

ИНСТИТУЦИОНАЛНОТО СЪТРУДНИЧЕСТВО КАТО ФАКТОР ЗА ДИГИТАЛНАТА ТРАНСФОРМАЦИЯ НА НАРОДНА БИБЛИОТЕКА „ИВАН ВАЗОВ“ – ПЛОВДИВ

Иван Крачанов, Димитър Минев
Народна библиотека „Иван Вазов“ – Пловдив

Резюме: Настоящият текст разглежда развитието на дигиталните ресурси и услуги на Народна библиотека „Иван Вазов“ – Пловдив в контекста на съвременните технологични и инфраструктурни промени в библиотечния сектор. Акцент е поставен върху изграждането и усъвършенстването на дигиталната библиотека на институцията, базирана на платформата CultIS, както и върху стремежа за по-висока достъпност, свързаност и по-ефективно управление на дигитализираното културно наследство. Разгледани са решения като пълнотекстово търсене, обработка на OCR текстове и оптимизиране на административните процеси.

Особено внимание е отделено на участието на библиотеката в проекта КЛАДА-БГ, който създава условия за сътрудничество с технологични и академични партньори в областта на езиковите технологии и изкуствения интелект. Анализирани са възможностите за подобряване на разпознаването и нормализацията на исторически текстове, за семантично обогатяване на съдържанието и за разширяване на услугите към потребителите. В този контекст е представено и значението на включването на библиотеката в проекта PressMint, чрез който българоезичните колекции от историческа периодика могат да се интегрират в общоевропейска изследователска среда. Така дигитализацията се очертава не само като средство за опазване, но и като стратегически инструмент за по-широка видимост, научна употреба и устойчиво присъствие на българското културно наследство в цифровата епоха.

Ключови думи: дигитализация; дигитална библиотека; културно наследство; изкуствен интелект; историческа периодика

ВЪВЕДЕНИЕ

През последните години развитието на дигиталните ресурси в Народна библиотека „Иван Вазов“ – Пловдив (НБИВ) се очертава като последователен преход от натрупване на дигитални ресурси към изграждане на среда за търсене, обработка, свързване и повторна употреба на предоставената информация. Този преход е обусловен както от вътрешните потребности на библиотеката да осигури по-удобен достъп до ценни и труднодостъпни фондове, така и от включването ѝ в по-широки национални и европейски инфраструктури. В този контекст дигитализацията не следва да се разбира единствено като консервационна дейност, а като процес по създаване на използвани цифрови ресурси, които могат да бъдат индексирани, анализирани и интегрирани в среди за приложна и научна работа (Luchev et al., 2021; Wilkinson et al., 2016).

Съществена роля за тази трансформация има участието на библиотеката в проектите PressMint (CLARIN, 2025a) и КЛаДА-БГ – национална изследователска инфраструктура, интегрирана в рамките на европейските инициативи CLARIN и DARIAH (CLaDA-BG, 2025). Чрез тези партньорства в НБИВ се установява специфичен модел на работа, взаимноизгодно партньорство, в което културни институции и технологични центрове си сътрудничат, обменят съдържание и експертиза и заедно създават услуги и инструменти. От една страна, библиотеката предоставя дигитализирани текстове и знание, а от друга, научните партньори разработват решения за по-ефективно оптично разпознаване на символи (OCR), езикова обработка, създаване и управление на метаданни, семантично обогатяване и др. Така библиотечната колекция започва да функционира не само като дигитален архив, но и като база за научна дейност и намира приложения в областта на обработката на естествен език.

Целта на доклада е да анализира как междуинституционалното проектно сътрудничество подпомага прехода в библиотеката от традиционна дигитализация към изграждане на колекции от дигитални ресурси, които са машинно обработваеми и оперативно съвместими в глобален мащаб. В този контекст изследването си поставя два основни въпроса:

1. Какви конкретни инфраструктурни, технологични и организационни резултати произтичат от участието на библиотеката в партньорства като КЛаДА-БГ?

