot tudi vseh podatkovnih virov (zdravstvene dokumentacije ipd.). Komunikacija se odvija neposredno. Vsak problem, ki se v procesu pojavi, podaljša proces, poleg tega pa blokira tudi vse potrebne resurse. Dodaten problem je ta, da se lahko, zaradi časovne omejenosti obiska, zgodi, da trpi kakovost izvajanja naloge (npr. nekatere napake so lahko spregledane).

Z uporabo EDC sistema lahko monitor označi problematične podatke v eCRF obrazcu in po koncu obiska še naprej sledi, kako se odvija odpravljanje napak v elektronskih obrazcih, ter popravljene obrazce verificira. To pomeni, da namesto direktne komunikacije EDC sistem omogoča indirektno komunikacijo, s katero se je mogoče izogniti blokadi resursov. Še več,



kakovost podatkov zbranih v elektronske obrazce je višja, ker lahko EDC sistem avtomatično opozarja uporabnike ob vnosu napačnih podatkov. Rezultat tega je manj napačnih podatkov in manj poizvedb (Bart, 2003).

Stroški PDC in EDC monitoring procesov so lahko predstavljeni kot:

$$\phi_{MP} = F_{M_{rev}^P} \times \varphi_{CRA} + F_{M_{cor}^P} \times (\varphi_I + \varphi_{IA} + \varphi_{CRA}) + F_{M_{sig}^P} \times \varphi_I + F_{M_{loc}^P} \times \varphi_{CRA} + F_{M_{snd}^P} \times \varphi_{CRA}$$

$$\phi_{ME} = F_{M_{rev}^E} \times \varphi_{CRA} + F_{M_{loc}^E} \times \varphi_{CRA} + F_{M_{cor}^E} \times (\varphi_I + \varphi_{IA}) + F_{M_{ver}^E} \times \varphi_{CRA}$$

V zgornjih enačbah se  $F_{MP}$  funkcije nanašajo na PDC naloge predstavljene na Sliki 8 in  $F_{ME}$  funkcije se nanašajo na EDC naloge na Sliki 11.

Med pregledovanjem obrazcev mora monitor zabeležiti vse problematične vnose, kar je različno izpeljano za PDC in EDC proces. Na primer, EDC sistem omogoča vnos opozorila raziskovalcu ob problematičnem podatku, medtem ko pri PDC procesu monitor na poseben seznam zapisuje problematične vnose skupaj z informacijami o tem, v katerih obrazcih so ugotovljeni. Zaradi tega lahko ocenimo, da je napor za zapis opozorila dvakrat večji pri PDC procesu kot pri EDC procesu ( $\mu_{M_{rev}^P} = 2\mu_{M_{rev}^E}$ ). Še več, število opozoril je odvisno od števila napak, ki jih raziskovalec naredi in je precej višje za PDC kot za EDC (kot je že predstavljeno v Metodah v poglavju »II.5 Ocenjevanje, Kakovost podatkov«).

$$F_{M_{rev}^P} = N_D \times \mu_{M_{chk}} + N_D \times \varepsilon_{mon}^P \times \mu_{M_{rev}^P}$$

$$F_{M_{rev}^E} = N_D \times \mu_{M_{chk}} + N_D \times \varepsilon_{mon}^E \times \mu_{M_{rev}^E}$$

Za oba, PDC in EDC, lahko pričakujemo, da je napor za končanje obiska enak in ga lahko izrazimo kot:

$$F_{M_{loc}^P} = F_{M_{loc}^E} = N_M \times \mu_{M_{loc}}$$

Način odpravljanja napak med podatki je za PDC in EDC proces tudi različen. Odpravljanje napak pri EDC procesu se odvija s pomočjo EDC sistema, pri čemer vsa komunikacija med uporabniki poteka prek sistema. Poleg tega je, zahvaljujoč se avtomatičnemu opozarjanju na nepravilnosti ob vnosu podatkov, delež napak med podatki precej nižji za EDC kot za PDC proces in po navedbah (Spink, 2002) znaša  $\varepsilon_{mon}^E = 20\%$   $\varepsilon_{mon}^P = 2\%$ . To pomeni, da bo potrebno vložiti več napora za odpravljanje napak pri vnosu podatkov v PDC procesu.

$$F_{M_{cor}^P} = N_D \times \varepsilon_{mon}^P \times \mu_{M_{cor}}, F_{M_{cor}^E} = N_D \times \varepsilon_{mon}^E \times \mu_{M_{cor}}$$

$$F_{M_{sig}^P} = N_{sec} \times \mu_{M_{sig}^P}$$

$$F_{M_{snd}^P} = N_{sec} \times \mu_{M_{snd}^P}$$

$$F_{M_{ver}^E} = N_{sec} \times \mu_{M_{ver}^E}$$

Na osnovi podatkov zbranih prek intervjujev smo naredili naslednjo oceno povprečnih naporov:

- PDC: zapisovanje enega opozorila -  $\mu_{M_{rer}^P} = 1 \frac{min}{data}$ ;
- EDC: zapisovanje enega opozorila -  $\mu_{M_{rer}^E} = 30 \frac{sec}{data}$ ;
- preverjanje podatka -  $\mu_{M_{chk}} = 30 \frac{sec}{data}$ ;
- zaključek obiska monitorja -  $\mu_{M_{loc}} = 2 \frac{min}{visit}$ ;
- odpravljanje napake -  $\mu_{M_{cor}^P} = \mu_{M_{cor}^E} = 1 \frac{min}{data}$ ;
- PDC: verifikacija papirnatega dokumenta -  $\mu_{M_{sig}^P} = 1 \frac{min}{doc}$ ;
- PDC: pošiljanje CRF dokumenta koordinatorju -  $\mu_{M_{snd}^P} = 2 \frac{min}{doc}$ ;
- EDC: verifikacija elektronskega dokumenta -  $\mu_{M_{ver}^E} = 1 \frac{min}{doc}$ .

Na osnovi teh ocen, z uporabo zgoraj navedenih enačb, pridemo do izračunov za stroške PDC in EDC monitoring procesa:

$$\phi_{MP} \approx 233,500 \text{ €}$$

$$\phi_{ME} \approx 140,000 \text{ €}$$

$$\Delta\phi_M \approx 93,500 \text{ €}.$$

#### Upravljanje s podatki

Kot smo omenili ob predlogu EDC procesa, podproces EDC Upravljanje s podatki ne zahteva nobenih naporov za digitalizacijo podatkov, ker so ti že digitalizirani ob zajemu podatkov v raziskovalnem centru. To ima za posledico, da je skupno število napak med podatki zmanjšano za število napak, ki nastanejo pri prepisovanju podatkov iz papirnatih obrazcev v bazo podatkov.

PDC sklop »Reševanje poizvedb« zahteva dodaten napor pri komunikaciji, izmenjavi dokumentov in sledenju statusov poizvedb. Rešene poizvedbe prihajajo nazaj h koordinatorju in narediti je treba morebitne spremembe podatkov v bazi ter jih dokumentirati. Med vsemi temi postopki podatkovni center po navadi uporablja dodatne računalniške rešitve, ki pomagajo slediti dokumentom, poizvedbam in spremembam. Ažurno je posodabljanje informacij v teh rešitvah.

Če izhajamo iz eEPC modela, prikazanega na Sliki 9 in Sliki 12, lahko napore za izvajanje upravljanja s podatki pri PDC in EDC procesu izrazimo z naslednjimi enačbami:

$$\text{PDC Upravljanje s podatki: } F_{DM^P} = F_{DM_{ENT}^P} + F_{DM_{ANL}^P} + F_{DM_{QRY}^P} + F_{DM_{CLO}^P}$$

EDC Upravljanje s podatki:  $F_{DM^E} = F_{DM_{ANL}^E} + F_{DM_{QRY}^E} + F_{DM_{CLO}^E}$

#### Vnos podatkov

EDC vnos podatkov v bazo izvajajo sami raziskovalci (oziroma raziskovalne sestre). Po drugi strani pa je pri PDC procesu potrebno opraviti več nalog v postopku vnosa podatkov. To posledično zvišuje stroške PDC Upravljanja s podatki.

$$F_{DM_{ENT}^P} = F_{DM_{erf}^P} + F_{DM_{dde}^P} + F_{DM_{adm}^P} + F_{DM_{add}^P} + F_{DM_{eda}^P} + F_{DM_{edc}^P} + F_{DM_{pde}^P}$$

PDC proces vnosa podatkov se začne z evidentiranjem prispelih obrazcev. Monitorji pošiljajo celotne mape izpolnjenih CRF obrazcev. Zato mora biti v primeru vzorčne študije evidentirano skupno  $N_{sec} = 10,000$  dokumentov (map).

$$F_{DM_{erf}^P} = N_{sec} \times \mu_{DM_{erf}^P}$$

Ob uporabi metode podvojenega vnosa podatkov (angl. »double data entry«) za minimizacijo števila napak, ki nastanejo zaradi prepisovanja podatkov, potrebujemo dvojni napor za vnos podatkov.

$$F_{DM_{dde}^P} = N_D \times \mu_{DM_{dde}^P} \times 2$$

Upoštevajoč, da je povprečen delež napak pri vnosu ( $\varepsilon_{ent}^P$ ) okoli 1 % na osebo (King & Lashley, 2000; Wahi et al., 2008; Kleinman, 2001; Kawado et al., 2003; Day et al., 1998), lahko pričakujemo dvakrat več neujemanj med podatki v vsaki CRF mapi, digitalizirani po metodi dvojnega vnosa podatkov.

$$F_{DM_{adm}^P} = N_D \times \varepsilon_{ent}^P \times 2 \times \mu_{DM_{adm}^P}$$

Pri vnosu podatkov vnašalci podatkov včasih ne znajo prebrati ročno zapisanih podatkov v papirnatih obrazcih, kar mora biti tudi evidentirano ( $F_{DM_{add}^P}$ ). Po drugi strani razkrije nekatere nepravilnosti računalniška analiza vnesenih podatkov in tudi te morajo biti evidentirane ( $F_{DM_{pde}^P}$ ). Dodaten korak ( $F_{DM_{edc}^P}$ ) za reševanje teh evidentiranih nepravilnosti je nujen za zniževanje deleža napak ob vnosu. Vse naštetje nepravilnosti, imenovane »neverodostojni vnosi« (angl. »plausibility discrepancies«), predstavljajo 80 % vseh poizvedb, ki jih naročnik (koordinator) generira ( $\varepsilon_{pla}^P = 80 \% \varepsilon_{qry}^P$ ; glej Metode, poglavje »II.5 Ocenjevanje, Kakovost podatkov«).

$$F_{DM_{add}^P} + F_{DM_{pde}^P} = N_D \times \varepsilon_{pla}^P \times \mu_{DM_{pde}^P}$$

V končni fazi, zahvaljujoč uporabi metode podvojenega vnosa podatkov, le manjši delež med vsemi napakami predstavlja napake ob vnosu podatkov. Zaradi tega lahko

zanemarimo napor za odpravljanje teh napak ( $F_{DM_{edc}^P} \approx 0$ ). Ob predpostavki uporabe kakovostnih računalniških rešitev je lahko napor, ki ga osebe vložijo v računalniško analizo podatkov, zanemarljiv in ga lahko tudi izpustimo iz izračunov ( $F_{DM_{eda}^P} \approx 0$ ).

Naše ocene povprečnih naporov so naslednje:

- PDC: evidentiranje sprejete mape -  $\mu_{DM_{erf}^P} = 1 \frac{min}{doc}$ ;
- PDC: vnos podatkov v bazo -  $\mu_{DM_{dde}^P} = 3 \frac{sec}{data}$ ;
- PDC: reševanje neujemanj -  $\mu_{DM_{ddm}^P} = 1 \frac{min}{data}$ ;
- PDC: evidentiranje poizvedbe -  $\mu_{DM_{pde}^P} = 1 \frac{min}{data}$ .

Tako lahko razliko v stroških PDC in EDC procesa vnosa podatkov ocenimo kot:

$$\phi_{DM_{ENT}^P} \approx F_{DM_{erf}^P} \times \varphi_C + F_{DM_{dde}^P} \times \varphi_{DE} + F_{DM_{ddm}^P} \times \varphi_{DM} + (F_{DM_{ddd}^P} + F_{DM_{pde}^P}) \times \varphi_{DM}$$

$$\Delta\phi_{DM_{ENT}} = \phi_{DM_{ENT}^P} \approx 26,000 \text{ €}$$

#### Računalniška analiza podatkov

Izhajajoč iz sklopa »Računalniška analiza podatkov« na Sliki 9, lahko napor za računalniško analizo podatkov v primeru PDC procesa izrazimo kot:

$$F_{DM_{ANL}^P} = F_{DM_{cda}^P} + F_{DM_{ade}^P} + F_{DM_{qpr}^P} + F_{DM_{qsn}^P}$$

Funkcionalnosti EDC sistema omogočajo, da je EDC proces poenostavljen in ne vključuje naporov za evidentiranje poizvedb in pošiljanje:

$$F_{DM_{ANL}^E} = F_{DM_{cda}^E} + F_{DM_{qpr}^E}$$

Na primer, po računalniški analizi koordinator v EDC aplikacijo vnese le poizvedbe in »odklene« ustrezne obrazce.

Če predvidimo, da je pri PDC in EDC procesu za samo računalniško analizo potreben približno enak napor, ter če z namenom poenostavitve izračunov predpostavimo, da je ta napor precej manjši od ostalih naporov v procesu, lahko napor za izvajanje same računalniške analize v naših izračunih zanemarimo.

$$F_{DM_{cda}^P} = F_{DM_{cda}^E} \approx 0$$

Po drugi strani pa različni delež podatkov, za katere se pojavijo poizvedbe pri PDC ( $\varepsilon_{qry}^P = 2\%$ ) in EDC ( $\varepsilon_{qry}^E = 0,2\%$ ) procesu, vpliva na razliko v naporih in stroških priprave poizvedb. Glede na dejstvo, da mora pri PDC procesu koordinator za vsako poizvedbo zapisati poizvedbo in informacijo o tem, na kateri obrazec in podatek se

poizvedba nanaša, predvidevamo, da to zahteva dvakrat več napora, kot vnos poizvedbe v EDC sistem, kjer koordinator poizvedbo vnaša neposredno zraven problematičnega podatka.

$$F_{DM_{ade}^P} = N_D \times (\varepsilon_{qry}^P - \varepsilon_{pla}^P) \times \mu_{DM_{ade}^P}$$

$$F_{DM_{qpr}^P} = N_D \times \varepsilon_{qry}^P \times \mu_{DM_{qpr}^P}$$

$$F_{DM_{qpr}^E} = N_D \times \varepsilon_{qry}^E \times \mu_{DM_{qpr}^E}$$

Končno, ker smo za našo vzorčno klinično študijo predvideli 10x24=240 monitoring obiskov in ker vsak med njimi sproži proces upravljanja s podatki, ki ima za posledico nastanek dokumentov s poizvedbami, smo v naših izračunih predvideli, da bo ustvarjenih 240 dokumentov s poizvedbami ( $N_{DQ} = N_M$ ).

$$F_{DM_{qsn}^P} = N_{DQ} \times \mu_{DM_{qsn}^P}$$

Na osnovi omenjenih ugotovitev smo v naših izračunih upoštevali naslednje ocene povprečnih naporov:

- PDC: evidentiranje neskladnosti -  $\mu_{DM_{ade}^P} = 1 \frac{min}{data}$ ;
- PDC: priprava poizvedbe -  $\mu_{DM_{qpr}^P} = 2 \frac{min}{data}$ ;
- EDC: priprava poizvedbe -  $\mu_{DM_{qpr}^E} = 1 \frac{min}{data}$ ;
- PDC: pošiljanje poizvedbe monitorju -  $\mu_{DM_{qsn}^P} = 2 \frac{min}{doc}$ .

Upoštevajoč omenjene predpostavke, znašajo izračuni stroškov PDC in EDC procesa računalniške analize podatkov:

$$\phi_{DM_{ANL}^P} = F_{DM_{ade}^P} \times \varphi_{DM} + F_{DM_{qpr}^P} \times \varphi_C + F_{DM_{qsn}^P} \times \varphi_C \approx 18,000 \text{ €}$$

$$\phi_{DM_{ANL}^E} = F_{DM_{qpr}^E} \times \varphi_C \approx 1,000 \text{ €}$$

$$\Delta\phi_{DM_{ANL}} \approx 17,000 \text{ €}$$

#### Reševanje poizvedb

Kot je prikazano na Sliki 9 in Sliki 12, lahko uporaba EDC rešitve bistveno zmanjša aktivnosti v procesu reševanja poizvedb. Ob uporabi EDC rešitve so aktivnosti pošiljanja informacij od enega do drugega akterja avtomatizirane. Zaradi manjšega števila dokumentov v procesu so tudi ostale aktivnosti manj zapletene. Tako lahko npr. EDC aplikacija prikaže poizvedbo neposredno zraven problematičnega podatka v eCRF obrazcu. Po drugi strani pa mora pri PDC procesu raziskovalec v lastnem arhivu poiskati

dokumentacijo (ustrezen CRF obrazec) in posamezen vnos, na katerega se nanaša poizvedba.

Na osnovi modelov na Sliki 9 in Sliki 12 (sklop »Reševanje poizvedb«) smo izpeljali naslednje enačbe:

$$F_{DM_{QRY}^P} = F_{DM_{qre}^P} + F_{DM_{qfw}^P} + F_{DM_{qin}^P} + F_{DM_{rfw}^P} + F_{DM_{qpe}^P} + F_{DM_{rsn}^P}$$

$$F_{DM_{QRY}^E} = F_{DM_{qre}^E} + F_{DM_{qin}^E} + F_{DM_{qpe}^E}$$

Oceno stroškov smo zasnovali na naslednjih predpostavkah. Prva predvideva, da je povprečen napor za pregledovanje in evidentiranje poizvedbe enak povprečnem naporu za vpogled in evidentiranje obdelane poizvedbe. To nam omogoča, da izračune poenostavimo na naslednji način:

$F_{DM_{qpe}^P} = F_{DM_{qre}^P}$  in  $F_{DM_{qpe}^E} = F_{DM_{qre}^E}$ , kjer je:

$$F_{DM_{qre}^P} = N_D \times \varepsilon_{qry}^P \times \mu_{DM_{qre}^P}$$

$$F_{DM_{qre}^E} = N_D \times \varepsilon_{qry}^E \times \mu_{DM_{qre}^E}$$

Druga predpostavka je ta, da v primeru EDC procesa sistem avtomatično skrbi za evidentiranje poizvedb in odgovorov na poizvedbe, zaradi česar je povprečen napor za pregledovanje poizvedb pri EDC procesu dvakrat manjši od povprečnega napora za pregledovanje in evidentiranje poizvedb pri PDC procesu ( $\mu_{DM_{qre}^P} = 2\mu_{DM_{qre}^E}$ ).

Tretja predpostavka je ta, da je povprečen napor za pošiljanje poizvedb iz raziskovalnega centra monitorju enak tistemu za pošiljanje monitorja raziskovalnemu centru.

$$F_{DM_{rfw}^P} = F_{DM_{qfw}^P}$$

$$F_{DM_{qfw}^P} = N_{DQ} \times \mu_{DM_{qfw}^P}$$

Glede na podatke pridobljene prek intervjujev je povprečen napor za reševanje poizvedbe v primeru PDC procesa ocenjen kot petkrat večji od napora za vnos podatka v CRF ( $\mu_{DM_{qin}^P} = 5\mu_{CRF^P}$ ). V primeru EDC procesa je ta napor dvakrat manjši ( $\mu_{DM_{qin}^E} = 2.5\mu_{CRF^P}$ ), ker ima raziskovalec vse informacije (razen virov podatkov) že prikazane v EDC aplikaciji.

$$F_{DM_{qin}^P} = N_D \times \varepsilon_{qry}^P \times \mu_{DM_{qin}^P}$$

$$F_{DM_{qin}^E} = N_D \times \varepsilon_{qry}^E \times \mu_{DM_{qin}^E}$$

Četrta in končna predpostavka je ta, da je napor monitorja za pošiljanje odgovorov na poizvedbe enak naporu koordinatorja za pošiljanje poizvedb monitorju ( $F_{DM_{rsn}^P} = F_{DM_{qsn}^P}$ ).

$$F_{DM_{rsn}^P} = N_{DQ} \times \mu_{DM_{qsn}^P}$$

Iz vsega omenjenega lahko pridemo do naslednjih enačb za izračun stroškov za EDC in PDC procese reševanja poizvedb:

$$\begin{aligned} \phi_{DM_{QRY}^P} &= F_{DM_{qre}^P} \times \varphi_{CRA} \times 2 + F_{DM_{qfw}^P} \times (\varphi_{CRA} + \varphi_I) \times 2 + F_{DM_{qin}^P} \times (\varphi_I + \varphi_{IA}) \\ &\quad + F_{DM_{rsn}^P} \times \varphi_{CRA} \end{aligned}$$

$$\phi_{DM_{QRY}^E} = F_{DM_{qre}^E} \times \varphi_{CRA} \times 2 + F_{DM_{qin}^E} \times (\varphi_I + \varphi_{IA})$$

V naših izračunih smo upoštevali naslednje ocene povprečnih naporov:

- PDC: pregledovanje in evidentiranje poizvedbe -  $\mu_{DM_{qre}^P} = 1 \frac{min}{data}$ ;
- EDC: pregledovanje poizvedbe -  $\mu_{DM_{qre}^E} = 15 \frac{sec}{data}$ ;
- PDC posredovanje poizvedb RC -  $\mu_{DM_{qfw}^P} = 2 \frac{min}{doc}$ ;
- PDC: reševanje poizvedbe -  $\mu_{DM_{qin}^P} = 1 \frac{min}{data}$ ;
- EDC: reševanje poizvedbe -  $\mu_{DM_{qin}^E} = 45 \frac{sec}{data}$ .

Na osnovi zgoraj omenjenih enačb in ocen lahko stroške reševanja poizvedb za vzorčno študijo ocenimo kot:

$$\phi_{DM_{QRY}^P} \approx 21,000 \text{ €}; \phi_{DM_{QRY}^E} \approx 1,000 \text{ €}; \Delta\phi_{DM_{QRY}} \approx 20,000 \text{ €}$$

#### Zaključevanje obrazcev

Zaključni korak upravljanja s podatki je »zaklepanje« podatkov v bazi. Ko so podatki posredovani iz raziskovalnih centrov preverjeni in shranjeni, mora biti baza zaklenjena za nadaljnje spremembe. Pri PDC procesu je treba odgovore vnesti v evidenco poizvedb ter narediti morebitne spremembe podatkov v bazi in šele potem v interni aplikaciji za sledenje dokumentom označiti obrazce kot obdelane (zaključene). Predvideni napori ob tem so naslednji:

$$F_{DM_{CLO}^P} = F_{DM_{rev}^P} + F_{DM_{dcr}^P} + F_{DM_{epf}^P}$$

$$F_{DM_{rev}^P} = N_D \times \varepsilon_{qry}^P \times \mu_{DM_{rev}^P}$$

$$F_{DM_{dcr}^P} = N_D \times \varepsilon_{cor}^P \times \mu_{DM_{dcr}^P}$$

$$F_{DM_{epf}^P} = N_{sec} \times \mu_{DM_{ecf}^P}$$

Po drugi strani je v primeru EDC procesa dovolj, da koordinator ob pregledovanju odgovorov na poizvedbe obrazce označi kot zaklenjene, ker so morebitne spremembe podatkov že narejene s strani uporabnikov v raziskovalnem centru.

$$F_{DM_{CLO}^E} = F_{DM_{loc}^E} = N_D \times \varepsilon_{qry}^E \times \mu_{DM_{loc}^E}$$

Ocene povprečnih naporov ob tem so naslednje:

- PDC: evidentiranje odgovora na poizvedbo -  $\mu_{DM_{rev}^P} = 1 \frac{min}{data}$ ;
- PDC: sprememba podatka v bazi -  $\mu_{DM_{dcr}^P} = 2 \frac{min}{data}$ ;
- PDC: evidentiranje obdelanega obrazca -  $\mu_{DM_{ecf}^P} = 1 \frac{min}{doc}$ ;
- EDC: zaklepanje obrazca -  $\mu_{DM_{loc}^E} = 15 \frac{sec}{data}$ .

Iz tega sledi, da lahko razliko med stroški PDC in EDC procesa zaključevanja obrazcev izračunamo kot:

$$\phi_{DM_{CLO}^P} = F_{DM_{rev}^P} \times \varphi_C + F_{DM_{dcr}^P} \times \varphi_{DM} + F_{DM_{epf}^P} \times \varphi_C \approx 19,000 \text{ €}$$

$$\phi_{DM_{CLO}^E} = F_{DM_{loc}^E} \times \varphi_C \ll \phi_{DM_{CLO}^P}; \phi_{DM_{CLO}^E} \approx 0 \text{ €}$$

$$\Delta\phi_{DM_{CLO}} \approx 19,000 \text{ €}$$

Upoštevajoč izračune stroškov vseh sklopov procesa upravljanja s podatki (vnos podatkov, računalniška analiza podatkov, reševanje poizvedb, zaključevanje obrazcev), znaša razlika v stroških PDC in EDC procesa upravljanja podatkov za našo vzorčno klinično študijo:

$$\Delta\phi_{DM} = \Delta\phi_{DM_{ENT}} + \Delta\phi_{DM_{ANL}} + \Delta\phi_{DM_{QRY}} + \Delta\phi_{DM_{CLO}}$$

$$\Delta\phi_{DM} \approx 82,000 \text{ €}$$

Končno, skupni stroški PDC procesa ( $\phi^P$ , ) in EDC procesa ( $\phi^E$ ) ter razlika med njimi ( $\Delta\phi$ ), ki je posledica uporabe EDC aplikacije in z njo povezanih sprememb procesov zajema podatkov, monitoringa in upravljanja s podatki, za našo vzorčno študijo znašajo:

$$\phi^P = \phi_{DG^P} + \phi_{M^P} + \phi_{DM^P} \approx 350,000 \text{ €}$$

$$\phi^E = \phi_{DG^E} + \phi_{M^E} + \phi_{DM^E} \approx 158,500 \text{ €}$$

$$\Delta\phi = \phi^P - \phi^E \approx 192,500 \text{ €}$$



Iz končnih rezultatov primerjave stroškov za vzorčno študijo je razvidno, da zaradi zmanjšanja obsega dela uporaba EDC rešitve prinaša 55 % prihrankov pri stroških dela na zbiranju podatkov klinične študije.

### *Stroški vzdrževanja sistema*

Vzdrževanje sistema v primeru zbiranja podatkov na papirju vključuje le stroške komunikacije oziroma tiskanja in pošiljanja dokumentov med akterji procesa. Za našo vzorčno študijo smo stroške vzdrževanja delovanja PDC sistema ocenili kot 15.800 € (Tabela 11).

Tabela 11. Stroški vzdrževanja PDC sistema (skupno število CRF obrazcev -  $N_{CRF} = 1.000$ ; število strani v CRF obrazcu -  $n_{p/CRF} = 50$  strani; delež podatkov za katere se pojavijo poizvedbe pri PDC -  $\varepsilon_{qry}^P = 2\%$ ; cena tiskanja ene strani -  $\varphi_p = 0,10$  €/stran; cena pošiljanja enega dokumenta -  $\varphi_s = 15$  €/doc;)

Strošek	Izračun	Vrednost
Poraba papirja za poizvedbe in odgovore	$N_{CRF} \times n_{p/CRF} \times \varepsilon_{qry}^P \times 2 \times \varphi_p$	200 €
Pošiljanje dokumentov (CRF obrazcev in poizvedb)	$N_{CRF} \times \varphi_s + N_{CRF} \times \varepsilon_{qry}^P \times 2 \times \varphi_s$	15.600 €
<b>SKUPAJ</b>		<b>15.800 €</b>

Po drugi strani je treba v primeru elektronskega zbiranja podatkov zagotoviti nemoteno delovanje EDC rešitve cel čas trajanja študije. Stroški vzdrževanja EDC sistema za vzorčno študijo smo ocenili na 8.000 € (Tabela 12).

Tabela 12. Stroški vzdrževanja EDC sistema

Strošek	Izračun	Vrednost
Vzdrževanje računalniške opreme	- 250 €/leto za podatkovni strežnik - 250 €/leto za spletni strežnik	1.000 €
Vzdrževanje programske opreme	- 250 €/leto za podatkovni strežnik - 250 €/leto za spletni strežnik	1.000 €
Vzdrževanje iEDC in podpora uporabnikom	3.000 €/leto	6.000 €
<b>SKUPAJ</b>		<b>8.000 €</b>

Iz zgornjih izračunov smo ugotovili, da so stroški vzdrževanja PDC sistema precej višji od stroškov vzdrževanja EDC sistema. V primeru vzorčne študije so ti stroški za PDC sistem dvakrat višji od stroškov za EDC sistem.

### III.5.2. Čas za končanje zbiranja podatkov

Kot smo predstavili v Metodah v poglavju »II.5 Ocenjevanje, Čas za končanje zbiranja podatkov«, je pri ocenjevanju časa potrebnega za končanje procesa zbiranja podatkov dovolj, če primerjamo le čas od zadnjega obiska monitorja do zaklepanja baze podatkov oziroma čas, potreben za zadnji cikel upravljanja s podatki (zadnjega preizkušanca).

Lahko predvidimo, da v primeru PDC procesa v povprečju potrebujemo en teden za vsako posredovanje informacij od enega uporabnika do drugega. Takšna posredovanja so

najmanj 3 (v primeru, da je podatkovni center zadovoljen z vsemi odgovori na poizvedbe), in sicer: CRF od monitorja do podatkovnega centra; poizvedbe od podatkovnega centra do monitorja in naprej do raziskovalnega centra; odgovor od raziskovalnega centra prek monitorja do podatkovnega centra. Za vso ostalo obdelavo podatkov (vnos v bazo, iskanje napak, popravila podatkov v bazi, analize in priprave odgovorov na poizvedbe) lahko predvidimo skupaj še en teden dela. Torej, v povprečju lahko pričakujemo, da bo pretekel najmanj en mesec od posredovanja podatkov zadnjega preizkušanca do zaklepanja baze podatkov. V literaturi se omenjajo še precej višje ocene, in sicer 2–3 mesece (Mary Ann Beyster et al., 2005; Oliveira & Salgado, 2006). Seveda je ta ocena zelo odvisna od števila podatkov v zadnjem ciklu, oddaljenosti raziskovalnega centra od naročnika študije kot tudi organizacije in hitrosti dela udeležencev v procesu.

