

Oznaka poročila: ARRS-RPROJ-ZP-2015/101



## ZAKLJUČNO POROČILO RAZISKOVALNEGA PROJEKTA

### A. PODATKI O RAZISKOVALNEM PROJEKTU

#### 1. Osnovni podatki o raziskovalnem projektu

<b>Šifra projekta</b>	J3-4246
<b>Naslov projekta</b>	Statistične metode za visoko razsežne omične podatke v povezavi s semantičnimi relacijami iz biomedicinske literature
<b>Vodja projekta</b>	8992 Janez Stare
<b>Tip projekta</b>	J Temeljni projekt
<b>Obseg raziskovalnih ur</b>	9630
<b>Cenovni razred</b>	A
<b>Trajanje projekta</b>	07.2011 - 06.2014
<b>Nosilna raziskovalna organizacija</b>	381 Univerza v Ljubljani, Medicinska fakulteta
<b>Raziskovalne organizacije - soizvajalke</b>	2451 Zavod Biomedicinski Raziskovalni Inštitut
<b>Raziskovalno področje po šifrantu ARRS</b>	3 MEDICINA 3.08 Javno zdravstvo (varstvo pri delu)
<b>Družbeno-ekonomski cilj</b>	07. Zdravje
<b>Raziskovalno področje po šifrantu FOS</b>	3 Medicinske vede 3.05 Druge medicinske vede

### B. REZULTATI IN DOSEŽKI RAZISKOVALNEGA PROJEKTA

#### 2. Povzetek raziskovalnega projekta<sup>1</sup>

SLO

Napredne tehnologije za merjenje izraženosti genov so sprožile raziskovanje in razvoj podobnih orodij na vseh "omičnih" področjih (genomika, proteomika...).

Glavna značilnost tako pridobljenih podatkov je njihova visoka razsežnost, kar pomeni, da je število merjenih spremenljivk mnogo večje od števila enot. V zadnjih desetih letih je produkcija in uporaba visokorazsežnih podatkov v biomedicinskem raziskovanju skokovito narasla in

razvitih je bilo več statističnih metod za pravilno analizo takšnih podatkov. Biomedicinske raziskave, ki uporabljajo visokorazsežne podatke, imajo tipično dva cilja (i) odkriti pravila, po katerih natančno napovemo razredno pripadnost novih enot (napovedovanje razreda) in (ii) identificirati spremenljivke, katerih porazdelitve se ločijo med v naprej določenimi razredi (primerjava razredov). Tretji cilj, ki pa se v zadnjem času vse bolj pojavlja, je (iii) integracija podatkov iz različnih virov z namenom, da dobimo bolj zanesljive rezultate glede prvih dveh ciljev. Viri podatkov pti tretjem cilju so lahko že objavljeni podatki o istem znanstvenem vprašanju ali pa meritve na istih vzorcih, ki pa so namenjene drugim "omičnim" področjem ali morda drugim kliničnim vidikom (na primer, ko so meritve genske izraženosti in analiza sprememb v številu kopij DNA zaporedij narejene na istih enotah in so dostopni tudi klinični podatki).

Namen tega projekta je bil nadaljevati naše metodološke raziskave na področju napovedovanja in primerjave razredov, evalvacija strategij in razvoj nekaterih novih in nujno potrebnih metod za integracijo visoko razsežnih podatkov iz različnih virov.

Dodatno smo k temu dodali semantične relacije, pridobljene iz biomedicinske literature. Biomedicinska literatura je še vedno glavni vir biomedicinskega znanja, vendar so zaradi njene obsežnosti za njeno procesiranje potrebne zahtevne računske metode. V naši raziskovalni skupini že dalj časa razvijamo programe, ki z uporabo procesiranja naravnega jezika, iščejo ustrezne semantične relacije. Te predstavljajo dejstva, objavljena v literaturi, in ta so predstavljena v obliki, ki je primerna tudi za nadaljno statistično analizo.

ANG

The high-throughput technologies developed in the late Nineties for the measurement of gene expression, triggered the research and development of similar tools in all the "omics" fields (genomics, proteomics, ...).

The main characteristic of the high-dimensional data produced using high-throughput technologies is that the number of measured variables greatly exceeds the number of samples included in the experiment.

Over the last ten years there has been an increasing production and use of high-dimensional data in biomedical research, and many statistical methods were developed to appropriately deal with them.

Two common scientific goals of biomedical studies that use high-dimensional data are (i) to develop rules to accurately predict the class membership of new samples (class prediction) and (ii) to identify the variables whose distributions differ between some pre-specified classes (class comparison). A third emerging goal is (iii) the integration of different data sources to obtain more reliable results from class prediction or class comparison studies. The different data sources can be either previously published data addressing the same scientific question, or measurements regarding different "omics" and possibly clinical aspects on the same samples (for example when gene expression measurements and copy number analysis are performed on the same subjects and clinical information is available).

The aim of this project was to continue our ongoing statistical methodological research for high-dimensional class prediction and class comparison studies, and to evaluate strategies and develop some new and much needed methods for the integration of high-dimensional data deriving from different sources.

Additionally, we integrated semantic relations extracted from the biomedical literature. The biomedical literature is still the largest source of biomedical knowledge. However, due to its huge volume computer-based methods are needed to process it and to make it available to other applications. We extracted semantic relations from the literature by using natural language processing methods. The extracted semantic relations represent the facts expressed in the literature in a machine-processable form, suitable to be integrated with our statistical methods.

### 3. Poročilo o realizaciji predloženega programa dela na raziskovalnem projektu<sup>2</sup>

Člani naše projektne skupine so v obdobju 2011-2014 objavili 117 izvirnih znanstvenih člankov, od tega 99 v revijah, ki so indeksirane v SCI ali SSCI.

V teku projekta smo dosegli veliko pomembnih metodoloških rezultatov na področju statistične

analize in odkrivanja zakonitosti iz literature z uporabo visokorazsežnih »omičnih« podatkov. Dosežene rezultate smo predstavili v objavljenih znanstvenih člankih, na mednarodnih konferencah in z vabljenimi predavanji. Razvili smo tudi nekaj prosto dostopnih programskih orodij, v katerih smo implementirali nove metode, ki smo jih razvili v teku projekta.

Aktivno smo sodelovali z raziskovalnimi skupinami s področja molekularne biologije, farmakologije in medicine, s katerimi smo sodelovali pri načrtovanju in pri analizi visokorazsežnih »omičnih« podatkov. Rezultati tovrstnih sodelovanj so bili predstavljeni v znanstvenih člankih in s predstavitvami na mednarodnih konferencah.

Tukaj na kratko povzememo najpomembnejše rezultate projekta. Letna poročila vsebujejo podrobnejše opise.

Na področju uvrščanja visokorazsežnih podatkov smo posvetili posebno pozornost neuravnoteženim podatkom. Ukvarjali smo se s pristopi, ki se trenutno uporabljajo za reševanje pristranosti pri uvrščanju neuravnoteženih podatkov. Najpomembnejše rezultate smo objavili v reviji BMC Bioinformatics.