2. Как участието в проекта PressMint разширява вече постигнатите резултати към по-високо равнище на корпусна обработка, семантично обогатяване и научна видимост на европейско равнище и какви са позите от тези нови перспективи за библиотеката?

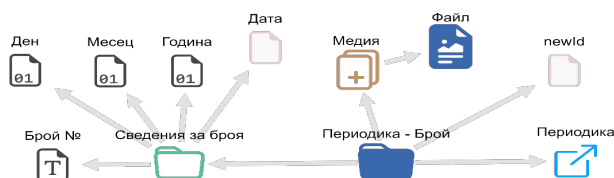
МЕТОДОЛОГИЯ

Методологическият дизайн на изследването е качествен и се основава на подхода „case study approach“ (анализ на казус). Като обект на анализ е избран конкретен институционален случай, чрез който се изследват изследователските въпроси/хипотези. Предмет на анализа са дигиталната библиотека на НБИВ и участието в проектите КЛаДА-БГ и PressMint. Проследява се развитието на дигитализацията в конкретна институция в нейния реален организационен и технологичен контекст. Предмет на анализа са процесите, чрез които дигитализираните културно-исторически ресурси се превръщат от цифрови копия в съвместими, машинно обработваеми и потенциално повторно използваеми данни. Документалният подход се прилага при разглеждането на проектна информация, функционални описания, публикации, технически решения и данни за дигитализираните ресурси. Примерите са подбрани по три критерия: пряка връзка с междуинституционално сътрудничество, наличие на конкретен технологичен или организационен резултат и значимост за достъпа, търсенето, обработката или повторната употреба на дигитализираното културно наследство.

РЕЗУЛТАТИ

Инфраструктурни резултати: КЛАДА-БГ, CultIS и изграждане на дигитална библиотека на НБИВ

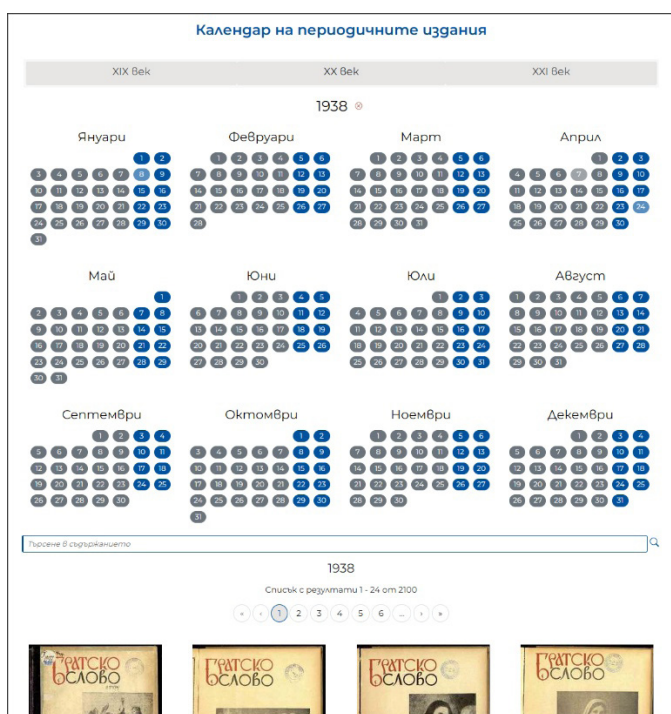
В институционален план един от най-видимите резултати от тази линия на развитие е новата Дигитална библиотека на НБИВ, изградена върху модифицирана версия на платформата CultIS (Goynov et al., 2024). Нейната значимост се състои в обединяването на няколко изисквания: представяне на разнообразни по тип дигитални обекти, пълнотекстово търсене в наличните ресурси, двуезична среда, различни нива на потребителски и административен достъп, възможност за по-ефективно управление на метаданните, оптимизирана и бързо работеща среда, визуално привлекателен интерфейс. Платформата е модулна и позволява адаптиране към спецификата на културното съдържание, което е особено важно за библиотеки, работещи едновременно с книги, периодика, ръкописи, изображения и аудио-визуални материали (фиг. 1).