Po drugi strani je v primeru EDC komunikacija skorajda instantna, tako da je čas potreben za zaklepanje podatkov le en teden upravljanja s podatki. V literaturi so te ocene še manjše in se gibljejo do vsega 3–4 dni (Mary Ann Beyster et al., 2005; Oliveira & Salgado, 2006). Če bi ta prihranjeni čas v primeru komercialne študije III faze poskušali preračunati v denar, upoštevajoč da en dan prodaje zdravila prinaša do 1 milijon € prihodkov, lahko bolj zgodnje pojavljanje zdravila (preverjenega prek klinične študije) na trgu pomeni tudi od 20 milijonov do skoraj 100 milijonov evrov prihrankov (oziroma prihodkov) (Brown, 2004). V našem primeru vzorčne študije bi prihranek zaradi zgodnjega časa končanja študije znašal okoli 20 milijonov €. Seveda je ta številka le hipotetična in pogojena s tem, da naročnik izkoristi čas prihranjen ob zbiranju podatkov klinične študije za zgodnejše pojavljanje zdravila na trgu, ter pod predpostavko, da prihodek od prodaje zdravila res lahko prinaša 1 milijon € na dan. Kakorkoli, omenjene številke kažejo na velik potencial uporabe tehnologije za zmanjšanje stroškov oziroma povečanje prihodkov iz razvoja in trženja novega zdravila.

## IV. Razprava

V Rezultatih smo predstavili celovit pristop k razvoju sistema za elektronsko zbiranje podatkov kliničnih študij. Naš sistem iEDC je prilagojen podpori procesa zbiranja podatkov, ki smo ga modelirali v eEPC tehnologiji. Sistem smo razvili, upoštevajoč metode načrtovanja uporabnosti. Na koncu smo izvedli primerjavo stroškov in časa, potrebnega za zbiranje podatkov za vzorčno klinično študijo v elektronski obliki in na papirju.

Kot netipično vlogo v procesu zbiranja podatkov smo vpeljali vlogo »internega monitorja«. Kot je razloženo v Rezultatih v poglavju »III.2.3 Predlog ključnih funkcionalnosti, **Error! Reference source not found.**«, je vloga internega monitorja, da nadomesti vlogo monitorja, kar je lahko primerno pri t.i. nekomercialnih kliničnih študijah. Pri takšnih študijah naročnik (pogosto sama klinična ustanova) študije ne dela z namenom registracije zdravila, ampak preverjanja nove metode zdravljenja (kombinacije zdravil ipd.). Denarna sredstva na razpolago so pogosto skromna, po drugi strani pa lahko izkoristi lastno osebje za doseganje večje kakovosti podatkov. Zato se naročnik lahko odloči, da med raziskovalce in njihove sodelavce, ki sodelujejo v študiji, razdeli naloge vnosa podatkov in preverjanja podatkov. Isti uporabnik lahko nekatera poglavja eCRF obrazcev izpolnjuje, druga pa preverja. Katera poglavja ali katere uporabnike bo kdo preverjal, lahko določi koordinator na začetku študije ali med samo študijo, glede na razpoložljivost in obremenjenost posameznih uporabnikov. Glede na dejstvo, da je praksa nekomercialnih študij pogosto takšna, da kakovost zbranih podatkov zaupajo samim raziskovalcem brez formalnega monitoringa, lahko uvajanje internega procesa monitoringa izboljša zanesljivost rezultatov študije.

Struktura elementov (mape in strani), ki smo jo predstavili v Rezultatih v poglavju »III.4.1 Model elektronskega CRF obrazca - eCRF, Opis strukturnih elementov«, omogoča izdelavo elektronskih obrazcev (eCRF) s sestavljanjem primerne strukture s pomočjo teh osnovnih gradnikov (različnih tipov map in strani). Takšna zasnova po eni strani pospešuje sestavljanje eCRF z uporabo administracijskega modula (Slika 17), po drugi strani pa omogoča ponovno uporabo gradnikov pri različnih študijah. Uporaba administracijskega modula uporabnika tudi približa informacijskemu pogledu na študijo. Ko študijo oblikujejo strokovnjaki s področja zdravstva, takšno orodje (administracijski modul) uporabnika (recimo koordinatorja) napelje k strukturirani obliki obrazcev, v kateri uporabnik prepozna ponavljajoče se elemente in jih zavestno uporablja v različnih korakih študije. Tako oblikovani strukturirani obrazci oziroma zbrani podatki so bolj enostavni za obdelavo in analizo. Aplikacijo se zato da lažje nadgraditi z dinamičnimi poročili in statistikami. Še več, administracijski modul bi lahko nadgradili tudi z možnostjo ustvarjanja vzorcev za dinamične statistike in poročila in s tem raziskovalcem še bolj približali delo informatikov.

Vseeno takšna funkcionalnost zahteva obsežno semantično analizo raziskovalnih pristopov v kliničnih študijah in ugotovitev vzorcev ter pravil pri zasnovi in analizi podatkov. To delo je precej zahtevno in predmet tako znanstvenih objav (Collada Ali et al., 2004; Dinakarandian, Lee, Vishwanath, & Lingambhotla, 2006; Rubin, Shah, & Noy, 2008; Ida Sim, Olasov, & Carini, 2004) kakor tudi strokovnih iniciativ, kot sta Trial Bank Project (<http://rctbank.ucsf.edu/>) in razvoj CDISC/ADaM standarda.

V Uvodu v poglavju »I.1 Opis področja, Veljavna zakonodaja in standardi« smo našli nekaj ključnih standardov na področju informatike v zdravstvu s posebnim poudarkom na CDISC naboru standardov in ODM (Operational Data Model) standardu, kot bistvenim za EDC sisteme. Med zasnovo našega sistema je bil ta standard še v fazi razvoja. Zato smo se sami lotili modeliranja in poskušali narediti čim bolj univerzalen strukturni model elementov. Pri rezultatih našega dela se je izkazalo, da je zelo podoben CDISC ODM standardu v pogledu strukturnih elementov. Zato smo mnenja, da bi naš sistem brez večjih težav prilagodili ODM standardu in s tem naš sistem naredili interoperabilen (v tej meri, kolikor je že ODM standard široko sprejet).

Izbor spletne tehnologije smo zasnovali na predpostavki, da je dostopnost do interneta dandanes podobna kot pri drugih komunikacijskih kanalih (pošta in telefonija), kar bo v prihodnjih letih zagotovo postalo dejstvo. Problemi se ne bodo pojavljali pri zagotavljanju internetnega dostopa, temveč pri zagotavljanju dostopa do spletne rešitve, ki je lahko oviran s strani lokalnih varnostnih politik različnih raziskovalnih centrov sodelujočih v klinični študiji. Ta problem ni tehnološke narave, ampak organizacijsko-politične narave, in glede na koristi, ki jih EDC lahko prinaša, verjamemo, tudi večinoma rešljiv. Sami smo pri uporabi prototipov vsakič naleteli na pozitiven odziv raziskovalnih centrov, kjer je vsak od njih zagotovil nemoten dostop do spletne EDC rešitve. Na splošno kaže razvoj spletnih rešitev na številnih področjih uporabe (bančništvo, spletne trgovine, socialne mreže, komunikacijske aplikacije, poizvedovanje itn.) na primerljivost, če že ne na dominantnost, spletnih tehnologij v pogledu prilagodljivosti, zanesljivosti in uporabnosti.

Primerjavo stroškov smo zasnovali na ocenah posameznih parametrov, ki se pojavljajo v modelih EDC in PDC procesov. Pri tem je jasno, da so točne vrednosti stroškov in trajanja teh procesov odvisne tako od samih modelov (procesov) kakor tudi od vrednosti upoštevanih parametrov. V naših izračunih smo upoštevali ocene deleža napak in poizvedb med podatki, ocene povprečnih naporov za izvajanje posameznih aktivnosti v procesih kot tudi cene dela akterjev v procesih. Ocene teh parametrov so zasnovane na podatkih pridobljenih v intervjujih kot tudi na podatkih objavljenih v strokovni literaturi. Kljub dejstvu, da te ocene niso najbolj natančne (bolj natančne bi bile eksperimentalne meritve posameznih parametrov) verjamemo, da naši rezultati dobro odražajo razlike med stroški in trajanjem PDC in EDC procesov. V nadaljevanju bomo diskutirali naše

predpostavke parametrov uporabljene v izračunih ter poskušali razumeti, kako variacije teh parametrov lahko vplivajo na izračune.

Nekatere med našimi ocenami deleža napak in poizvedb med podatki temeljijo na podatkih pridobljenih v intervjujih. Med njimi sta delež napak, ki jih odkrije monitor v PDC procesu in znaša okoli 10 %, ter delež poizvedb pri PDC, ki je okoli 2 % in je podoben podatkom objavljenim v strokovni literaturi (Spink, 2002). Oceno deleža napak in poizvedb za EDC smo podali na osnovi naših ocen za PDC in objavljenih študij s primerjavo vrednosti teh parametrov za PDC in EDC procese (Bart, 2003; Day et al., 1998; Kawado et al., 2003; King & Lashley, 2000; Kleinman, 2001; Spink, 2002; Wahi et al., 2008). Vse te študije potrjujejo, da elektronsko zbiranje podatkov zmanjšuje število napak med podatki in število poizvedb. Izboljšave so odvisne od kakovosti EDC rešitve, ki mora biti premišljeno zasnovana s ciljem doseganja uporabniku čim bolj prijazne rešitve. V kolikor je EDC rešitev slabo zasnovana, lahko pričakujemo večje število napak in posledično tudi poizvedb, kar zmanjšuje razliko med PDC in EDC stroški. Na primer, z dvakrat večjim deležem napak in poizvedb (v primerjavi s tistimi upoštevanimi v izračunih, predstavljenimi v Rezultatih v poglavju »III.5 Ocene, Stroški izvajanja procesa zbiranja podatkov«) EDC stroški narastejo s 158.500 € na 173.000 €, kar je 8,5 % povečanje. Posledično se skupni prihranki za našo vzorčno študijo ob uporabi EDC sistema zmanjšajo s 55 % (kot je predstavljeno v Rezultatih) na 45 %. Še vedno takšni prihranki niso zanemarljivi.

Velikost klinične študije vpliva tudi na stroške zbiranja podatkov. Na primer, za dvakrat »manjšo« študijo od naše vzorčne študije, v kateri bi sodelovalo 5 raziskovalnih centrov, vsak s 50 preizkušanci in 500 zbranimi podatki na preizkušanca, bi bili stroški PDC zbiranja podatkov znižani s 350.000 € na le 44.000 €. Kljub temu, da so prihranki EDC procesa še vedno 55 %, znesek 24.000 € prihrankov verjetno ne bi zadostoval za upravičenost investicije v EDC rešitev.

Verjamemo, da je za bolj natančno ugotovitev prihrankov ključnega pomena merjenje povprečnih naporov za posamezne naloge in njihovo upoštevanje v izračunih. Na primer v primeru, da so povprečni napor EDC nalog podobni povprečnim naporom PDC nalog (npr. vnos podatkov v elektronske obrazce zahteva enako časa kot vnos podatkov v papirnate obrazce), so izračuni nekoliko drugačni. In sicer, EDC stroški v tem primeru narastejo za 13 %, s 158.500 € na 179.000 €, kar znižuje prihranke s 55 % na 50 %. Kljub temu so takšni prihranki še vedno ugodni. Vsekakor pa za čim natančnejšo oceno predlagamo meritve posameznih parametrov in njihovo upoštevanje pri izračunih.

V Rezultatih v poglavju »III.1 Opis PDC procesa« in poglavju »III.2 Funkcionalni okvir rešitve, Predlog EDC procesa« smo predstavili eEPC modele PDC in EPC procesov. Čeprav smo pri zasnovi modelov upoštevali najbolj uveljavljene prakse, se lahko specifičnosti posameznih študij odražajo tudi na samih modelih. Ni redkost, da se naročnik študije

## IV. Razprava

odloči za zniževanje stroškov monitoringa na ta način, da omeji pregledovanje podatkov le na določen delež podatkov, namesto vseh podatkov vnesenih v obrazce. Na primer, to lahko stori s preverjanjem le kritičnih podatkov, kot so podatki o terapiji, rezultatih terapije in neželenih učinkih. Če so ti podatki npr. polovica vseh podatkov, potem v naših izračunih pridemo do zmanjšanja PDC stroškov za 32 % (s 350.000 € na 238.500 €) in EDC stroškov za 43 % (s 158.500 € na 90.000 €). Torej, v tem primeru bodo prihranki EDC procesa zmanjšani s 192.500 € na 148.500 €, čeprav se bo stopnja prihrankov povečala s 55 % na 62 %.

Stroški in prihranki za vse zgoraj naštetih scenarije so prikazani v spodnjih tabelah. Poleg stroškov in prihrankov za standardne vrednosti parametrov predstavljenih v Metodah in Rezultatih, tabele prikazujejo izračune še za:

- dvakratno vrednost deleža napak/poizvedb pri EDC;
- manjšo študijo (12 mesecev, 5 raziskovalnih centrov s po 50 preizkušanci in 500 podatki na preizkušanca);
- enake povprečne napore za PDC in EDC;
- le 50 % podatkov je pregledanih ob monitoringu.

Tabela 13. Stroški zajema podatkov in monitoringa za različne vrednosti parametrov ( $\phi_{DGP}$  - stroški PDC zajema podatkov;  $\phi_{DGE}$  - stroški EDC zajema podatkov;  $\Delta\phi_{DG}$  - razlika med stroški PDC in EDC zajema podatkov;  $\phi_{MP}$  - stroški PDC monitoringa;  $\phi_{ME}$  - stroški EDC monitoringa;  $\Delta\phi_M$  - razlika med stroški PDC in EDC monitoringa)

scenarij	$\phi_{DGP}$	$\phi_{DGE}$	$\Delta\phi_{DG}$	$\phi_{MP}$	$\phi_{ME}$	$\Delta\phi_M$
standardne vrednosti predstavljene v Metodah in Rezultatih	33.660 €	16.500 €	17.160 €	233.559 €	140.093 €	93.466 €
dvakratna vrednost deleža napak in poizvedb pri EDC	33.660 €	16.500 €	17.160 €	233.559 €	152.612 €	80.947 €
manjša študija (12 mesecev; 5 RC x 50 CRF x 500 podatkov)	4.290 €	2.063 €	2.228 €	29.210 €	17.526 €	11.683 €
enaki povprečni napor za PDC in EDC	33.660 €	33.000 €	660 €	233.559 €	142.592 €	90.967 €
le 50% podatkov je pregledano ob monitoringu	33.660 €	16.500 €	17.160 €	120.984 €	71.358 €	49.626 €

Tabela 14. Stroški upravljanja s podatki za različne vrednosti parametrov ( $\phi_{DM_{ENT}^P}$  - stroški PDC vnosa podatkov;  $\phi_{DM_{ANL}^P}$  - stroški PDC računalniške analize podatkov;  $\phi_{DM_{ANL}^E}$  - stroški EDC računalniške analize podatkov;  $\Delta\phi_{DM_{ANL}}$  - razlika med stroški PDC in EDC računalniške analize podatkov;  $\phi_{DM_{QRY}^P}$  - stroški PDC reševanja poizvedb;  $\phi_{DM_{QRY}^E}$  - stroški EDC reševanja poizvedb;  $\Delta\phi_{DM_{QRY}}$  - razlika med stroški PDC in EDC reševanja poizvedb)

scenarij	$\phi_{DM_{ENT}^P}$	$\phi_{DM_{ANL}^P}$	$\phi_{DM_{ANL}^E}$	$\Delta\phi_{DM_{ANL}}$	$\phi_{DM_{QRY}^P}$	$\phi_{DM_{QRY}^E}$	$\Delta\phi_{DM_{QRY}}$
standardne vrednosti predstavljene v Metodah in Rezultatih	26.099 €	18.034 €	835 €	17.199 €	20.713 €	1.000 €	19.713 €
dvakratna vrednost deleža napak in poizvedb pri EDC	26.099 €	18.034 €	1.670 €	16.364 €	20.713 €	2.000 €	18.713 €
manjša študija (12 mesecev; 5 RC x 50 CRF x 500 podatkov)	3.262 €	2.279 €	104 €	2.175 €	2.673 €	125 €	2.548 €
enaki povprečni napor za PDC in EDC	26.099 €	18.034 €	1.670 €	16.364 €	20.713 €	1.500 €	19.213 €
le 50% podatkov je pregledano ob monitoringu	26.099 €	18.034 €	835 €	17.199 €	20.713 €	1.000 €	19.713 €

#### IV. Razprava

Tabela 15. Stroški upravljanja s podatki za različne vrednosti parametrov - nadaljevanje ( $\phi_{DM^{P_{CLO}}}$  – stroški PDC zaključevanja podatkov;  $\phi_{DM^{E_{CLO}}}$  – stroški EDC zaključevanja podatkov;  $\Delta\phi_{DM^{CLO}}$  – razlika med stroški PDC in EDC zaključevanja podatkov;  $\phi_{DM^P}$  – stroški PDC upravljanja s podatki;  $\phi_{DM^E}$  – stroški EDC upravljanja s podatki;  $\Delta\phi_{DM}$  – razlika med stroški PDC in EDC upravljanja s podatki)

scenarij	$\phi_{DM^{P_{CLO}}}$	$\phi_{DM^{E_{CLO}}}$	$\Delta\phi_{DM^{CLO}}$	$\phi_{DM^P}$	$\phi_{DM^E}$	$\Delta\phi_{DM}$
standardne vrednosti predstavljene v Metodah in Rezultatih	19.125 €	209 €	18.917 €	83.971 €	2.044 €	81.928 €
dvakratna vrednost deleža napak in poizvedb pri EDC	19.125 €	417 €	18.708 €	83.971 €	4.087 €	79.884 €
manjša študija (12 mesecev; 5 RC x 50 CRF x 500 podatkov)	2.391 €	26 €	2.365 €	10.605 €	255 €	10.350 €
enaki povprečni napor za PDC in EDC	19.125 €	417 €	18.708 €	83.971 €	3.587 €	80.384 €
le 50% podatkov je pregledano ob monitoringu	19.125 €	209 €	18.917 €	83.971 €	2.044 €	81.928 €

Tabela 16. Skupni stroški in prihranki za različne vrednosti parametrov ( $\phi^P$  - PDC stroški;  $\phi^E$  - EDC stroški;  $\Delta\phi$  - razlika med PDC in EDC stroški)

scenarij	$\phi^P$	$\phi^E$	$\Delta\phi$	EDC prihranki
standardne vrednosti predstavljene v Metodah in Rezultatih	351.190 €	158.637 €	192.554 €	55%
dvakratna vrednost deleža napak in poizvedb pri EDC	351.190 €	173.199 €	177.991 €	51%
manjša študija (12 mesecev; 5 RC x 50 CRF x 500 podatkov)	44.105 €	19.844 €	24.261 €	55%
enaki povprečni napor za PDC in EDC	351.190 €	179.179 €	172.011 €	49%
le 50% podatkov je pregledano ob monitoringu	238.615 €	89.902 €	148.713 €	62%

Na osnovi izračunov prikazanih v Tabelah 13-16 lahko ugotovimo, da EDC proces prinaša prihranke v primerjavi z PDC procesom v vseh petih scenarijih. Največ prihrankov je posledica manjšega deleža napak in poizvedb, ki posledično zahtevajo manj naporov za odpravljanje nepravilnosti med podatki. Drugi pomemben dejavnik, ki vpliva na stroške zbiranja podatkov je velikost študije (število raziskovalnih centrov, preizkušancev, zbranih podatkov). Iz naših primerjav sledi, da so prihranki samega procesa zbiranja podatkov nezadostni, da bi upravičili stroške investicije v EDC sistem. Kakorkoli, z uporabo naših modelov naročnik študije lahko izračuna ocene stroškov svoje študije v obeh primerih zbiranja podatkov (PDC in EDC), kar mu lahko pomaga pri izboru primernega načina zbiranja podatkov in upravičevanju investicij v EDC sistem.

Modeli, ki smo jih v naših izračunih upoštevali, zajemajo le procese zbiranja podatkov klinične študije. Pomembno se je zavedati, da prehod na elektronsko zbiranje podatkov za naročnika študije pomeni tudi spremembe v organizaciji in izvajanju same študije. EDC pomembno zmanjšuje stroške uporabe papirja. Pri upravljanju študije si lahko koordinator pomaga z ažurnimi podatki, ki jih zagotavlja on-line EDC sistem. Tako lahko naroča obiske monitorjev le takrat, ko je zadostno število obrazcev izpolnjeno, namesto obiskov po vnaprej določenem časovnem planu, ali pa opozori kakšnega od centrov na zamude pri izpolnjevanju podatkov. Resni nezaželeni učinki (angl. »Serious Adverse Events« - SAE) evidentirani v EDC sistemu lahko avtomatično sprožijo mehanizme za SAE poročanje in tako izboljšajo varnost študije. Končno, bistvene prihranke naročniku študije lahko prinese skrajšanje časa od zaključka zadnjega preizkušanca do zaklepanja podatkovne baze študije, ki (ob uspešnem zaključku študije) pripomore k hitrejšemu začetku trženja zdravila.

Pomembno je omeniti tudi pomanjkljivosti EDC pristopa v primerjavi s PDC pristopom. Na primer, uporaba EDC sistema prinaša stroške v zvezi z vzdrževanjem sistema in podporo uporabnikom. Morebitne strojne ali omrežne omejitve (npr. požarni zid, delna ali popolna nedostopnost do spleta, ali počasna spletna povezava v nekaterih od raziskovalnih centrov itn.) lahko zahtevajo dodatne stroške za njihovo reševanje. Uporaba EDC sistema je v določenih situacijah otežena ali nemogoča, kot na primer pri obiskih pri bolniku na domu, ali med samim kliničnim posegom. Dodatne težave se lahko pojavijo zaradi morebitnih zakonskih omejitev vezanih na elektronske vire podatkov (angl. »eSource«) (The eClinical Forum & PhRMA EDC/eSource Taskforce, 2006). Končno, prehod s papirnega na elektronski proces zbiranja podatkov za naročnika študije in udeležence študije lahko predstavlja velik organizacijski izziv. Bolj poglobljeno razpravo na to temo je pred kratkim v svojem prispevku objavil J.A.Welker (Welker, 2007), v katerem poudarja, da je v večjih podjetjih brez kakovostnega tima, ki bo proces implementacije rešitve izvedel, uvajanje EDC rešitve, kljub najboljši kakovosti same rešitve, obsojeno na propad.

V naši nalogi smo obravnavali dva »skrajna« pristopa – popolnoma papirni proces in popolnoma elektronski proces. Pogosto se zgodi, da znotraj ene študije potrebujemo mešanico različnih pristopov – nekateri centri zbirajo podatke na papirnatih obrazcih, drugi na elektronskih. Podatkovni center mora v tem primeru združiti in obdelati podatke iz različnih virov (papirnatih in elektronskih), zaradi česar so tudi modeli procesov bolj zapleteni. Posledično se lahko zgodi, da prihranki ob uporabi EDC sistema ne upravičujejo vzdrževanja tako kompleksnega hibridnega PDC/EDC procesa.

Vse predstavljene pomanjkljivosti prehoda na EDC sistem morajo biti uravnotežene s prednostmi in prihranki uporabe EDC sistema pri posamezni študiji. Le na ta način lahko upravičimo prehod s PDC procesa na EDC proces. V naši nalogi smo pokazali, da uporaba modelov procesov skupaj z natančnimi ocenami parametrov procesov pripomore k identifikaciji in oceni prihrankov ob prehodu na EDC proces. Pri tem je pomembno poudariti, da naši modeli ne vključujejo vseh aspektov organizacije in implementacije klinične študije, na katere vpliva sprememba načina zbiranja podatkov. Upoštevajoč dejstvo, da samo zbiranje podatkov ene študije faze III lahko predstavlja 30 % vseh stroškov študije (glede na primer Fast Track Systems 16.000.000 \$ od skupaj 58.400.000 \$ (BIO-IT World, 2003)), je optimizacija tega procesa z uporabo EDC sistema z namenom zniževanja stroškov zelo smiselna. Še več, glede na podatke objavljene v »PhRMA Membership Annual Report 2008« (Pharmaceutical Research and Manufacturers of America, 2008) stroški izvajanja kliničnih študij predstavljajo 47,9 % celotne investicije v raziskavo in razvoj novega zdravila (faza III – 28,1 %, faza II – 13,1 %, faza I – 6,7 %), kar pomeni, da zniževanje stroškov kliničnih študij lahko pomembno zniža stroške investicije v novo zdravilo.



## V. Sklep

Pomemben segment organizacije klinične študije je zbiranje podatkov. Iz zgodovinskih, operativnih in tehnoloških razlogov v velikem številu kliničnih študij zbiranje podatkov še vedno poteka na papirju. Alternativa zbiranju podatkov na papirju (PDC) je elektronsko zbiranje podatkov (EDC). To prinaša številne prednosti predvsem v pogledu upravljanja s podatki. Za odločitev o prehodu na elektronsko zbiranje podatkov je treba upoštevati več različnih dejavnikov, med katerimi so ključni: uporabnost rešitve za elektronsko zbiranje podatkov (EDC rešitev), stroški elektronskega zbiranja podatkov v primerjavi z zbiranjem podatkov na papirju, ter tehnološke in organizacijske omejitve in prednosti uporabe EDC rešitve.

S popisom in analizo procesov in podprocesov zbiranja podatkov je mogoče ugotoviti aktivnosti, katerim je mogoče zagotoviti podporo s primerno EDC rešitvijo. S pristopom, ki upošteva metode načrtovanja uporabnosti, je mogoče razviti uporabniku prijazno EDC rešitev. Poleg tega omogoča modeliranje procesov izračune stroškov procesov in simulacijske študije, ki pripomorejo k razumevanju ključnih parametrov, ki vplivajo na stroške, in jih je mogoče upoštevati pri odločanju o ustreznem pristopu k zbiranju podatkov (PDC ali EDC) za posamezno klinično študijo.

V okviru disertacije smo ob upoštevanju metod načrtovanja uporabnosti zasnovali spletno rešitev za zbiranje podatkov kliničnih študij (iEDC), namenjeno predvsem multicentričnim kliničnim preskušanjem. Ob tem smo predlagali prenavo procesov zbiranja podatkov iz PDC procesov v EDC procese. Prenovljene procese smo zasnovali tako, da je mogoča zamenjava oziroma podvajanje določenih vlog v sistemu (raziskovalec in monitor), z namenom reduciranja upravljanja s podatki in stroškov zbiranja podatkov.

V naši nalogi smo pokazali, kako prehod s papirnega na spletno zasnovano elektronsko zbiranje podatkov vpliva na stroške zbiranja podatkov in njegovih podprocesov. Za ta namen smo razvili eEPC (Extended Event-driven Process Chains) modele PDC in EDC procesov in podprocesov (zajem podatkov, monitoring in upravljanje s podatki) in jih poenostavili, da bi lahko izračunali ocene stroškov. Podali smo ocene vrednosti parametrov, ki se v modelih pojavljajo, med katerimi so: velikost klinične študije, delež napak in poizvedb med podatki, povprečni napor za izvajanje posameznih nalog ter cene osebja. Pri tem smo naše ocene zasnovali na objavljeni literaturi in informacijah, pridobljenih v intervjujih.

Naši rezultati kažejo, da do večine prihrankov ob uporabi EDC sistema prihaja zaradi zmanjšanja stroškov monitoringa in upravljanja s podatki. Točne vrednosti prihrankov so

odvisne od parametrov, ki se v izračunih pojavljajo. Na primer, velikost klinične študije pomembno vpliva na prihranke, ki jih EDC sistem prinaša, medtem ko so stroški manj občutljivi na spremembe vrednosti povprečnih naporov za izvajanje posameznih nalog. Takšni rezultati niti niso presenetljivi, ampak naš pristop omogoča, da jih kvantificiramo. Z natančnejšimi ocenami parametrov za posamezno študijo lahko pridemo do zanesljivih ocen prihrankov ob prehodu s PDC na EDC sistem, zaradi katerih se lažje odločamo za takšen prehod. Z enakim pristopom, kot smo ga predstavili v naši nalogi, je mogoče narediti tudi celotno oceno stroškov klinične študije, z modeliranjem klinične študije kot enotnega poslovnega procesa.

## VI. Prispevki k znanosti

V doktorski nalogi smo predstavil naslednje prispevke k znanosti:

- **Spletno zasnovan sistem za zbiranje podatkov kliničnega preskušanja z reduciranim upravljanjem podatkov**

Razvili smo sistem za zbiranje podatkov klinične študije zasnovan na spletni tehnologiji (iEDC), ki v celoti podpira proces zbiranja podatkov klinične študije. Sistem je prilagodljiv za različne študije in omogoča razdelitev vlog med različnimi uporabniki sistema (raziskovalec, monitor, koordinator), pri čemer je z vpeljavo vloge internega monitoringa omogočena redukcija upravljanja podatkov. Pri zasnovi sistema smo upoštevali obstoječe zakonske akte in priporočila na področju kliničnih preskušanj in elektronskega zbiranja podatkov. Z uporabo metod načrtovanja uporabnosti smo razvili uporabniku prijazno iEDC rešitev. Skupaj z novim procesom vnosa in preverjanja podatkov iEDC sistem prispeva k boljši kakovosti zbranih podatkov in h krajšemu trajanju študije glede na klasično zbiranje podatkov.

- **Model procesov in podprocesov za zbiranja podatkov kliničnih študij v elektronski obliki**

Na osnovi analize pregleda literature, spletne ankete in intervjujev smo s pomočjo tehnike eEPC (Extended-driven Process Chains) naredili model procesa zbiranja podatkov klinične študije na papirju (PDC proces). PDC proces smo razbili na podproces upravljanja klinične študije in tri podprocese zbiranja podatkov, in sicer: PDC Zajem podatkov, PDC Monitoring in PDC Upravljanje s podatki. Vsak od teh podprocesov je podrobno opisan z zaporedjem aktivnosti (funkcij) in dogodkov ter dokumenti in vlogami, ki nastopajo pri posameznih aktivnostih procesa.

Izhajajoč iz modela PDC procesa smo za vsakega od identificiranih podprocesov predlagali nov podproces (EDC proces), ki vključuje enotno računalniško podporo celotnemu procesu zbiranja podatkov klinične študije. Predlog EDC procesa smo podali v obliki modelov v eEPC tehniki, in sicer po enega za vsak posamezen podproces elektronskega zbiranja podatkov: EDC Zajem podatkov, EDC Monitoring in EDC Upravljanje s podatki. Ker je predlagan EDC proces fleksibilen glede števila udeležencev v procesu, ga je mogoče uporabljati tudi v primeru majhnega števila raziskovalcev v raziskovalnem centru.

▪ **Model procesov za analizo stroškov procesa zbiranja podatkov za izvajanje kliničnih študij**

Z namenom ocenjevanja prednosti in pomanjkljivosti prehoda z zbiranja podatkov na papirju (PDC) na elektronsko zbiranje podatkov (EDC) smo predlagali izvirno metodo za oceno stroškov PDC in EDC procesov. Prvi korak v predlagani metodi je izdelava modela PDC in EDC procesov v eEPC tehniki. Zatem je potrebno za vsako od aktivnosti (funkcij) v procesu določiti po eno enačbo za izračun stroškov. Te funkcije praviloma vsebujejo določene parametre procesa, kot so denimo cene dela, povprečni napori za izvajanje aktivnosti in kakovost podatkov. Z natančno oceno parametrov procesov je mogoče priti do natančnih izračunov stroškov in trajanja celotnega procesa. Z variacijami parametrov je mogoče narediti simulacijsko študijo, ki lahko pokaže, kateri parametri bistveno vplivajo na ocene stroškov EDC oziroma PDC procesa.