V prvem članku smo preučevali uspešnost zelo popularne metode SMOTE (Synthetic Minority Oversampling TEchique), ki pridobi uravnoteženo učno množico tako, da generira sintetične enote iz manjšega razreda. Pokazali smo, da je uporabnost metode pri uvrščanju visokorazsežnih podatkov zelo omejena, saj je metoda neučinkovita pri zmanjšanju pristranosti uvrščanja v prid večjemu razredu. V veliki meri je bilo mogoče obnašanje metode SMOTE pojasniti s teoretičnimi ugotovitvami, ki smo jih skupaj z empiričnimi in simulacijskimi rezultati predstavili v članku. Rezultate smo predstavili tudi na mednarodni konferenci »11th International Conference on Machine Learning and Applications« (ICMLA 2012), na Floridi, ZDA, kjer je prispevek dobil nagrado za najboljši članek na delavnici o neuravnoteženih podatkih.

V drugem članku smo se ukvarjali s PAM (Prediction of microarrays), metoda najbližjega skrčenega centroida za uvrščanje visokorazsežnih podatkov, kjer je enota uvrščena v tisti razred, pri katerem je razdalja do skrčenega centroida najmanjša. Raven krčenja centroidov določa prag krčenja. V praksi prag krčenja minimizira skupno napako, ki je ocenjena s prečnim preverjanjem (cross-validation). Pokazali smo, da je ta pristop pomanjkljiv, ko so podatki neuravnoteženi in visokorazsežni: uvrščanje je pristrano v prid prevladujočemu razredu. Predlagali smo rešitev, kjer ocenimo optimalni prag krčenja z minimizacijo G-povprečja (ang. G-means, geometrično povprečje napovedne točnosti posameznih razredov). Rezultati na pravih in simuliranih podatkih so pokazali, da smo s tem pristopom uspeli zmanjšati pristranost. Napovedna točnost za manjši razred se je v primerjavi z učenjem z izvorno različico metode PAM izboljšala in je bila približno enaka napovedni točnosti prevladujočega razreda

Poleg tega pripravljamo dva dodatna članka, ki opisujejo rezultate, ki smo jih dosegli s preučevanjem lastnosti metod za združevanje klasifikatorjev. Predvsem smo se osredotočili na metode ojačevanja (boosting), pristope, kjer se združi rezultate učenja, pridobljene na več učnih podmnožicah, kjer je dana večja teža enotam, ki so bile napačno uvrščene v prejšnjih korakih. Rezultate tega sklopa projekta smo že predstavili na mednarodnih konferencah »Applied Statistics« v Ribnem in »International Society for Clinical Biostatistics« na Dunaju. Predlagali smo popravek za najbolj pogosto uporabljeno metodo ojačevanja, algoritem AdaBoost.M1, kjer napako na učni množici ocenimo z notranjim prečnim preverjanjem. Naša metoda izboljšuje osnovni algoritem, ko se le-tapreprilega podatkom. Preprileganje je v nekaterih primerih mogoče zmanjšati tudi z uporabo šibkejših oz. bolj enostavnih osnovnih klasifikatorjev, ali z uporabo metode stohastičnega premičnega ojačevanja (Stochastic Gradient boosting), ki uporablja naključno izbrano podmnožico učnih podatkov in s tem zmanjšuje preprileganje.

Razvili smo knjižnico v jeziku R, kjer smo vključili nove metode in na novo napisali funkcije za nekatere druge že obstoječe metode ojačevanja, ki so bile prepočasne, ali celo neuporabne za visokorazsežne podatke. Knjižnica je prosto dostopna na spletu (<https://github.com/llaarraa/cartHD>).

Izmed znanstvenih sodelovanj, kjer smo analizirali visokorazsežne "omične" podatke, na tem mestu omenjamo zgolj sodelovanje s skupino dr. Manuela Gariboldi iz nacionalnega onkološkega inštituta v Milanu, s katero smo izvedli integrirano analizo izraženosti genov in mikroRNA za 60 bolnic z rakom dojk. Skupina v Milanu je eksperimentalno ovrednotila nekatere rezultate o interakciji med enim mikroRNA in enim genom, ki smo jih identificirali z analizo podatkov. Rezultati so bili povzeti v članku, ki je bil objavljen v reviji PLOS One.

Tudi na področju generiranja raziskovalnih hipotez in odkrivanja znanja z analizo strokovne literature smo dosegli dobre znanstvene rezultate. Nadaljevali smo tudi z razvojem lastne programske opreme, ki jo vse več uporabljamo za praktične raziskave.

V tem delu projekta smo sodelovali z raziskovalci iz Nacionalne knjižnice za medicino, Bethesda, ZDA. Sodelovanje je bilo vezano na dve področji. Na prvem področju smo vpeljali nov pristop "pregledovanja odkritij", ki temelji na teoriji grafov z uporabo semantičnih relacij iz biomedicinske literature. Novo metodo smo uporabili na primeru depresije in odkrili, da je za to bolezen pomembna interakcija med vnetjem, cirkadnimi značilnostmi ter neurotransmiterjem Norepinephrinom. Rezultate smo predstavili na mednarodni konferenci AMIA.

Na drugem področju smo uporabili zaprti način generiranja hipotez zato, da bi pojasnili opažanje, da kvaliteta spanja na splošno pada s staranjem, še zlasti pri moških. Naša nova hipoteza pojasnjuje padec kvalitete spanca s kombiniranjem dejstva, da testosteron izboljšuje spanje z inhibiranjem kortisola, z dejstvom da je raven testosterona zmanjšana pri starejših moških. Rezultate smo objavili v reviji Sleep, metodološke rezultate teh raziskav pa bomo v nadaljevanju uporabili na področju generiranja raziskovalnih hipotez..

Na področju odkrivanja znanja z analizo strokovne literature smo se lotili iskanja novih načinov zdravljenja neovaskularizacije pri diabetični retinopatiji. Pri tem smo uporabili orodje SemBT, ki smo ga razvili na našem inštitutu ter podatke raznih relevantnih »omičnih« raziskav. Z uporabo tega pristopa smo uspeli neodvisno odkriti nekatera zdravila, ki se trenutno uporabljajo pri kliničnih študijah ter postaviti nekaj hipotez za nove načine zdravljenja. Opravili smo pregled in primerjavo metod in dostopnih računalniških orodij, še zlasti iz zornega kota odkrivanja novih terapevtskih pristopov. Rezultate smo objavili v letu 2013 v reviji Biomed Res Int in v reviji Cardiovasc Hematol Agents Med Chem.

Na področju odkrivanja zakonitosti iz literature smo se ukvarjali tudi s preučevanjem topoloških lastnosti kompleksnih omrežij. Znano je, da se kompleksna omrežja ponašajo z zanimivimi topološkimi lastnostmi, kot so npr. majhen premer omrežja, visok koeficient zgoščanja, porazdelitev vozlišč, ki sledi potenčni porazdelitvi in modularnost. Sestavili in analizirali smo omrežje biomedicinskih terminov iz geslovnika MeSH. Analiza je pokazala, da omrežje MeSH terminov sledi modelu majhnega sveta, je močno modularno ter odporno tako na ciljane kot slučajne napade. Ugotovili smo, da potenčna porazdelitev ni ustrezen model za opisovanje porazdelitve stopenj vozlišč.

Poleg tega smo se ukvarjali tudi z razvojem metodologije za napovedovanje relacij v velikih kompleksnih omrežjih. Pokazali smo, da za proces odkrivanja zakonitosti iz literature lahko, z ustreznimi modifikacijami, uporabimo obstoječo metodologijo za napovedovanje povezav. V ta namen smo analizirali podatkovno bazo semantičnih predikatov iz bibliografske zbirke MEDLINE. Za napovedovanje povezav smo uporabili različne mere podobnosti med vozlišči, kot so npr. št. Skupnih sosedov, Jaccardov koeficient in koeficient prednostnega povezovanja. Rezultati so pokazali visoko napovedno točnost.

