

ОСВЕДОМЕНОСТ НА УЧЕНИТЕ И ПРАКТИКИ ЗА УПРАВЛЕНИЕ НА ИЗСЛЕДОВАТЕЛСКИ ДАННИ В БЪЛГАРСКИ УНИВЕРСИТЕТИ

Таня Тодорова, Магдалена Гарванова

Университет по библиотекознание и информационни технологии

Резюме: Целта на изследването е да се проучат осведомеността на учените и практиките за управление на изследователски данни в български университети. То е част от международно проучване на тема „Компетентност по управление на изследователски данни“, инициирано през 2016 г. В България анкетното изследване се провежда в периода 2017 – 2018 г. в УниБИТ, Химикотехнологичния и металургичен университет и Шуменския университет „Еп. Константин Преславски“. Общият брой на ефективно анкетираните респонденти е 108, като 57,4% от тях са представители на социалните и хуманитарните науки, а 42,6% – на природните и инженерните науки. Анализът откроява прилики и разлики в осведомеността и практиките на учените съобразно научната област.

Ключови думи: управление на изследователски данни, отворен достъп, висше образование, България.

Въведение

През януари 2021 г. стартира Българският портал за отворена наука (Bulgarian Open Science Cloud – BOSC) в съответствие с ангажимента на страната ни за участие в инициативата на Европейската комисия „Европейски облак за отворена наука“¹. Това е и важна цел в два актуални стратегически документа – „Национална стратегия за развитие на научните изследвания в Република България 2017 – 2030: По-добра наука за по-добра България“² и „Национална пътна карта на България за научноизследователска инфраструктура 2017 – 2023“³. Учените имат изключително значение за осъществяването на инициативата на ЕК за отворена наука. В представения Национален план за развитие на инициативата за отворена наука в Република България (2021 – 2025 г.)⁴ специално внимание се отделя на развитието на умения и компетентности на изследователския състав, свързани с методите на отворената наука, включително за управление на изследователски данни с отворен достъп. В този документ отворената наука се дефинира като нов подход към научния процес, основан на съвместна работа и нови начини за разпространение на знания чрез използване на цифрови

технологии и на нови инструменти за сътрудничество. „Отвореността“ се разбира като свободен и безплатен достъп до резултатите от публично финансирани научни изследвания, но и в по-широк аспект – като начин на извършване на цялостното научно изследване, включително споделянето на ценна научна информация на ранен етап в научноизследователския процес. Образователни програми с такава насоченост се препоръчва да залегнат като приоритет в стратегиите на институционално ниво и при подготовката на докторанти във висшите училища.

Методология на изследването

Компетентността на учените в областта на управлението на изследователските данни е от решаващо значение за постигане на качество и ефективно използване на изследователските данни при изпълнението на научноизследователски проекти, както и за дълготрайното съхраняване, споделяне и повторна употреба на данните. Това откроява актуалността на настоящото изследване, което *цели да проучи осведомеността на учените и практиките за управление на изследователски данни на преподавателския състав в български университети*. То е част от международно проучване на тема „Компетентност по управление на изследователски данни“, инициатирано през 2016 г. [1]. Управлението на изследователските данни (Research Data Management – RDM) е свързано с цялостната организация на данните: от тяхното включване в изследователския цикъл до разпространението и архивирането на получени завършени и стойностни резултати [2]. Състои се от редица процеси, като: планиране, проектиране, проучване и създаване на данни, дейности по съхраняване, опазване и сигурност на данните, както и по осигуряването на възможности за споделяне и повторно използване на създаваните изследователски данни [3].

В България онлайн базираното анкетно изследване се провежда в Университета по библиотекознание и информационни технологии (УниБИТ) през 2017 г. [4], а в периода март – май 2018 г. – в Химикотехнологичния и металургичен университет (ХТМУ) и в Шуменския университет „Еп. Константин Преславски“ (ШУ). Общият брой на ефективно анкетираните респонденти – преподаватели и докторанти, е 108.

Оригиналната анкетна карта „Компетентност по управление на изследователски данни“ (достъпна от: <https://inlitas.org/survey/index.php/493475?newtest=Y&lang=en>) съдържа 26 въпроса, свързани с: демографска информация; осведоменост по въпросите за управление на изследователски данни; практики за създаване,

съхраняване и обмен на данни; използване на метаданни; отворен достъп до изследователски данни; предпочитания и притеснения при споделянето на данни; дългосрочно съхраняване и публичен достъп до изследователски данни; институционални политики; наличие в университетите на план за управление на изследователски данни и на програми за обучение на преподавателския състав в областта на управлението на изследователските данни.