## VII. Literatura

- Al-Shahi, R., & Warlow, C. (2000). Using patient-identifiable data for observational research and audit. *BMJ (Clinical Research Ed.)*, 321(7268), 1031-1032.
- Atoji, C. (2007, September). Nextrials' Founder Discusses the (Hoped-for) Collision Between EDC and EHRs. *Digital HealthCare & Productivity.com*. Retrieved from <http://www.digitalhcp.com/hitw/newsletters/2007/09/11/costello/>
- Bart, T. (2003). Comparison of Electronic Data Capture with Paper Data Collection – Is There Really an Advantage? In *Business Briefing: PharmaTech 2003*, Business Briefing: PharmaTech (5th ed.). Business Briefing Ltd. Retrieved from <http://www.touchbriefings.com/cdps/cditem.cfm?NID=17>
- Beuscart-Zéphir, M., Anceaux, F., Menu, H., S, Guerlinger, R., Watbled, L., & Evrard, F. (2005). User-centred, multidimensional assessment method of Clinical Information Systems: a case-study in anaesthesiology. *International Journal of Medical Informatics*, 74(2-4), 179-189. doi:10.1016/j.ijmedinf.2004.07.005
- BIO-IT World. (2003, May 9). Where the Millions Go. Retrieved from [http://www.bio-itworld.com/archive/050903/data\\_sidebar\\_2450.html](http://www.bio-itworld.com/archive/050903/data_sidebar_2450.html)
- Bleicher, P. (2003). Clinical trial technology: at the inflection point. *BIOSILICO*, 1(5), 163-168. doi:10.1016/S1478-5382(03)02373-4
- Brandt, C. A., Nadkarni, P., Marenco, L., Karras, B. T., Lu, C., Schacter, L., Fisk, J. M., et al. (2000). Reengineering a Database for Clinical Trials Management: Lessons for System Architects. *Controlled Clinical Trials*, 21(5), 440-461. doi:10.1016/S0197-2456(00)00070-2
- Brown, E. G. (2004, September). Clinical Trials' EDC Endgame. Forrester Research ,Inc. Retrieved from <http://www.forrester.com/Research/Document/Excerpt/0,7211,33344,00.html>
- Christiansen, D. H., Hosking, J. D., Dannenberg, A. L., & Williams, O. D. (1990). Computer-assisted data collection in multicenter epidemiologic research. The Atherosclerosis Risk in Communities Study. *Controlled Clinical Trials*, 11(2), 101-115. doi:2161309
- Collada Ali, L., Fazi, P., Luzi, D., Ricci, F. L., Serbanati, L. D., & Vignetti, M. (2004). Toward a Model of Clinical Trials. In *Biological and Medical Data Analysis* (pp. 299-312). Retrieved from <http://www.springerlink.com/content/y3yxru6pxhfpw015>
- Day, S., Fayers, P., & Harvey, D. (1998). Double data entry: what value, what price? *Controlled Clinical Trials*, 19(1), 15-24. doi:9492966
- Dinakarpanian, D., Lee, Y., Vishwanath, K., & Lingambhotla, R. (2006). MachineProse: an Ontological Framework for Scientific Assertions. *J Am Med Inform Assoc*, 13(2), 220-232. doi:10.1197/jamia.M1910
- DZ RS. (2004). Zakon o varstvu osebnih podatkov (ZVOP-1). *Uradni list RS*, (86/2004).

- Eich, H. P., & Ohmann, C. (1999). Generalisation and extension of a web-based data collection system for clinical studies using Java and CORBA. *Studies in Health Technology and Informatics*, 68, 568-72. doi:10724953
- Evropska komisija. (2005). Direktiva Komisije 2005/28/ES o načelih in podrobnih smernicah za dobro klinično prakso v zvezi z zdravili v preskušanju za humano uporabo ter o zahtevah za pridobitev dovoljenja za proizvodnjo ali uvoz takšnih izdelkov. *Uradni list Evropske Skupnosti, UL L 91*, 13-19.
- Evropski parlament in Svet. (2001). Direktiva 2001/20/ES o približevanju zakonov in drugih predpisov držav članic v zvezi z izvajanjem dobre klinične prakse pri kliničnem preskušanju zdravil za ljudi. *Uradni list Evropske Skupnosti, UL L 121*(posebna izdaja v slovenščini poglavje 13 zvezek 26), 299 - 309.
- FDA. (1997). FDA Title 21 CFR Part 11:Electronic Records; Electronic Signatures; Final Rule. FDA. Retrieved from [http://www.fda.gov/ora/compliance\\_ref/part11/frs/background/11cfr-fr.htm](http://www.fda.gov/ora/compliance_ref/part11/frs/background/11cfr-fr.htm)
- FDA. (2007, May). Guidance for Industry Computerized Systems Used in Clinical Investigations. FDA. Retrieved from <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070266.pdf>
- Gerritsen, M. G., Sartorius, O. E., vd Veen, F. M., & Meester, G. T. (1993). Data management in multi-center clinical trials and the role of a nation-wide computer network. A 5 year evaluation. *Proceedings / the ... Annual Symposium on Computer Application [sic] in Medical Care. Symposium on Computer Applications in Medical Care*, 659-62. doi:8130557
- Hollingsworth, R. A., Hay, C., & Richards, B. (1999). An Internet implementation of an international clinical study. *Studies in Health Technology and Informatics*, 68, 528-31. doi:10724944
- Johnson, C. M., Johnson, T. R., & Zhang, J. (2005). A user-centered framework for redesigning health care interfaces. *Journal of Biomedical Informatics*, 38(1), 75-87. doi:10.1016/j.jbi.2004.11.005
- Jørgensen, C. K., & Karlslose, B. (1998). Validation of automated forms processing. A comparison of Teleform with manual data entry. *Computers in Biology and Medicine*, 28(6), 659-67. doi:9878977
- Kalra, D., Gertz, R., Singleton, P., & Inskip, H. M. (2006). Confidentiality of personal health information used for research. *BMJ (Clinical Research Ed.)*, 333(7560), 196-198. doi:10.1136/bmj.333.7560.196
- Karlberg, J. P. (2008). Industry Clinical Testing of New Medicinal Products Requires 95,000 Study Sites and 1,300,000 Subjects Annually. *Clinical Trial Magnifier*, 1(6), 101-109.
- Kawado, M., Hinotsu, S., Matsuyama, Y., Yamaguchi, T., Hashimoto, S., & Ohashi, Y. (2003). A comparison of error detection rates between the reading aloud method and the double data entry method. *Controlled Clinical Trials*, 24(5), 560-569. doi:10.1016/S0197-2456(03)00089-8

- Kelly, M., & Oldham, J. (1997). The Internet and randomised controlled trials. *International Journal of Medical Informatics*, 47(1-2), 91-99. doi:10.1016/S1386-5056(97)00091-9
- King, D. W., & Lashley, R. (2000). A Quantifiable Alternative to Double Data Entry. *Controlled Clinical Trials*, 21(2), 94-102. doi:10.1016/S0197-2456(00)00042-8
- Kiuchi, T., & Kaihara, S. (1997). Automated generation of a World Wide Web-based data entry and check program for medical applications. *Computer Methods and Programs in Biomedicine*, 52(2), 129-138. doi:10.1016/S0169-2607(96)01793-2
- Kiuchi, T., Ohashi, Y., Konishi, M., Bandai, Y., Kosuge, T., & Kakizoe, T. (1996). A world wide web-based user interface for a data management system for use in multi-institutional clinical trials--Development and experimental operation of an automated patient registration and random allocation system. *Controlled Clinical Trials*, 17(6), 476-493. doi:10.1016/S0197-2456(96)00104-3
- Kleinman, K. (2001). Adaptive Double Data Entry: A Probabilistic Tool for Choosing Which Forms to Reenter. *Controlled Clinical Trials*, 22(1), 2-12. doi:10.1016/S0197-2456(00)00116-1
- Koyani, S. J., Bailey, R. W., & Nall, J. R. (2004). *Research-Based Web Design & Usability Guidelines*. Computer Psychology.
- Krischer, J. P., Hurley, C., Pillalamarri, M., Pant, S., Bleichfeld, C., Opel, M., & Shuster, J. J. (1991). An automated patient registration and treatment randomization system for multicenter clinical trials. *Controlled Clinical Trials*, 12(3), 367-77. doi:1651208
- Kush, R. (2006). The Future for Electronic Data Capture. *Drug Development*, 46-47.
- Kushniruk, A. W., Patel, C., Patel, V. L., & Cimino, J. J. (2001). 'Televaluation' of clinical information systems: an integrative approach to assessing Web-based systems. *International Journal of Medical Informatics*, 61(1), 45-70. doi:10.1016/S1386-5056(00)00133-7
- Kushniruk, A. W., & Patel, V. L. (2004). Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of Biomedical Informatics*, 37(1), 56-76. doi:10.1016/j.jbi.2004.01.003
- Lecerof, A., & Paterno, F. (1998). Automatic support for usability evaluation. *Software Engineering, IEEE Transactions on*, 24(10), 863-888. doi:10.1109/32.729686
- Leiner, F., Gaus, W., Haux, R., Knaup-Gregori, P., & Wagner, G. (2003). *Medical Data Management* (1st ed.). Springer.
- Mary Ann Beyster, David H. Hardison, & Gary M. Lubin. (2005, April). Improving Clinical Trials by Implementing Information Technology (IT): Where will you be in three years? Science Applications International Corporation. Retrieved from <http://www.saic.com/life%2Dsciences/pdfs/mcv.pdf>
- Mastro, D. (2007). Survey Results: EDC or Paper, the Internet or the Post Office, or All of the Above. *The Monitor (ACRP)*, 81-83.
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.

- Nielsen, J. (2000, March 19). Why You Only Need to Test with 5 Users. *Usability Testing with 5 Users (Jakob Nielsen's Alertbox)*. Retrieved August 13, 2009, from <http://www.useit.com/alertbox/20000319.html>
- Oliveira, A. G., & Salgado, N. C. (2006). Design aspects of a distributed clinical trials information system. *Clinical Trials (London, England)*, 3(4), 385-96. doi:3/4/385
- Paul, J., Seib, R., & Prescott, T. (2005). The Internet and Clinical Trials: Background, Online Resources, Examples and Issues. *Journal of Medical Internet Research*, 7(1), e5.
- Pavlovic, I., & Miklavcic, D. (2007). Web-Based Electronic Data Collection System to Support Electrochemotherapy Clinical Trial. *Information Technology in Biomedicine, IEEE Transactions on*, 11(2), 222-230. doi:10.1109/TITB.2006.879581
- Pavlović, I. (2004). *Web-based medical record system for clinical trials* (M.S. thesis). University of Ljubljana, Slovenia.
- Pepine, C. J., Handberg-Thurmond, E., Marks, R. G., Conlon, M., Cooper-DeHoff, R., Volkens, P., & Zellig, P. (1998). Rationale and design of the International Verapamil SR/Trandolapril Study (INVEST): an Internet-based randomized trial in coronary artery disease patients with hypertension. *Journal of the American College of Cardiology*, 32(5), 1228-37. doi:9809930
- Pharmaceutical Research and Manufacturers of America. (2008, March). Pharmaceutical Industry Profile 2008. PhRMA, Washington, DC. Retrieved from <http://www.phrma.org/files/2008%20Profile.pdf>
- Prokscha, S. (1999). *Practical Guide to Clinical Data Management* (1st ed.). CRC.
- Rubin, D., Shah, N., & Noy, N. (2008). Biomedical ontologies: a functional perspective. *Brief Bioinform*, 9(1), 90, 75.
- Scheer, A. (2000). *ARIS - Business Process Modeling* (3rd ed.). Springer.
- Schriver, K. (1989). Evaluating text quality: the continuum from text-focused to reader-focused methods. *Professional Communication, IEEE Transactions on*, 32(4), 238-255. doi:10.1109/47.44536
- Shneiderman, B., & Plaisant, C. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (4th ed.). Addison Wesley.
- Sim, I., Owens, D. K., Lavori, P. W., & Rennels, G. D. (2000). Electronic trial banks: a complementary method for reporting randomized trials. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 20(4), 440-450.
- Sim, I., Olasov, B., & Carini, S. (2004). An ontology of randomized controlled trials for evidence-based practice: content specification and evaluation using the competency decomposition method. *Journal of Biomedical Informatics*, 37(2), 108-119. doi:10.1016/j.jbi.2004.03.001
- Spink, C. (2002, March). Electronic Data Capture (EDC) as a means for e-clinical trial success. IBM Global Services, Pharmaceutical Clinical Development.



- Strang, W. N., Cucherat, M., Yzebe, D., & Boissel, J. (2000). Trial summary software. *Computer Methods and Programs in Biomedicine*, 61(1), 49-60. doi:10.1016/S0169-2607(98)00094-7
- Strobl, J., Cave, E., & Walley, T. (2000). Data protection legislation: interpretation and barriers to research. *BMJ (Clinical Research Ed.)*, 321(7265), 890-892.
- Svet Evrope. (1965). Direktiva 93/39/EGS evropskega parlamenta in Sveta. *UL*, 22(1), 65.
- The eClinical Forum, & PhRMA EDC/eSource Taskforce. (2006, September 14). The Future Vision of Electronic Health Records as eSource for Clinical Research. The eClinical Forum and PhRMA EDC/eSource Taskforce. Retrieved from <http://www.eclinicalforum.com/content/Knowledge/Articles/Future%20EHR-CR%20Environment%20Version%201.pdf>
- Virzi, R. A. (1992). Refining the test phase of usability evaluation: how many subjects is enough? *Hum. Factors*, 34(4), 457-468.
- Vlada RS. (2006). Zakon o zdravilih. *Uradni list RS*, (31), 54. člen, stav 1.
- van Waes, L. (2000). Thinking aloud as a method for testing the usability of Websites: the influence of task variation on the evaluation of hypertext. *Professional Communication, IEEE Transactions on*, 43(3), 279-291. doi:10.1109/47.867944
- Wahi, M. M., Parks, D. V., Skeate, R. C., & Goldin, S. B. (2008). Reducing Errors from the Electronic Transcription of Data Collected on Paper Forms: A Research Data Case Study. *J Am Med Inform Assoc*, 15(3), 386-389. doi:10.1197/jamia.M2381
- Weber, G., Kassai, M., Csontos, Z., Czuczor, C., & Horváth, P. O. (1999). First Hungarian, internet-based prospective, multicenter study: the hernia-project. *Acta Chirurgica Hungarica*, 38(2), 219-20. doi:10596334
- Welker, J. A. (2007). Implementation of electronic data capture systems: Barriers and solutions. *Contemporary Clinical Trials*, 28(3), 329-336. doi:10.1016/j.cct.2007.01.001
- Wübbelt, P., Fernandez, G., & Heymer, J. (2000). Clinical trial management and remote data entry on the Internet based on XML case report forms. *Studies in Health Technology and Informatics*, 77, 333-7. doi:11187568



## VIII. Seznam kratic

CDISC - Clinical Data Interchange Standards Consortium, ključna iniciativa za standardizacijo izmenjave podatkov kliničnih študij

CRF - obrazci namenjeni zbiranju podatkov pacienta v okvirju klinične študije (angl. »Case Report Forms«)

CRO – pogodbeno organizacija za organizacijo in nadzor kliničnih študij (angl. »Contract Research Organization«, »Clinical Research Organization«)

eCRF - elektronski CRF obrazec

EDC - elektronsko zbiranje podatkov (angl. »Electronic Data Collection«)

eEPC - Extended Event-driven Process Chains, široko sprejeta tehnika za opisovanje procesov

GCP - dobra klinična praksa, nabor navodil za dobro organizacijo kliničnega preskušanja (angl. »Good Clinical Practice«)

HL7 - Health Level 7, nabor široko sprejetih standardov za izmenjavo elektronskih sporočil v zdravstvenih sistemih

ODM - Operational Data Model, bistveni CDISC standard za EDC sisteme

PDC - proces zbiranja podatkov na papirju (angl. »Paper Data Collection«)

RDC - oddaljeno zbiranje podatkov (angl. »Remote Data Capture«)



## IX. Priloge

Anketna vprašanja in rezultati anket, ki sem jih upošteval ob načrtovanju EDC rešitve

- Priloga A – Anketna vprašanja in rezultati ankete za farmacevtska podjetja
- Priloga B – Anketna vprašanja in rezultati ankete za CRO organizacije
- Priloga C – Anketna vprašanja in rezultati ankete za klinične ustanove

Objavljeni članki, katerih vsebina vključuje rezultate raziskovalnega dela opisanega v doktorski disertaciji

- Pavlovic, I., & Miklavcic, D. (2007). Web-Based Electronic Data Collection System to Support Electrochemotherapy Clinical Trial. *Information Technology in Biomedicine, IEEE Transactions on*, 11(2), 222-230. doi: 10.1109/TITB.2006.879581.
- Pavlović, I., Kern, T., & Miklavcic, D. (2009). Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study. *Contemporary Clinical Trials*, 30(4), 300-316. doi: 10.1016/j.cct.2009.03.008.



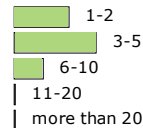
## PRILOGA A

### **Anketna vprašanja in rezultati ankete za farmacevtska podjetja**

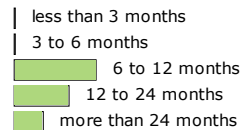




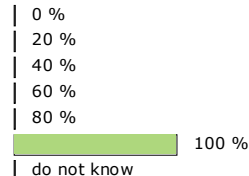
On average, how many clinical trials are started by your organization yearly (in the last two years)?



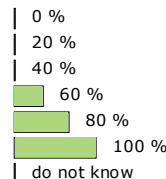
On average, what is the duration of clinical trials that are organized by your organization?



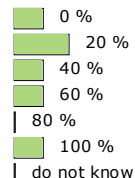
What is the percentage of multicentric clinical trials among all the trials that are organized by your organization?



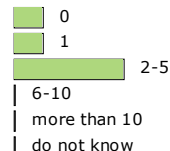
What is the percentage of international clinical trials among all the trials that are organized by your organization?



What is the percentage of the clinical trials organized by CRO among all the trials that are sponsored by your company?

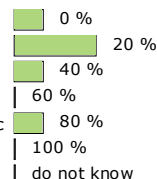


In the last five years, how many different CROs has your organization cooperated with?

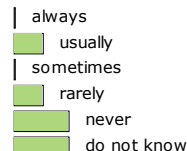


What is the percentage of clinical trials that use EDC (Electronic Data Capture) among all the trials that are sponsored by your organization?

[EDC (Electronic Data Capture) is the way of data collecting where the investigators/researchers enter data directly into the computer (using electronic forms) by themselves. If necessary, investigator can prior to that also collect data on the paper forms or print the completed paper forms afterwards.]

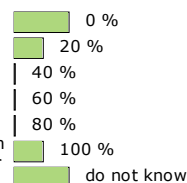


In the case of outsourcing clinical trial to the CRO agency, does the CRO agency offers the EDC usage?



What is the percentage of online EDC solutions among all EDC solutions used by your organization in the last two years?

[Offline EDC solutions enable data entry even if the user's (investigator's) computer is not directly connected to the central database. In that case data is stored locally on the user's computer. On the other hand, an online EDC system requires that the user is connected to the central database (through Internet or by some kind of connection).]



What are the main benefits that you expect from electronic data collection?

[List at least three benefits.]

-----  
 Costs Availability of data Speed  
 -----  
 hitrost, direktan komunikacija, manj podvajanja dela

-----  
manj napak, manj dela pri monitoriranju,  
sprotna kontrola e-sistema ob vnosu podatkov  
s strani investigatorja  
-----  
manj papirja, hitrejši pogled v aktualno stanje,  
večja točnost podatkov  
-----  
Odsotnost arhiviranja na papirju. Monitoriranje  
z uporabo EDC. Manj napak pri vnosu  
podatkov, posledično manj vprašanj in  
hitrejša končna analiza.  
-----  
točnost preglednost hitra dostopnost podatkov

**Does your organization need a single (uniform) EDC solution that could be used for all your clinical trials?**

YES  
 NO  
 do not know

**What are the main reasons for your answer in the previous question?**

[List the reasons.]

-----  
About to start with EDC - no experience yet  
-----  
enostavnost uporabe, izkušnje pri uporabi  
sistema skrajšajo čas učenja pri novo začeti  
raziskavi  
-----  
ker je pomembno imeti enoten sistem v  
podjetju  
-----  
poenotenje vnosa podatkov mednarodno in  
možnost uporabe v statistični analizi, hitrejši  
dostop HQ do podatkov in možnost nadzora  
(ne da bi potovali)  
-----  
Raznolikost študij, za novejšje že globalno  
obstaja EDC sistem.  
-----  
za monitorje je lažje če je enotna EDC rešitev

**In your opinion, what are the main limitations of using a single EDC solution in your organization?**

[List the limitations.]

-----  
Could be it would limit the protocol  
-----  
Kot navedeno v zgornjem odgovoru.  
-----  
N.A. - to rešujejo na sedežu podjetja  
-----  
ne vidim jih  
-----  
vse težje doseganje odobritev kliničnih  
preizkušanj in vedno manjše število klin.  
preizku., ki jih lahko zvedemo, ni centralnega  
budgeta za klinična preizkušanja, predrage  
storitve  
-----  
zelo različno zasnovane raziskave, zaradi zelo  
raznovrstnih preskušanih zdravil

**In your opinion, in the case of introduction of an EDC solution, who is the person that will use it the most?**

[State the role of the person (e.g. "investigator", or "monitor", or "statistician",...)]

-----  
monior, statistik  
-----  
monitor  
-----  
monitor in statistik  
-----  
preskuševalec  
-----  
They will all use it as much as required  
-----  
Vodja kliničnega preskušanja in monitor.

**In your opinion, in the case of introduction of EDC solution, who is the one that will benefit the most?**

[State the role of the person (e.g. "investigator", or "monitor", or "statistician",...)]

-----  
monitor  
-----  
monitor in statistik  
-----  
Največ monitor in statistik, precej tudi  
preskuševalec.  
-----  
Perhaps the monitor who can monitor some  
from home  
-----  
vsi  
-----  
vsi udeleženi

**Does your organization have an internal IT department?**

YES  
 NO  
 do not know

**Does your company already have an in-house EDC solution?**

YES  
NO  
do not know

**Does your company have plans on developing an in-house EDC solution?**

YES  
NO  
do not know

**Does your organization use different commercial EDC solutions?**

YES  
NO  
do not know

**In your opinion, what is the reasonable investment in a single EDC solution for your company?**

less than 30.000 €  
30.000 € - 60.000 €  
60.000 € - 100.000 €  
100.000 € - 200.000 €  
200.000 € - 500.000 €  
500.000 € - 1.000.000 €  
more than 1.000.000 €  
do not know

**In your opinion, in the case of introduction of EDC who is the person who decides on investment?**

[State the role of the person (e.g. "general manager", or "chief monitor", or "chief information officer (CIO)",...)]

-----  
direktor  
-----  
direktor na sedežu firme  
-----  
Direktor.  
-----  
General manager  
-----  
skupina na sedežu podjetja (v HQ)  
-----  
strokovno vodstvo podjetja na evropski ravni

**Adding new trials with simple CRF**

[EDC system enables adding new trials in the trials database. The CRF for the new trial must be simple. Simple CRF consists of one or more pages which contain basic data input fields (e.g. text, number, date or selection between several options). Simple CRF do not contain more complex structures like interactive maps (e.g. human map for marking tumor locations), tables and other more complex structures.]

5 - very important  
4 - important  
3 - partially important  
2 - less important  
1 - unimportant  
do not know

**Adding new trials with complex CRF**

[EDC system enables adding new trials in the trials database. The CRF for the new trial can be of any complexity and content. Complex CRF may contain interactive maps (e.g. human map for marking tumor locations), tables and more complex structures.]

5 - very important  
4 - important  
3 - partially important  
2 - less important  
1 - unimportant  
do not know

**Supervision on course of clinical trials**

[EDC system let "coordinator" (a person who coordinates trial activities) access to statistics on trial activities (e.g. number of trials started in particular time period, number of completed trials, number of involved patients, average duration of the trial, summaries on data entry activities for each study, number of entry errors, etc.) as well as insight into data collected during each trial.]

5 - very important  
4 - important  
3 - partially important  
2 - less important  
1 - unimportant  
do not know

**Local accessibility**

[User can access the EDC system from any computer in the local network (e.g. inside the institution/company).]

5 - very important  
4 - important  
3 - partially important  
2 - less important  
1 - unimportant  
do not know

**World wide accessibility**

[User can access the EDC system from any computer connected to the Internet (e.g. from other institution/company network, home, or airport).]

5 - very important  
4 - important  
3 - partially important  
2 - less important  
1 - unimportant  
do not know

**Adding new (distant) centers after the trial starts - multicentric trials**

[During the clinical trial a new trial center can be added to the trial and users from the new center can contribute data to the trial and access that data.]

5 - very important  
4 - important  
3 - partially important  
2 - less important  
1 - unimportant

| do not know

### **Adding/removing users from the EDC system after the trial starts**

[During the clinical trial, as necessary, users can be added or removed from the system or their roles/privileges can be changed.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Localized (non-English) user interface**

[User interface of the EDC system can be localized in any world language (otherwise in English).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Multilanguage support**

[User interface of the EDC system can be localized in any world language for each different trial and each different center included in the trial (if multicentric trial).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Digital images submission**

[Support for submitting digital images to the trial database and reviewing submitted images.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Digital images import**

[Connectivity of EDC system with the existing local digital imaging systems (e.g. CT, roentgen MRI, ultrasound, etc.).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Electronic data interchange**

[Connectivity of EDC system with the existing local digital data sources (e.g. medical record system, laboratory data etc.).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Non-formatted eCRF print-out**

[Completed electronic CRF can be printed where the printed copy is visually identical to the electronic copy (each data occupies one row just like in the electronic CRF) ]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Formatted eCRF print-out**

[Completed electronic CRF can be printed where the design of the printed copy is predefined and different from the electronic copy (each data does not necessarily occupy one row like in the electronic CRF) ]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Statistics on clinical trial activities and states**

[A person responsible for the clinical trial has the insight into statistics on trial activities and the state of the trial. A coordinator of the clinical trials activities has the insight into statistics on trial activities and state for each clinical trial registered in the EDC system. ]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### **Statistics on collected data**

[A person responsible for the clinical trial has the insight into statistics on collected data (e.g. response to the therapy, average drug dose, number of male and female patients, average patient age, etc.). A coordinator of the clinical trials' activities has the insight into statistics on collected data for each trial registered in the EDC system. ]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

**Common statistics on clinical trials activities and states**

[Coordinator of the clinical trials' activities has the insight into common statistics on the activities and statuses of all the clinical trials registered in the system. ]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

**Cross statistics on data from all the clinical trials**

[Coordinator of the clinical trials activities has the insight into cross statistics on data collected in all the clinical trials registered in the system (e.g. number of male and female patients, average patient age, average BSA, etc.). ]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know



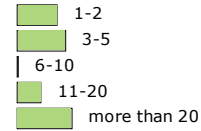
PRILOGA B

**Anketna vprašanja in rezultati ankete za CRO organizacije**

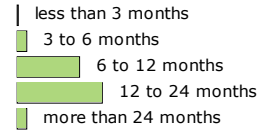




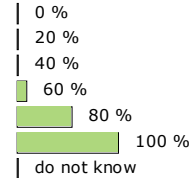
**On average, how many clinical trials are started by your organization yearly (in the last two years)?**



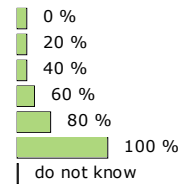
**On average, what is the duration of clinical trials that are organized by your organization?**



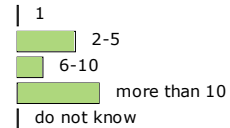
**What is the percentage of multicentric clinical trials among all the trials that are organized by your organization?**



**What is the percentage of international clinical trials among all the trials that are organized by your organization?**

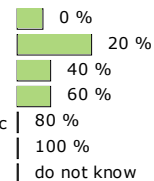


**In the last five years, how many different sponsors has your organization cooperated with?**



**What is the percentage of clinical trials that use EDC (Electronic Data Capture) among all the trials that are organized by your organization?**

[EDC (Electronic Data Capture) is the way of data collecting where the investigators/researchers enter data directly into the computer (using electronic forms) by themselves. If necessary, investigator can prior to that also collect data on the paper forms or print the completed paper forms afterwards.]

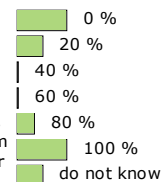


**In case of using EDC, is this required by the sponsor of the clinical trial and does the sponsor provide the support with its own EDC?**

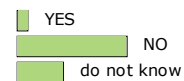


**What is the percentage of online EDC solutions among all EDC solutions used by your organization in the last two years?**

[Offline EDC solutions enable data entry even if the user's (investigator's) computer is not directly connected to the central database. In that case data is stored locally on the user's computer. On the other hand, an online EDC system requires that the user is connected to the central database (through Internet or by some kind of connection).]



**Does your organization need a single (uniform) EDC solution that could be used for all your clinical trials?**



**What are the main reasons for your answer in the previous question?**

[List the reasons.]

- 
- CRO need to adapt to Sponsors' requirements
- 
- Different sponsors want to use different EDC packages
- 
- Different studies Different information required
- 
- have yet to find one I like enough
- 
- I have a small oneman company with which I support clinical trials for pharmaceutical

companies. I don't need an EDC solution myself, the pharmaceutical companies decide whether or not EDC is used and which solution they need.

I work as freelance for other companies. I work according to their practices and procedures.

I'm working as an independent CRA. I only have a very small company.

naročnik je ta ki določa

Need to have flexibility to meet client needs

Sometimes sponsor has already a preferred provider for EDC

Standardisation Training

We are subcontracted by global CROs or pharma/biotech and they use their own EDC if needed. Not our decision.

We do not have experience.

**In your opinion, what are the main limitations of using a single EDC solution in your organization?**

[List the limitations.]

Client preference

denarne

Depends on its flexibility.

Different studies Different information required

have not found one that is better than paper

I think there are no limitations in using a single EDC system

jih ni

My company is too small.

N/A

no experience.

not applicable

Same as above

Sponsor dependence

Sponsors want us to use their selected vendor.

Sponsors were not asking it.

**In your opinion, in the case of introduction of an EDC solution, who is the person that will use it the most?**

[State the role of the person (e.g. "investigator", or "monitor", or "statistician",...)]

investigator

Investigator and monitor

investigator site data management monitor

Investigator, CRA and DM

Investigator.

monitor

monitor, raziskovalec

No simple answer to this. Basically all listed in the question

pharmaceutical company

preskuševalec, monitor, statistik, vsi enako

statistician

Study coordinator/ Study nurse and the monitor.