#### 4. Ocena stopnje realizacije programa dela na raziskovalnem projektu in zastavljenih raziskovalnih ciljev<sup>3</sup>

S stopnjo realizacije našega programa smo zelo zadovoljni. V teku projekta smo objavili metodološke članke v uglednih statističnih revijah in revijah informacijskih in komunikacijskih tehnologij, ki so nastali po programu projekta. Sodelovanje s skupino z italijanskega onkološkega inštituta, ki smo jo napovedali v proietku, se je izkazalo kot zelo uspešno in predvidevamo, da bomo s sodelovanjem nadaljevali in da bo to doprineslo k dodatnim skupnim objavam. V teku projekta smo tudi vzpostavili stike z drugimi raziskovalnimi skupinami z Univerze v Ljubljani, ki se ukvarjajo z visokorazsežnimi podatki. Sodelovanje smo že konkretizirali s skupnimi objavami in predstavitevami na mednarodnih konferencah. Pričakujemo, da bomo dodatno objavili nekatere rezultate, ki smo jih dosegli v teku projekta in so že bili predstavljeni na mednarodnih konferencah. Zadovoljni smo tudi, da smo razvili prosto dostopno programsko orodje, ki bo omogočilo uporabo naših metodoloških raziskav drugim raziskovalcem.

#### 5. Utemeljitev morebitnih sprememb programa raziskovalnega projekta oziroma sprememb, povečanja ali zmanjšanja sestave projektne skupine<sup>4</sup>

Ni bilo sprememb.

#### 6. Najpomembnejši znanstveni rezultati projektne skupine<sup>5</sup>

Znanstveni dosežek			
1.	COBISS ID	30458841	Vir: COBISS.SI
Opis	Naslov	SLO	Izboljšane metode najbližjega skrčenega centroida za visoko-razsežne neuravnotežene podatke
		ANG	Improved shrunken centroid classifiers for high-dimensional class-imbalanced data
		SLO	PAM, metoda najbližjega skrčenega centroida (NSC), je popularna metoda za uvrščanje visoko-razsežnih podatkov. ALP in AHP sta tudi NSC algoritma, ki sta bila predlagana kot izboljšanje metode PAM. Vse NSC metode temeljijo na skrčenih centroidih; nivo skrčenja se v praksi določi tako, da se minimizira napako, ki jo določimo s prečnim preverjanjem. V članku pokažemo, da so v primeru neuravnoteženih podatkov vse tri metode skrčenega centroida pristrane v prid večjemu razredu. Pristranost je večja, ko je število spremenljivk veliko ali nivo neravnotežja večji ali pa, ko so razlike med razredoma manjše. Da bi se izognili pristranosti predlagamo, da se nivo skrčenja določi na podlagi maksimizacije geometrijskega povprečja napovednih točnosti za posamezen razred, ki se jih oceni s prečnim preverjanjem (g-povprečje). Naši rezultati kažejo, da tovrstni pristop deluje bolje od obstoječega. Pomembno, število spremenljivk, ki se jih upošteva pri učenju, je v primeru uporabe našega pristopa manjše, kot če se uporablja prvotni pristop. Te ugotovitve so podprte s simuliranimi, kot tudi pravimi visoko-razsežnimi neuravnoteženimi podatki.
	ANG	Background PAM, a nearest shrunken centroid method (NSC), is a popular classification method for high-dimensional data. ALP and AHP are NSC algorithms that were proposed to improve upon PAM. The NSC methods base their classification rules on shrunken centroids; in practice the amount of shrinkage is estimated minimizing the overall cross-validated (CV) error rate. Results We show that when data are class-imbalanced the three NSC classifiers are biased towards the majority class. The bias is larger when the number of variables or class-imbalance is larger and/or the differences between classes are smaller. To diminish the class-imbalance problem of the NSC classifiers we propose to estimate the amount of shrinkage by	

		maximizing the CV geometric mean of the class-specific predictive accuracies (g-means). Conclusions The results obtained on simulated and real high-dimensional class-imbalanced data show that our approach outperforms the currently used strategy based on the minimization of the overall error rate when NSC classifiers are biased towards the majority class. The number of variables included in the NSC classifiers when using our approach is much smaller than with the original approach. This result is supported by experiments on simulated and real high-dimensional class-imbalanced data.
	Objavljeno v	BioMed Central; BMC bioinformatics; 2013; Vol. 14; str. [1-27], 64; Impact Factor: 2.672; Srednja vrednost revije / Medium Category Impact Factor: 1.873; A': 1; WoS: CO, DB, MC; Avtorji / Authors: Blagus Rok, Lusa Lara
	Tipologija	1.01 Izvirni znanstveni članek
2.	COBISS ID	30528217 Vir: COBISS.SI
	Naslov	<i>SLO</i> Metoda SMOTE za visoko-razsežne neuravnotežene podatke
		<i>ANG</i> SMOTE for high-dimensional class-imbalanced data
	Opis	<p><i>SLO</i> Uvrščanje, kjer se uporablja neuravnotežene podatke, je pristrano v prid večjemu razredu. Pristranost je v primeru visoko-razsežnih podatkov, kjer število spremenljivk presega število enot, še večja. Pristranost lahko zmanjšamo z uporabo zmanjšanja večjega razreda (s slučajnim izborom enot, ki jih uporabljamo pri učenju) ali pa povečanjem manjšega razreda (s pomočjo ponovnega vzorčenja). V splošnem zmanjšanje večjega razreda pomaga pri zmanjšanju pristranosti, medtem ko metoda slučajnega povečanja manjšega razreda, v primeru visoko-razsežnih podatkov, ne deluje. Metoda SMOTE (sintetično generiranje novih enot) je zelo popularna metoda povečanja manjšega razreda, za katero je bilo pokazano, da deluje bolje od slučajnega povečanja manjšega razreda, vendar pa njeno delovanje v primeru visoko-razsežnih podatkov še ni bilo raziskano. V tem članku analiziramo teoretične in empirične lastnosti metode SMOTE, kjer uporabljamo prave in simulirane visoko-razsežne podatke. Kljub temu, da metoda SMOTE uspešno zmanjša pristranost v primeru nizko-razsežnih podatkov, ugotavljamo, da v primeru visoko-razsežnih podatkov metoda ne deluje dobro za večino uporabljenih klasifikatorjev. Metoda SMOTE je uporabna zgolj v kombinaciji z metodami najbližjega soseda, vendar samo, če pred učenjem zmanjšamo razsežnost podatkov z metodami za izbiro spremenljivk; pokažemo in razložimo, zakaj je v primeru, ko se razsežnost podatkov pred učenjem ne zmanjša, uvrščanje z metodami najbližjega soseda pristrano v prid manjšemu razredu. Teoretično pokažemo tudi, da z metodo SMOTE ne spremenimo pričakovane vrednosti manjšega razreda, vendar zmanjšamo njegovo variabilnost in v učno množico uvajamo odvisnost med enotami. Pokažemo kako te teoretične ugotovitve vplivajo na uvrščanje.</p> <p><i>ANG</i> Background Classification using class-imbalanced data is biased in favor of the majority class. The bias is even larger for high-dimensional data, where the number of variables greatly exceeds the number of samples. The problem can be attenuated by undersampling or oversampling, which produce class-balanced data. Generally undersampling is helpful, while random oversampling is not. Synthetic Minority Oversampling TEchnique (SMOTE) is a very popular oversampling method that was proposed to improve random oversampling but its behavior on high-dimensional data has not been thoroughly investigated. In this paper we investigate the properties of SMOTE from a theoretical and empirical point of view, using simulated and real high-dimensional data. Results While in most cases SMOTE seems beneficial with low-dimensional data, it does not attenuate the bias towards the classification in the majority class for most classifiers when data are high-dimensional, and it is less effective than random</p>