В предходна публикация са интерпретирани резултати и констатации от три групи въпроси, свързани със: сътрудничеството, обмена на данни между изследователите и практиките за отворен достъп; метаданните и използването на стандарти за метаданни; създаването и прилагането на план за управление на изследователски данни [5].

За целите на настоящия доклад се разглеждат резултатите от целево подбрани въпроси относно осведомеността на учените по управление на изследователски данни и относно съществуващите практики за управление на изследователски данни в научната работа на преподавателския състав. Анализът се стреми да открие прилики и разлики в осведомеността, практиките и възприятията на респондентите съобразно тяхната научна област – между представителите на социалните и хуманитарните науки (СХН – УниБИТ и ШУ) и на природните и инженерните науки (ПИН – ХТМУ).

IBM SPSS Statistics 19 е използван за обработката на данните, приложени са описателни статистически методи и хи-квадрат тест с установено ниво на значимост 0.05 между представителите на СХН и ПИН. Графиката е създадена с помощта на MATLAB.

Резултати и обобщения

1. Демографска информация

Общият брой на ефективно анкетираните респонденти е 108, като 57,4% от тях са представители на социалните и хуманитарните науки (СХН), а 42,6% – на природните и инженерните науки (ПИН). 72,2% са жени, а 26,9% са мъже. 1 респондент посочва опцията „Не желая да отговоря“. Преобладаващата част от участвалите в изследването са на възраст 36 – 45 години (32,4%). Следват респонденти от следните възрастови групи: 56 – 65 години (25,9%), 46 – 55 години (22,2%), 26 – 35 години (13,9%) и 3,7% от анкетираните са на възраст 65 години. 2 респонденти не отговарят на този въпрос. По отношение на изследователския опит, сред обхванатите участници в изследването преобладават тези с над 20-годишна научна и публикационна практика (28,7%), следвани от изследователи с над 5-годишен опит

(22,2%), 5 – 10 години (17,6%), 16 – 20 години (15,7%) и 11 – 15 години (14,8%), като 1 респондент посочва, че „не се е занимавал с изследователска дейност до момента“.

2. Изводи относно осведомеността на учените и практиките за управление на изследователски данни в български университети

Изследването има за цел да установи *какво е типичното поведение на изследователите, представители на социалните и хуманитарните науки (УниБИТ и ШУ) и на природните и инженерните науки (ХТМУ), в контекста на техните предпочитания и подходи за управление на изследователските данни в процеса на научна дейност*. Въпросите акумулират информация относно:

- типа файлове с данни и обема на данните, използвани при изследователската дейност;
- откъде учените вземат/набавят данни за своето изследване и как използват данните от външни източници;
- какъв тип файлове с данни и какъв обем от данни създават в резултат от изследователската си дейност;
- практики за съхранение на изследователските данни;
- практики относно използването на метаданни, на включването на съпроводителна/допълнителна информация към създаваните авторски изследователски данни в резултат от научните изследвания.

Резултатите показват, че в професионалната си дейност почти всички от анкетираните изследователи преобладаващо използват стандартни офис документи (текст, електронни таблици, презентации и др. под.) (90,7%), следвано от работа с интернет и уеббазирани данни (уебстраници, електронна поща, блогове, данни от социални мрежи и др. под.) (75,9%). Установяват се статистически значими разлики между представителите на СХН и тези на ПИН по отношение на предпочитанията към типа файлове с данни за изследователска работа. Изследователите от ПИН използват много по-интензивно, отколкото колегите си от СХН, графични файлови формати (JPEG, GIF, TIFF, PNG и др. под.), $\chi^2_{(1)} = 10.123$, $p = 0.001$; архивирани данни (ZIP, RAR, ZAR и др. под.), $\chi^2_{(1)} = 5.577$, $p = 0.018$; и софтуерни приложения (инструменти за моделиране, редактори, компилатори и др. под.) $\chi^2_{(1)} = 9.134$, $p = 0.003$. Респондентите посочват, че най-често обемът на данните, който използват при научно изследване, достига до GB (гигабайт) (57,4%), понякога до TB (терабайт) (3,7%).