There is no single person that uses an EDC system most. Each person has to use the system according to his role in the study. Otherwise the EDC system doesn't work.

**In your opinion, in the case of introduction of EDC solution, who is the one that will benefit the most?**

[State the role of the person (e.g. "investigator", or "monitor", or "statistician",...)]

Data management

DM

maybe the monitor

monitor

monitor data management

monitor, raziskovalec  
 -----  
 pharmaceutical company (all roles there)  
 -----  
 preskuševalec  
 -----  
 sponsor  
 -----  
 Statistician  
 -----  
 Statistician (no data entry needed, some DCFs avoided by built-in logical checks, etc.) Monitor (progress and some of the data checks can be done remotely)  
 -----  
 Statistician.  
 -----  
 Statistician/data management  
 -----  
 There is no single person that benefits most. Depending on the role within a clinical study, each level may benefit from using EDC.

**In your opinion, in the case of introduction of EDC who is the person who decides on investment?**

[State the role of the person (e.g. "general manager", or "chief monitor", or "chief information officer (CIO)",...)]

-----  
 CEO  
 -----  
 CEO and CFO  
 -----  
 director level  
 -----  
 direktor  
 -----  
 DM  
 -----  
 general manager  
 -----  
 General manager (expensive)  
 -----  
 General manager.  
 -----  
 global head of monitoring and study management  
 -----  
 It will be a general decision making process, involving different levels of staff.  
 -----  
 N/A  
 -----  
 Sponsor decision

**Does your organization have plans on developing an in-house EDC solution?**

YES  
 NO  
 do not know

**Does your organization have an internal IT department?**

YES  
 NO  
 do not know

**Does your organization already have its own EDC solution?**

YES  
 NO  
 do not know

**What are the average yearly IT investments in your organization in the last five years?**

less than 10.000 €  
 10.000 € - 30.000 €  
 30.000 € - 60.000 €  
 60.000 € - 100.000 €  
 more than 100.000 €  
 do not know

**Does your organization plan any additional IT investments in the next five years?**

YES  
 NO  
 do not know

**If your organization plans additional IT investments in the next five years, what is the the estimated yearly amount of the planned investments?**

less than 10.000 €  
 10.000 € - 30.000 €  
 30.000 € - 60.000 €  
 60.000 € - 100.000 €  
 more than 100.000 €  
 do not know

**Adding new trials with simple CRF**

[EDC system enables adding new trials in the trials database. The CRF for the new trial must be simple. Simple CRF consists of one or more pages which contain basic data input fields (e.g. text, number, date or selection between several options). Simple CRF do not contain more complex structures like interactive maps (e.g. human map for marking tumor locations), tables and

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant

other more complex structures.]

do not know

### Adding new trials with complex CRF

[EDC system enables adding new trials in the trials database. The CRF for the new trial can be of any complexity and content. Complex CRF may contain interactive maps (e.g. human map for marking tumor locations), tables and more complex structures.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Supervision on course of clinical trials

[EDC system let "coordinator" (a person who coordinates trial activities) access to statistics on trial activities (e.g. number of trials started in particular time period, number of completed trials, number of involved patients, average duration of the trial, summaries on data entry activities for each study, number of entry errors, etc.) as well as insight into data collected during each trial.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Local accessibility

[User can access the EDC system from any computer in the local network (e.g. inside the institution/company).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### World wide accessibility

[User can access the EDC system from any computer connected to the Internet (e.g. from other institution/company network, home, or airport).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Adding new (distant) centers after the trial starts - multicentric trials

[During the clinical trial a new trial center can be added to the trial and users from the new center can contribute data to the trial and access that data.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Adding/removing users from the EDC system after the trial starts

[During the clinical trial, as necessary, users can be added or removed from the system or their roles/privileges can be changed.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Localized (non-English) user interface

[User interface of the EDC system can be localized in any world language (otherwise in English).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Multilanguage support

[User interface of the EDC system can be localized in any world language for each different trial and each different center included in the trial (if multicentric trial).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Digital images submission

[Support for submitting digital images to the trial database and reviewing submitted images.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Digital images import

[Connectivity of EDC system with the existing local digital imaging systems (e.g. CT, roentgen MRI, ultrasound, etc.).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Electronic data interchange

[Connectivity of EDC system with the existing local digital data sources (e.g. medical record system, laboratory data etc.).]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Non-formatted eCRF print-out

[Completed electronic CRF can be printed where the printed copy is visually identical to the electronic copy (each data occupies one row just like in the electronic CRF) ]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Formatted eCRF print-out

[Completed electronic CRF can be printed where the design of the printed copy is predefined and different from the electronic copy (each data does not necessarily occupy one row like in the electronic CRF) ]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Statistics on clinical trial activities and states

[A person responsible for the clinical trial has the insight into statistics on trial activities and the state of the trial. A coordinator of the clinical trials activities has the insight into statistics on trial activities and state for each clinical trial registered in the EDC system. ]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Statistics on collected data

[A person responsible for the clinical trial has the insight into statistics on collected data (e.g. response to the therapy, average drug dose, number of male and female patients, average patient age, etc.). A coordinator of the clinical trials' activities has the insight into statistics on collected data for each trial registered in the EDC system. ]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Common statistics on clinical trials activities and states

[Coordinator of the clinical trials' activities has the insight into common statistics on the activities and statuses of all the clinical trials registered in the system. ]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Cross statistics on data from all the clinical trials

[Coordinator of the clinical trials activities has the insight into cross statistics on data collected in all the clinical trials registered in the system (e.g. number of male and female patients, average patient age, average BSA, etc.). ]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know



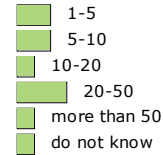
PRILOGA C

**Anketna vprašanja in rezultati ankete za klinične ustanove**



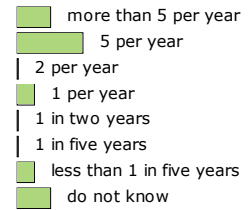


On average, how many clinical trials (experimental studies) are started in your institution yearly (in the last five years)?

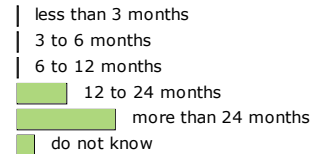


On average, how many observational studies are started in your institution?

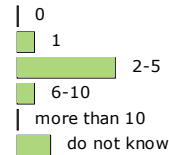
[]



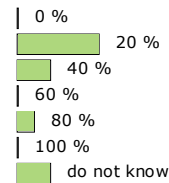
On average, what is the duration of clinical trials that are in progress your institution?



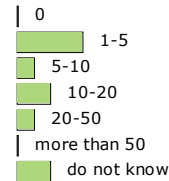
How many observational studies are currently in progress in your institution?



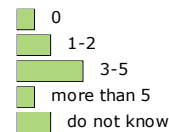
What is the percentage of clinical trials sponsored/organized by your institution among all the trials that are in progress in your institution?



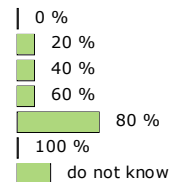
How many clinical trials are intended to start in your institution next year?



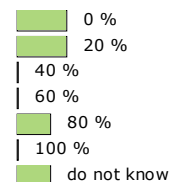
How many observational studies are intended to start in your institution next year?



What is the percentage of multicentric clinical trials among all the trials that are in progress in your institution?



What is the percentage of multicentric clinical trials among all the trials that are sponsored/organized by your institution?



**In the case when you are not a sponsor/organizer of a clinical trial, how often do you reuse the collected data also for your own reasearch activities?**

- always
- usually
- sometimes
- rarely
- never
- do not know

**What are the main reasons for your answer in the previous question?**

[List the reasons.]

-----  
Do by the company  
-----  
double blind edsign - data not unblinded until the study stops  
-----  
manjsi delež bolnikov prepoved od glavnega raziskovalca  
-----  
Samostojna obdelava podatkov  
-----  
strokovna odlicnost, relevantnost nasega dela, zanimivo  
-----  
we are not the data owner.

**How oftarn do you make a cross analysis of data collected from different studies?**

- always
- usually
- sometimes
- rarely
- never
- do not know

**What are the main reasons for your answer in the previous question?**

[List the reasons.]

-----  
""  
-----  
always on our own studies rarely on data from other  
-----  
lack of time to do it more often  
-----  
ne smemo  
-----  
Our hospital has recently merged with another hospital. I think it's difficult to answer the questions above because I do not know that much about the amount of trials at the other hospital.  
-----  
Podrobnjejša analiza  
-----  
This is not scientific

**Does your institution have a Data Center - a department for IT support to the scientific-research activities?**

[Beside the other, Data Center takes care of electronic data collection, storage and processing of clinical data.]

- YES
- NO
- do not know

**How many employees are there in your Data Center?**

- 1-5
- 6-10
- 11-20
- 21-50
- more than 50
- do not know

**Currently, do you see any need for further improvements of your Data Center in order to improve the studies management and the research results?**

[E.g. additional employees, new activities, or infrastructure improvements.]

- YES
- NO
- do not know

**Does your institution plan any additional investments in Data Center in the next five years?**

- YES
- NO
- do not know

**If your organization plans additional investments in Data Center in the next five years, what is the the estimated yearly amount of the planned investments?**

- less than 10.000 €
- 10.000 € - 30.000 €
- 30.000 € - 60.000 €
- 60.000 € - 100.000 €
- more than 100.000 €
- do not know

**Is there a wireless network access (WLAN) in your institution?**

[Wireless LAN (WLAN)]

- YES
- NO
- do not know

**What kind of IT support you usually provide for your own studies?**

[List typical examples, e.g.: "we enter data into Excel", or "we develop database and user interface for data entry for each study separately", or "for each study we buy a separate commercial EDC solution", etc. ]

- EDC Informed MACRO
- excel
- Full EDC support, inhouse developed
- Običajno imajo študije svoje podatkovne baze
- Professional Electronic data capture
- uporabljamo pisne protokole. Statistik jih vnese v e verzijo
- We use Teleform to design CRFs that we scan into SLQ databases or directly into SPSS.

**What is the average budget that you assign to IT support for your studies?**

- less than 5.000 €
- 5.000 € - 10.000 €
- 10.000 € - 30.000 €
- 30.000 € - 60.000 €
- more than 60.000 €
- do not know

**What is the percentage of trials that use EDC (Electronic Data Collection) among all the clinical trials that are in progress in your institution?**

[EDC (Electronic Data Collection) is the way of data collecting where the investigators/researchers enter data directly into the computer (using electronic forms) by themselves. If necessary, investigator can prior to that also collect data on the paper forms or print the completed paper forms afterwards.]

- 0 %
- 20 %
- 40 %
- 60 %
- 80 %
- 100 %
- do not know

**What is the percentage of studies that use EDC (Electronic Data Collection) among all the observational studies that are in progress in your institution?**

- 0 %
- 20 %
- 40 %
- 60 %
- 80 %
- 100 %
- do not know

**What is the percentage of online EDC solutions among all EDC solutions currently used by your institution?**

[Offline EDC solutions enable data entry even if the user's (investigator's) computer is not directly connected to the central database. In that case data is stored locally on the user's computer. On the other hand, an online EDC system requires that the user is connected to the central database (through Internet or by some kind of connection).]

- 0 %
- 20 %
- 40 %
- 60 %
- 80 %
- 100 %
- do not know

**Does your institution already have its own EDC solution?**

- YES
- NO
- do not know

**Does your organization need a single (uniform) EDC solution that could be used for all your studies?**

[Such solution would enable following of the course of study, as well as data entry and data review.]

- YES
- NO
- do not know

**What are the main reasons for your answer in the previous question?**

[List the reasons.]

- It would simplify the design process using one solution and it would make it easier for users.
- centralna baza podatkov
- different study designs, different needs
- investment costs, training costs
- neodvisno delovanje
- we use available EDC

**In your opinion, what are the main limitations of using a single EDC solution in your institution?**

[List the limitations.]

-----  
 medical personal can not find an agreement  
 -----  
 ne vem  
 -----  
 ni omejitev  
 -----  
 none  
 -----  
 Not so friendly  
 -----  
 see above  
 -----  
 We are also running studies for the industry and in these studies we have to use the companies's EDC solutions.

**In your opinion, in the case of implementation of an EDC solution, who is the person that will use it the most?**

[State the role of the person (e.g. "investigator", or "study coordinator", or "statistician",,...)]

-----  
 clinical research assistants, statisticians  
 -----  
 clinician data manager statistician  
 -----  
 raziskovalne sestre  
 -----  
 Research nurses, study coordinators and data managers  
 -----  
 statistik  
 -----  
 study coordinator statistician  
 -----  
 study nurse

**In your opinion, in the case of introduction of EDC solution, who is the one that will benefit the most?**

[State the role of the person (e.g. "investigator", or "study coordinator", or "statistician",,...)]

-----  
 all  
 -----  
 everybody  
 -----  
 Investigator initiated trials - the hospital! I believe an EDC solution will improve data quality and data may be analyzed at an earlier stage.  
 -----  
 Invevtigator  
 -----  
 nobody  
 -----  
 preiskusevalec  
 -----  
 raziskovalne sestre in zdravniki  
 -----  
 statistician, sponsor, datamanager

**Does your organization have plans on developing an in-house EDC solution?**

| YES  
 [Bar chart showing YES is significantly higher than NO and do not know]

**In your opinion, what is the reasonable cost of implementation of a single EDC solution in your institution?**

[Bar chart showing 'less than 30.000 €' is the most selected option]

- less than 30.000 €
- 30.000 € - 60.000 €
- 60.000 € - 100.000 €
- 100.000 € - 200.000 €
- 200.000 € - 500.000 €
- 500.000 € - 1.000.000 €
- more than 1.000.000 €
- do not know

**Adding new studies with simple CRF**

[EDC system enables adding new studies to the studies database. The CRF for the new study must be simple. Simple CRF consists of one or more pages which contain basic data input fields (e.g. text, number, date or selection between several options). Simple CRF do not contain more complex structures like interactive maps (e.g. human map for marking tumor locations), tables and other more complex structures.]

[Bar chart showing '5 - very important' is the highest rating]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

**Adding new studies with complex CRF**

[EDC system enables adding new studies to the studies database. The CRF for the new study can be of any complexity and content. Complex CRF may contain interactive maps (e.g. human map for marking tumor locations), tables and more complex structures.]

[Bar chart showing '5 - very important' is the highest rating]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

**Supervision on course of clinical studies**

[EDC system lets "coordinator" (a person who coordinates study activities) access to statistics on study activities (e.g. number of studies started in particular time period, number of completed studies, number of involved patients, average duration of the study, summaries on data entry activities for

[Bar chart showing '5 - very important' is the highest rating]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant

each study, number of entry errors, etc.) as well as insight into data collected during each study.]

do not know

### Local accessibility

[User can access the EDC system from any computer in the local network (e.g. inside the institution/company).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### World wide accessibility

[User can access the EDC system from any computer connected to the Internet (e.g. from other institution/company network, home, or airport).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Adding new (distant) centers after the study starts - multicentric studies

[During the clinical study a new study center can be added to the study and users from the new center can contribute data to the study and access that data.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Adding/removing users from the EDC system after the study starts

[During the clinical study, as necessary, users can be added or removed from the system or their roles/privileges can be changed.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### CRF page edit

[During the clinical study, as necessary, data fields can be added or deleted.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### CRF structure edit

[During the clinical study, as necessary, CRF structure can be slightly changed (e.g. one CRF page can be split in two pages)]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Localized (non-English) user interface

[User interface of the EDC system can be localized in any world language (otherwise in English).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Multilanguage support

[User interface of the EDC system can be localized in any world language for each different study and each different center included in the study (if multicentric study).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Digital images submission

[Support for submitting digital images to the study database and reviewing submitted images.]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Digital images import

[Connectivity of EDC system to the existing local digital imaging systems (e.g. CT, roentgen MRI, ultrasound, etc.).]

5 - very important  
 4 - important  
 3 - partially important  
 2 - less important  
 1 - unimportant  
 do not know

### Data import

[Connectivity of EDC system to the existing local digital data sources (e.g. medical record system, laboratory data etc.).]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Electronic data interchange

[E.g. a possibility of data export into sponsor's EDC system in order to avoid double data entry: into local EDC system and sponsor's EDC system.]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Non-formatted eCRF print-out

[Completed electronic CRF can be printed where the printed copy is visually identical to the electronic copy (each data occupies one row just like in the electronic CRF)]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Formatted eCRF print-out

[Completed electronic CRF can be printed where the design of the printed copy is predefined and different from the electronic copy (each data does not necessarily occupy one row like in the electronic CRF)]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Statistics on clinical trial activities and states

[A person responsible for the clinical trial has the insight into statistics on trial activities and the state of the trial. A coordinator of the clinical trials activities has the insight into statistics on trial activities and state for each clinical trial registered in the EDC system.]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Statistics on collected data

[A person responsible for the clinical study has the insight into statistics on collected data (e.g. response to the therapy, average drug dose, number of male and female patients, average patient age, etc.). A coordinator of the clinical studies' activities has the insight into statistics on collected data for each study registered in the EDC system.]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Common statistics on clinical studies' activities and states

[Coordinator of the clinical studies' activities has the insight into common statistics on the activities and statuses of all the clinical studies registered in the system.]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

### Cross statistics on data from all the clinical studies

[Coordinator of the clinical studies' activities has the insight into cross statistics on data collected in all the clinical studies registered in the system (e.g. number of male and female patients, average patient age, average BSA, etc.).]

- 5 - very important
- 4 - important
- 3 - partially important
- 2 - less important
- 1 - unimportant
- do not know

OBJAVLJENI ČLANKI

**Pavlovic, I., & Miklavcic, D. (2007). Web-Based Electronic Data Collection System to Support Electrochemotherapy Clinical Trial. Information Technology in Biomedicine, IEEE Transactions on, 11(2), 222-230. doi: 10.1109/TITB.2006.879581.**

**Pavlović, I., Kern, T., & Miklavcic, D. (2009). Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study. Contemporary Clinical Trials, 30(4), 300-316. doi: 10.1016/j.cct.2009.03.008.**





# Web-Based Electronic Data Collection System to Support Electrochemotherapy Clinical Trial

Ivan Pavlović and Damijan Miklavčič

**Abstract**—Many branches of the healthcare industry are being influenced by information and communication technology (ICT). Clinical trials are not an exception. Despite this fact, more than 75% of clinical trials data are being collected on paper records. Recent ICT advances, such as broad acceptance of Internet Technology which are rapidly improving electronic data collection (EDC) tools, however, may soon reduce this percentage of “paper” supported clinical trials. In this paper, we present our Web-based EDC system designed to support a small-scale research-oriented clinical trial for establishing standard operating procedures (SOP) for electrochemotherapy with a new medical device, named Cliniporator. The definition of the SOP can only be based on a comprehensive analysis of collected data and results of clinical trial. Therefore, it is necessary to record treatment efficiency and, in this respect, to carefully follow and collect treatment parameters. We thus established central database and the Web application for filling database with data submitted by users from distant medical centers across Europe. Also, we enabled transmitting of data stored on the local Cliniporator medical devices to the central database as well as submitting of tumor images and marking of tumor nodules on interactive human map developed in Macromedia Flash. We provided users with dynamically generated basic statistics, and, several times during data collection process, we performed statistical data analysis. In order to assure high quality of data in a database, we included several mechanisms: automatic data validation, digital signatures, the form completeness notification system, e-mail alerting of completed forms, and “check tables.” After 13 months of using the systems, we performed a simple usability evaluation of the system by asking users to answer to a questionnaire, and here we present the results. With this paper, we try to share our experience and encourage others to exploit Internet and Web technologies to improve clinical trials data collection, follow up, and data analysis.

**Index Terms**—Clinical trial, computer applications, electronic data collection, Internet.

## I. INTRODUCTION

SOCIETY is continuously being influenced by modern information and communication technology (ICT). Each and every economic branch is somehow affected by the use of ICT. Healthcare has already profited extensively from these developments [1]. With the term “healthcare” we describe the industry associated with the provision of medical care to individuals. Probably one of the strongest breakthroughs of ICT in the healthcare was introduction of electronic patient records (EPR). This is

the main reason for the increase of resources that the healthcare industry allocates for the ICT from 2% of its revenues during the 1990s to 5%–7% in recent years [2].

EPRs bring significant improvements in collecting and processing clinical trials data. However, data stored in EPR are not suitable or do not include all the required data for the purpose of clinical trial. Therefore, collection of clinical-trial data usually requires specific data-collection software. Several vendors offer electronic data collection (EDC) tools to support clinical trials. Even more, they are expanding their EDC offers into broader eClinical suites [3]. Those involve: EDC, clinical-data-management systems (CDMS), drug-safety monitoring, trial management, site management, vocabulary management, eSubmission, electronic patient diaries, interactive voice-response systems, and investigator management.

Despite the fact that EDC tools have been available for more than two decades, clinical trials are still mainly conducted using paper data collection as the primary tool (over 75% [4]). The reason for this can be partially ascribed to the fact that standards for exchange of clinical data among different parties still have to be developed, and that these standards should be extended to facilitate data collection at the investigation site. Present technological applications often do not have adequate functionality to meet current needs. In addition, available commercial applications are usually economically acceptable only for large-scale trials. And, even then, they can hardly meet specific trial needs, like integration of other data sources and data from medical devices. It is inherent to the trial process that it can be modified during the study, and thus EDC system should be flexible to support these modifications. With respect to all the requirements listed above, it is still hardly possible to expect a single system that could meet the variety of different trial needs.

In this paper, we present the EDC system, which we developed in support of multicenter clinical trial performed in the frame of European Commission 5th framework demonstration project European Standard Operating Procedures of Electrochemotherapy and Electrogenotherapy (ESOPE). Following are the starting points for developing our EDC system:

- two years of clinical trial using new medical device;
- trial is running on different locations (medical centers) scattered around Europe;
- allowing collection of data and follow-up of progress in the project;
- intermediate and final analysis of collected data for a purpose of defining and validating standard operating procedures for electrochemotherapy;
- maintenance during the project and at least for additional two years (preferably at no cost).

Manuscript received April 26, 2005; revised February 21, 2006 and March 30, 2006. This work was supported in part by the ESOPE (QLK3-02002-2003) project funded by the European Commission and in part by the Slovenian Research Agency.

The authors are with the Faculty of Electrical Engineering, University of Ljubljana, SI-1000 Ljubljana, Slovenia (e-mail: ivan@ibk.fe.uni-lj.si; damijan@svarun.fe.uni-lj.si).

Digital Object Identifier 10.1109/TITB.2006.879581

Easy deploying, maintenance, and upgrading of the EDC system, followup of the progress in the project, and scalability of the system were set as our major goals to be achieved. As per these goals and starting points, following is our strategy for developing our EDC system:

- centralized database;
- development of an application for easy data submission to the central database;
- computer- and platform-independent solution;
- remote-application maintenance and upgrading;
- application must work or be accessible inside the clinical network;
- respecting the “privacy” and ownership of each medical center’s data.

The system we built is based on Internet technologies to allow for mobility, to enable multimedia content, and to achieve wide user acceptance. We participated in the analysis of collected clinical data, and here we present the complexity of that process. After 13 months, we performed a basic-usability evaluation of the system, which was done through the Web-questionnaire filled by the users, and in this paper, we present a summary of their responses. Additionally, we built another Web application to enable supervision of device performance and disposable consumables use for the industrial partner supplying medical centers with the medical devices used during the study.

## II. ESOPE PROJECT

Recently, as part of the Cliniporator project funded by the European Commission (QLK-1999-00484), a new medical device, called Cliniporator, was developed. It was specifically designed for clinical electrochemotherapy and electrogenetherapy. This device was designed to deliver trains of short high-voltage pulses and longer low-voltage pulses in order to electroporate the tissue and enhance drug including DNA delivery into cells. During the pulse application, the control unit measures voltage and current, which allows monitoring and immediate control of the treatment performed, if necessary. After the pulse application, voltage and current measurements are stored in the local database on the device.

After the Cliniporator project, the medical device Cliniporator was in the stage of testing in clinical environment. The testing was performed in four medical centers in Europe within the ESOPE project (QLK3-02002-2003) funded by the European Commission. The aim of the project was to define and validate standard operating procedures (SOPs) for electrochemotherapy and electrogenetherapy. The definition and validation of the SOP is based on comprehensive intermediate and final analysis of collected data and results of clinical trial. It was, therefore, necessary to record treatment efficiency and, in this respect, to carefully follow and collect treatment parameters including the drug used, anesthesia and sedation, pulse applications, tumor type and location, and response to the treatment.

For collection of data acquired in electrochemotherapy clinical trial, standard paper forms (case report forms—CRFs) were prepared. The CRF consists of a number of subforms, the extent of which depends on the number of treated tumors and the num-

ber of sessions required to treat each tumor. The CRF includes patient’s general data, his/her medical history, tumor-treatment data, and tumor response. Tumor treatment can be repeated, if necessary. For each patient, medical personnel have to fill in approximately 40 pages of forms, on average.

The paper-based forms were intended to be used for documentation purposes, and for clinical-data analysis. We established a unified database of all collected data in medical centers involved in the study. This central database is intended to offer a better overview of trial data and easier follow up of the progress on the project, and to serve as a source for comprehensive analysis of collected data and results during and also after the end of the project.

The central database was planned to be daily filled with data from different locations around Europe. For this purpose and to fulfill expectations, we built a Web-based EDC application (Cliniporator Web-Recorder) that provides a user-friendly interface for remote filling of central database, so the medical personnel can easily carry it out. Each user from the medical centers involved in the study is able to submit his/her patient- and treatment-related data directly to the central database.

Due to the distances between the centers, the EDC application has to be easy to deploy, maintain, and debug. These were the reasons to develop a Web-application. EDC application has to provide two main services:

- digital CRF—digital forms for submitting CRF data to the central database;
- local database upload—a module for transferring data from the local Cliniporator-device databases to the central database.

These two services are the core of our EDC system. The schema of the system is presented in Fig. 1. However, the advantage of the proposed Web-based solution is that any other service can easily be added, which we also demonstrated. During the trial process, we extended the initial application by adding two additional services. One was the statistic service that offers dynamically generated statistics charts for already submitted data, which allowed the coordinator to follow the progress of the project. The other one was added to be used by the device manufacturer for tracking the device functioning and electrode supplies and is fully described in [18].

## III. CENTRAL DATABASE

The central database is a relational database established on the Microsoft SQL Server 2000. At this moment (after 32 months of continuous use) its size is 1.5 GB (including more than 1800 tumor images). It consists of more than 30 tables. Most of the tables are related to the CRF data, but there are also tables dedicated to storing images, data uploaded from local databases and users’ data, and permissions and activities.

Following is the trial data stored in the central database:

- patient data (demography, medical history, physical examination, etc.);
- therapy data (sessions, evaluation visits, follow-up visits, etc.);

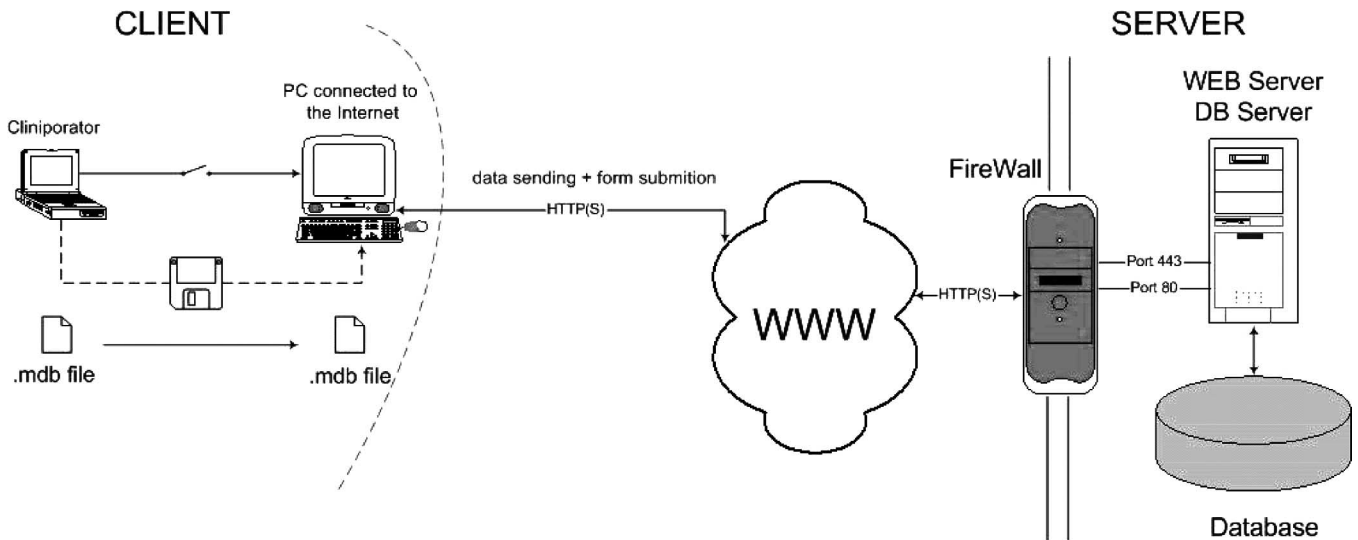


Fig. 1. Schema of the EDC system. On the client side, the user's computer needs just an Internet browser to access the server in order to add and edit the data in the central database. All the processing is performed on the server side, except the data validation, which is performed by the client browser.

- data from the local databases of Cliniporator medical devices;
- digital images of tumor nodules in different phases of therapy and follow up.

Data uploaded from the local Cliniporator devices have to be transferred from the Microsoft Access database to the MS SQL Server 2000 central database. This process is automated through the data transformation service (DTS) package. The DTS package extracts data from the tables in the MDB file which the user uploads to the server, does necessary transformations, and stores transformed data in the corresponding tables in the central database.

In order to prevent a hazardous loss of data, two copies of backup of the database are automatically created once a week: one on the local hard disk, and the other on the tape. In addition, two copies of backup of transaction log are stored once a day. The scheduled jobs executed on the recurring basis are created for these purposes.

#### IV. WEB-BASED EDC APPLICATION

The web-based application for electronic data collection, which controls the users, interface to the system, is the core of the system. We named it Cliniporator Web-Recorder. It is a set of active server pages (ASP), which are running on the Microsoft Internet Information Server 6 (IIS 6) Web server. It processes users' requests, queries the database, generates web pages, and runs other processes like storing data from submitted local databases to the central database, storing submitted images, online generating of data export tables or basic statistics, or sending e-mail alerts to the coordinator or other people in charge of following the progress, the control of data entries, etc.

The Cliniporator Web-Recorder currently offers several functionalities:

- web forms (digital CRF) for data entry;
- upload of images;

The screenshot shows the Electronic CRF page. It features a patient list on the left, a main content area with laboratory results, and a submit button at the bottom. The page is annotated with navigation elements: 'main navigation', 'pages navigation', 'top navigation (sections)', 'logout button', and 'submit button'.