		undersampling. SMOTE is beneficial for k-NN classifiers for high-dimensional data if the number of variables is reduced performing some type of variable selection; we explain why, otherwise, the k-NN classification is biased towards the minority class. Furthermore, we show that on high-dimensional data SMOTE does not change the class-specific mean values while it decreases the data variability and it introduces correlation between samples. We explain how our findings impact the class-prediction for high-dimensional data. Conclusions In practice, in the high-dimensional setting only k-NN classifiers based on the Euclidean distance seem to benefit substantially from the use of SMOTE, provided that variable selection is performed before using SMOTE; the benefit is larger if more neighbors are used. SMOTE for k-NN without variable selection should not be used, because it strongly biases the classification towards the minority class.
	Objavljeno v	BioMed Central; BMC bioinformatics; 2013; Vol. 14; str. [1-19], 106; Impact Factor: 2.672; Srednja vrednost revije / Medium Category Impact Factor: 1.873; A': 1; WoS: CO, DB, MC; Avtorji / Authors: Blagus Rok, Lusa Lara
	Tipologija	1.01 Izvirni znanstveni članek
3.	COBISS ID	677804 Vir: COBISS.SI
	Naslov	<i>SLO</i> Uporaba odkrivanja zakonitosti z analizo strokovne literature za identifikacijo novih pristopov zdravljenja
		<i>ANG</i> Using literature-based discovery to identify novel therapeutic approaches
	Opis	<i>SLO</i> Predstavljamo obetavno in silico paradigmo, ki ji rečemo odkrivanje zakonitosti z analizo strokovne literature (ang. Literature-based Discovery, kratica LBD) in opisujemo njen potencial pri identificiranju novih farmakoloških pristopov zdravljenja bolezni. Cilj LBD je generirati nove hipoteze z analizo biomedicinske literature. Dodatni viri znanja, kot so ontologije ter specializirane podatkovne zbirke, se pogosto uporabljajo kot dopolnilo objavljene literature. MEDLINE, največja in najpomembnejša biomedicinska bibliografska zbirka, se najpogosteje uporablja pri LBD. Obstajata dve inačici LBD, odprto odkrivanje in zaprto odkrivanje. Z odprtim odkrivanjem, lahko poiščemo nov terapevtski pristop za obstoječo bolezen, ali pa poiščemo nove terapevtske aplikacije za obstoječa zdravila. Z zaprtim odkrivanjem lahko poiščemo razlago za relacijo med koncepti. Kot primer, če že imamo hipotezo o tem da je določeno zdravilo koristno za določeno bolezen, z zaprtim odkrivanjem lahko identificiramo mehanizme skozi katere zdravilo lahko terapevtsko učinkuje na bolezen. Na kratko opisujemo LBD metodologijo in pojasnimo bolj natančno trenutno razpoložljiva LBD orodja; omenimo na kratko tudi nekatera, ki niso več na razpolago. Potem predstavimo nekaj primerov v katerih je pristop LBD uporabljen za identifikacijo novih terapevtskih pristopov. Za zaključek, LBD je zmogljiv pristop, ki lahko nadopolni tradicionalne metode odkrivanja zdravil, še zlasti pri odkrivanju novih tarč za zdravila in pri repositioniranju obstoječih zdravil.
		<i>ANG</i> We present a promising in silico paradigm called literature-based discovery (LBD) and describe its potential to identify novel pharmacologic approaches totreating diseases. The goal of LBD is to generate novel hypotheses by analyzing the vast biomedical literature. Additional knowledge resources, suchas ontologies and specialized databases, are often used to supplement the published literature. MEDLINE, the largest and most important biomedical bibliographic database, is the most common source for exploiting LBD. There are two variants of LBD, open discovery and closed discovery. With open discovery we can, for example, try to find a novel therapeutic approach for a given disease, or find new therapeutic applications for an existing drug. Withclosed discovery we can find an

		<p>explanation for a relationship between two concepts. For example, if we already have a hypothesis that a particular drug is useful for a particular disease, with closed discovery we can identify the mechanisms through which the drug could have a therapeutic effect on the disease. We briefly describe the methodology behind LBD and then discuss in more detail currently available LBD tools; we also mention in passing some of those no longer available. Next we present several examples in which LBD has been exploited for identifying novel therapeutic approaches. In conclusion, LBD is a powerful paradigm with considerable potential to complement more traditional drug discovery methods, especially for drug target discovery and for existing drug relabeling.</p>
	Objavljeno v	Bentham Science Publishers; Cardiovascular & hematological agents in medicinal chemistry; 2013; Vol. 11, no. 1; str. 14-24; Avtorji / Authors: Hristovski Dimitar, Rindfleisch Thomas C., Peterlin Borut
	Tipologija	1.01 Izvirni znanstveni članek
4.	COBISS ID	31259609 Vir: COBISS.SI
	Naslov	<p><i>SLO</i> Integracija podatkov omičnih študij z odkrivanjem, ki temelji na literaturi - k identifikaciji novih načinov zdravljenja neovaskularizacije pri diabetični retinopatiji</p> <p><i>ANG</i> Integration of data from omic studies with the literature-based discovery towards identification of novel treatments for neovascularization in diabetic retinopathy</p>
	Opis	<p><i>SLO</i> Diabetična retinopatija (DR) je sekundarni zaplet sladkorne bolezni, povezan z neovaskularizacijo mrežnice in predstavlja najpogostejši vzrok slepote odraslega prebivalstva v razvitem svetu. Kljub raziskovalnih prizadevanj, narava patogenetskih procesov, ki vodijo k DR še vedno ni znana kar zavira razvoj novih učinkovitih načinov zdravljenja. Napredek omske tehnologije zdaj ponujaja neprimerljiv vpogled v globalne molekularne spremembe pri DR, vendar identifikacija novih načinov zdravljenja, ki temeljijo na ogromnih količin podatkov, pridobljenih z omskimi študijami, še vedno predstavlja velik izziv. Zaradi tega razloga smo poskušali olajšati odkrivanje novih zdravil za DR z dopolnjevanjem rezultatov omskih raziskav z podatki, ki obstajajo v objavljeni literaturi z metodo odkrivanja, ki temelji na literaturi (ang. LBD - Literature-based Discovery). Da bi to dosegli, smo zbrali podatke transkriptomskih študij, opravljenih na tkivu mrežnice živalskih modelih DR, smo opravili meta analizo teh podatkovnih baz in ugotovljenih spremenjenih genov in poti. Z uporaba LBD orodja SemBT smo ugotovili, katere terapije bi lahko stabilizirali spremembe genske ekspresije pri DR. Pokazali smo, da z uporabo tega pristopa, ne samo da smo ponovno odkrili zdravila, ki so trenutno v uporabi ali v kliničnih raziskavah, temveč smo tudi odkrili nove smernice zdravljenja neovaskularizacije pri DR.</p> <p><i>ANG</i> Diabetic retinopathy (DR) is a secondary complication of diabetes associated with retinal neovascularization and represents the leading cause of blindness in the adult population in the developed world. Despite research efforts, the nature of pathogenetic processes leading to DR is still unknown, making development of novel effective treatments difficult. Advances in omic technologies now offer unprecedented insight into global molecular alterations in DR, but identification of novel treatments based on massive amounts of data generated in omic studies still represents a considerable challenge. For this reason, we attempted to facilitate discovery of novel treatments for DR by complementing the interpretation of omic results using the vast body of information existing in the published literature with the literature-based discovery (LBD) approaches. To achieve this, we collected data from transcriptomic studies performed on retinal tissue from animal models of DR, performed a meta-</p>