Фиг. 1. Модул „Периодика – Брой“. Специфичното тук е, че при добавяне на нов брой е необходимо само да се въведат номерът на броя, датата на издаване и от спускащо меню да се избере заглавието на периодичното издание. Всички останали метаданни, свързани със заглавието, са предварително въведени еднократно в друг модул – „Периодика“. По този начин се елиминира нуждата от въвеждане на пълен набор от метаданни за всеки отделен брой (Goynov et al., 2024).

Потребителската част на системата е насочена към по-добра ориентация, висока скорост на извеждане на резултати и по-нисък праг на достъп до ресурсите. Важен акцент е съчетаването на визуална яснота и функционалност, тъй като качеството на уебсредата влияе върху начина, по който потребителите оценяват надеждността и полезността на дигиталната услуга (Flavián et al., 2009). В този смисъл решения като ясно структурирана начална страница, адаптивен изглед за различни устройства, удобен преглед на PDF ресурси, тъмна/светла основна тема, възможност за преглеждане на ресурсите в цвят или в скала на сивото целят да направят използването на наличните ресурси по-удобно и същевременно да има определена степен на персонализация според желанието на потребителите. Особено важна е възможността периодиката да бъде разглеждана не само чрез списъци и филтри, а и по хронологичен модел, чрез календар (фиг. 2), който съответства по-добре на логиката на историческите проучвания.

Също толкова съществено е, че административната част на системата е проектирана така, че да намали ръчния труд при добавяне на ново съдържание. Повторното използване на вече въведени данни, автоматичното попълване на свързани полета, историята на редакциите, проследимостта на промените и помощните функции за транслитерация и превод подпомагат администраторите при управлението на дигиталните обекти. Подобни решения имат непосредствен ефект както върху производителността, така и върху качеството и консистентността на метаданните. Значение има и стремежът към по-голяма инфраструктурна автономия чрез хостване на собствен сървър и бекъп на данните в партньорската среда на проекта КЛАДА-БГ, тъй като устойчивостта на дигиталната библиотека зависи не само от софтуера, но и от модела на съхранение и контрол върху цифровите активи (Goynov et al., 2024).



Фиг. 2. Календар на периодичните издания
(National Library Ivan Vazov, 2026)

Технологични резултати: OCR, речникови ресурси и търсене по стари правописни форми

Централният технологичен проблем, който свързва текущите три главни линии на развитие – локалната Дигитална библиотека, проектите КЛАДА-БГ и PressMint, са обработката, разпространението и употребата на исторически

текстове. В случая на НБИВ значителна част от ценните материали са създадени преди правописната реформа от 1945 г., съдържат остарели буквени символи, словоформи и се характеризират с типографски особености, които затрудняват OCR (Kratchanov et al., 2020). За да стане един документ по-използваем в цифрова среда, е необходимо текстът да бъде достатъчно точно разпознат, също така да бъде индексирани и поставен в контекст, който допуска откриване и анализ. При масовата дигитализация ръчната корекция на грешки при OCR е с проблематично приложение като основен метод, тъй като изисква значителен човешки ресурс и забавя публикуването на ново съдържание (Balk, Ploeger, 2009; Jain et al., 2021). Поради това НБИВ развива модел, основан на баланс между автоматизация и експертен контрол, който цели максимално качествен резултат без твърде голямо забавяне на процеса от сканирането на оригинала до превръщането му в дигитален обект.