Case Number	First Name	Last Name	Visit
001	TEST		
002	TEST		
003	TEST		
004	blan		

Section	Item	Value	Units	Abnormal	Significance
LABORATORY	COMPLETE BLOOD COUNT	12.2	g/100ml		
	Hemoglobin	295	x10 <sup>9</sup> /L		
	White Blood Cell Count	6.1	x10 <sup>9</sup> /L		
BIOCHEMISTRY	Creatinine	70	umol/L		
	Urea Nitrogen	1.04	mmol/L		
HOMEOSTASIS	Pulse/minute	89	/min		
	Partial thromboplastin time	21	Sec		
	DKR	1.04			

Fig. 2. Electronic CRF page. The users enter the data to the central database through the set of Web forms similar to the paper-based CRF.

- interactive human map for marking tumor locations;
- local MDB database upload from the Cliniporator medical device;
- basic online statistics;
- interactive tumor-history graph.

##### A. Web-Forms (Digital CRF) for Data Entry

Web-forms (digital Clinical Report Forms—CRFs) are form-like Web pages (Fig. 2). Through digital CRF, users submit patient and treatment data to the central database. A digital CRF has the same structure and design as the paper-based CRF. Digital CRF are organized in sections (prestudy visit, sessions,

evaluation visit day 15 and day 30, follow-up visits, end of study, concomitant medications, and adverse events), and some of these sections are further divided into several pages.

Digital CRF contains automatic data validation. This entry checking mechanism warns the user if he/she mistypes or enters meaningless data. This is performed by client-side JavaScript code embedded into the Web pages. As the validation routine is performed on the client computer, there is no need for additional client—server communication, which reduces the Internet traffic and improves the application speed.

At the end of every section, the user has the opportunity to “digitally sign” the section, by filling/entering his/her password in the appropriate signature-input field. By signing a section, the corresponding forms are “locked” and all further modifications are disabled. In this way, the user confirms that the data in the forms are complete and valid. Signing of the section triggers an e-mail alert. The data manager and the project coordinator receive an e-mail, which informs them that the forms are completed, so they can include data in the data analysis. If any mistake in the signed data is detected, a user who signed the data has to send the request to unlock the section to the project coordinator, who approves the changes and forwards the request to the database administrator. The database administrator is the only one who is able to access directly to the central database and edit the data. A copy of the unlock request is kept as part of the project documentation.

We integrated additional mechanism for improving quality of data in the database: the form-completeness notification. This mechanism is there to notify users of yet incomplete forms. Namely, when the user completes any particular section or part of the section, he/she signs it. Automatically all the corresponding links to the pages in the page navigation bar are changed from normally black color to “gold.” Therefore, in order to complete the forms, the user has just to follow the black links to the pages that are not yet filled. Similarly, button links to the sections in the section navigation get “gold stamp” if the whole section is signed. And also, if all the forms of the particular patient are signed, the “gold stamp” appears above the edit icon in the patient navigation. This enables users to follow nonstamped links and quickly find incomplete forms and fill them.

### B. Upload of Images

Image upload enables storing of digital tumor images into the central database. Images, captured with digital camera, are uploaded in their original size. During the upload process, a smaller image, suitable for displaying, as well as a thumbnail of the image, are dynamically generated and also stored in the database. All the images are stored in the JPEG format. Users can add a caption and a description to each image. Images can be added at all phases of the treatment (prestudy, sessions, follow up, etc.). In the image gallery, the user can review all the uploaded images of a single patient. Thumbnails of images are grouped by tumor nodules, and sorted by the date of capture. By selecting appropriate thumbnail, the user can see the large image subtitled with submitted comments. The user can also download the originally submitted image. Image gallery proved

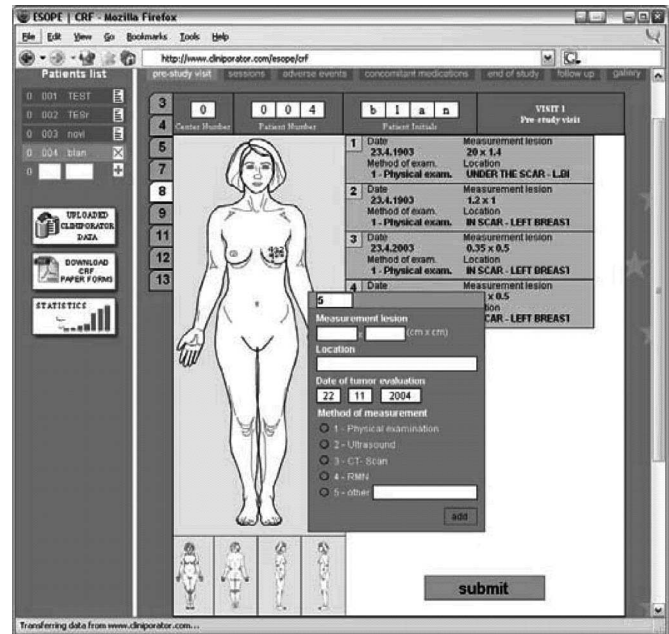


Fig. 3. Interactive human map. Users mark the tumor nodules and enter basic data on the digital human map developed in Macromedia Flash.

to be very useful for visual inspection and observation of tumor changes.

### C. Interactive Human Map for Marking Tumor Locations

The purpose of the interactive human map is visual presentation of the tumor locations on the human body. According to the patient’s sex selected within the CRF, appropriate body map is displayed. Users can switch between four views: front, rear, left, and right. By simply clicking on the map, the user can mark the location of the tumor, and then submit basic data about the tumor (such as location and measurement of lesion) and upload corresponding images. During the sessions, the user can select on the map which of the tumors, identified and recorded during the prestudy visit, are treated. The map was developed in Macromedia Flash Player 6, and it exchanges XML-structured data with the Web server. The interactive human map is shown in Fig. 3.

### D. Local MDB Database Upload From the Cliniporator Medical Device

Upload of local Cliniporator device databases is also performed through the Internet browser. The user simply selects the local database file, fills in upload comment, and selects the submit button. The rest of the process is automatic: application saves uploaded file on the server; stores upload information (date and time, user ID, name of the file, and comment) in the central database; and then copies data from the uploaded file to the central database. A DTS package, which we developed, takes care of data copying and overwriting. At the end of the upload process, the user is informed of the upload success/failure.

In the list of the uploaded data, the user can check all the data uploaded from his/her center. Data are sorted by patients,

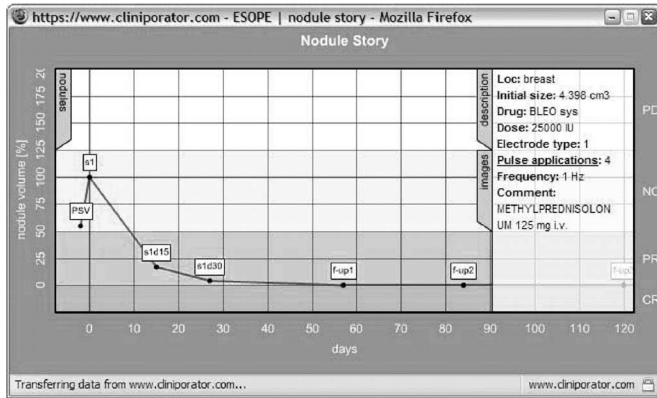


Fig. 4. Interactive tumor-history graph. On the graph, the user can see the timeline of the nodule size and therapy data as well as tumor images and a list of all the patient nodules (not presented here).

sessions, nodules, and pulse applications. It is possible to check data for every particular pulse application. A pop-up window with pulse description contains a table with the pulse data (like device settings and pulse measurements), and a pulse figure dynamically generated from the recorded voltage and current measurements.

#### E. Basic Online Statistics

Basic statistics, which allow to follow up the progress on the project, are dynamically generated from the data in the central database. Information about the number of treated patients per center as well as the number of ended therapies, number of tumor treatments and uploaded corresponding local databases, and distribution of applications of different electrode types and different drugs used is available online. All centers have access to these statistics and can compare their activities with others. Some of the statistics (usually local statistics) can be dedicated to a particular center and, therefore, hidden to others. Statistics are presented in Macromedia Flash 3-D charts.

#### F. Interactive Tumor-History Graph

Users have access to the tumor history for each of the patients submitted from the same center. Tumor history is a record of changes of tumor-nodule sizes through the evaluation and follow-up period after treatment. We designed interactive Flash presentation of the tumor history for each patient in the database (Fig. 4). This presentation contains the list of the tumor nodules for a particular patient and graph with timeline of nodule size for each of the nodules. Users can show or hide any of the timelines to clear the graph. Each timeline is accompanied by data of tumor initial size and location, therapy, drug, and pulse parameters, as well as tumor images. The interactive tumor-history graph is dynamically generated from the data stored in the database on demand.

### V. SECURITY MEASURES

The data security is assured on three levels: server level; data transmission level; and application level.

TABLE I  
DATABASE ACCESS PERMISSIONS

	User	Read-Only User	Guest User	Device Manufacturer
Cliniporator Web-Recorder <b>CRF</b>	r + w	R	r + w	X
Cliniporator Web-Recorder <b>Basic Statistics</b>	R	R	R	X
Cliniporator Web-Recorder <b>Data Extraction Tables</b>	r	R	r	X
Cliniporator Service Module <b>Data Extraction Tables</b>	X	X	X	R
Cliniporator Service Module <b>Querying the Database</b>	X	X	X	R

Access to the database is granted differently for different user types and can be assigned according to the needs and agreements. (r – read data from the same center; w – edit data from the same center; R – read data from all centers; X – no access)

On the server level we have two main security measures: the firewall and the database backup. The computer, which runs the web server and the database server, is set behind the firewall. The firewall examines all the data traffic coming in and out of the server computer and blocks the traffic which does not meet the specified security criteria. The data stored in the central database are frequently backed up according to the policy described earlier, in the “Central database” section.

On the data-transmission level, all the communication between the client and the server is encrypted through the secure socket layer (SSL) protocol. SSL is an encryption protocol that encodes data sent over the World Wide Web, rendering it unreadable to anyone intercepting the transmission. When an SSL session is started, the server sends its public key to the user’s browser, which the browser uses to send a randomly-generated private key back to the server in order to have a secret key exchange for that session. Developed by Netscape Communications Corporation and RSA Data Security, Inc., SSL is the leading security protocol used on the Web.

On the application level, we perform authentication and authorization of users. The client authenticates him/herself by username and password. The application checks the user identity in the central database where the table of users and their permissions are stored. Based on the user’s identity, authorization allows appropriate access to the resources (e.g., it enables access only to data uploaded by his/her medical center—see also Table I). However, the coordinator has the right to review the data of all the centers. Contrastingly, the industrial partner has access to limited subset of data, of all medical centers, that are related to Cliniporator device performance and electrode usage. Accessing whole data or specific subset depends on the agreement between parties involved and can easily be granted and changed, if necessary. For the moment, data-access permissions are granted as presented in Table I.

### VI. DATA QUALITY AND DATA ANALYSIS

The main aim of the ESOPE project, and thus also the main purpose of the Cliniporator Web-Recorder, was to aid

in establishing SOPs for electrochemotherapy and electrogenotherapy with the Cliniporator device. This aim requires comprehensive analysis of the collected data at different times during the clinical trial.

Analysis of the medical-trial data usually may not be done automatically, especially when the trial is running on different locations and performed by different personnel. The human interaction in this process is absolutely necessary. An expert can easily interpret different artifacts and, if necessary, exclude some data from the analysis. Direct statistical analysis of the data in the database can lead to erroneous conclusions. It is too optimistic to expect that all the data are entered correctly and all the trial is performed perfectly. However, data analysis is necessary and useful. It shows correlations between different treatment parameters, and proves or denies our hypotheses, leading to better understanding of the biological processes.

Data quality control appeared to be very demanding and essential. In order to improve data quality source verification we created “check tables”—on the fly-generated tables of data extracted from the central database, which were available to all the users through the Cliniporator Web-Recorder. These tables contained only the data related to the several data analyses essential for the ESOPE project. In these tables, users were more easily checking for missing data or unusual values and then correcting these data in the Web forms. The coordinator also had access to “check tables” and he was assisting in the correction process.

We enabled form-completeness-notification system and e-mail alerting of completed forms. In this way, we tried to move most of the data controlling process from coordinator to users themselves, and to shift control of form completeness and data quality control from serial to parallel process. With this solution, data review can begin as soon as data have been entered and the quality of data in the central database can be high even during the data entry, without locking the database. This high confidence of data in the database during the study enables frequent data analysis and, thus, better directing of the study.

In the central database, data are organized in numerous tables. But, for the statistical analysis, the most suitable form of data representation is a “flat table” (one single table with all the required data included in it). Therefore, we usually have to extract data from a set of tables in central database and to organize them in a flat table. This extraction of data requires complex queries on a set of tables. However, once derived, extraction can be automatic for repetitive analysis. The result, i.e., the “flat table,” serves as a source for the data analysis, graphical presentation, and statistical analysis. Graphical presentation sometimes requires further transformations and reorganization of data in the result table. Also, some data analysis and statistical evaluation, specially the ones which are performed frequently and/or regularly, can be automated and repeated as the number of patients included is increasing.

## VII. USABILITY EVALUATION

As the human–computer interaction is increasing its role in everyday processes, the importance of usability engineering [5] has become obvious. This resulted in development of

new methodologies for usability studies and improvement of the existing ones [7]–[13]. However, performing a complex usability study can be a very expensive process, and, therefore, not always applicable. However, some authors [14], [15] claim that gathering feedback information from four to five users of the system is enough to discover 80% of the usability problems of the system. We decided to perform a basic usability evaluation with a simple structured questionnaire, with having in mind all the weaknesses of this approach [16]. In this way, we believe that we could find some of the features of the system that were necessary to improve in order to reach better usability of the system. In addition, through constant interaction in meetings and direct contacts with users and their feedback we were discovering advantages and drawbacks of the system. In this way, we believe that we could find some of the features of the system that were necessary to improve in order to reach better usability.

After 13 months of running the system, we designed a structured questionnaire to obtain users’ opinions about different aspects of the interface and their use of it. Following the usability evaluation done by Chadwick *et al.* [17], we divided the 46 questions under six heads: 1) attitudes toward paper-based system; 2) perceived benefits of electronic records; 3) user satisfaction and ease of use of electronic interface; 4) problems when using the system; 5) likely future usage; and 6) attitude toward data security.

The questionnaire was Web based and available to each user of the system. As there was only one registered user per each medical center involved in the study, we received four completed questionnaires. Even though the number of users and responses is low, most of the obtained answers clearly pointed to some of the advantages and disadvantages of the EDC system. The summary of the results of the questionnaire is presented below. (The complete results of the questionnaire are presented in [18].)

### A. Attitudes Toward Paper-Based System

Most of the users were satisfied with the paper CRF. They did not find it too complex, but some of them expressed to have difficulties in searching for data in paper records. Moreover, all of them agreed that the analysis of data collected in the paper forms is a demanding process.

### B. Perceived Benefits of Electronic Records

Users expressed great understanding of benefits that the electronic records bring. They all shared the opinion that the system like this brings optimization of the overall process, and that it improves the analysis of collected data. They mostly had a positive attitude toward complete substitution of paper records with electronic records.

### C. User Satisfaction and Ease of Use of Electronic Interface

This portion of questions was the most important for the evaluation of the Cliniporator Web-Recorder application. We expected here, on one side, to see the flaws in solutions we built, and on the other side to get a confirmation for our ideas that we integrated into the system.

The overall evaluation of the application was positive. Users found the application fast enough for their needs. They were completely satisfied with the interface design and navigation concept. In their opinion, entering and searching for information was simple. The idea of having Web forms similar to the paper forms and organized into many short pages instead of few long-scrolling pages was well accepted. Users considered the interactive human map extremely useful. The automatic check of filled data was useful, as well as the signing of the forms.

Other application features, like image gallery, upload of the local-database data, and presentation of dynamic statistics were also positively accepted. The users considered useful both storing the images into the central database and accessing them in the image gallery. The upload of local-database data was convenient, and the overview of uploaded data useful. In their opinion, beside the electric-pulse data, the presentation of the electric-pulse figure was also important. Also, they were positive of putting basic statistics on the Web.

Apart from one user, all others had positive impression of the speed of fixing bugs in the system. However, they all agreed that the process of upgrading the system was convenient and not disturbing their work.

#### D. Problems When Using the System

In this portion, we tried to identify the weakest points in the system. As expected, filling of both paper and electronic CRF was inconvenient practice. This is something that will probably stay for considerable time, and all we can do is to make these two processes as similar as possible.

Our concern that the paper forms are more suitable for use during the medical procedure was confirmed. This is something that is mostly a question of technology advancement, and it is probably just a matter of time when this problem will be overcome. For example, use of tablet personal computer (PC) or pocket PC with wireless local area network (WLAN) connection would enable filling of electronic records even during the medical procedure. This procedure, however, needs to become acceptable for patients and medical doctors.

The three most important questions in this portion, from our point of view, were pointing the system accessibility and reliability. The users appraised the system as reliable, which is good. All of them, except one user, found the login procedure convenient. The main problem in the system was connectivity, as two users emphasized that it was happening occasionally that they could not access the system. We believe that the reason for this lies in significant system improvements that we introduced (establishing of secure connection and migration to a new server operating system) a few weeks before asking users to fill out the questionnaire, which coincided with the time of intensive data-verification activities. Unfortunately, this caused frequent system restarting in the days of the highest user activity. This, however, could be avoided by proper planning of activities and their harmonization with users' activities.

#### E. Likely Future Usage

The aim of this portion was to evaluate the present use and to highlight the direction of future improvements. The outlook for likely future use is good with each of the users stating that they would like to use the system even after the end of the project. The users estimated the overall system usefulness as positive, and expressed the wish to use similar EDC systems for other clinical trials. They did not agree on the advantages of Web-based medical-record system over the usual (non-Web based) computer medical-record system. Nevertheless, all the users agreed that an obvious advantage of this system was the support of multimedia content. One of the users was, however, concerned over the use of medical applications over the Internet.

Despite good evaluation results of system performances, users pointed out that most of the performances should be enhanced. They mostly expected that the application speed and reliability should be improved. One of the users clearly expressed that the application design should be improved. Another wished that the navigation be improved. Still another expected more dynamic statistics to be introduced.

#### F. Attitudes Toward Data Security

The last portion of the questionnaire referred to the system security. The general attitude of users was that the security measures were important in medical applications on the Web. They were willing to accept decrease in the system performances with an improvement in the overall system security. And finally, users expressed trust in the experts who implemented the system and thus showed no concern for the implemented security measures (log in with password and encrypted connection).

## VIII. DISCUSSION

Data presented by Paul *et al.* [19] show that in the last ten years, there has been a noticeable increase in the number of reported online clinical trials. This increase is driven by improvements in computer technologies with a special contribution of Internet technologies, which enabled simple solutions for many trial management and communication problems. Since the introduction of the Internet-based solutions for electronic data collection in medical applications, many of them were offering partial or complex solutions for clinical trials. In the beginning, authors were focused on developing generic systems that could enable quick and generic development of medical forms [20], [21]. Some authors were focused on different technological aspects and solutions like Java [21], XML [22], security aspects [23]–[25], or data validation [26]. However, according to our review of numerous available international journals, there are very few articles about noncommercial systems to support multicenter trials that integrate electronic data collection, data management including data validation, trial management, data reports (online statistical analysis), and strong security measures (including authentication, authorization, data encryption through SSL, and servers' protection) [27]–[32]. However, Brown gives an overview of the EDC market and commercial vendors offering software for clinical trials [3].

Based on the Internet technologies, our EDC system enables efficient collection of clinical-trial data within the ESOPE project from distant medical centers, thus easing and empowering trial management and result analysis. It enables data storage and backups, as well as control of entries, timeliness, and completeness of the data. In our opinion, the key feature of every EDC system is assuring high data quality. There are several approaches in achieving high data quality including: real-time data validation, automated e-mail notifications [32], presentation of complete medical-text summary [26], [31], different types of users with different roles in the process (e.g., user, investigator, monitor, and administrator) [28], and digital/electronic signatures [26], [28]. We tried to surpass this usually costly process by empowering user's autocorrection of submitted data through the use of "check tables" and completeness notification system, and by enabling parallel submission and monitoring of entries. We enabled different data-access permissions for different user types in order to support different roles that participants have in the trial.

This system was not only easy to implement but was also easy to maintain, which were among the initial goals. The system is scalable and flexible as it allows adding new functionalities and changing the existing ones. In two years of running the EDC system, we enlarged the initial set of functionalities with generation of dynamic basic statistics, image gallery, pulse-description window, and few smaller improvements for automatic data validation. Recently we added two important features: e-mail alert of submitting concluded form, and form-completeness-status notification.

The described EDC system can be easily expanded with new interfaces. We have already added a Cliniporator Service Module interface, which enables manufacturers of the Cliniporator device to query the central database for data essential to device functioning and electrode supplies [18].

The EDC system that we developed does not require advanced computer knowledge on the part of the medical personnel who use it. It supports multimedia content and interactivity. The implementation of Macromedia Flash animations for medical applications has been extremely well received by users. This along with the image-gallery feature and integration of data from different sources (data entered through the Web-forms and Cliniporator device local databases) makes our system one of a few noncommercial EDC systems incorporating these features.

Our EDC system includes security measures necessary for medical applications. Finally, as assessed by the questionnaire, our EDC system is well appraised and accepted by users themselves. The users showed their interest to continue using the system even after the end of the project. They also confirmed their need for similar systems for other clinical trials.

During two years of trial and system operation, the system was stable and reliable. Users were filling medical forms, submitting images, and sending local-database data. It successfully passed challenges like migrating server side of the system to the new operating system (from Windows Server 2000 to Windows Server 2003) and upgrading security measures by introducing security socket layer. A lot of smaller modifications were in-

stantly performed. At users' requests, few new features were added without disturbing their normal work.

## IX. CONCLUSION

In this paper, we presented a custom Web-based EDC system that we developed to support ESOPE clinical trials and also the usability evaluation of the system done through the questionnaire answered by users after using the system for 13 months. This EDC system had to enable data collection and frequent intermediate data analysis during at least two years of multicenter clinical trial.

Based on our experience, Web-based applications are suitable for distributed medical instrumentation service and monitoring. Another advantage of using Internet technologies is the support of multimedia content. Easy implementation of image, sound, video, or interactive animation content brings many possibilities for different practical solutions. Our image acquisition, interactive human map, and tumor-history graph are just some of them.

The Cliniporator Web-Recorder addresses an important challenge to healthcare: how to exploit Internet and Web technologies to improve clinical trials. The nature of the research process is such that it does not stand constraints that commercial EDC software and fixing of trial procedure brings. This is why we believe that many of the trials need dedicated EDC systems, extended with different features useful for the particular study. The Internet technology fills the gap between application development and deployment. This shifts up a gear in research process. Users/researchers can promptly shape the application design and functionalities. The usual two-way process between users/researchers and system developers can bring the maximum from otherwise usually rigid EDC systems. We want to share our experience and encourage others to go on this path.

## REFERENCES

- [1] S. N. Laxminarayan, "Editorial: Information technology in biomedicine: Maturational insights," *IEEE Trans. Inf. Technol. Biomed.*, vol. 6, no. 1, pp. 1-7, Mar. 2002.
- [2] K. Siau, "Health care informatics," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 1, pp. 1-7, Mar. 2003.
- [3] E. G. Brown. (2004, Sep). *Clinical trials' EDC endgame*. Forrester Research, Inc. [Online]. Available <http://www.forrester.com/go?docid=33344>
- [4] L. Alschuler, L. Bain, and R. D. Kush. (2004, Mar.). Improving data collection for patient care and clinical trials. Health Level 7 and Clinical Data Interchange Standards Consortium. [Online]. Available: <http://nextwave.sciencemag.org/cgi/content/full/2004/03/24/8>
- [5] J. Nielsen, *Usability Engineering*. New York: Academic, 1993.
- [6] X. Ferré, N. Juristo, H. Windl, and L. Constantine, "Usability basics for software developers," *IEEE Softw.*, vol. 18, no. 1, pp. 22-29, Jan./Feb. 2001.
- [7] A. W. Kushniruk and V. L. Patel, "Cognitive and usability engineering methods for the evaluation of clinical information systems," *J. Biomed. Inf.*, vol. 37, pp. 56-76, 2004.
- [8] C. M. Johnson, T. R. Johnson, and J. Zhang, "A user-centered framework for redesigning health care interfaces," *J. Biomed. Inf.*, vol. 38, pp. 75-87, 2005.
- [9] K. Schriver, "Evaluating text quality: The continuum from text-focused to reader-focused methods," *IEEE Trans. Prof. Commun.*, vol. 32, no. 4, pp. 238-255, Dec. 1989.
- [10] M. C. Beuscart-Zephir, F. Anceaux, H. Menu, S. Guerlinger, L. Watbled, and F. Evrard, "User-centered, multidimensional assessment method of clinical information systems: A case study in anesthesiology," *Int. J. Med. Inf.*, vol. 74, pp. 179-189, 2005.



- [11] A. Lecerof and F. Paternó, "Automatic support for usability evaluation," *IEEE Trans. Softw. Eng.*, vol. 24, no. 10, pp. 863–888, Oct. 1998.
- [12] A. W. Kushniruk, C. Patel, V. L. Patel, and J. L. Cimino, "'Televaluation' of clinical information systems: An integrative approach to assessing Web-based systems," *Int. J. Med. Inf.*, vol. 61, pp. 45–70, 2001.
- [13] L. V. Vaes, "Thinking aloud as a method for testing the usability of web-sites: The influence of task variation on the evaluation of hypertext," *IEEE Trans. Prof. Commun.*, vol. 43, no. 3, pp. 279–291, Sep. 2000.
- [14] J. Nielsen. (2000, Mar. 19), "Why you only need to test with 5 users," Alertbox. [Online]. Available: <http://www.useit.com/alertbox/20000319.html>
- [15] R. Virzi, "Refining the test phase of usability evaluation: How many subjects is enough?," *Hum. Factors*, vol. 34, pp. 457–468, 1992.
- [16] A. Woolrych and G. Cockton, "Why and when five test users aren't enough," in *Proc. IHM-HCI 2001*, Lille, France, Sep. 10–14, 2001, vol. 2, pp. 105–108.
- [17] D. W. Chadwick, D. Mundy, and J. New, "Experiences of using a PKI to access a hospital information system by high street opticians," *Comput. Commun.*, vol. 26, pp. 1893–1903, Oct. 2003.
- [18] I. Pavlović, "Web-based medical record system for clinical trials" M.S. thesis, Faculty Electr. Eng., Univ. Ljubljana, Ljubljana, Slovenia, 2004.
- [19] J. Paul, R. Seib, and T. Prescott, "The Internet and clinical trials: background, online resources, examples and issues," *J. Med. Internet Res.* [Online] 2005. 7, no. 1. e5 Available <http://www.jmir.org/2005/1/e5/>
- [20] T. Kiuchi and S. Kaihara, "Automated generation of a World Wide Web-based data entry and check program for medical applications," *Comput. Methods Programs Biomed.*, vol. 52, no. 2, pp. 129–138, Feb. 1997.
- [21] H. Sippel and C. Ohmann, "A web-based collection system for clinical studies using Java," *Med. Inf.*, vol. 23, no. 3, pp. 223–229, 1998.
- [22] P. Wübbelt, G. Fernandez, and J. Heymer, "Clinical trial management and remote data entry on the Internet based on XML case report forms," *Stud. Health Technol. Inf.*, vol. 77, pp. 333–337, 2000.
- [23] S. Gritzalis, J. Iliadis, D. Gritzalis, D. Spinellis, and S. Katsikas, "Developing secure Web-based medical applications," *Med. Inf.*, vol. 24, no. 1, pp. 75–90, 1999.
- [24] H. M. Chao, C. M. Hsu, and S. G. Miaou, "A data-hiding technique with authentication, integration, and confidentiality for electronic patient records," *IEEE Trans. Inf. Technol. Biomed.*, vol. 6, no. 1, pp. 46–53, Mar. 2002.
- [25] M. van der Haak, A. C. Wolf, R. Bradner, P. Drings, M. Wannenmacher, and T. Wetter, "Data security and protection in cross-institutional electronic patient records," *Int. J. Med. Inf.*, vol. 70, pp. 117–130, 2003.
- [26] R. G. Marks, "Validating electronic source data in clinical trials," *Controlled Clin. Trials*, vol. 25, pp. 437–446, 2004.
- [27] M. Formica, K. Kabbara, R. Clark, and T. McAlindon, "Can clinical trials requiring frequent participant contact be conducted over the Internet? Results from an online randomized controlled trial evaluating a topical ointment for herpes labialis," *J. Med. Internet Res.*, vol. 6, no. 1, 2004.
- [28] C. D. Lallas, G. M. Preminger, M. S. Pearle, R. J. Leveillee, J. E. Lingeman, and J. P. Schwoppe, "Internet based multi-institutional clinical research: A Convenient and secure option," *J. Urol.*, vol. 171, no. 5, pp. 1880–1885, May 2004.
- [29] C. H. Elsner *et al.*, "Open source or commercial products for electronic data capture in clinical trials? A scorecard comparison," *Comput. Cardiol.*, vol. 30, pp. 371–373, 2003.
- [30] S. J. Rangel, B. Narasimhan, N. Geraghty, and R. L. Moss, "Development of an internet-based protocol to facilitate randomized clinical trials in pediatric surgery," *J. Pediatr. Surg.*, vol. 37, no. 7, pp. 990–994, 2002.
- [31] R. Marks, H. Bristol, M. Conlon, and C. J. Pepine, "Enhancing clinical trials on the Internet: Lessons from INVEST," *Clin. Cardiol.*, vol. 24, suppl. V, V17–V23, 2001.
- [32] R. A. Hollingsworth, C. Hay, and B. Richards, "An Internet implementation of an international clinical study," *Stud. Health Technol. Inf.*, vol. 68, pp. 528–531, 1999.



**Ivan Pavlović** was born in Belgrade, Serbia and Montenegro, in 1973. He received the B.Sc. degree from Belgrade University, Belgrade, in 1999, and the M.S. degree in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, in 2004.

Currently, he is a Research Associate with the Faculty of Electrical Engineering, University of Ljubljana. He is active in the field of biomedical engineering. His current research interests include Web-based computer systems for clinical trial electronic data collection.