Учените бяха поканени да споделят откъде обикновено вземат данни за своите изследователски проекти. Подходът на три четвърти от анкетираните (75,9%) е винаги да се набавят данни от множество

добре познати и надеждни източници. Следва предпочитанието да се работи с данни на собствения изследователски екип или научна група в университета заедно със създаването на нови данни за съответното изследване (54,6%) (вж. таблица 1). Установява се различие в нагласата на изследователите, представители на ПИН, спрямо тези на СХН – $\chi^2_{(1)} = 7.212$, $p = 0.007$, по отношение на по-активно споделяне и използване на данни на собствения изследователски екип или научна група от съответния университет.

Таблица 1. Предпочитани източници на данни за научните изследвания

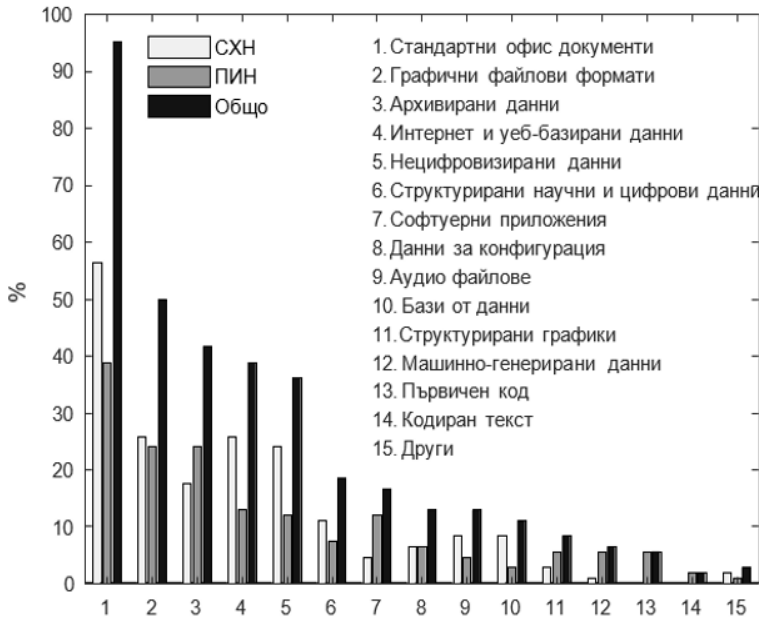
Откъде обикновено вземате данни за Вашите изследвания?	СХН		ПИН		Общо	
	N	%	N	%	N	%
Създавам нови данни.	31	28,7	28	25,9	59	54,6
От данни на собствен изследователски екип/научна група от съответния университет.	27	25	32	29,6	59	54,6
От данни на собствена изследователска мрежа (или лични/професионални контакти).	27	25	17	15,7	44	40,7
Винаги от един добре познат и надежден източник.	6	5,6	4	3,7	10	9,3
Винаги от множество добре познати и надеждни източници.	50	46,3	32	29,6	82	75,9
Търся данни от външни източници.	3	2,8	2	1,9	5	4,7

Всички изследователи, участници в проучването, заявяват, че винаги търсят и използват данни от външни източници. В тази си част проучването се стреми да установи каква е практиката на учените за работа с данни от външни източници. Резултатите показват, че 83,3% от изследователите отделят време, полагат усилия и модифицират данните от външни източници, за да ги направят използвани за целите на своя научен проект. Това е особено характерно за представителите на СХН – 50,9% (вж. таблица 2). Практики в изследователската дейност, при която данните от външни източници се използват „направо във вида, който имат, без никакви модификации и проблеми“ или чрез „полагане на малко усилия за редакция и/или модификации“, са в по-голяма степен характерни за представителите на ПИН. Статистически значими са разликите с другата група от СХН, съответно $\chi^2_{(1)} = 4.035$, $p = 0.045$ и $\chi^2_{(1)} = 4.031$, $p = 0.045$.