В рамките на проекта КЛАДА-БГ и сътрудничеството с Института по информационни и комуникационни технологии на БАН (ИИКТ-БАН) се разработват речникови и езикови ресурси, които подпомагат разпознаването на стари български текстове и улесняват търсенето в различни правописни варианти (Kratchanov et al., 2020; Simov, 2019). Пример за работата в тази насока е разработеният речник CLADABG-MODEL – морфологичен ресурс, който включва словоформи от основния речников състав на българския език, преобразувани според по-стари правописни норми, като при създаването му са отчетени буквени замени, позиционен контекст и граматични характеристики и неговата „състарена“ версия съдържа 1 121 872 словоформи. Подобряването на точността на OCR е важна, защото резултантният текст се счита за начална стъпка към автоматизация и усъвършенстване на методите за откриване на информация, анотация и извличане на знание (Hinze et al., 2016). Друг пример е подпомагането на търсещата машина на Дигиталната библиотека с речник от състарени словоформи, чрез който потребителят може да достигне до релевантни резултати, без да познава историческите правописни норми. Ако се потърси думата „вяра“, ще бъдат изведени резултати, които съдържат както „вяра“, така и остарялата словоформа „вѣра“. Създават се методи за автоматично нормализиране на остарели текстове, които създават условия върху историческите документи да бъдат прилагани езикови инструменти, разработени за съвременния български език (Simov et al., 2025).

Взаимодействието между библиотеките и технологичните институти във връзка с усъвършенстването на езиковите инструменти е двупосочно и взаимно обогатяващо. От една страна, технологичните партньори предоставят на библиотеката нови възможности за подобрен достъп до културното наследство, семантично обогатяване, по-ефективно пълнотекстово търсене и управление на дигиталните колекции. От друга страна, библиотеката може да предостави автентичен и богат езиков и културно-исторически материал, който е необходим за обучението, тестването и усъвършенстването на модели за обработка на естествен език. В този смисъл библиотеката не е само ползвател на нови технологии, а и активен участник в тяхното развитие.

Организационни резултати: нов модел на сътрудничество между НБИБ и технологичните партньори

Естествено продължение на развитието на дигиталната екосистема на НБИБ е участието на библиотеката в проекта PressMint – инициатива на CLARIN за изграждане на многоезикови, съпоставими и оперативно съвместими корпуси от исторически вестници, в която библиотеката участва заедно с ИИКТ-БАН. В проекта се включват институции от общо 17 държави. Значимостта му за НБИБ е стратегическа, защото измества фокуса от локалното публикуване към включване в общеевропейска среда за обмен на данни, инструменти и изследователски практики. В подобен контекст дигитализираната периодика се превръща в свързан, анотиран корпус, компонентите на който могат да бъдат сравнявани, превеждани, използвани за изследвания в дигиталната хуманитаристика, лингвистиката и др. (CLARIN, 2025a).

PressMint е особено важен и поради това, че поставя работата с исторически вестници в рамката на FAIR принципите. За културното наследство това означава подготовка на данните така, че да бъдат откриваеми, достъпни, оперативно съвместими и повторно използваеми (Hermon, Niccolucci, 2021; Wilkinson et al., 2016). PressMint стъпва върху вече натрупаната методологическа и технологична експертиза от по-ранни проекти на CLARIN, като особено важен ориентир е ParlaMint (CLARIN, 2025b) – проект, насочен към изграждането на стандартизирани, взаимно съпоставими и анотирани корпуси от парламентарни дебати, главно от европейски държави. Данните в PressMint са организирани по стандарта TEI и следват специализирания модел на ParlaMint, което осигурява висока степен на оперативна съвместимост на данните и устойчивост при дългосрочното им използване.

В рамките на PressMint приносят на НБИБ е в няколко взаимносвързани направления. Библиотеката предвижда включване на значителна част от своята дигитализирана историческа периодика (около 150 хил. страници), като се цели представителност на регионалния и националния печат. От библиотеката се очаква да осигури машинно четим текст с възможно най-високо качество за предоставения масив историческа периодика. Съвместно с ИИКТ-БАН се предвижда създаването на инструменти за обучение и оценка на модели за корекция на текста след OCR, включително чрез двойки от некоригиран и ръчно коригиран текст. Именно такива паралелни набори са особено ценни за разработване на по-точни автоматични решения, защото позволяват моделите да се учат от реални, а не изкуствено симулирани грешки (Booth et al., 2024). Паралелно с работата по корекция на OCR се създават инструменти за автоматизирано сегментиране на текстовите полета според тяхното предназначение (заглавни редове, статии, реклами и др.). Предвидено е създаването на нормализирани версии на старите текстове, така че корпусът да бъде по-лесно използваем както за подобро откриване на информация, така и за автоматичен езиков анализ и превод. Включено е също обогатяването на документите с метаданни, ключови думи и именуванни същности, които ги правят по-добре свързани в мрежи от знание и с други корпуси.