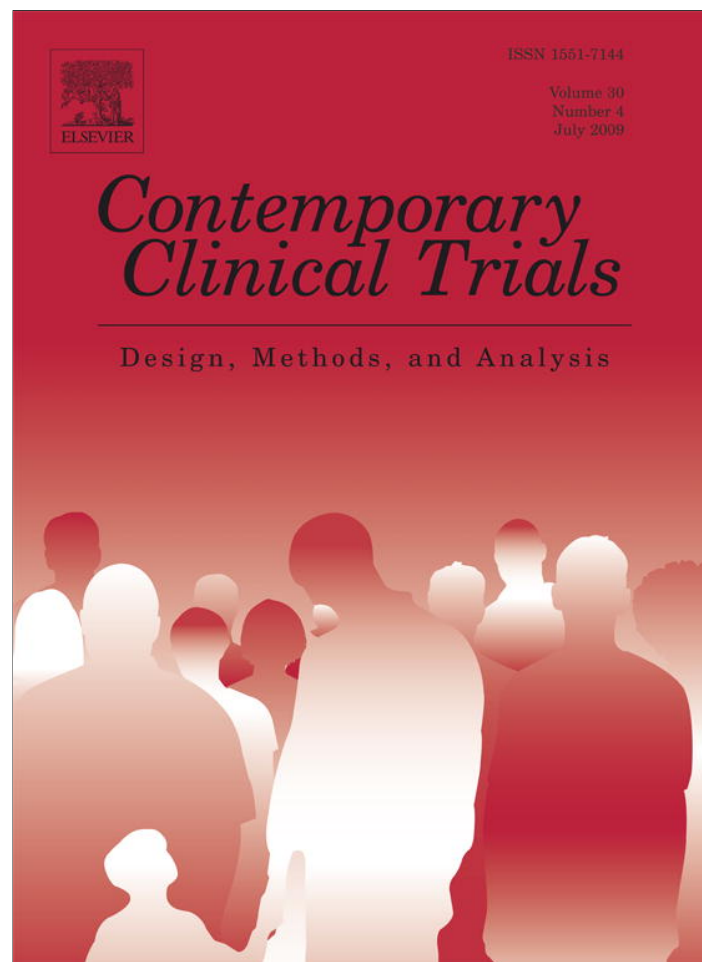
**Damijan Miklavčič** was born in Ljubljana, Slovenia, in 1963. He received the Ph.D. degree in electrical engineering from the University of Ljubljana, Ljubljana.

Currently, he is a Professor with the Faculty of Electrical Engineering, University of Ljubljana, and the Head of the Laboratory of Biocybernetics. He is active in the field of biomedical engineering. His current research interests include electroporation-assisted drug delivery, including cancer treatment by means of electrochemotherapy, tissue oxygenation,

and modeling.



Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

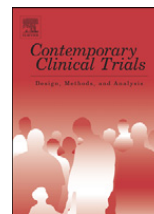
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

# Contemporary Clinical Trials

journal homepage: [www.elsevier.com/locate/conclintrial](http://www.elsevier.com/locate/conclintrial)

## Comparison of paper-based and electronic data collection process in clinical trials: Costs simulation study

Ivan Pavlovič<sup>a,b</sup>, Tomaž Kern<sup>c</sup>, Damijan Miklavčič<sup>a,\*</sup><sup>a</sup> University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, SI-1000 Ljubljana, Slovenia<sup>b</sup> Institute for Project Management and Information Technology, Kotnikova 30, SI-1000 Ljubljana, Slovenia<sup>c</sup> University of Maribor, Faculty of Organizational Sciences, Kidričeva cesta 55a, SI-4000 Kranj, Slovenia

### ARTICLE INFO

#### Article history:

Received 25 November 2008

Accepted 22 March 2009

#### Keywords:

Clinical trial

Data collection

Electronic data collection

EDC

Internet

Business process

Process models

Costs

Simulation study

### ABSTRACT

An alternative to clinical trial paper-based data collection (PDC) is internet based electronic data collection (EDC), where the investigators over the internet enter data directly in the electronic database by themselves. In our study we considered clinical trial as a business process. Our objective was to model PDC and EDC process and to estimate the difference of the costs of PDC and EDC process for a sample clinical trial based on these models.

We used Extended Event-driven Process Chains (eEPC) modeling technique to model PDC and EDC process. In order to evaluate the costs of the processes we assigned costs functions to each process function which appears in the model. The parameters which appear in these functions include efforts, staff prices and data quality parameters. We estimated the values of all these parameters and performed costs calculations for a sample clinical trial.

Through an analysis and modeling efforts we identified sub-processes which contain main differences affecting duration and costs of the PDC and EDC process: data gathering at the research center; monitoring; and data management. The most significant model difference between PDC and EDC process appeared in data management sub-process. For the sample clinical trial considered in our simulation study and our parameters estimations the EDC process decreased data collection costs for 55%. For different scenarios of parameters variations we show that the EDC process may bring from 49% to 62% of savings when compared to PDC process.

© 2009 Elsevier Inc. All rights reserved.

### 1. Introduction

Clinical trial (CT) is “any investigation in human subjects intended to discover or verify the clinical, pharmacological and/or other pharmaco-dynamic effects of one or more medicinal product(s)” (for full definition see [1]). CT can be carried out in either one or multiple sites. One of the core documents in CT is the Case Report Form (CRF). The CRF is a form where the investigator enters all patients' clinical and non-clinical data related to the trial. This is like a CT dossier of

the patient. Data collected in the CRF are both: data related to the patient health parameters; and data related to the medical procedure performed in the trial.

The data collected in the paper forms have to be entered in the electronic database in order to perform computer data analysis. For this purpose investigators usually send the copies of paper CRF to the data center where data managers enter these data into the database. This paper-based routine has many disadvantages which result in erroneous data in the database and longer duration of clinical trial (especially for the large multi-centre clinical trials) [2,22].

An alternative to paper data collection is internet based electronic data collection (EDC), where the investigators over the internet enter data in the electronic database by themselves. In this way the errors from copying data from paper forms to electronic database by the person who did not

\* Corresponding author. Fakulteta za elektrotehniko, LBK, Tržaška 25, 1000 Slovenia. Tel.: +386 1 4768 456; fax: +386 1 4264 658.

E-mail addresses: [ivan.pavlovic@fe.uni-lj.si](mailto:ivan.pavlovic@fe.uni-lj.si) (I. Pavlovič), [tomaz.kern@fov.uni-mb.si](mailto:tomaz.kern@fov.uni-mb.si) (T. Kern), [damijan.miklavcic@fe.uni-lj.si](mailto:damijan.miklavcic@fe.uni-lj.si) (D. Miklavčič).

collect the data are avoided. Also, electronic Case Report Forms (eCRF) usually contain check routines which reduce erroneous data entry. Another advantage of EDC is that the data managers have continuous insight into the data and the data collection process and thus can manage the whole CT process better. Good overviews of the advantages of electronic data collection were provided by Welker [3] and Brandt et al. [4].

Despite the fact that the EDC tools have been available for more than two decades, clinical trials however are still mainly conducted using paper data collection as the primary tool (according to [5] in 2004 it was still over 75%). The reason for this was partially ascribed to the fact that technological applications often did not have adequate functionality to improve the data collection process as a whole and that applied technology was expensive to introduce and maintain. In recent years, a development of internet technologies and its availability reduced many technological obstacles. Many authors namely report successful implementation of internet based EDC solutions [6–14]. However, without being able to predict the costs, duration and quality of processes performed, a change from PDC to EDC is still a risky decision.

In order to facilitate decision between PDC and EDC use in clinical trial (CT), with respect to quality and cost of the process, in our study we considered CT as a business process. The CT as a business process is complex and includes different sub-processes and activities like: protocol development; protocol use and implementation in the CT experimentation; data collection; and the evaluation of the CT results. Each sub-process and activity has different objectives and is performed in different environments, carried out by its own agents and resources, and governed by specific rules. Some results of the efforts to develop a comprehensive model of the entire CT process were already published by Luzi et al. [15,16]. More

specific models which deal only with CT data flow were published by PhRMA (Pharmaceutical Research and Manufacturers of America) [17]. The technological improvements may bring major changes to CT data collection process through the change from paper data collection (PDC) process to electronic data collection (EDC) process. Therefore, we focus our work on data collection process with its sub-processes (data gathering, monitoring and data management). In this paper we present the results of our modeling of PDC and EDC process as well as the estimations of the difference of the costs of PDC and EDC process for sample clinical trial.

## 2. Methods

There are several business process modeling techniques. Among others, these include eEPC (Extended Event-driven Process Chains), BPEL (Business Process Execution Language) and UML (Unified Modeling Language). We decided to use eEPC modeling technique because it is “user perspective” and allows to model both IT (Information Technology) supported functions as well as “manual” functions.

On Fig. 1 we present the main elements of an eEPC graph, which we used to model EDC and PDC process. This example shows that each function in the process is triggered by some event and must end with another event (which usually triggers another function). Different human or organizational resources may be involved in different ways in completing the function. Also some (input or output) documents as well as supporting computer systems may appear in the process. Beside the elements presented on Fig. 1, an eEPC graph also may contain some branching or merging elements like “^” (AND), “v” (OR) or “x” (XOR).

A detail description of eEPC model elements can be found in ARIS documentation [18].

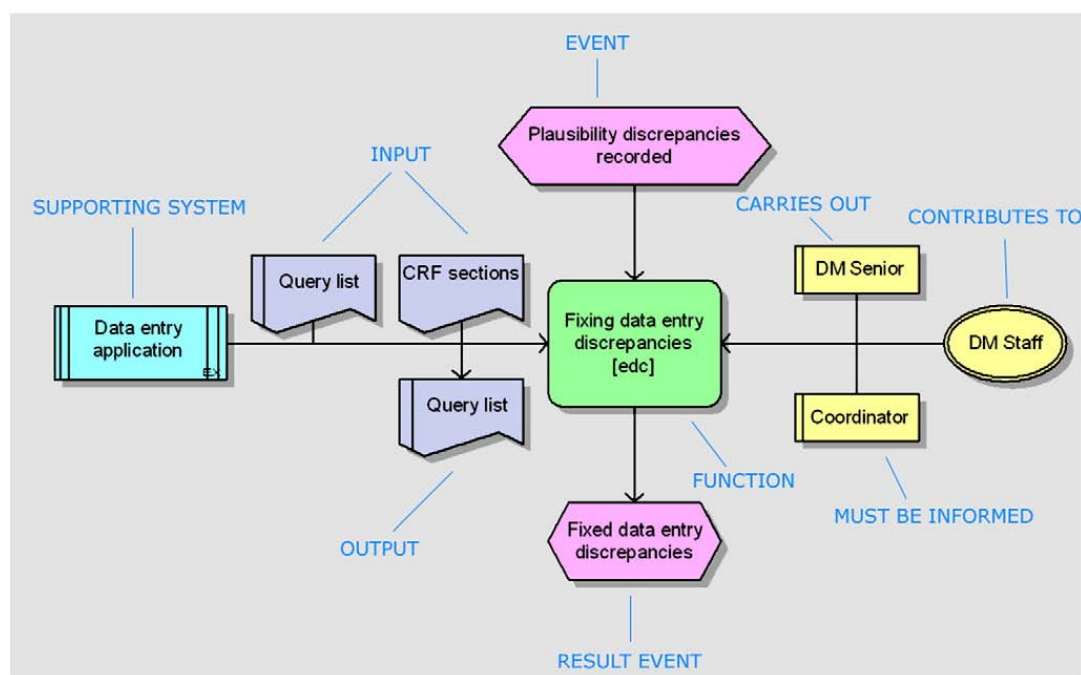


Fig. 1. eEPC graph elements.

2.1. Process models

We decided to model EDC and PDC process, because simple analysis would only yield specific results (for specific

case). The model however gives us the opportunity to insert different values in a parameterization and sensitivity study.

We are aware of the fact that every clinical trial has its own characteristics and that each CT sponsor might have different

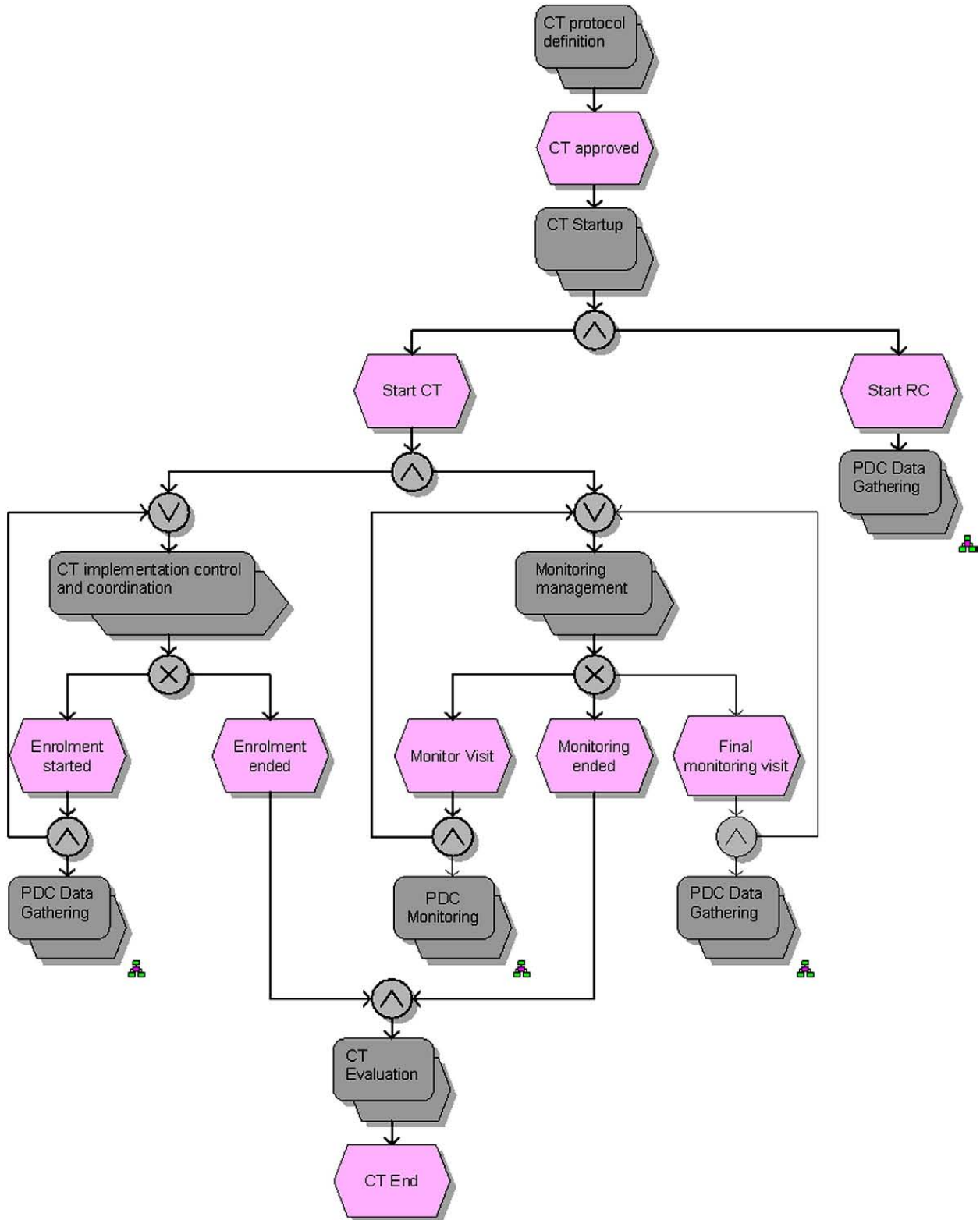


Fig. 2. PDC CT Management model. EDC model is the same as the presented PDC model, being different only in sub-processes. (CT – clinical trial, RC – research center).

approach to data collection. For paper-based data collection these include sending Case Report Forms (CRF) to sponsor by investigator after monitor (also known as Clinical Research Associate – CRA) verifies data, or sending CRF by monitor, or sending CRF by fax prior to monitoring, etc. However, we decided to model PDC and EDC processes which are widely accepted, having in mind that PDC and EDC processes both have to be designed to reach the same data quality.

We therefore made an overview of the literature and we performed a set of interviews with key participants in PDC and

EDC process (investigator, monitor, study coordinator, data manager) followed by e-mail correspondence. In the models we included functions which assure data quality and improve performances of each of the two approaches (PDC and EDC).

Here we present eEPC (Extended Event-driven Process Chains) models that we developed in order to compare two approaches to clinical trial (CT) data collection: paper data collection (PDC) model and electronic data collection (EDC) model. For both data collection approaches CT management can be presented by the same model as presented in Fig. 2.

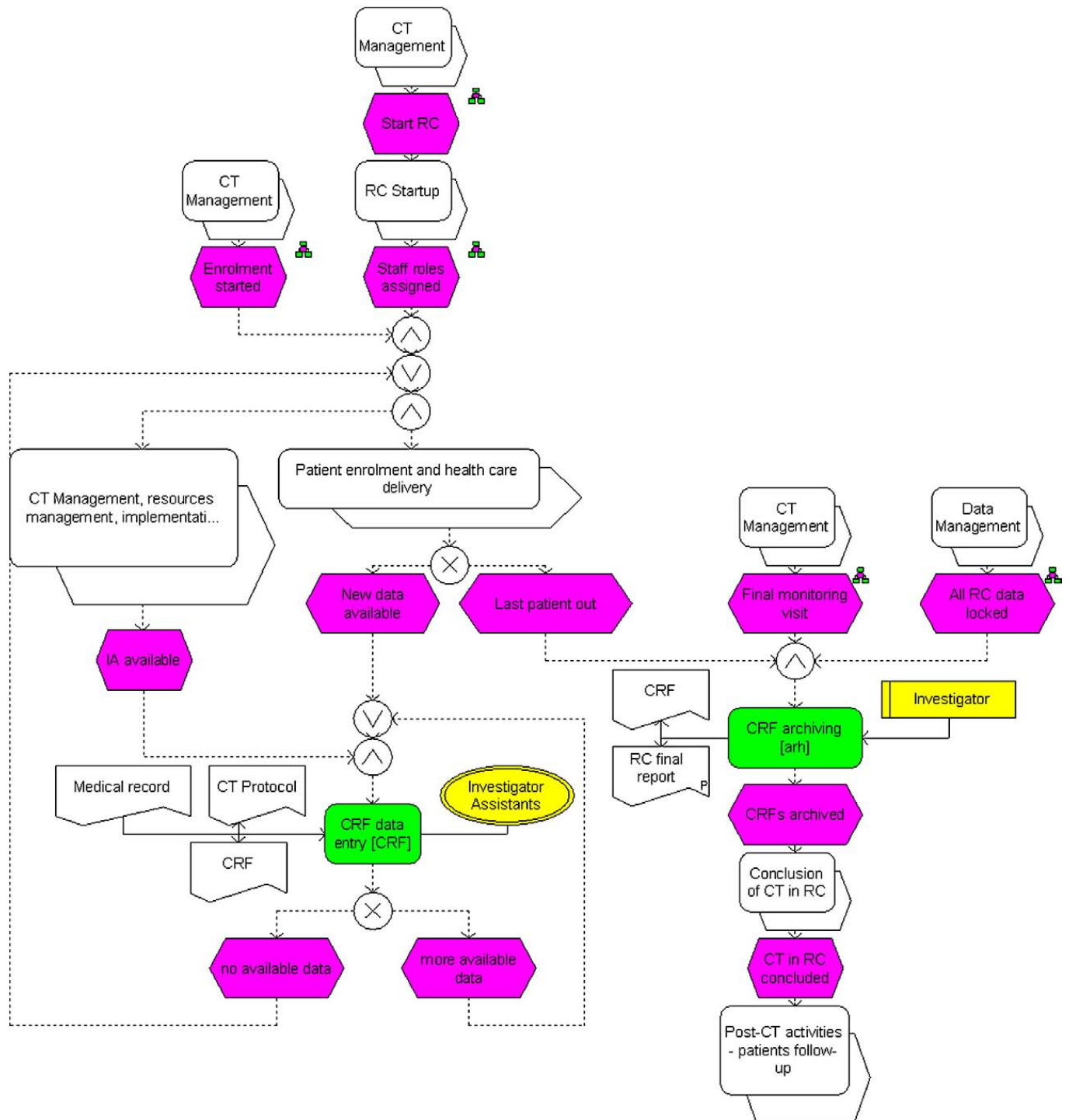


Fig. 3. PDC Data Gathering model (CT – clinical trial, RC – research center, CRF – Case Report Form, IA – investigator assistant).

In our study we focused on sub-processes which contain main differences affecting duration and costs of the processes. These are:

- data gathering at the research center (PDC Data Gathering vs. EDC Data Gathering);
- monitoring (PDC Monitoring vs. EDC Monitoring); and
- data management (PDC Data Management vs. EDC Data Management).

In the following subsections these sub-processes are described in details.

2.1.1. Data gathering at the research center

One of the key roles of the research center is to provide all the important information as required by CT Protocol and associated documentation. Data have to be gathered with high accuracy and delivered to the sponsor in a timely manner. With high level of generalization we may present the process of data gathering at the research site as depicted in Fig. 3 for PDC and Fig. 4 for EDC.

2.1.2. Monitoring

In order to improve the quality of data at the research center a sponsor assigns monitors to visit research centers

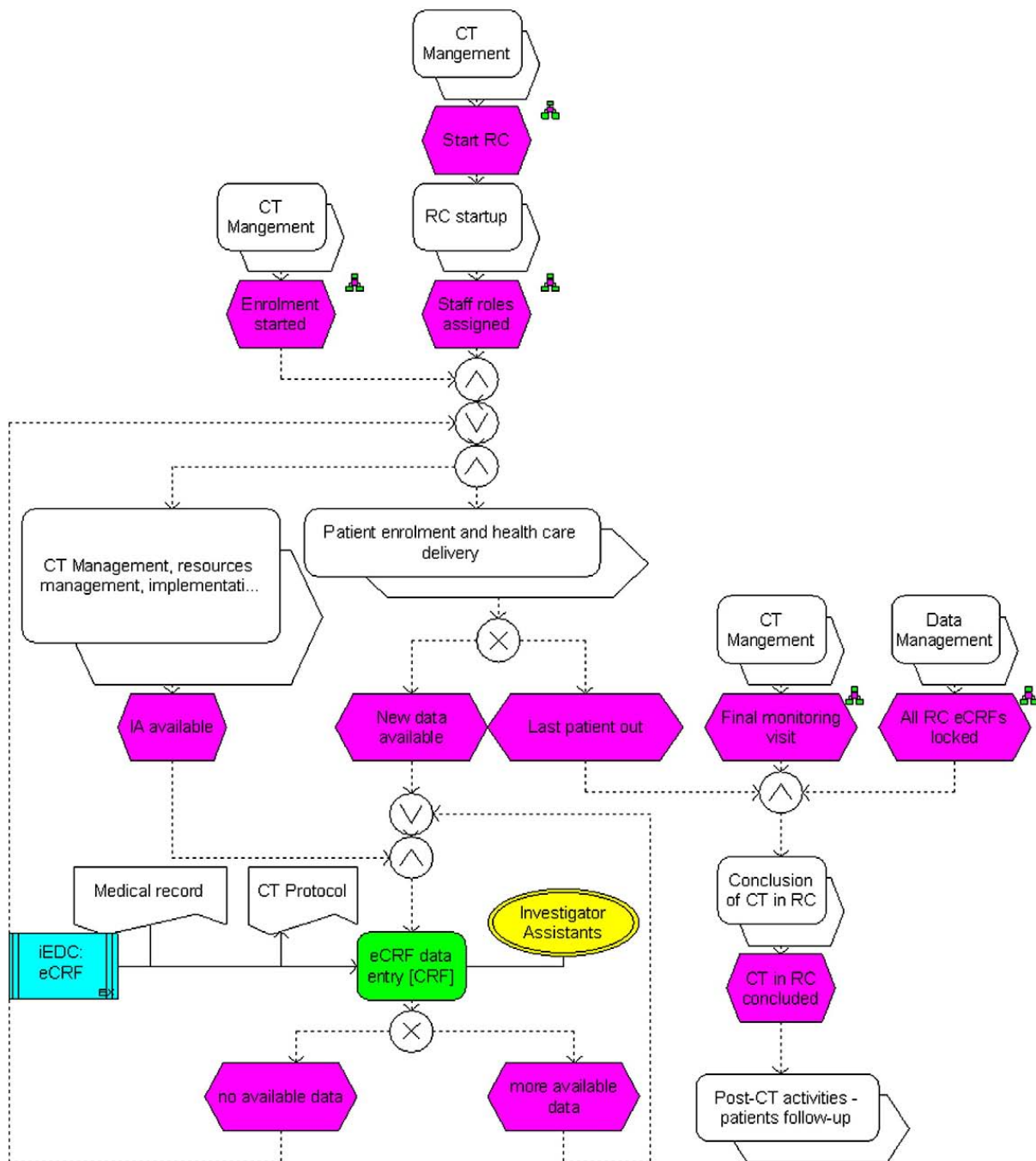


Fig. 4. EDC Data Gathering model (CT – clinical trial, RC – research center, eCRF – electronic Case Report Form, iEDC – electronic data collection application).



intermittently and, in addition to other tasks, check how the data collection is progressing with the aim of improving data quality. The change in the way data are collected (paper vs. electronic) also has considerable impact on monitoring process itself (Figs. 5 and 6).

### 2.1.3. Data management

Data gathered at the research center and verified by monitor are provided to data center where data management staff takes care of data provided by all the participating research centers and perform all the necessary data analysis. In order to be able

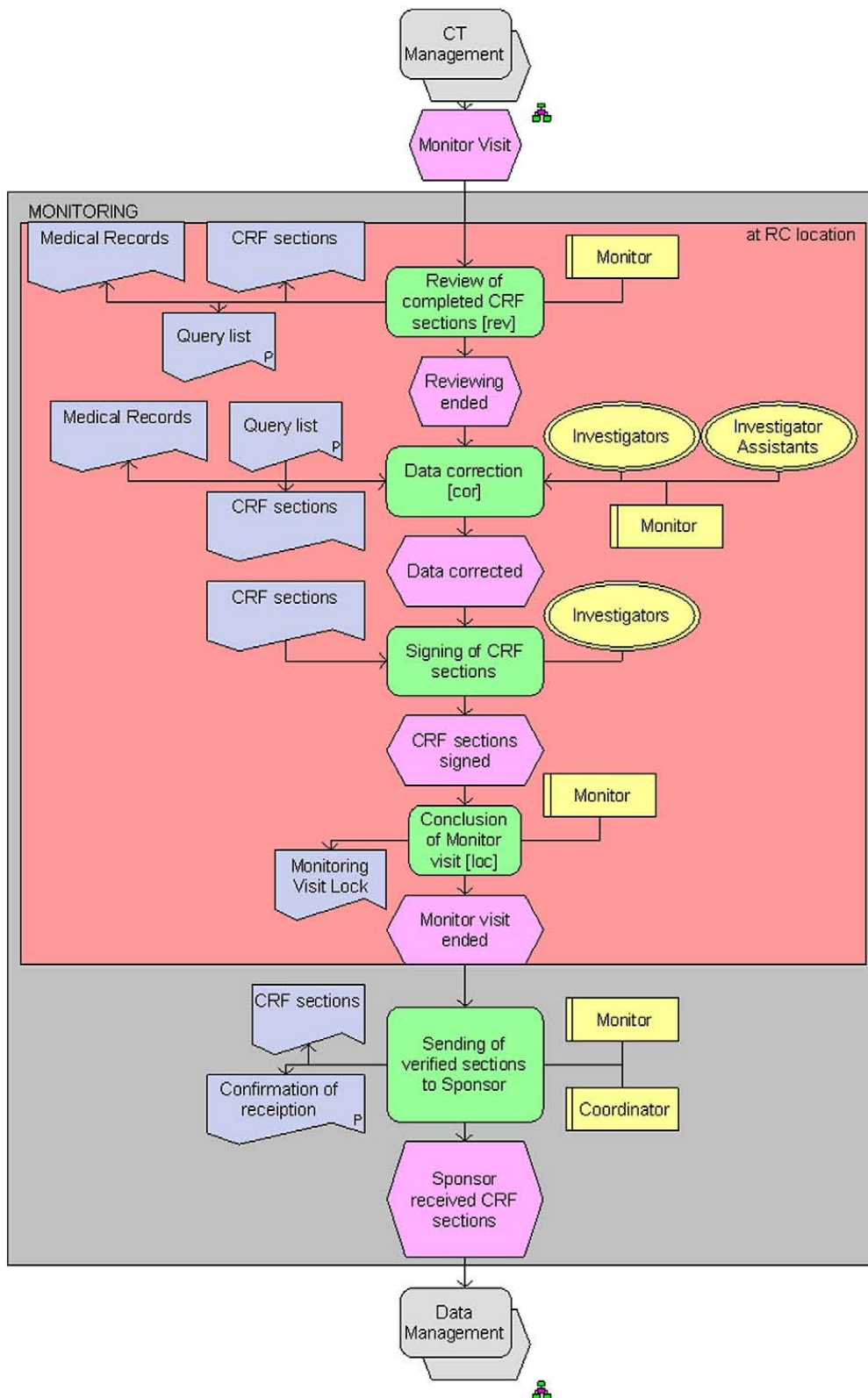


Fig. 5. PDC Monitoring model (CT – clinical trial, CRF – Case Report Form).

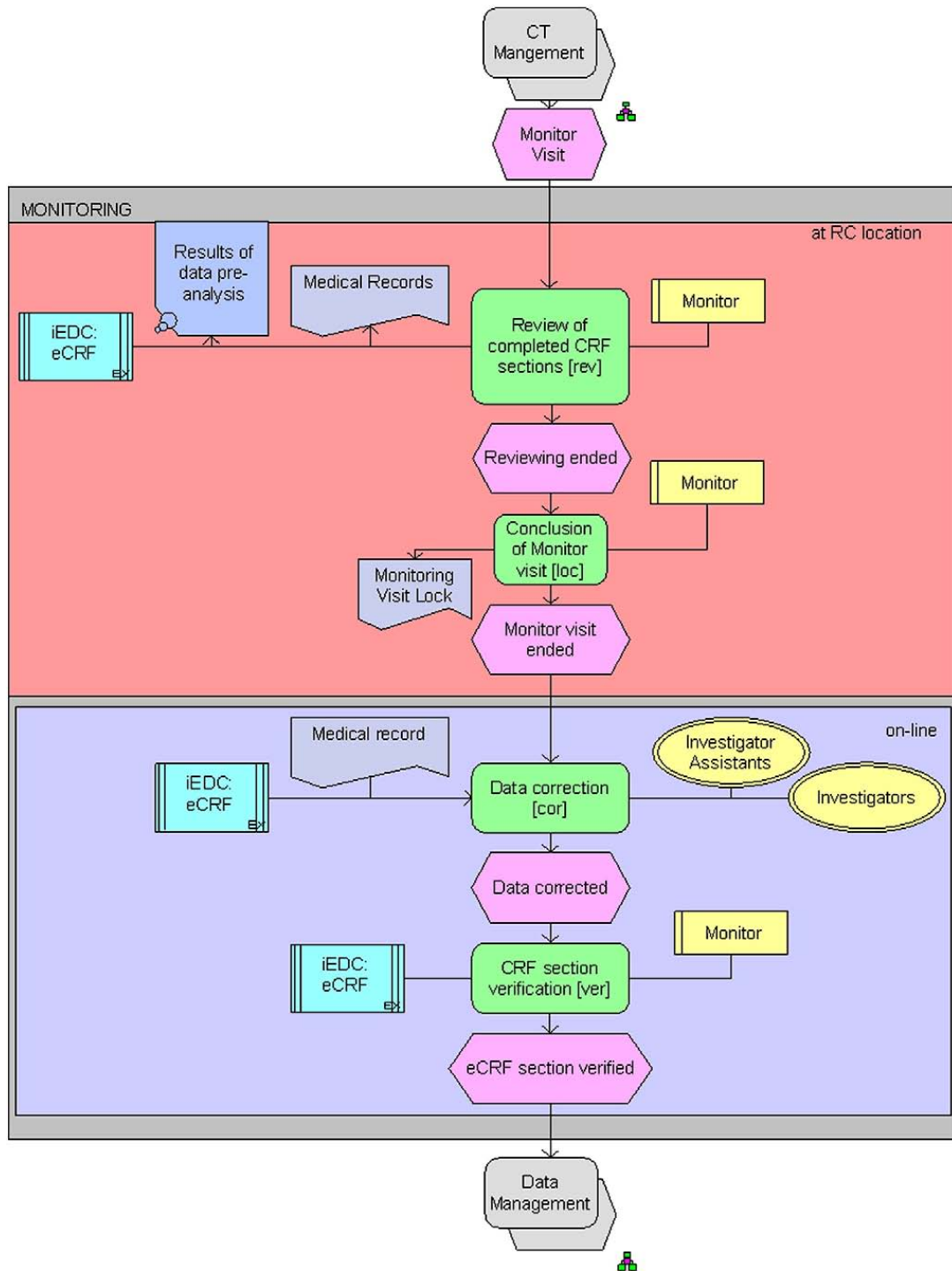


Fig. 6. EDC Monitoring model (CT – clinical trial, eCRF – electronic Case Report Form, iEDC – electronic data collection application).

to perform data analysis data have to be computerized, and are most often stored in central database. Therefore, PDC Data Management process must also include digitalization of data provided on paper forms while EDC Data Management may focus exclusively on data cleaning. The entire data management process is presented in Fig. 7 for PDC and in Fig. 8 for EDC.