		analysis of these datasets and identified altered genes and pathways. Using the SemBT LBD framework, we have determined which therapies could regulate perturbed pathways or that could stabilize the gene expression alterations in DR. We show that by using this approach, we not only could reidentify drugs currently in use or in clinical trials, but also could indicate novel treatment directions for ameliorating neovascularization processes in DR.
	Objavljeno v	Hindawi Pub. Co.; BioMed research international; 2013; Vol. 2013; Impact Factor: 2.706; Srednja vrednost revije / Medium Category Impact Factor: 2.947; WoS: DB, QA; Avtorji / Authors: Maver Aleš, Hristovski Dimitar, Rindflesch Thomas C., Peterlin Borut
	Tipologija	1.01 Izvirni znanstveni članek
5.	COBISS ID	2048311059 Vir: COBISS.SI
	Naslov	<i>SLO</i> Struktura omrežja sopojavnosti MeSH terminov: Statistična analiza makroskopskih lastnosti
		<i>ANG</i> Large-scale structure of a network of co-occurring MeSH terms
	Opis	<i>SLO</i> <p>Mrežo konceptov lahko predstavimo kot omrežje, v katerem množica vozlišč predstavlja koncepte, množica povezav pa relacije med koncepti. Kompleksna omrežja se ponašajo z nekaterimi topološkimi lastnostmi, kot so npr. majhen premer omrežja, visok koeficient zgoščanja, porazdelitev stopenj vozlišč, ki sledi potenčni porazdelitvi in visoka modularnost. V članku se ukvarjamo s proučevanjem topoloških lastnosti omrežij sopojavnosti biomedicinskih konceptov, konstruiranih na osnovi množice vseh MeSH deskriptorjev oz. glavnih MeSH deskriptorjev. Redukcijo omrežja smo opravili s pomočjo filtriranja parov povezav na osnovi Pearsonovega hi-kvadrat testa za neodvisnost. Za opis in predstavitev topoloških lastnosti smo uporabili premer omrežja, povprečno dolžino poti, koeficient zgoščanja in porazdelitev stopenj vozlišč. Povprečna dolžina poti v omrežju z vsemi MeSH deskriptorji znaša 1.95, premer omrežja znaša tri povezave, koeficient zgoščanja pa 0.26. Na osnovi Kolmogorov-Smirnovega testa lahko zavrnamo model potenčne porazdelitve kot ustrezen model za opisovanje porazdelitve stopenj vozlišč. Povprečna dolžina poti v omrežju z glavnimi MeSH deskriptorji znaša 2.63, s premerom sedmih povezav in koeficientom zgoščanja 0.15. Tudi v tem primeru na osnovi Kolmogorov-Smirnovega testa zavrnamo model potenčne porazdelitve stopenj vozlišč. Iz obeh omrežij je razvidno, da se vozlišča z nižjimi stopnjami ponašajo z večjim zgoščanjem kot vozlišča z višjimi stopnjami. Pri simuliranem napadu na vozlišča, kjer smo odstranili 10 % vozlišč z najvišjo stopnjo, je ogromna komponenta pri obeh omrežjih ohranila okrog 90 % vseh vozlišč. Zaradi majhne povprečne dolžine poti in visoke stopnje zgoščanja obe omrežji sledita modelu majhnega sveta. Ugotavljamo, da model potenčne porazdelitve ni ustrezen za opisovanje porazdelitve stopenj vozlišč. Omrežji sta poleg tega močno modularni, močno odporni na tarčne in slučajne napade ter se ponašata z minimalno disortativnostjo.</p> <p>Concept associations can be represented by a network that consists of a set of nodes representing concepts and a set of edges representing their relationships. Complex networks exhibit some common topological features including small diameter, high degree of clustering, power-law degree distribution, and modularity. We investigated the topological properties of a network constructed from co-occurrences between MeSH descriptors in the MEDLINE database. We conducted the analysis on two networks, one constructed from all MeSH descriptors and another using only major descriptors. Network reduction was performed using the Pearson's chi-square test for independence. To characterize topological properties of the network we adopted some specific measures, including diameter, average path length, clustering coefficient, and degree distribution. For the full</p>

	ANG	MeSH network the average path length was 1.95 with a diameter of three edges and clustering coefficient of 0.26. The Kolmogorov-Smirnov test rejects the power law as a plausible model for degree distribution. For the major MeSH network the average path length was 2.63 edges with a diameter of seven edges and clustering coefficient of 0.15. The Kolmogorov-Smirnov test failed to reject the power law as a plausible model. The power-law exponent was 5.07. In both networks it was evident that nodes with a lower degree exhibit higher clustering than those with a higher degree. After simulated attack, where we removed 10% of nodes with the highest degrees, the giant component of each of the two networks contains about 90% of all nodes. Because of small average path length and high degree of clustering the MeSH network is small-world. A power-law distribution is not a plausible model for the degree distribution. The network is highly modular, highly resistant to targeted and random attack and with minimal assortativity.
Objavljeno v		Public Library of Science; PloS one; 2014; Vol. 9, iss. 7; Impact Factor: 3.534; Srednja vrednost revije / Medium Category Impact Factor: 2.663; A': 1; WoS: RO; Avtorji / Authors: Kastrin Andrej, Rindfleisch Thomas C., Hristovski Dimitar
Tipologija		1.01 Izvirni znanstveni članek

### 7. Najpomembnejši družbeno-ekonomski rezultati projektne skupine<sup>6</sup>

	Družbeno-ekonomski dosežek	
1.	COBISS ID	30802905 Vir: COBISS.SI
	Naslov	SLO Uvajanje farmakogenetike v klinično prakso ANG Translating pharmacogenetics to clinical practice
	Opis	SLO Vabljen predavanje na delavnici: Casym Training Event Systems Medicine of Multifactorial Disorders Workshop & Tutorial and 8th CFGBC Symposium, Ljubljana, 12-15 junij 2013 ANG Invited lecture at the workshop: Vabljen predavanje na delavnici: Casym Training Event Systems Medicine of Multifactorial Disorders Workshop & Tutorial and 8th CFGBC Symposium, Ljubljana, 12-15 June 2013.
	Šifra	B.04 Vabljen predavanje
	Objavljeno v	Faculty of Medicine; Book of abstracts; 2013; Str. 24; Avtorji / Authors: Dolžan Vita, Lusa Lara
	Tipologija	1.10 Objavljeni povzetek znanstvenega prispevka na konferenci (vabljen predavanje)
2.	COBISS ID	30783961 Vir: COBISS.SI
	Naslov	SLO Uvrščanje z neravnoteženimi in visoko-razsežnimi podatki ANG Class-imbalanced class prediction for high-dimensional data
	Opis	SLO Vabljen predavanje na delavnici: Did the genomic data flood overrun the statistical levee? (International Biometric Society - avstrijska in švicarska sekcija) ANG Invited lecture at the workshop: Did the genomic data flood overrun the statistical levee? - organized by the Austrian and Swiss section of the International Biometric Society, Wien, 4th of July 2013.
	Šifra	B.04 Vabljen predavanje
	Objavljeno v	[s. n.]; Did the genomic data flood overrun the statistical levee?; 2013; Avtorji / Authors: Lusa Lara, Blagus Rok