Таблица 2. Изследователски практики относно използването на данни от външни/други източници

Как обикновено използвате данните, които вземате от външни/други източници?	СХН		ПИН		Общо	
	N	%	N	%	N	%
Направо във вида, който имат, без никакви модификации и проблеми.	6	5,6	11	10,2	17	15,8
Полагам малко усилия за редакция и/или модификации.	10	9,3	15	13,9	25	23,2
Използвам данните, след като отделя време и полага усилия, за да ги направя използваеми за настоящия проект.	55	50,9	35	32,4	90	83,3

Група от въпроси в анкетната карта целят да установят *какъв тип файлове с данни и обем на данни* създават учените в резултат от своята изследователска дейност. Преобладаващият тип данни от изследванията са стандартни офис документи (текст, електронни таблици, презентации и др. под.) (95,4%), последвани от графични файлови формати (JPEG, GIF, TIFF, PNG и др. под.) (50%), архивирани данни (ZIP, RAR, ZAR и др. под.) (41,7%) (вж. фиг. 1). Логично произтича изводът, че представителите на ПИН, в по-голяма степен от колегите си от СХН, генерират като резултат от научната си дейност архивирани данни ($\chi^2_{(1)} = 7.275$, $p = 0.007$) и софтуерни приложения ($\chi^2_{(1)} = 7.755$, $p = 0.005$).



Фиг. 1. Типове файлове с данни – резултат от изследователска дейност

Респондентите заявяват, че най-често обемът на създадените изследователски данни е МВ (мегабайт) (52,8%), особено характерен за проектите в областта на социалните и хуманитарните науки. Съответно GB (гигабайт) е посочен от 44,4%, а ТВ (терабайт) – от 1,9% от учените.

Отговорите на въпроса „Къде обикновено съхранявате данните, които създавате като резултат от Вашата изследователска дейност?“ открояват предпочитанието на 92,7% от изследователите да архивират получените данни от научните изследвания на собствено устройство (персонален компютър, таблет, външна преносима памет и др. под.). Това е по-изявено като поведение при специалистите от СХН (57,9%) спрямо представителите на ПИН (39,3%), като се установява статистически значима разлика – $\chi^2_{(1)} = 5.599$, $p = 0.018$. Други практики за съхраняване на данни, като: съхраняване в облак (44,9%), сървъри или институционални хранилища на университета (24,3%) или във външни хранилища (11,2%), са в по-малка степен избор на представителите и от двете научни области.

Както е видно от таблица 3, само 13,9% от изследователите не прилагат съпроводителна/допълнителна информация, не

присъединяват метаданни към генерираните изследователски данни в процеса на научната дейност.

Таблица 3. Съпроводителна/допълнителна информация (метаданни) към изследователските данни

Прилагате ли съпроводителна/допълнителна информация (метаданни) към Вашите изследователски данни?	СХН		ПИН		Общо	
	N	%	N	%	N	%
Да, административна информация (например: автор, дата на създаване, име на файл, условия за достъп/ограничения и др. под.).	46	42,6	31	28,7	77	71,3
Да, специфична информация (например: автор/създател, финансиращ орган, наименование на проект, номер на проект, ключови думи и др. под.).	32	29,6	20	18,5	52	48,1
Да, техническа информация (например: файлов формат, размер на файл, софтуер/ хардуер, необходими за използване на данните и др. под.).	13	12	14	13	27	25
Да, описание на базата данни (например: методика, структура и описание на файловете, възможности за приложение и др. под.).	17	15,7	11	10,2	28	25,9
Не, аз не прилагам съпроводителна/допълнителна информация към създаваните изследователски данни.	6	5,6	9	8,3	15	13,9

Установява се, че 71,3% от учените имат практика да дават допълнителна информация с административен характер и относно правата и условията на достъп до данните. 48,1% от респондентите включват специфична информация за авторския колектив, проекта, финансиращия орган, ключови думи и др. под., което е по-характерно за представителите на СХН спрямо колегите им от ПИН. От всички анкетирани лица около 25% предоставят допълнителни описания и технически характеристики за базата данни и за нейните възможности за използване.