### 2.2. Model parameters

The key property of PDC and EDC process which we wanted to compare was the cost of the process. The costs of the business

process are difficult to evaluate as they are affected by the staff efforts, the price of the staff, the price and amortization of the other resources used (such as equipment, materials, etc...), and numerous other economical parameters.

In our analysis we do not consider the costs of the clinical trial process which are not directly related to data collection process (e.g. patient recruitment, or health delivery). Furthermore, we do not consider the costs of the technical resources used in the processes (e.g. the price of the EDC system) or the amortization costs. We focus only on the staff costs and try to recognize how these depend on the process

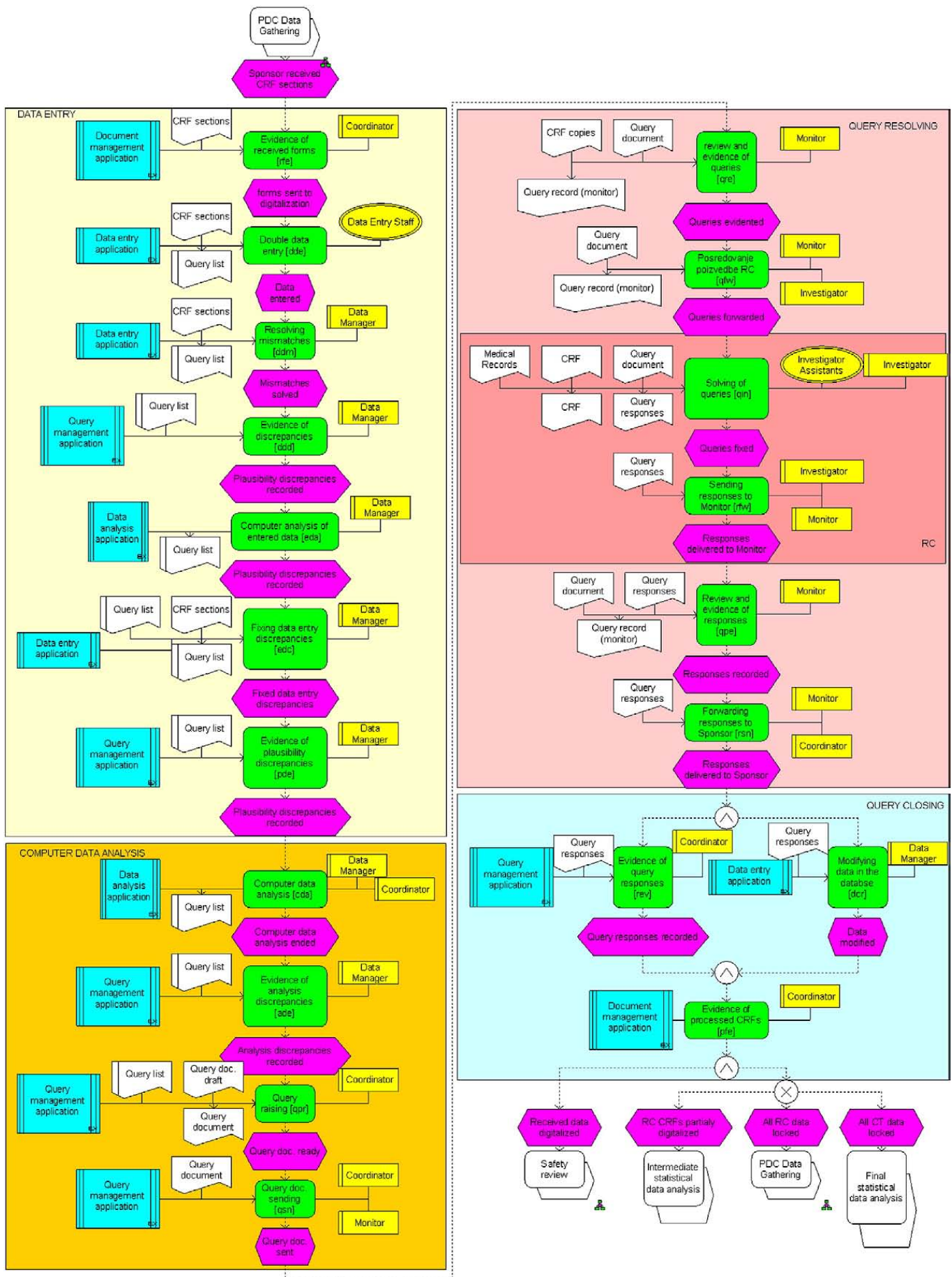


Fig. 7. PDC Data Management model (CT – clinical trial, CRF – Case Report Form, RC – research center).

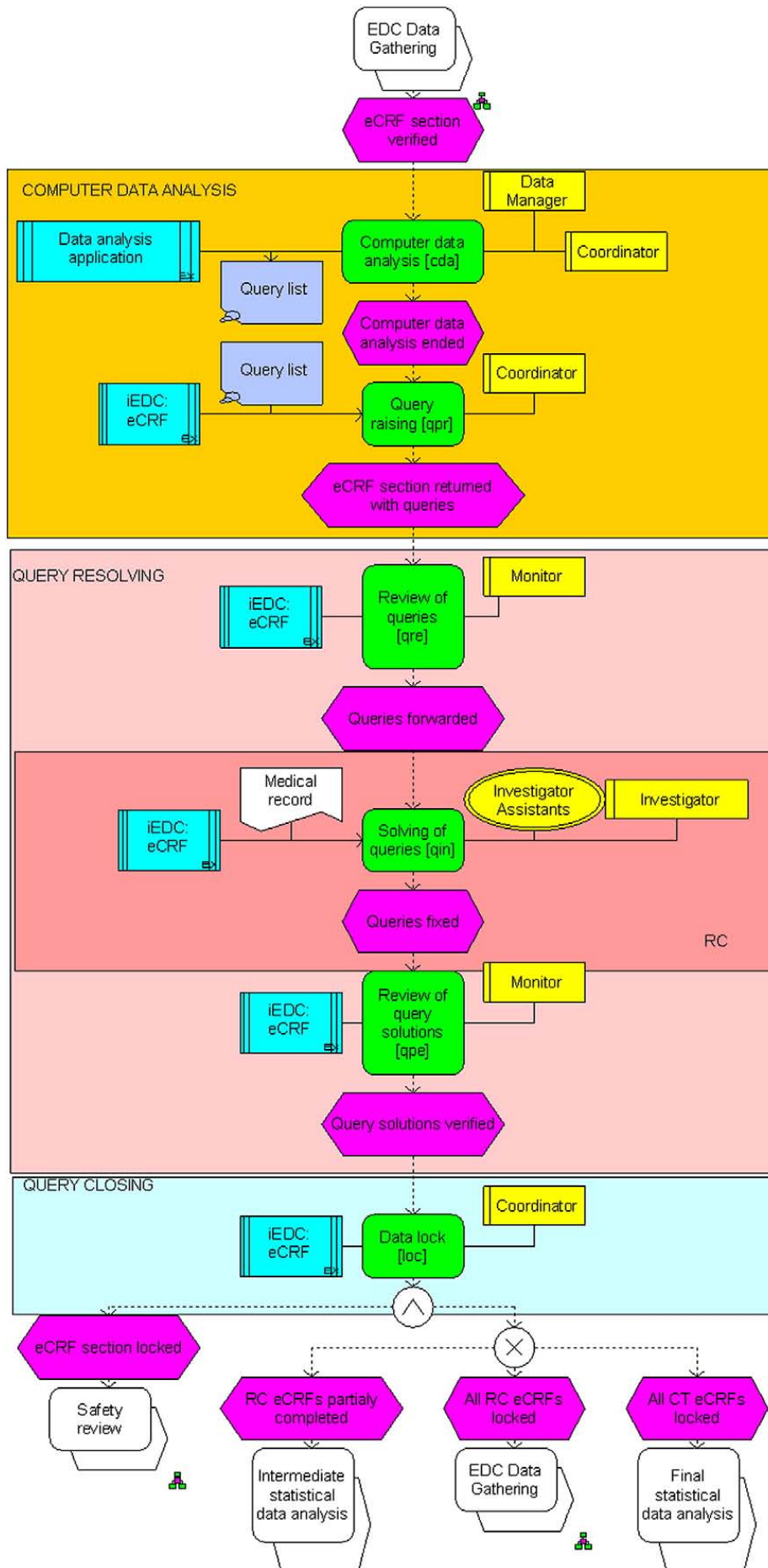


Fig. 8. EDC Data Management model (CT – clinical trial, eCRF – electronic Case Report Form, RC – research center, iEDC – electronic data collection application).

we choose, PDC or EDC. The staff costs depend on staff prices and the efforts required by each of the tasks included in data collection process. Therefore, in order to evaluate the costs we consider the following groups of parameters: efforts, staff prices and data quality parameters.

### 2.2.1. Efforts

The task (function) effort is the time that human resources (participants in the process) have to dedicate to the task in order to complete it. If we exclude all waiting times, this time includes: the preparation time; the execution time; and the conclusion time. In this analysis we simplify the calculation by considering all these three times together as task duration and trying to assess it for each task separately. In our equations the efforts are represented with the symbol  $F$ .

For each of the tasks, when calculating task duration, we took in consideration if any of the parameters of the clinical trial affects the task duration (e.g. the number of data in the CRF, or the number of CRF documents, etc.). Therefore, we introduced average efforts as a measure of time needed for task execution of one instance of the related parameter (e.g. time needed for writing down one data into paper CRF). In our equations we represent the average efforts with the symbol  $\mu$ .

**2.2.1.1. Data gathering at the research center.** From the information obtained in interviews, we estimate that the average data entry speed is 5 data per minute for PDC and 10 data per minute for EDC. Faster EDC data fill in is possible due to the fact that EDC usually offers lists of values (from look-up tables/vocabularies/code lists) and can auto-calculate some fields and, in general, it is quicker for average computer user to type the text then to hand-write it. Therefore, we use the following estimations in our calculations:

- writing down data in CRF –  $\mu_{DC_{CRF}^P} = 12 \frac{s}{data}$
- entering data in eCRF –  $\mu_{DC_{CRF}^E} = 6 \frac{s}{data}$
- archiving paper CRF –  $\mu_{DC_{arh}^P} = 12 \frac{min}{doc}$

**2.2.1.2. Monitoring.** During data verification monitor has to note down all doubtful, or erroneous data and this is done differently for PDC and EDC (e.g. EDC application enables entering comment/warning next to doubtful data field, while in PDC monitor has to write down on a paper a notice which contains comment/warning together with the reference to the field in the form). Therefore we can consider that the effort for taking a note of possible error is twice bigger for PDC than for EDC ( $\mu_{M_{er}}^P = 2\mu_{M_{er}}^E$ ).

Based on information obtained from interviews we use the following average efforts estimation in our calculations:

- noting down an error (PDC) –  $\mu_{M_{er}}^P = 1 \frac{min}{data}$
- noting down an error (EDC) –  $\mu_{M_{er}}^E = 30 \frac{s}{data}$
- checking a data –  $\mu_{M_{chk}} = 30 \frac{s}{data}$
- locking monitor visit –  $\mu_{M_{loc}} = 2 \frac{min}{visit}$
- correcting data –  $\mu_{M_{cor}}^P = \mu_{M_{cor}}^E = 1 \frac{min}{data}$
- signing CRF section (include verifying) –  $\mu_{M_{sig}}^P = 1 \frac{min}{doc}$
- sending CRF section by monitor to sponsor –  $\mu_{M_{snd}}^P = 2 \frac{min}{doc}$
- verifying CRF section –  $\mu_{M_{ver}}^E = 1 \frac{min}{doc}$

### 2.2.1.3. Data management

**2.2.1.3.1. Data entry.** We estimate average efforts needed for each step in data entry process as follows:

- evidence of received form –  $\mu_{DM_{erf}^P} = 1 \frac{min}{doc}$
- entering data in database –  $\mu_{DM_{dde}^P} = 3 \frac{s}{data}$
- resolving mismatches –  $\mu_{DM_{ddm}^P} = 1 \frac{min}{data}$
- evidence of query –  $\mu_{DM_{pde}^P} = 1 \frac{min}{data}$

**2.2.1.3.2. Computer data analysis.** We estimate that average effort for generating query document is twice higher for PDC than EDC, because, in the case of PDC, the coordinator has to write down both the question and to which field it is related. In the case of EDC the application enables the coordinator to enter the query right beside the field to which it relates.

Therefore, we estimate the average efforts as follows:

- evidence of query –  $\mu_{DM_{ade}^P} = 1 \frac{min}{data}$
- building query document (PDC) –  $\mu_{DM_{qpr}^P} = 2 \frac{min}{data}$
- building query document (EDC) –  $\mu_{DM_{qpr}^E} = 1 \frac{min}{data}$
- sending query document by sponsor to monitor –  $\mu_{DM_{qsn}^P} = 2 \frac{min}{doc}$

**2.2.1.3.3. Query resolving.** In the case of EDC no effort is needed to make evidence of queries and responses, as the EDC application takes care by itself of handling this information. Thus we consider that on average less effort is needed for review of each query in EDC process than in PDC (e.g.  $\mu_{DM_{qre}^E} = 2\mu_{DM_{qre}^P}$ ).

According to the interviewers the average effort for resolving one query is estimated as five times higher than effort needed to fill in data into CRF ( $\mu_{DM_{qin}^E} = 5\mu_{CRF}^P$ ). When EDC application is used this can be estimated to be twice less ( $\mu_{DM_{qin}^E} = 2.5\mu_{CRF}^P$ ) as the investigator has all the information, apart from the data source documents, already available in the EDC application.

In our calculations we estimate the average efforts as follows:

- review and evidence of query (PDC) –  $\mu_{DM_{qre}^P} = 1 \frac{min}{data}$
- review of query (EDC) –  $\mu_{DM_{qre}^E} = 15 \frac{s}{data}$
- forwarding query list by monitor to RC –  $\mu_{DM_{qfw}^P} = 2 \frac{min}{doc}$
- solving a query (PDC) –  $\mu_{DM_{qin}^P} = 1 \frac{min}{data}$
- solving a query (EDC) –  $\mu_{DM_{qin}^E} = 45 \frac{s}{data}$

**2.2.1.3.4. Query closing.** The related average efforts can be estimated as:

- evidence of query response –  $\mu_{DM_{rev}^P} = 1 \frac{min}{data}$
- correcting data in the database –  $\mu_{DM_{dcr}^P} = 2 \frac{min}{data}$
- evidence of processed form –  $\mu_{DM_{ecf}^P} = 1 \frac{min}{doc}$
- review of query response (EDC) –  $\mu_{DM_{loc}^E} = 15 \frac{s}{data}$

### 2.2.2. Staff prices

The costs of the process depend on the staff prices. In our analysis we estimated staff prices as follows:

- Investigator:  $\varphi_I = 20 \frac{€}{h}$
- Investigator assistant:  $\varphi_{IA} = 10 \frac{€}{h}$
- Monitor (Clinical Research Associate – CRA):  $\varphi_{CRA} = 15 \frac{€}{h}$
- Data entry staff:  $\varphi_{DE} = 6 \frac{€}{h}$
- Data manager (senior):  $\varphi_{DM} = 20 \frac{€}{h}$
- Study coordinator:  $\varphi_C = 25 \frac{€}{h}$

The prices can vary from center to center, from country to country. As needed, the above estimations can be replaced with specific prices for particular clinical trial.

### 2.2.3. Data quality

For most clinical trials data quality is set as a requirement. Certain level of data quality (low proportion of erroneous data) has to be assured in order to have reliable study results. There are several factors which affect the number of errors and these include:

- Error rate discovered by monitor – according to our interviewers for PDC  $\varepsilon_{\text{mon}}^{\text{P}} = 10\%$ , and it is reported in [24] that for EDC  $\varepsilon_{\text{mon}}^{\text{E}}$  can be estimated as 20%  $\varepsilon_{\text{mon}}^{\text{P}}$ , which gives  $\varepsilon_{\text{mon}}^{\text{E}} = 2\%$ .
- Error rate introduced by data entry staff (entry error rate –  $\varepsilon_{\text{ent}}^{\text{P}}$ ) – according to literature [19–23] can be around 1%.
- Overall query rate ( $\varepsilon_{\text{qry}}^{\text{P}}$ ) – according to our interviewers for PDC  $\varepsilon_{\text{qry}}^{\text{P}} = 2\%$ ; different sources (mentioned in [25]) pride EDC on 80% to 95% reduction on queries which let us estimate  $\varepsilon_{\text{qry}}^{\text{E}} = 0.2\%$ .
- Plausibility discrepancies (data out of range, missing values, invalid combinations, etc.) – according to our interviewers and Spink [24] for PDC it represents 80% of all queries ( $\varepsilon_{\text{pla}}^{\text{P}} = 80\% \varepsilon_{\text{qry}}^{\text{P}}$ ).
- Analysis discrepancies (such as protocol violations or erroneous data, discovered by data analysis application and DM staff) –  $\varepsilon_{\text{ana}}^{\text{P}} = 20\% \varepsilon_{\text{qry}}^{\text{P}}$ .
- Rate of data which have to be corrected after query resolving – according to our interviewers this can be estimated as half of all the queries  $\varepsilon_{\text{cor}}^{\text{P}} = 50\% \varepsilon_{\text{qry}}^{\text{P}}$ .

The rates listed above are further discussed in the Results chapter and used to calculate the process costs.

### 2.3. Sample clinical trial

For the purpose of costs evaluation we decided to conceive a sample clinical trial. The characteristics of our sample clinical trial are:

- 10 research centers (RC)
- 100 patients per center (all together 1000 patients)
- CRF contains 10 sections
- Entire CRF contains 1000 data
- 24 months study duration
- Monitor visits each RC once a month.

Therefore we used the following values in our calculations:

- Total number of CRF forms (equal to the number of patients) –  $N_{\text{CRF}} = 1000$
- Total number of CRF sections –  $N_{\text{sec}} = 10,000$
- Total number of collected data –  $N_{\text{D}} = 1,000,000$
- Total number of monitor visits –  $N_{\text{M}} = 240$ .

## 3. Results

### 3.1. Calculations

#### 3.1.1. Data gathering at the research center

The difference in data capturing function (PDC: CRF data entry; EDC: eCRF data entry) and archiving function (PDC: CRF

archiving) brings the differences in efforts and costs. If we generalize (for the purpose of estimation) that these processes are equal among all the participating research centers (RCs) for the sample clinical trial presented in Methods, a difference in costs of data gathering can be presented as:

$$\Delta\Phi_{\text{DG}} = \Phi_{\text{DG}^{\text{P}}} - \Phi_{\text{DG}^{\text{E}}}$$

$$\Phi_{\text{DG}^{\text{P}}} = F_{\text{DG}_{\text{CRF}}^{\text{P}}} \times \varphi_{\text{IA}} + F_{\text{DG}_{\text{arh}}^{\text{P}}} \times \varphi_{\text{I}}$$

$$\Phi_{\text{DG}^{\text{E}}} = F_{\text{DG}_{\text{CRF}}^{\text{E}}} \times \varphi_{\text{IA}}$$

Total efforts for writing down data on paper CRF is presented as  $F_{\text{DG}_{\text{CRF}}^{\text{P}}} = N_{\text{D}} \times \mu_{\text{DG}_{\text{CRF}}^{\text{P}}}$ , where  $\mu_{\text{DG}_{\text{CRF}}^{\text{P}}}$  is an average effort needed for writing down one data item. Total efforts needed for archiving all the paper CRF documents is presented as  $F_{\text{DG}_{\text{arh}}^{\text{P}}} = N_{\text{CRF}} \times \mu_{\text{DG}_{\text{arh}}^{\text{P}}}$ , where  $\mu_{\text{DG}_{\text{arh}}^{\text{P}}}$  is an average effort needed for archiving one paper CRF.

On the other hand, total efforts for entering data into eCRF is presented as  $F_{\text{DG}_{\text{CRF}}^{\text{E}}} = N_{\text{D}} \times \mu_{\text{DG}_{\text{CRF}}^{\text{E}}}$ , where  $\mu_{\text{DG}_{\text{CRF}}^{\text{E}}}$  is an average effort needed for entering one data.

We can calculate cost difference in PDC and EDC Data Gathering for our sample clinical trial as:

$$\Phi_{\text{DG}^{\text{P}}} \approx 33,500 \text{ €}$$

$$\Phi_{\text{DG}^{\text{E}}} \approx 16,500 \text{ €}$$

$$\Delta\Phi_{\text{DG}} \approx 17,000 \text{ €}.$$

#### 3.1.2. Monitoring

The costs for PDC and EDC Monitoring processes can be estimated as:

$$\begin{aligned} \Phi_{\text{M}^{\text{P}}} &= F_{\text{M}_{\text{rev}}^{\text{P}}} \times \varphi_{\text{CRA}} + F_{\text{M}_{\text{cor}}^{\text{P}}} \times (\varphi_{\text{I}} + \varphi_{\text{IA}} + \varphi_{\text{CRA}}) + F_{\text{M}_{\text{sig}}^{\text{P}}} \\ &\quad \times \varphi_{\text{I}} + F_{\text{M}_{\text{loc}}^{\text{P}}} \times \varphi_{\text{CRA}} + F_{\text{M}_{\text{snd}}^{\text{P}}} \times \varphi_{\text{CRA}} \end{aligned}$$

$$\begin{aligned} \Phi_{\text{M}^{\text{E}}} &= F_{\text{M}_{\text{rev}}^{\text{E}}} \times \varphi_{\text{CRA}} + F_{\text{M}_{\text{loc}}^{\text{E}}} \times \varphi_{\text{CRA}} + F_{\text{M}_{\text{cor}}^{\text{E}}} \times (\varphi_{\text{I}} + \varphi_{\text{IA}}) \\ &\quad + F_{\text{M}_{\text{ver}}^{\text{E}}} \times \varphi_{\text{CRA}}. \end{aligned}$$

In the equations above the  $F_{\text{M}^{\text{P}}}$  functions correspond to the PDC functions presented on Fig. 5 and  $F_{\text{M}^{\text{E}}}$  functions correspond to the EDC functions presented on Fig. 6.

Furthermore, the number of notes that have to be taken depends on the number of errors discovered by monitor and this is considerably higher for PDC than EDC (as previously discussed in subsection: Methods: Model parameters: Data Quality).

$$F_{\text{M}_{\text{rev}}^{\text{P}}} = N_{\text{D}} \times \mu_{\text{M}_{\text{chk}}} + N_{\text{D}} \times \varepsilon_{\text{mon}}^{\text{P}} \times \mu_{\text{M}_{\text{fer}}^{\text{P}}}$$

$$F_{\text{M}_{\text{rev}}^{\text{E}}} = N_{\text{D}} \times \mu_{\text{M}_{\text{chk}}} + N_{\text{D}} \times \varepsilon_{\text{mon}}^{\text{E}} \times \mu_{\text{M}_{\text{fer}}^{\text{E}}}$$

We however consider that monitor visit lock effort is the same for PDC and EDC and can be presented as:

$$F_{\text{M}_{\text{loc}}^{\text{P}}} = F_{\text{M}_{\text{loc}}^{\text{E}}} = N_{\text{M}} \times \mu_{\text{M}_{\text{loc}}}.$$

The way data correction is performed and documented is also different for PDC and EDC. EDC data correction can be done through the EDC system which supports raising queries and

sending messages with queries to investigators or their assistants. Furthermore, due to the validation routines on data entry when using EDC system, a number of errors among data is considerably lower for EDC process than PDC process and is reported in [24] to be  $\varepsilon_{\text{mon}}^E = 20\%$   $\varepsilon_{\text{mon}}^E = 2\%$ . This means that there will be less effort needed for EDC data correction than PDC.

$$F_{M_{\text{cor}}^P} = N_D \times \varepsilon_{\text{mon}}^P \times \mu_{M_{\text{cor}}^P}, F_{M_{\text{cor}}^E} = N_D \times \varepsilon_{\text{mon}}^E \times \mu_{M_{\text{cor}}^E}$$

$$F_{M_{\text{sig}}^P} = N_{\text{sec}} \times \mu_{M_{\text{sig}}^P}$$

$$F_{M_{\text{snd}}^P} = N_{\text{sec}} \times \mu_{M_{\text{snd}}^P}$$

$$F_{M_{\text{ver}}^E} = N_{\text{sec}} \times \mu_{M_{\text{ver}}^E}$$

By putting our estimations into the equations above we get the difference in costs between PDC and EDC Monitoring processes to be:

$$\Phi_{M^P} \approx 233,500 \text{ €}$$

$$\Phi_{M^E} \approx 140,000 \text{ €}$$

$$\Delta\Phi_M \approx 93,500 \text{ €}.$$

### 3.1.3. Data management

Considering that the Data Management process for PDC is like on Fig. 7 and for EDC is like on Fig. 8, efforts for these two processes might be presented with the following equations:

$$\text{PDC Data Management: } F_{DM^P} = F_{DM_{\text{ENT}}^P} + F_{DM_{\text{ANL}}^P} + F_{DM_{\text{QRY}}^P} + F_{DM_{\text{CLO}}^P}$$

$$\text{EDC Data Management: } F_{DM^E} = F_{DM_{\text{ANL}}^E} + F_{DM_{\text{QRY}}^E} + F_{DM_{\text{CLO}}^E}.$$

**3.1.3.1. Data entry.** EDC data entry into the database is done by investigators (and their assistants). On the other hand, several data entry functions performed by data management staff are required in the case of PDC (Fig. 7). Thus, for PDC, Data Management costs considerably increase.

$$F_{DM_{\text{ENT}}^P} = F_{DM_{\text{eff}}^P} + F_{DM_{\text{dde}}^P} + F_{DM_{\text{ddm}}^P} + F_{DM_{\text{ddd}}^P} + F_{DM_{\text{eda}}^P} + F_{DM_{\text{ede}}^P} + F_{DM_{\text{pde}}^P}$$

The PDC data entry process begins with the evidence of received forms. RCs/monitors are sending entire CRF sections after being completed. Therefore all together  $N_{\text{sec}} = 10,000$  documents have to be evidenced.

$$F_{DM_{\text{eff}}^P} = N_{\text{sec}} \times \mu_{DM_{\text{eff}}^P}$$

When double data entry approach is used to minimize data entry errors, data entry staff has to enter data twice.

$$F_{DM_{\text{dde}}^P} = N_D \times \mu_{DM_{\text{dde}}^P} \times 2$$

Considering that average data entry error rate ( $\varepsilon_{\text{ent}}^P$ ) is around 1% [19–23] for each staff member, this gives us nearly

twice more mismatches in entered data produced by data entry staff.

$$F_{DM_{\text{ddm}}^P} = N_D \times \varepsilon_{\text{ent}}^P \times 2 \times \mu_{DM_{\text{ddm}}^P}$$

Furthermore, data entry staff faces problems in reading some data written in CRFs and these have to be evidenced ( $F_{DM_{\text{ddd}}^P}$ ). On the other hand some other data discrepancies are discovered by computer analysis of entered data and also evidenced ( $F_{DM_{\text{pde}}^P}$ ). Additional step ( $F_{DM_{\text{edc}}^P}$ ) to resolve these discrepancies needs to be introduced in the process in order to lower the error level. All these discrepancies we call plausibility discrepancies and we consider that they represent 80% of all the queries raised by sponsor ( $\varepsilon_{\text{pla}}^P = 80\% \varepsilon_{\text{qry}}^P$ ; see **Methods: Data quality** section).

$$F_{DM_{\text{ddd}}^P} + F_{DM_{\text{pde}}^P} = N_D \times \varepsilon_{\text{pla}}^P \times \mu_{DM_{\text{pde}}^P}$$

Among plausibility discrepancies, thanks to double data entry, only few of these are eventually caused by data entry staff and in our calculation the effort for correcting these errors might be omitted ( $F_{DM_{\text{edc}}^P} \approx 0$ ). Also the effort for computer data analysis is relatively low comparing to all the other efforts and related costs as well and can be omitted from our calculation ( $F_{DM_{\text{eda}}^P} \approx 0$ ).

Considering average efforts estimations the difference in the costs of data entry process between PDC and EDC can thus be calculated as:

$$\Phi_{DM_{\text{ENT}}^P} \approx F_{DM_{\text{eff}}^P} \times \varphi_C + F_{DM_{\text{dde}}^P} \times \varphi_{DE} + F_{DM_{\text{ddm}}^P} \times \varphi_{DM} + (F_{DM_{\text{ddd}}^P} + F_{DM_{\text{pde}}^P}) \times \varphi_{DM}$$

$$\Delta\Phi_{DM_{\text{ENT}}} = \Phi_{DM_{\text{ENT}}^P} \approx 26,000 \text{ €}.$$

**3.1.3.2. Computer data analysis.** The PDC data analysis efforts can be presented as follows:

$$F_{DM_{\text{ANL}}^P} = F_{DM_{\text{cda}}^P} + F_{DM_{\text{ade}}^P} + F_{DM_{\text{qpr}}^P} + F_{DM_{\text{qsn}}^P}.$$

In the case of EDC data analysis process is simpler as it does not require special efforts for query evidence and sending (this is done automatically through the EDC application):

$$F_{DM_{\text{ANL}}^E} = F_{DM_{\text{cda}}^E} + F_{DM_{\text{qpr}}^E}.$$

For example, after computer data analysis, the coordinator simply raises queries in the EDC application and unlocks the doubtful forms.