	Tipologija	1.10 Objavljeni povzetek znanstvenega prispevka na konferenci (vabljeni predavanja)
3.	COBISS ID	Vir: vpis v poročilo
	Naslov	<i>SLO</i> Nagrada za najboljši prispevek na delavnici na ICMLA konferenci
		<i>ANG</i> Best paper workshop award at ICMLA 2012
	Opis	<i>SLO</i> Nagrada za najboljši prispevek na delavnici "Class imbalance: past, present and future" na 11. konferenci "Conference on Machine Learning and Applications (ICMLA 2012)", Florida, ZDA. Priznanje je dostopno na spletni strani konference (povezava: <a href="http://www.icmla-conference.org/icmla12/links/best.ws-Class%20Imbalances.pdf">http://www.icmla-conference.org/icmla12/links/best.ws-Class%20Imbalances.pdf</a> ). Članek je objavljen kot znanstveni prispevek na konferenci (BLAGUS, Rok, LUSA, Lara. Evaluation of SMOTE for high-dimensional class-imbalanced microarray data. V: TAO, Dacheng (ur.). ICMLA 2012. [S. l.]: Institute of Electrical and Electronics Engineers, 2012, str. 89-94. [COBISS.SI-ID 30373849])
		<i>ANG</i> Best paper award at the workshop "Class imbalances: past, present and future" at the 11th International Conference on Machine Learning and Applications (ICMLA 2012), held in Florida, U.S.A. The award is published at the conference web site ( <a href="http://www.icmla-conference.org/icmla12/links/best.ws-Class%20Imbalances.pdf">http://www.icmla-conference.org/icmla12/links/best.ws-Class%20Imbalances.pdf</a> ). The paper is published in the conference proceedings (BLAGUS, Rok, LUSA, Lara. Evaluation of SMOTE for high-dimensional class-imbalanced microarray data. V: TAO, Dacheng (ur.). ICMLA 2012. [S. l.]: Institute of Electrical and Electronics Engineers, 2012, str. 89-94. [COBISS.SI-ID 30373849]).
	Šifra	E.02 Mednarodne nagrade
	Objavljeno v	<a href="http://www.icmla-conference.org/icmla12">http://www.icmla-conference.org/icmla12</a>
	Tipologija	1.08 Objavljeni znanstveni prispevek na konferenci
4.	COBISS ID	Vir: vpis v poročilo
	Naslov	<i>SLO</i> Izvlečki in program
		<i>ANG</i> Abstracts and program
	Opis	<i>SLO</i> Janez Stare je od leta 2004 naprej (vsako leto) predsednik programskega odbora mednarodne konference Applied Statistics (glej <a href="http://conferences.nib.si/AS2013/default.htm">http://conferences.nib.si/AS2013/default.htm</a> ). Skupaj z Laro Luso urejata zbornik konference.
		<i>ANG</i> Janez Stare has been chair of the Programme committee of the Applied Statistics (see <a href="http://conferences.nib.si/AS2013/default.htm">http://conferences.nib.si/AS2013/default.htm</a> ) since 2004. He and Lara Lusa they are editors of the conference proceedings.
	Šifra	B.02 Predsedovanje programskemu odboru konference
	Objavljeno v	Statistical Society of Slovenia; 2013; 83 str.; Avtorji / Authors: Lusa Lara, Stare Janez
	Tipologija	2.25 Druge monografije in druga zaključena dela
5.	COBISS ID	Vir: vpis v poročilo
	Naslov	<i>SLO</i> Metodološki zvezki
		<i>ANG</i> Advances in Methodology and Statistics
	Opis	<i>SLO</i> Lara Lusa je od leta 2013 sourednica revije Metodološki zvezki
		<i>ANG</i> Since 2013 Lara lusa is the co-editor of the journal Advances in methodology and statistics
	Šifra	C.04 Uredništvo mednarodne revije
	Objavljeno v	Metodološki zvezki - Advances in Methodology and Statistics

Tipologija	4.00	Sekundarno avtorstvo
------------	------	----------------------

## 8. Drugi pomembni rezultati projektne skupine<sup>7</sup>

Število čistih citatov znanstvenih del v zadnjih 10 letih (2005-2014) je 4155.  
 Število objavljenih izvirnih znanstvenih člankov v teku projekta (2011-2014) je 117, od tega 99 v revijah, ki so indeksirane v SCI ali SSCI.  
 Dva vabljeni predavanja na temo projekta.

## 9. Pomen raziskovalnih rezultatov projektne skupine<sup>8</sup>

### 9.1. Pomen za razvoj znanosti<sup>9</sup>

SLO

Člani naše raziskovalne skupine so imeli dolgo tradicijo raziskovanja na področju analize preživetja, kar se kaže v številnih znanstvenih objavah, ki so dokumentirane v naši bibliografiji. V zadnjih letih smo bili aktivni tudi pri raziskovanju visokorazsežnih podatkov; izvajali smo analize ter preučevali metodološke značilnosti uporabljenih metod. Razvili smo tudi javno dostopna orodja za analizo „omičnih“ podatkov, kot tudi za pomoč pri interpretaciji rezultatov. S tem projektom smo lahko združili različna znanja naše skupine in nadaljevali raziskovanje na tem hitro rastočem področju.

Projekt je veliko doprinesel k pravilni statistični analizi visokorazsežnih podatkov, tako metodološko kot pri uporabi teh metod v praksi.. Naš cilj je bil boljše razumevanje lastnosti obstoječih metod in razvoj novih, ustrežnejših metod za analizo visokorazsežnih „omičnih“ podatkov in orodij za interpretacijo rezultatov iz tovrstnih analiz, s čimer smo povečali vpliv in uporabnost naših odkritij. Menimo, da smo dosegli ta cilj, saj smo prispevali metodološke rezultate, razvili nova programska orodja in uspostavili stike z raziskovalnimi biomedicinskimi skupinami, v sodelovanju s katerimi smo načrtovali in analizirali visokorazsežne „omične“ podatke.

ANG

The members of our research group have been working in the fields of time-to-event data, producing many scientific contributions documented in our bibliography. In the last few years we have also been working on methodological and applied projects with high-dimensional data. Furthermore, we developed publicly available tools for the analysis of genomic data and for their interpretation. This project gave us the opportunity to put together the different know-how of our group and to continue our research in this fast developing field.

We believe that, as expected, this project contributed many important results from the methodological and from the applied point of view. Our aim was to provide a better understanding of the properties of some existing methods and to develop new methods and tools for the analysis and interpretation of high-dimensional "omics" data. We think that we achieved this goal because we contributed methodologically innovative results and developed new publicly available tools for the analysis and interpretation of high-dimensional "omics" data. Moreover, we collaborated with biomedical researchers, jointly planning and analyzing their high-dimensional "omics" data.

### 9.2. Pomen za razvoj Slovenije<sup>10</sup>

SLO

Tehnologija, ki se uporablja v „omičnem“ raziskovanju se izredno hitro razvija, kar ustvarja nov tip podatkov, ki terjajo razvoj novih statističnih metod in novih orodij za njihovo analizo in interpretacijo. Take podatke generira čedalje večje število slovenskih raziskovalcev na področju

biomedicine.