Изводи

Анализът на констатациите от проучването относно осведомеността и съществуващите практики за управление на изследователски данни на преподавателския състав, представители на социалните и хуманитарните науки (СХН – УниБИТ и ШУ) и на природните и инженерните науки (ПИН – ХТМУ), откроява някои специфични според принадлежността към съответната научна област, както следва:

- В научната си работа изследователите от ПИН използват и създават по-често спрямо представителите на СХН графични файлови формати, архивирани данни и софтуерни приложения;
- Изследователите от ПИН са по-отворени към споделяне на данни на собствения изследователски екип или на научна група от съответния университет;
- Близки са нагласите и поведението на изследователите от СХН и ПИН по отношение на интензитета и значението на използването на данни, получени от външни източници. Преобладаващата част от учените (83,3%) адаптират данните към спецификата на научния си проект. При изследователите от ПИН се практикува и директно използване или след малка обработка на данните, получавани от външни източници;
- Преобладаващият обем на създаваните изследователски данни в процеса на научноизследователски проект е МВ (мегабайт) (52,8%). Това е често срещан обем при проектите в областта на социалните и хуманитарните науки;
- Анкетираният преподавателски състав и докторанти от УниБИТ, ХТМУ и ШУ предпочитат да архивират получените данни от научните си изследвания на собствено устройство (97,2%), което е по-изявено при представителите на социалните и хуманитарните науки;
- С малки изключения (13,9%), изследователите присъединяват стандартни метаданни към генерираните изследователски данни от процеса на научна дейност. Преобладава включването на допълнителна административна информация – 71,3%, докато само при 25% се добавя и техническа информация. Сред учените от СХН е по-популярна практиката да се предоставя специфична информация за проекта, колектива, финансиращия орган и др. под., отколкото сред представителите на ПИН.

В предходна публикация бяха докладвани изводи за осведомеността на изследователския състав, участвал в проучването

от УниБИТ, ХТМУ и ШУ, относно отворения достъп, метаданните и използването на стандарти за метаданни и за прилагането на план за управление на изследователските данни [5].

Обобщено, проучването на тема „Компетентност по управление на изследователски данни“ установи незадоволително ниво на компетентност в областта на управлението на изследователските данни на анкетирания преподавателски състав. Откриват се пропуски в осведомеността относно: отворения достъп; добрите практики за описване, съхраняване, верифициране, разпространяване, използване и повторно използване на изследователските данни; метаданните и използването на стандарти за метаданни, както и относно създаването и прилагането на план за управление на изследователските данни, включително за дългосрочното съхраняване на данните. Резултатите показват някои различия във възприятията и практиките съобразно научната област на респондентите – между представителите на СХН (УниБИТ и ШУ) и на ПИН (ХТМУ), открити при съответния анализ.

Проведеното проучване доказва необходимост от създаването на програми за продължаващо обучение в областта на компетентността по управление на изследователски данни в университетите, както и за установяването на форми за предоставяне на консултации и подкрепа на изследователите, особено на младите учени, асистенти и докторанти. Предложените констатации ще бъдат полезни при подготовката на учебно съдържание и методология на обучение, съответстващи на научната област и специфичните изисквания на изследователите.

Изпълнението на Националния план за развитие на инициативата за отворена наука в Република България (2021 – 2025 г.) изисква надграждане на капацитета на научноизследователския състав в областта на управлението на изследователските данни и популяризиране на културата на споделяне в системата на висшето образование и науката.

Благодарности: Докладът е разработен по НИП „Синтез на динамичен модел за оценка на психологически и физически въздействия от прекомерна употреба на смарт технологии“, Договор № КП-06-Н 32/4/07.12.2019 с ФНИ на МОН, Конкурс за финансиране на фундаментални научни изследвания – 2019 г.

Бележки

¹ **Balgarskiyat** portal za otvorena nauka. <<https://bpos.bg/>> (2 juli 2021).

[**Българският** портал за отворена наука. <<https://bpos.bg/>> (5 юли 2021).]

² **Natsionalna** strategia za razvitie na nauchnite izsledvania v Republika Bulgaria 2017 – 2030: Po-dobra nauka za po-dobra Bulgaria. <<http://horizon2020.mon.bg/?h=downloadFile&fileId=4366>> (25 юли 2021).

[**Национална** стратегия за развитие на научните изследвания в Република България 2017 – 2030: По-добра наука за по-добра България. <<http://horizon2020.mon.bg/?h=downloadFile&fileId=4366>> (25 юли 2021).]

³ **Natsionalna** patna karta na Bulgaria za nauchnoizsledovatelaska infrastruktura 2017 – 2023. <<http://horizon2020.mon.bg/?h=downloadFile&fileId=438>> (5 юли 2021).

[**Национална** пътна карта на България за научноизследователска инфраструктура 2017 – 2023. <<http://horizon2020.mon.bg/?h=downloadFile&fileId=438>> (5 юли 2021).]