We can consider that the same efforts are put in computer data analysis for both PDC and EDC and for the simplification purposes we might estimate that these efforts are considerably low and might be omitted from our calculations.

$$F_{DM_{\text{cda}}^P} = F_{DM_{\text{cda}}^E} \approx 0$$

On the other hand, different query rates for PDC ( $\epsilon_{\text{qry}}^{\text{P}} = 2\%$ ) and EDC ( $\epsilon_{\text{qry}}^{\text{P}} = 0.2\%$ ) lead to different efforts and costs for generating query documents.

$$F_{\text{DM}_{\text{ade}}^{\text{P}}} = N_{\text{D}} \times (\epsilon_{\text{qry}}^{\text{P}} - \epsilon_{\text{pla}}^{\text{P}}) \times \mu_{\text{DM}_{\text{ade}}^{\text{P}}}$$

$$F_{\text{DM}_{\text{qpr}}^{\text{P}}} = N_{\text{D}} \times \epsilon_{\text{qry}}^{\text{P}} \times \mu_{\text{DM}_{\text{qpr}}^{\text{P}}}$$

$$F_{\text{DM}_{\text{qpr}}^{\text{E}}} = N_{\text{D}} \times \epsilon_{\text{qry}}^{\text{E}} \times \mu_{\text{DM}_{\text{qpr}}^{\text{E}}}$$

Finally, as our sample clinical trial has all together  $10 \times 24 = 240$  monitoring visits and each of them triggers data management process which results in query document, we considered that in principle 240 query documents ( $N_{\text{DQ}} = N_{\text{M}}$ ) would be generated within the entire sample clinical trial.

$$F_{\text{DM}_{\text{qsn}}^{\text{P}}} = N_{\text{DQ}} \times \mu_{\text{DM}_{\text{qsn}}^{\text{P}}}$$

Considering average efforts estimations the difference in costs for PDC and EDC data analysis can be calculated as:

$$\Phi_{\text{DM}_{\text{ANL}}^{\text{P}}} = F_{\text{DM}_{\text{ade}}^{\text{P}}} \times \varphi_{\text{DM}} + F_{\text{DM}_{\text{qpr}}^{\text{P}}} \times \varphi_{\text{C}} + F_{\text{DM}_{\text{qsn}}^{\text{P}}} \times \varphi_{\text{C}} \approx 18,000 \text{ €}$$

$$\Phi_{\text{DM}_{\text{ANL}}^{\text{E}}} = F_{\text{DM}_{\text{qpr}}^{\text{E}}} \times \varphi_{\text{C}} \approx 1000 \text{ €}$$

$$\Delta\Phi_{\text{DM}_{\text{ANL}}} \approx 17,000 \text{ €}.$$

**3.1.3.3. Query resolving.** Different organizations have different approaches to query resolving. This can be done in direct communication between sponsor (coordinator) and investigators, or may include monitor in this process. We decided to take the approach where coordinator communicates only with monitor. Monitor then forwards queries (those which she/he cannot resolve by her/himself) to corresponding investigators. Corresponding investigators resolve these queries, and send the responses back to monitor who collects all the queries and respective answers and sends them to the coordinator. In this way monitor is aware of all eventual data changes.

From the models on Figs. 7 and 8 we obtain the following equations:

$$F_{\text{DM}_{\text{QRY}}^{\text{P}}} = F_{\text{DM}_{\text{qre}}^{\text{P}}} + F_{\text{DM}_{\text{qfw}}^{\text{P}}} + F_{\text{DM}_{\text{qin}}^{\text{P}}} + F_{\text{DM}_{\text{rfw}}^{\text{P}}} + F_{\text{DM}_{\text{qpe}}^{\text{P}}} + F_{\text{DM}_{\text{rsn}}^{\text{P}}}$$

$$F_{\text{DM}_{\text{QRY}}^{\text{E}}} = F_{\text{DM}_{\text{qre}}^{\text{E}}} + F_{\text{DM}_{\text{qin}}^{\text{E}}} + F_{\text{DM}_{\text{qpe}}^{\text{E}}}.$$

Further estimation of costs is based on the following assumptions. The first is that the average efforts needed for review and evidence of query are equal to those needed for review and evidence of responses to queries. This allows us to simplify our calculations considering that  $F_{\text{DM}_{\text{qpe}}^{\text{P}}} = F_{\text{DM}_{\text{qre}}^{\text{P}}}$  and  $F_{\text{DM}_{\text{qpe}}^{\text{E}}} = F_{\text{DM}_{\text{qre}}^{\text{E}}}$ , where:

$$F_{\text{DM}_{\text{qre}}^{\text{P}}} = N_{\text{D}} \times \epsilon_{\text{qry}}^{\text{P}} \times \mu_{\text{DM}_{\text{qre}}^{\text{P}}}$$

$$F_{\text{DM}_{\text{qre}}^{\text{E}}} = N_{\text{D}} \times \epsilon_{\text{qry}}^{\text{E}} \times \mu_{\text{DM}_{\text{qre}}^{\text{E}}}.$$

It is hard to predict to which research centers the queries raised by sponsor will belong. For the purpose of calculating  $F_{\text{DM}_{\text{qfw}}^{\text{P}}}$  efforts we can consider that there are some queries

raised after each monitor visit resulting in one query document for each center. Therefore the number of query documents that monitor has to forward is equal to  $N_{\text{DQ}} = 240$ .

$$F_{\text{DM}_{\text{qfw}}^{\text{P}}} = N_{\text{DQ}} \times \mu_{\text{DM}_{\text{qfw}}^{\text{P}}}$$

Third assumption is that the same effort is needed on the RC side to send the responses back to monitor ( $F_{\text{DM}_{\text{rfw}}^{\text{P}}} = F_{\text{DM}_{\text{qfw}}^{\text{P}}}$ ).

$$F_{\text{DM}_{\text{qin}}^{\text{P}}} = N_{\text{D}} \times \epsilon_{\text{qry}}^{\text{P}} \times \mu_{\text{DM}_{\text{qin}}^{\text{P}}}$$

$$F_{\text{DM}_{\text{qin}}^{\text{E}}} = N_{\text{D}} \times \epsilon_{\text{qry}}^{\text{E}} \times \mu_{\text{DM}_{\text{qin}}^{\text{E}}}$$

Fourth and finally, we consider that effort needed by monitor to send responses back to coordinator is equal to efforts needed by coordinator to send the queries to monitor ( $F_{\text{DM}_{\text{rsn}}^{\text{P}}} = F_{\text{DM}_{\text{qsn}}^{\text{P}}}$ ).

$$F_{\text{DM}_{\text{rsn}}^{\text{P}}} = N_{\text{DQ}} \times \mu_{\text{DM}_{\text{qsn}}^{\text{P}}}$$

Therefore, the costs for PDC and EDC query resolving process are calculated as:

$$\Phi_{\text{DM}_{\text{QRY}}^{\text{P}}} = F_{\text{DM}_{\text{qre}}^{\text{P}}} \times \varphi_{\text{CRA}} \times 2 + F_{\text{DM}_{\text{qfw}}^{\text{P}}} \times (\varphi_{\text{CRA}} + \varphi_{\text{I}}) \times 2$$

$$+ F_{\text{DM}_{\text{qin}}^{\text{P}}} \times (\varphi_{\text{I}} + \varphi_{\text{IA}}) + F_{\text{DM}_{\text{rsn}}^{\text{P}}} \times \varphi_{\text{CRA}}$$

$$\Phi_{\text{DM}_{\text{QRY}}^{\text{E}}} = F_{\text{DM}_{\text{qre}}^{\text{E}}} \times \varphi_{\text{CRA}} \times 2 + F_{\text{DM}_{\text{qin}}^{\text{E}}} \times (\varphi_{\text{I}} + \varphi_{\text{IA}}).$$

Considering average efforts estimations the difference in costs of PDC and EDC query resolving is:

$$\Phi_{\text{DM}_{\text{QRY}}^{\text{P}}} \approx 21,000 \text{ €}; \Phi_{\text{DM}_{\text{QRY}}^{\text{E}}} \approx 1000 \text{ €}; \Delta\Phi_{\text{DM}_{\text{QRY}}} \approx 20,000 \text{ €}.$$

**3.1.3.4. Query closing.** The last step in data management leads to locking the data in the database. Data are considered to be clean and reliable, and therefore should be locked to further changes.

When PDC queries are resolved and responses sent back to the data management team, responses have to be evidenced and, if necessary, data in the database have to be corrected. Finally, all the processed forms (CRF sections) have to be evidenced as well. For these the efforts are:

$$F_{\text{DM}_{\text{CLO}}^{\text{P}}} = F_{\text{DM}_{\text{rev}}^{\text{P}}} + F_{\text{DM}_{\text{dcr}}^{\text{P}}} + F_{\text{DM}_{\text{epf}}^{\text{P}}}$$

$$F_{\text{DM}_{\text{rev}}^{\text{P}}} = N_{\text{D}} \times \epsilon_{\text{qry}}^{\text{P}} \times \mu_{\text{DM}_{\text{rev}}^{\text{P}}}$$

$$F_{\text{DM}_{\text{dcr}}^{\text{P}}} = N_{\text{D}} \times \epsilon_{\text{cor}}^{\text{P}} \times \mu_{\text{DM}_{\text{dcr}}^{\text{P}}}$$

$$F_{\text{DM}_{\text{epf}}^{\text{P}}} = N_{\text{sec}} \times \mu_{\text{DM}_{\text{epf}}^{\text{P}}}.$$

On the other hand, in the case of EDC, if necessary, investigators have already entered any data changes and clarifications to queries directly in the EDC application.



Therefore the coordinator has just to look through the responses and lock the database.

$$F_{DM_{CLO}^E} = F_{DM_{loc}^E} = N_D \times \varepsilon_{qry}^E \times \mu_{DM_{loc}^E}$$

Considering average efforts estimations we can calculate the difference in costs for PDC and EDC query closing as:

$$\Phi_{DM_{CLO}^P} = F_{DM_{rev}^P} \times \varphi_C + F_{DM_{dgr}^P} \times \varphi_{DM} + F_{DM_{epf}^P} \times \varphi_C \approx 19,000 \text{ €}$$

$$\Phi_{DM_{CLO}^E} = F_{DM_{loc}^E} \times \varphi_C \approx 0 \text{ €}$$

$$\Delta\Phi_{DM_{CLO}} \approx 19,000 \text{ €}.$$

Finally, considering the calculations above the entire difference in costs between PDC and EDC Data Management processes for our sample clinical trial is:

$$\Delta\Phi_{DM} = \Delta\Phi_{DM_{ENT}} + \Delta\Phi_{DM_{ANL}} + \Delta\Phi_{DM_{QRY}} + \Delta\Phi_{DM_{CLO}}$$

$$\Delta\Phi_{DM} \approx 82,000 \text{ €}.$$

In conclusion, the total PDC costs ( $\Phi^P$ ), EDC costs ( $\Phi^E$ ) and costs differences ( $\Delta\Phi$ ) between PDC and EDC process, affected by implementation of EDC application and related changes of data gathering, monitoring and data management processes, are estimated as:

$$\Phi^P = \Phi_{DG^P} + \Phi_{M^P} + \Phi_{DM^P} \approx 350,000 \text{ €}$$

$$\Phi^E = \Phi_{DG^E} + \Phi_{M^E} + \Phi_{DM^E} \approx 158,500 \text{ €}$$

$$\Delta\Phi = \Phi^P - \Phi^E \approx 192,500 \text{ €}.$$

From the results presented above we can see that just due to the savings in overall staff efforts (without consideration of savings due to shortening of data collection process duration) the EDC process costs are 55% lower than PDC costs.

### 3.2. Sensitivity analysis

We modeled paper data collection (PDC) and electronic data collection (EDC) processes in order to evaluate the difference in costs of these two approaches in clinical trial. Furthermore, we estimated the parameters of our models that affect the costs and calculated PDC and EDC costs as well as costs differences for a sample clinical trial. The exact values depend both on the models (processes) and the values of the parameters used to calculate the costs. Here we mainly considered different estimations of error/query rates and average efforts for each of the process elements, as well as

staff prices. However, in this section we will discuss our estimations of parameter and try to understand how variations of these may affect our results.

Our estimations of parameters are based on available literature and the information obtained in interviews. Despite the fact that these estimations are not highly reliable (more reliable would be experimental measurements of these parameters) we believe that our results reflect the PDC and EDC costs differences well.

Some of our estimations of error and query rates are based on information provided by interviewers. These include PDC error rate discovered by monitor which is around 10% and PDC query rate which is 2%. The last is also supported by some other studies [24]. The EDC error and query rates were estimated by taking in consideration PDC rates and published studies [19–25] on comparison between PDC and EDC error and query rates. These definitely prove that electronic data collection reduces number of errors and queries. The reduction is highly dependent on the quality of EDC application which must be designed carefully and user friendly. If the EDC application is under-designed and EDC error and query rates are thus higher, this will lower the difference in costs. However, if we double the EDC error and query rates this will raise EDC costs from 158,500 € to 173,000 €, which is 8.5% increase. As a result the overall costs savings of switching to EDC process decrease from 55% to 45%. This however is still a considerable saving.

The size of clinical trial also affects the savings considerably. Namely if we consider twice smaller sample clinical trial having 5 research centers with 50 patients each and 500 data per CRF the entire costs of PDC decrease from 350,000 € to only 44,000 €. Despite the fact that the savings ratio is still 55% the 24,000 € of EDC savings would probably not justify the investment into EDC application.

We believe that it is of key importance to measure the average efforts for all the process tasks and then recalculate the equations and get more reliable costs. For example, if we consider that the average efforts for EDC tasks are equal to the PDC tasks (e.g. that entering data in to eCRF is as fast as writing down data into paper CRF) the results slightly change. In that case the EDC costs rise for 13% from 158,500 € to 179,000 €, which results in EDC savings lowering from 55% to 50%. However, this might be still considered as a considerable reduction in costs. We intend to measure the parameters and use measured parameters to make more reliable calculations. We also encourage others to try to use their estimations of average efforts and error rates to calculate the costs differences according to our process models (PDC and EDC).

We set PDC and EDC models that might be challenged. For example, it is not so unusual that the sponsor decides to decrease monitoring costs by verifying just a sample of

**Table 1**

Data gathering and monitoring costs for different estimations of parameters ( $\Phi_{DG^P}$  – PDC Data Gathering costs;  $\Phi_{DG^E}$  – EDC Data Gathering costs;  $\Delta\Phi_{DG}$  – difference between PDC and EDC Data Gathering costs;  $\Phi_{M^P}$  – PDC Monitoring costs;  $\Phi_{M^E}$  – EDC Monitoring costs;  $\Delta\Phi_M$  – difference between PDC and EDC Monitoring costs).

Scenario	$\Phi_{DG^P}$	$\Phi_{DG^E}$	$\Delta\Phi_{DG}$	$\Phi_{M^P}$	$\Phi_{M^E}$	$\Delta\Phi_M$
Standard values as presented in Methods and Results	33.660 €	16.500 €	17.160 €	233.559 €	140.093 €	93.466 €
Doubled EDC error/query rates	33.660 €	16.500 €	17.160 €	233.559 €	152.612 €	80.947 €
Smaller clinical trial (12 months; 5 RC×50 CRF×500 data)	4.290 €	2.063 €	2.228 €	29.210 €	17.526 €	11.683 €
Equal EDC and PDC average efforts	33.660 €	33.000 €	660 €	233.559 €	142.592 €	90.967 €
Only 50% data are checked by monitor	33.660 €	16.500 €	17.160 €	120.984 €	71.358 €	49.626 €

**Table 2**

Data management costs for different estimations of parameters ( $\phi_{DM_{ENT}^P}$  – PDC data entry costs;  $\phi_{DM_{ANL}^P}$  – PDC data analysis costs;  $\phi_{DM_{ANL}^E}$  – EDC data analysis costs;  $\Delta\phi_{DM_{ANL}}$  – difference between PDC and EDC data analysis costs;  $\phi_{DM_{QRY}^P}$  – PDC query resolving costs;  $\phi_{DM_{QRY}^E}$  – EDC query resolving costs;  $\Delta\phi_{DM_{QRY}}$  – difference between PDC and EDC query resolving costs;  $\phi_{DM_{CLO}^P}$  – PDC query closing costs;  $\phi_{DM_{CLO}^E}$  – EDC query closing costs;  $\Delta\phi_{DM_{CLO}}$  – difference between PDC and EDC query closing costs;  $\phi_{DM}^P$  – PDC Data Management costs;  $\phi_{DM}^E$  – EDC Data Management costs;  $\Delta\phi_{DM}$  – difference between PDC and EDC Data Management costs).

Scenario	$\phi_{DM_{ENT}^P}$	$\phi_{DM_{ANL}^P}$	$\phi_{DM_{ANL}^E}$	$\Delta\phi_{DM_{ANL}}$	$\phi_{DM_{QRY}^P}$	$\phi_{DM_{QRY}^E}$	$\Delta\phi_{DM_{QRY}}$	$\phi_{DM_{CLO}^P}$	$\phi_{DM_{CLO}^E}$	$\Delta\phi_{DM_{CLO}}$	$\phi_{DM}^P$	$\phi_{DM}^E$	$\Delta\phi_{DM}$
Standard values as presented in Methods and Results	26.099 €	18.034 €	835 €	17.199 €	20.713 €	1.000 €	19.713 €	19.125 €	209 €	18.917 €	83.971 €	2.044 €	81.928 €
Doubled EDC error/query rates	26.099 €	18.034 €	1.670 €	16.364 €	20.713 €	2.000 €	18.713 €	19.125 €	417 €	18.708 €	83.971 €	4.087 €	79.884 €
Smaller clinical trial (12 months; 5 RC × 50 CRF × 500 data)	3.262 €	2.279 €	104 €	2.175 €	2.673 €	125 €	2.548 €	2.391 €	26 €	2.365 €	10.605 €	255 €	10.350 €
Equal EDC and PDC average efforts	26.099 €	18.034 €	1.670 €	16.364 €	20.713 €	1.500 €	19.213 €	19.125 €	417 €	18.708 €	83.971 €	3.587 €	80.384 €
Only 50% data are checked by monitor	26.099 €	18.034 €	835 €	17.199 €	20.713 €	1.000 €	19.713 €	19.125 €	209 €	18.917 €	83.971 €	2.044 €	81.928 €

gathered data instead of all the gathered data. For example, this could be done by checking only critical data (e.g. applied therapy, responses, and adverse events). Therefore, if we recalculate the costs with the assumption that monitor checks only half of all the gathered data the costs of PDC will decrease for 32% (from 350,000 € to 238,500 €) as well as the costs of EDC for 43% (from 158,500 € to 90,000 €). The savings amount will decrease from 192,500 € to 148,500 €, but this will actually raise the saving rate from 55% to 62%.

The costs and savings from all the different scenarios and estimations mentioned above are presented in the Tables 1–3. Namely, we present the costs for standard values as presented in Methods and Results and the following variations:

- doubled EDC error/query rates;
- smaller clinical trial (12 months, 5 research centers with 50 patients in each center contributing 500 data per patient);
- equal EDC and PDC average efforts; and
- only 50% data checked by monitor.

**4. Discussion**

From the results presented in Tables 1–3 we can conclude that the EDC process brings savings when compared to PDC process for all five scenarios. The savings primarily are due to lower error and query rates which reduce work that have to be done on data cleaning. Another important parameter that influences CT costs is the size of CT (number of centers, patients and collected data). It is obvious that for smaller CT savings might be insufficient to justify the investment in EDC application. However, by using our models and own parameters' assessments the sponsor can calculate the savings of EDC comparing to PDC process and consider this

information when making a decision on investment in an EDC application.

The models which we presented in this paper are limited to the data collection related processes. However, switch to EDC brings many changes in clinical trial organization and implementation. EDC remarkably reduces spending on paper. Furthermore, CT management may be driven by up-to-date information from EDC application. For example, monitoring visits might be appointed when sufficient data is submitted by research center, instead of having a priori scheduled visits. Serious Adverse Events (SAE) can automatically trigger SAE reporting mechanisms and enhance patients' safety. Finally, greatest savings EDC brings through shortening of last patient out to database lock time which eventually decreases drug's time to market.

There are also disadvantages which EDC brings, comparing to PDC. For example there is a need for extended costs due to hotline and maintenance of the EDC system. Also there are hardware and network constraints (e.g. firewall, low or no network availability, or slow network at some research centers, etc.) which might appear and which may also require additional costs for solving these issues. Another disadvantage of EDC is inconvenience of electronic data entry during outpatient visit, as well as during some demanding clinical activities. Further problems may appear due to the regulations related to eSource [26]. The last, but not the least, a switch from paper-based process to electronic process may be a demanding organizational challenge as well. An extended discussion on advantages and disadvantages of EDC over PDC was recently presented by Welker [3].

In our study, we considered two “extreme” approaches – fully paper-based data collection and fully electronic data collection. In many cases, a number of scenarios may be necessary for the same trial. In the same clinical trial, some

**Table 3**

Total costs and savings for different estimations of parameters ( $\phi^P$  – PDC costs;  $\phi^E$  – EDC costs;  $\Delta\phi$  – difference between PDC and EDC costs).

Scenario	$\phi^P$	$\phi^E$	$\Delta\phi$	EDC savings
Standard values as presented in Methods and Results	351.190 €	158.637 €	192.554 €	55%
Doubled EDC error/query rates	351.190 €	173.199 €	177.991 €	51%
Smaller clinical trial (12 months; 5 RC × 50 CRF × 500 data)	44.105 €	19.844 €	24.261 €	55%
Equal EDC and PDC average efforts	351.190 €	179.179 €	172.011 €	49%
Only 50% data are checked by monitor	238.615 €	89.902 €	148.713 €	62%

centers might perform PDC and other EDC. Data center should handle data coming from both sources, making data collection models more complex. It may appear that EDC savings cannot justify all the complexity that handling hybrid PDC/EDC clinical trial brings.

All the disadvantages mentioned above have to be balanced with the advantages of EDC and savings that EDC process may bring. Namely only in this way the switch from PDC to EDC can be justified. Nevertheless, our study shows that modeling of processes together with accurate estimations of process parameters clearly identifies where in the data collection process the costs savings come from. Having in mind that data management (site setup, monitoring and closing including CRF processing) costs may be estimated as almost 30% of Phase III clinical trial costs (according to published Fast Track Systems example \$16,000,000 of \$58,400,000 [28]), optimization of processes and introduction of IT solutions in order to reduce the costs of clinical trial may be a reasonable decision. Considering that clinical trial costs represent 47.9% (Phase III – 28.1%, Phase II – 13.1%, Phase I – 6.7%) of the entire drug R&D investments (according to PhRMA Membership Annual Report 2008 [27]) we can conclude that lowering clinical trial costs consequently reduce drug-to-market costs.

## 5. Conclusion

In our report we demonstrated an example of how a change from paper-based data collection (PDC) to internet based electronic data collection (EDC) affects the costs of data collection and its sub-processes. We developed eEPC (Extended Event-driven Process Chains) models for PDC and EDC sub-processes (data gathering, monitoring and data management) and simplified them to the extent that allowed us to calculate the related costs. We estimated the values of all the parameters which appear in the models, such as clinical trial size, error and query rates, average efforts and staff prices. We based these estimations on available literature and information obtained in interviews. The results show that most benefit comes from reducing monitoring and data management costs. The exact value depends on estimation of parameters which affect the calculations. For example, the variations in clinical trial size significantly affect the savings that EDC brings, while the costs are not that sensible to changes in average efforts for particular tasks. These results are not surprising, but with our approach we offer the way to quantify them. With more reliable estimation of average costs and error and query rates and considering specific values we can get more reliable results and use them to decide on switching from paper data collection to electronic data collection. We however have to emphasize that our models do not include all the aspects of organization and implementation of clinical trial which are affected by the change in data collection approach. The same approach that we presented in our study can however be used in assessing the costs of the clinical trial through modeling of the entire clinical trial as a business process.

## Acknowledgments

This research was in part supported by the Slovenian Research Agency and Slovenian Ministry of the Economy.

## References

- [1] European Parliament and the Council of the European Union. Directive 2001/20/EC. Off J Eur Communities May 1 2001;121:34–44.
- [2] Paul J, Seib R, Prescott T. The internet and clinical trials: background, online resources, examples and issues. *J Med Internet Res* Mar 16 2005;7(1):e5.
- [3] Welker JA. Implementation of electronic data capture systems: barriers and solutions. *Contemp Clin Trials* 2007;28:329–36.
- [4] Brandt CA, Argraves S, Money R, Ananth G, Trocky NM, Nadkarni PM. Informatics tools to improve clinical research study implementation. *Contemp Clin Trials* 2006;27:112–22.
- [5] Alschuler L, Bain L, Kush RD. Improving data collection for patient care and clinical trials. *Sci Career Mag* Mar 26 2004 [http://sciencecareers.sciencemag.org/career\\_magazine/previous\\_issues/articles/2004\\_03\\_26/noDOI.5622907321165187916](http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2004_03_26/noDOI.5622907321165187916). Archived at: <http://www.webcitation.org/5cGF6Kmf8>.
- [6] Pavlovic I, Miklavcic D. Web-based electronic data collection system to support electrochemotherapy clinical trial. *IEEE Trans Inf Technol Biomed Mar* 2007;11(2):222–30.
- [7] Edwards RL, Edwards SL, Bryner J, Cunningham K, Rogers A, Slattery ML. A computer-assisted data collection system for use in a multicenter study of American Indians and Alaska Natives: SCAPES. *Comput Methods Programs Biomed* Apr 2008;90(1):38–55.
- [8] Proctor SJ, Wilkinson J. A web-based study concept designed to progress clinical research for 'orphan' disease areas in haematological oncology in the elderly: the SHIELD programme. *Crit Rev Oncol/Hematol* 2007;61:79–83.
- [9] Formica M, Kabbara K, Clark R, McAlindon T. Can clinical trials requiring frequent participant contact be conducted over the internet? Results from an online randomized controlled trial evaluating a topical ointment for herpes labialis. *J Med Internet Res* 2004;6(1):e6.
- [10] Avidan A, Weissman C, Sprung CL. An internet web site as a data collection platform for multicenter research. *Anesth Analg* 2005;100:506–11.
- [11] Lopez-Carrero C, Arriaza E, Bolanos E, Ciudad A, Municio M, Ramos J, et al. Internet in clinical research based on a pilot experience. *Contemp Clin Trials* 2005;26:234–43.
- [12] Lallas CD, Preminger GM, Pearle MS, Leveillee RJ, Lingeman JE, Schwoppe JP. Internet based multi-institutional clinical research: a convenient and secure option. *J Urol* May 2004;171(5):1880–5.
- [13] Rangel SJ, Narasimhan B, Geraghty N, Moss RL. Development of an internet-based protocol to facilitate randomized clinical trials in pediatric surgery. *J Pediatr Surg* 2002;37(7):990–4.
- [14] Marks R, Bristol H, Conlon M, Pepine CJ. Enhancing clinical trials on the internet: lessons from INVEST. *Clin Cardiol* 2001;24(supplV):V17–23.
- [15] Collada AL, Fazi P, Luzi D, Ricci FL, Serbanati LD, Vignetti M. Toward a model of clinical trials. Proceedings of the 5th international symposium ISBMDA; Nov 18–19 2004. p. 299–312. Barcelona, Spain, <http://www.springerlink.com/content/y3yxru6pxhfpw015/>.
- [16] Luzi D, Ricci FL, Serbanati LD. E-clinical trials supported by a service-oriented architecture. Proceedings of Mednet 2006: 11th world congress on the internet in medicine – MEDNET; Oct 13–26 2006. Society for the Internet in Medicine, Toronto, Canada, [http://www.mednetcongress.org/fullpapers/MEDNET-137\\_DanielaLuziA4\\_e.pdf](http://www.mednetcongress.org/fullpapers/MEDNET-137_DanielaLuziA4_e.pdf). Archived at: <http://www.webcitation.org/5clrbshXk>.
- [17] Clinical Trial Electronic Data Capture Task Group. PhRMA Biostatistics and Data Management Technical Group. US PhRMA's EDC position paper, revision 1. PhRMA eClinical Forum; May 2005. <http://www.eclinicalforum.com/content/Knowledge/Articles/EDC%20Revision%201%20-%20Final%20Version.pdf>. Archived at: <http://www.webcitation.org/5cGFm0gTX>.
- [18] Scheer AW. ARIS – business process modeling. Berlin, Germany: Springer-Verlag; 2000.
- [19] King DW, Lashley R. A quantifiable alternative to double data entry. *Control Clin Trials* 2000;21:94–102.
- [20] Wahi MM, Parks DV, Skate RC, Goldin SB. Reducing errors from the electronic transcription of data collected on paper forms: a research data case study. *J Am Med Inform Assoc* 2008;15:386–9.
- [21] Kleinman K. Adaptive double data entry: a probabilistic tool for choosing which forms to reenter. *Control Clin Trials* 2001;22:2–12.
- [22] Kawado M, Hinotsu S, Matsuyama Y, Yamaguchi T, Hashimoto S, Ohashi Y. A comparison of error detection rates between the reading aloud method and the double data entry method. *Control Clin Trials* 2003;24:560–9.
- [23] Day S, Fayers P, Harvey D. Double data entry: what value, what price? *Control Clin Trials* 1998;19:15–24.
- [24] Spink C. Electronic Data Capture (EDC) as a means for e-clinical trial success. IBM Global Services. Pharmaceutical Clinical Development; 2002 (Mar).

- [25] Bart T. Comparison of electronic data capture with paper data collection – is there really an advantage? *Bus Brief Pharmatech* 2003:1–4.
- [26] The eClinical Forum and PhRMA EDC/eSource Taskforce. The future vision of electronic health records as eSource for clinical research. The eClinical Forum and PhRMA EDC/eSource Taskforce; Sep 14 2006. <http://www.eclinicalforum.com/content/Knowledge/Articles/Future%20EHR-CR%20Environment%20Version%201.pdf>. Archived at: <http://www.webcitation.org/5cTvNZOVC>.
- [27] PhRMA. Pharmaceutical industry profile 2008. Washington, DC: PhRMA; 2008 (Mar), <http://www.phrma.org/files/2008%20Profile.pdf>. Archived at: <http://www.webcitation.org/5f0RbcG14>.
- [28] BIO-IT World. Where the millions go.; May 9 2003. The YGS Group, 1808 Colonial Village Lane, Lancaster, PA. [http://www.bio-itworld.com/archive/050903/data\\_sidebar\\_2450.html](http://www.bio-itworld.com/archive/050903/data_sidebar_2450.html). Archived at: <http://www.webcitation.org/5f0UtE4yS>.