Splošno znano je, da veliko znanstvenih člankov, ki uporabljajo nove tehnologije, uporablja napačne statistične metode za načrtovanje in analizo poskusov, saj je poznavanje problemov, ki jih prinaša nov tip podatkov, zelo omejeno.

Statistiki in bioinformatiki, ki so sodelovali pri projektu, so v teku projekta nudili podporo raziskovalnemu in strokovnemu delu v slovenski biomedicini. Analizirali smo visokorazsežne podatke iz "omičnega" področja, ki so jih generirali slovenski raziskovalci ter pomagali pri načrtovanju raziskav in pri interpretaciji rezultatov. Prepričani smo, da je metodološko znanje, ki smo ga pridobili z izvedbo projekta, bistveno prispevalo k boljšemu sodelovanju s slovenskimi raziskovalci in na izboljšanje kakovosti izdelkov, ki so nastali tekom sodelovanja. To predstavlja eno od osnovnih usmeritev našega inštituta, ki je nuditi najkvalitetnejšo podporo raziskovalnemu delu v slovenski medicini.

ANG

The technologies used in the "omics" research are rapidly evolving, producing data that require the development of new statistical methods and of new tools for their analysis and interpretation.

The number of biomedical researchers that use "omics" data has been rapidly increasing.

It is well known that many of the research papers that use novel technologies have many flaws in the design and analysis of the data, because the understanding of problems related to a new type of data is initially very limited.

The statisticians and bioinformaticians involved in this project provided support to the slovenian researchers. We analyzed high-dimensional "omics" data that were generated by slovenian researchers and helped them in the planning of the experiments and with the interpretation of the obtained results. We are convinced that the methodological knowledge that we acquired during the project significantly improved the quality of the collaborations and of the scientific outputs. This constitutes one of the missions of our institute, which is to provide the highest quality support to the research in medicine in Slovenia.

#### 10. Samo za aplikativne projekte in podoktorske projekte iz gospodarstva!

**Označite, katerega od navedenih ciljev ste si zastavili pri projektu, katere konkretne rezultate ste dosegli in v kakšni meri so doseženi rezultati uporabljeni**

Cilj	
<b>F.01</b>	<b>Pridobitev novih praktičnih znanj, informacij in veščin</b>
	Zastavljen cilj <input type="radio"/> DA <input type="radio"/> NE
	Rezultat <input type="text"/>
	Uporaba rezultatov <input type="text"/>
<b>F.02</b>	<b>Pridobitev novih znanstvenih spoznanj</b>
	Zastavljen cilj <input type="radio"/> DA <input type="radio"/> NE
	Rezultat <input type="text"/>
	Uporaba rezultatov <input type="text"/>
<b>F.03</b>	<b>Večja usposobljenost raziskovalno-razvojnega osebja</b>
	Zastavljen cilj <input type="radio"/> DA <input type="radio"/> NE
	Rezultat <input type="text"/>
	Uporaba rezultatov <input type="text"/>
<b>F.04</b>	<b>Dvig tehnološke ravni</b>
	Zastavljen cilj <input type="radio"/> DA <input type="radio"/> NE
	<input type="text"/>

	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.05</b>	<b>Sposobnost za začetek novega tehnološkega razvoja</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.06</b>	<b>Razvoj novega izdelka</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.07</b>	<b>Izboljšanje obstoječega izdelka</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.08</b>	<b>Razvoj in izdelava prototipa</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.09</b>	<b>Razvoj novega tehnološkega procesa oz. tehnologije</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.10</b>	<b>Izboljšanje obstoječega tehnološkega procesa oz. tehnologije</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.11</b>	<b>Razvoj nove storitve</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.12</b>	<b>Izboljšanje obstoječe storitve</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.13</b>	<b>Razvoj novih proizvodnih metod in instrumentov oz. proizvodnih procesov</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>

<b>F.14</b>	<b>Izboljšanje obstoječih proizvodnih metod in instrumentov oz. proizvodnih procesov</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.15</b>	<b>Razvoj novega informacijskega sistema/podatkovnih baz</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.16</b>	<b>Izboljšanje obstoječega informacijskega sistema/podatkovnih baz</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.17</b>	<b>Prenos obstoječih tehnologij, znanj, metod in postopkov v prakso</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.18</b>	<b>Posredovanje novih znanj neposrednim uporabnikom (seminarji, forumi, konference)</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.19</b>	<b>Znanje, ki vodi k ustanovitvi novega podjetja ("spin off")</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.20</b>	<b>Ustanovitev novega podjetja ("spin off")</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.21</b>	<b>Razvoj novih zdravstvenih/diagnostičnih metod/postopkov</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.22</b>	<b>Izboljšanje obstoječih zdravstvenih/diagnostičnih metod/postopkov</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.23</b>	<b>Razvoj novih sistemskih, normativnih, programskih in metodoloških rešitev</b>	

	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.24</b>	<b>Izboljšanje obstoječih sistemskih, normativnih, programskih in metodoloških rešitev</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.25</b>	<b>Razvoj novih organizacijskih in upravljavskih rešitev</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.26</b>	<b>Izboljšanje obstoječih organizacijskih in upravljavskih rešitev</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.27</b>	<b>Prispevek k ohranjanju/varovanju naravne in kulturne dediščine</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.28</b>	<b>Priprava/organizacija razstave</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.29</b>	<b>Prispevek k razvoju nacionalne kulturne identitete</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.30</b>	<b>Strokovna ocena stanja</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.31</b>	<b>Razvoj standardov</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text" value=""/>
	Uporaba rezultatov	<input type="text" value=""/>
<b>F.32</b>	<b>Mednarodni patent</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE



	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.33</b>	<b>Patent v Sloveniji</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.34</b>	<b>Svetovalna dejavnost</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
<b>F.35</b>	<b>Drugo</b>	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>

**Komentar**

--

**11.Samo za aplikativne projekte in podoktorske projekte iz gospodarstva!**

Označite potencialne vplive oziroma učinke vaših rezultatov na navedena področja

	Vpliv	Ni vpliva	Majhen vpliv	Srednji vpliv	Velik vpliv	
<b>G.01</b>	<b>Razvoj visokošolskega izobraževanja</b>					
G.01.01.	Razvoj dodiplomskega izobraževanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.01.02.	Razvoj podiplomskega izobraževanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.01.03.	Drugo: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>G.02</b>	<b>Gospodarski razvoj</b>					
G.02.01	Razširitev ponudbe novih izdelkov/storitev na trgu	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.02.	Širitev obstoječih trgov	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.03.	Znižanje stroškov proizvodnje	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.04.	Zmanjšanje porabe materialov in energije	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.05.	Razširitev področja dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.06.	Večja konkurenčna sposobnost	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.07.	Večji delež izvoza	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.08.	Povečanje dobička	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.09.	Nova delovna mesta	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.10.	Dvig izobrazbene strukture zaposlenih	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.11.	Nov investicijski zagon	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.12.	Drugo: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