⁴ **Natsionalen** plan za razvitie na iniitsiativata za otvorena nauka v Republika Bulgaria (2021 – 2025). <[file:///C:/Users/UNIBIT/Downloads/plan-otvorena-nauka_130121%20\(1\).pdf](file:///C:/Users/UNIBIT/Downloads/plan-otvorena-nauka_130121%20(1).pdf)> (25 юли 2021).

[**Национален** план за развитие на инициативата за отворена наука в Република България (2021 – 2025). <[file:///C:/Users/UNIBIT/Downloads/plan-otvorena-nauka_130121%20\(1\).pdf](file:///C:/Users/UNIBIT/Downloads/plan-otvorena-nauka_130121%20(1).pdf)> (25 юли 2021).]

References/Литература

1. **Chowdhury, G., G. Walton, S. Kurbanoglu, Y. Ünal, J. Boustany.** Information practices for sustainability: Information, data and environmental literacy. The Fourth European Conference on Information Literacy (ECIL): Abstracts (S. Špiranec et al. eds.). Prague, Association of Libraries of Czech Universities, 2016, p. 22.
2. **Whyte, A., J. Tedds.** Making the Case for Research Data Management, DCC Briefing Papers. Edinburgh, Digital Curation Center, 2011, pp. 1 – 8.
3. **Cox, A. M., S. Pinfield.** Research data management and libraries: Current activities and future priorities. – In: *Journal of Librarianship and Information Science*, 46(4), 2017, pp. 299 – 316.
4. **Todorova, T., R. Krasteva, E. Tsvetkova.** Data Literacy and Research Data Management: The Case at ULSIT, Information Literacy in Everyday Life. Sixth European Conference on Information Literacy: Proceedings (S. Kurbanoglu et al. eds.). Cham: Springer International Publishing, 2018, pp. 535 – 544.
5. **Todorova, T., M. Garvanova.** Nablyudenia ot sravnitelno prouchvane na kompetentnostta po upravlenie na izsledovatelски данни v balgarski universiteti. Obshtestvoto na znaniето i humanizmat na XXI vek: XII natsionalna nauchna konferentsia s mezhdunarodno uchastie, 2019, s. 476 – 485.

[**Тодорова, Т., М. Гарванова.** Наблюдения от сравнително проучване на компетентността по управление на изследователски данни в български университети. Обществото на знанието и хуманизмът на XXI век: XII национална научна конференция с международно участие, 2019, с. 476 – 485.]

За авторите

Таня Тодорова е професор, доктор на науките, ръководител на катедра „Библиотечен мениджмънт и архивистика“ и заместник-ръководител на ЮНЕСКО катедра „ИКТ в библиотечните науки, образованието и културното наследство“ при УниБИТ.

За контакт с автора: t.todorova@unibit.bg

Магдалена Гарванова е доцент, преподавател в катедра „Информационни системи и технологии“ при УниБИТ.

За контакт с автора: m.garvanova@unibit.bg

RESEARCH DATA MANAGEMENT AWARENESS AND PRACTICES OF BULGARIAN UNIVERSITY RESEARCHERS

Tania Todorova, Magdalena Garvanova

University of Library Studies and Information Technologies

Abstract: The aim of this study is to examine the data literacy (DL) and research data management (RDM) awareness and practices of Bulgarian university researchers. In 2016 the international research group for implementation of the Research Data Management Survey (RDMS) were established. RDMS was conducted in three Bulgarian universities, namely: University of Library Studies and Information Technologies, University of Chemical Technology and Metallurgy and Konstantin Preslavsky University of Shumen. The general aggregation consists of 108 effectively surveyed lecturers and doctoral students. We aimed our analysis at obtaining information on the differences in practices and perceptions across the subject disciplines of respondents.

Keywords: research data management, higher education, open access, Bulgaria.

About the authors

Tania Todorova, Prof. D.Sc. is Head of the Library Management and Archival Studies Department and Deputy Head of UNESCO Chair “ICT at Library Science, Education and Cultural Heritage” at ULSIT.

To contact the author: t.todorova@unibit.bg

Magdalena Garvanova, Assoc. Prof. PhD is a lecturer in Information Systems and Technologies Department at ULSIT.

To contact the author: m.garvanova@unibit.bg