ДИСКУСИЯ

Представените резултати показват, че дигиталната трансформация на НБИБ разчита на нещо повече от натрупване на дигитални копия или внедряване на отделна софтуерна платформа. По-същественото е постепенното формиране на институционален модел, при който дигитализацията акцентира върху съвместимостта на ресурсите, пълнотекстово търсене, езикова обработка, инфраструктурна устойчивост и възможност за повторна научна употреба на съдържанието. В този смисъл участието в проектни партньорства като КЛаДА-БГ не е само външен източник на технологична подкрепа, а механизъм за преход от традиционна дигитална колекция към по-сложна екосистема от ресурси, експертни знания и споделени стандарти.

Важен е акцентът за трансформацията на вече постигнатите резултати в бъдещи възможности. Към първата група могат да се отнесат изграждането и усъвършенстването на Дигиталната библиотека, въвеждането на функционалности за по-ефективно администриране на периодични издания, пълнотекстовото търсене, разширените възможности за преглед и достъп до дигиталните обекти.

Втората група включва дейностите, свързани с проекта PressMint, които следва да се разглеждат преди всичко като следващ етап от развитието на процесите по дигитализация в библиотеката. Значението на PressMint се състои в това, че поставя локално натрупаните дигитални ресурси в по-широка рамка на междуинституционално сътрудничество в европейски контекст, като акцентът е върху корпусната обработка, оперативна съвместимост и международна научна употреба. Поради това участието на библиотеката в подобна инициатива изисква не само предоставяне на дигитализирано съдържание, но и активно участие в процеси по оценка на качеството, подготовка на текстови данни, описание на колекциите и съгласуване със стандарти за корпусно представяне.

От институционална гледна точка най-същественият извод е, че проектното сътрудничество променя ролята на библиотеката. Тя не функционира единствено като пазител и доставчик на съдържание, а постепенно се включва като партньор в създаването на езикови ресурси, данни, корпуси и инструменти за анализ. Тази промяна е тенденция в развитие и нейната устойчивост зависи от няколко условия: наличие на техническа инфраструктура, поддържане на експертен капацитет, ясни стандарти за съвместимост, дългосрочно съхранение на данните и продължаващо сътрудничество с академични и технологични партньори.

Ограничение на настоящия анализ е, че той разглежда един институционален случай и не претендира за универсална валидност за всички библиотеки. Изводите са приложими преди всичко към институции, които разполагат със значими дигитализирани колекции, имат потребност от пълнотекстово търсене и могат да участват в изследователски инфраструктури. За по-нататъшно потвърждаване на резултатите биха били полезни допълнителни емпирични данни: статистика за използването на дигиталната библиотека, анализ на потребителското поведение, както и оценка на ефективността на бъдещите инструменти за пост-OCR корекция и семантично обогатяване. Така настоящият доклад може да послужи като ос-

нова за следващ етап на изследване, в който институционалният анализ да бъде допълнен с количествени показатели за ефективност, качество и научна употреба на дигиталните ресурси.

ЗАКЛЮЧЕНИЕ

Разгледаният институционален случай показва, че дигиталната трансформация на НБИБ се развива като последователен процес, в който дигитализацията, технологичната инфраструктура, езиковата обработка и проектното сътрудничество се свързват в обща стратегическа посока. Изграждането на Дигиталната библиотека, използването на платформата CultIS, участието в КЛаДА-БГ и включването в PressMint очертават преход от създаване и публикуване на дигитални копия към формиране на комплексна среда за обработка, обогатяване и повторна употреба на културно-историческо съдържание.

Основният извод е, че междуинституционалното сътрудничество има съществена роля за развитието на библиотечните дигитални ресурси. То осигурява достъп до технологична експертиза, стимулира въвеждането на стандарти за съвместимост, подобряване на достъпността на дигиталните копия и метаданните и създава условия за включване на местните колекции в национални и европейски изследователски инфраструктури.