<b>G.03</b>	<b>Tehnološki razvoj</b>					
G.03.01.	Tehnološka razširitev/posodobitev dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.02.	Tehnološko prestrukturiranje dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.03.	Uvajanje novih tehnologij	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.04.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>G.04</b>	<b>Družbeni razvoj</b>					
G.04.01	Dvig kvalitete življenja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.02.	Izboljšanje vodenja in upravljanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.03.	Izboljšanje delovanja administracije in javne uprave	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.04.	Razvoj socialnih dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.05.	Razvoj civilne družbe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.06.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>G.05.</b>	<b>Ohranjanje in razvoj nacionalne naravne in kulturne dediščine in identitete</b>					
<b>G.06.</b>	<b>Varovanje okolja in trajnostni razvoj</b>					
<b>G.07</b>	<b>Razvoj družbene infrastrukture</b>					
G.07.01.	Informacijsko-komunikacijska infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.02.	Prometna infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.03.	Energetska infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.04.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>G.08.</b>	<b>Varovanje zdravja in razvoj zdravstvenega varstva</b>					
<b>G.09.</b>	<b>Drugo:</b>					

**Komentar**

--

**12.Pomen raziskovanja za sofinancerje<sup>11</sup>**

	Sofinancer			
1.	Naziv			
	Naslov			
	Vrednost sofinanciranja za celotno obdobje trajanja projekta je znašala:		EUR	
	Odstotek od utemeljenih stroškov projekta:		%	
	Najpomembnejši rezultati raziskovanja za sofinancerja			Šifra
	1.			
	2.			
3.				

	4.	
	5.	
Komentar		
Ocena		

### 13. Izjemni dosežek v letu 2014<sup>12</sup>

#### 13.1. Izjemni znanstveni dosežek

#### 13.2. Izjemni družbeno-ekonomski dosežek

## C. IZJAVE

Podpisani izjavljam/o, da:

- so vsi podatki, ki jih navajamo v poročilu, resnični in točni
- se strinjamo z obdelavo podatkov v skladu z zakonodajo o varstvu osebnih podatkov za potrebe ocenjevanja ter obdelavo teh podatkov za evidence ARRS
- so vsi podatki v obrazcu v elektronski obliki identični podatkom v obrazcu v pisni obliki
- so z vsebino zaključnega poročila seznanjeni in se strinjajo vsi soizvajalci projekta

### Podpisi:

*zastopnik oz. pooblaščen oseba  
raziskovalne organizacije:*

in

*vodja raziskovalnega projekta:*

Univerza v Ljubljani, Medicinska  
fakulteta

Janez Stare

**ŽIG**

Kraj in datum:

Ljubljana

12.3.2015

**Oznaka poročila: ARRS-RPROJ-ZP-2015/101**

<sup>1</sup> Napišite povzetek raziskovalnega projekta (največ 3.000 znakov v slovenskem in angleškem jeziku) [Nazaj](#)

<sup>2</sup> Napišite kratko vsebinsko poročilo, kjer boste predstavili raziskovalno hipotezo in opis raziskovanja. Navedite ključne ugotovitve, znanstvena spoznanja, rezultate in učinke raziskovalnega projekta in njihovo uporabo ter sodelovanje s tujimi partnerji. Največ 12.000 znakov vključno s presledki (približno dve strani, velikost pisave 11). [Nazaj](#)

<sup>3</sup> Realizacija raziskovalne hipoteze. Največ 3.000 znakov vključno s presledki (približno pol strani, velikost pisave 11) [Nazaj](#)

<sup>4</sup> V primeru bistvenih odstopanj in sprememb od predvidenega programa raziskovalnega projekta, kot je bil zapisan v predlogu raziskovalnega projekta oziroma v primeru sprememb, povečanja ali zmanjšanja sestave projektne skupine v zadnjem letu izvajanja projekta, napišite obrazložitev. V primeru, da sprememb ni bilo, to navedite. Največ 6.000 znakov vključno s presledki (približno ena stran, velikost pisave 11). [Nazaj](#)

<sup>5</sup> Navedite znanstvene dosežke, ki so nastali v okviru tega projekta. Raziskovalni dosežek iz obdobja izvajanja projekta (do oddaje zaključnega poročila) vpišete tako, da izpolnite COBISS kodo dosežka – sistem nato sam izpolni naslov objave, naziv, IF in srednjo vrednost revije, naziv FOS področja ter podatek, ali je dosežek uvrščen v A" ali A'. [Nazaj](#)

Navedite družbeno-ekonomske dosežke, ki so nastali v okviru tega projekta. Družbeno-ekonomski rezultat iz obdobja izvajanja projekta (do oddaje zaključnega poročila) vpišete tako, da izpolnite COBISS kodo dosežka – sistem nato sam izpolni naslov objave, naziv, IF in srednjo vrednost revije, naziv FOS področja ter podatek, ali je dosežek uvrščen v A" ali A'.

Družbeno-ekonomski dosežek je po svoji strukturi drugačen kot znanstveni dosežek. Povzetek znanstvenega dosežka je praviloma povzetek bibliografske enote (članka, knjige), v kateri je dosežek objavljen.

Povzetek družbeno-ekonomskega dosežka praviloma ni povzetek bibliografske enote, ki ta dosežek dokumentira, ker je dosežek sklop več rezultatov raziskovanja, ki je lahko dokumentiran v različnih bibliografskih enotah. COBISS ID zato ni enoznačen, izjemoma pa ga lahko tudi ni (npr. prehod mlajših sodelavcev v gospodarstvo na pomembnih raziskovalnih nalogah, ali ustanovitev podjetja kot rezultat projekta ... - v obeh primerih ni COBISS ID). [Nazaj](#)

<sup>7</sup> Navedite rezultate raziskovalnega projekta iz obdobja izvajanja projekta (do oddaje zaključnega poročila) v primeru, da katerega od rezultatov ni mogoče navesti v točkah 6 in 7 (npr. ni voden v sistemu COBISS). Največ 2.000 znakov, vključno s presledki. [Nazaj](#)

<sup>8</sup> Pomen raziskovalnih rezultatov za razvoj znanosti in za razvoj Slovenije bo objavljen na spletni strani: <http://sicris.izum.si/> za posamezen projekt, ki je predmet poročanja [Nazaj](#)

<sup>9</sup> Največ 4.000 znakov, vključno s presledki [Nazaj](#)

<sup>10</sup> Največ 4.000 znakov, vključno s presledki [Nazaj](#)

<sup>11</sup> Rubrike izpolnite / prepisite skladno z obrazcem "izjava sofinancerja" <http://www.arrs.gov.si/sl/progproj/rproj/gradivo/>, ki ga mora izpolniti sofinancer. Podpisan obrazec "Izjava sofinancerja" pridobi in hrani nosilna raziskovalna organizacija – izvajalka projekta. [Nazaj](#)

<sup>12</sup> Navedite en izjemni znanstveni dosežek in/ali en izjemni družbeno-ekonomski dosežek raziskovalnega projekta v letu 2014 (največ 1000 znakov, vključno s presledki). Za dosežek pripravite diapozitiv, ki vsebuje sliko ali drugo slikovno gradivo v zvezi z izjemnim dosežkom (velikost pisave najmanj 16, približno pol strani) in opis izjemnega dosežka (velikost pisave 12, približno pol strani). Diapozitiv/-a priložite kot priložko/-i k temu poročilu. Vzorec diapozitiva je objavljen na spletni strani ARRS <http://www.arrs.gov.si/sl/gradivo/>, predstavitev dosežkov za pretekla leta pa so objavljena na spletni strani <http://www.arrs.gov.si/sl/analize/dosez/>. [Nazaj](#)

Obrazec: ARRS-RPROJ-ZP/2015 v1.00a

10-6F-DD-5D-9A-B1-51-7B-A0-68-18-15-A8-B0-53-66-70-AA-DF-E3