Научният принос на доклада се състои в систематизирането на проектното сътрудничество като фактор за дигитална трансформация в библиотечна среда. Чрез анализа на конкретен институционален случай се показва как дигитализацията може да надхвърли функцията на опазване и онлайн достъп и да се развие в модел, ориентиран към съвместимост в глобален контекст, с акцент върху пълнотекстовото търсене, езиковата обработка, семантичното обогатяване и корпусната употреба на исторически текстове. Докладът откроява връзката между инфраструктурните решения, езиково-технологичните ресурси и международните изследователски мрежи като условие за по-висока видимост, устойчивост и повторна научна употреба на българското културно наследство.

Благодарности: Изследването, представено в този доклад/работа, е извършено в рамките на CLaDA-BG, българската национална изследователска инфраструктура за ресурси и технологии за езиково, културно и историческо наследство, интегрирана в рамките на CLARIN EU и DARIAH EU, финансирана от Министерството на образованието и науката на Република България (подкрепа за българската Национална пътна карта за научноизследователска инфраструктура).

REFERENCES

- Balk, H., Ploeger, L. (2009). IMPACT: Working together to address the challenges involving mass digitization of historical printed text. *OCLC Systems & Services: International Digital Library Perspectives*, 25(4), 233–248. <https://doi.org/10.1108/10650750911001824>
- Booth, C. W., Thomas, A., Gaizauskas, R. (2024). BLN600: A parallel corpus of

- machine/human transcribed nineteenth century newspaper texts. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 2440–2446).
- CLaDA-BG. (2025). *E-infrastruktura za balgarskoto ezikovo i kulturno nasledstvo* [E-infrastructure for Bulgarian language and cultural heritage]. <https://clada-bg.eu>
- CLARIN. (2025a). *PressMint: Interoperable corpora of historical newspapers*. <https://www.clarin.eu/pressmint>
- CLARIN. (2025b). *ParlaMint: Comparable and Interoperable Parliamentary Corpora*. 2025. <https://www.clarin.eu/parlamint>
- Flavián, C., Gurrea, R., Orús, C. (2009). Web design: A key factor for the website success. *Journal of Systems and Information Technology*, 11(2), 168–184. <https://doi.org/10.1108/13287260910955129>
- Goynov, M., Luchev, D., Paneva-Marinova, D., Senka, G., et al. (2024). *CultIS: Web-based platform for intelligent cultural content management*. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 14, 19–36. <https://doi.org/10.55630/dipp.2024.14.1>
- Hermon, S., Niccolucci, F. (2021). FAIR data and cultural heritage: Special issue editorial note. *International Journal on Digital Libraries*, 22(3), 251–255. <https://doi.org/10.1007/s00799-021-00309-8>
- Hinze, A., Bainbridge, D., Cunningham, S. J., Downie, J. S. (2016). Low-cost semantic enhancement to digital library metadata and indexing: Simple yet effective strategies. In *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries* (pp. 93–102). <https://doi.org/10.1145/2910896.2910910>
- Jain, P., Taneja, K., Taneja, H. (2021). Which OCR toolset is good and why: A comparative study. *Kuwait Journal of Science*, 48(2). <https://doi.org/10.48129/kjs.v48i2.9589>
- Kratchanov, I., Laskova, L., Simov, K. (2020). Towards improving OCR accuracy with Bulgarian language resources. *DHNB Publications*, 3(3), 115–123. <https://doi.org/10.5617/dhnbpub.11226>
- Luchev, D., Goynov, M., Paneva-Marinova, D., Stoykov, J., et al. (2021). Synergy of national cultural heritage and technology. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 11, 281–286. <https://doi.org/10.55630/dipp.2021.11.26>
- National Library Ivan Vazov. (2026). *Digitalna biblioteka: Kalendar na periodichnite izdania* [Digital Library: Periodical Issues Calendar]. <https://digital.libplovdiv.com/bg/calendar?page=1&year=1938>
- Simov, K. (2019). Integrated language and knowledge resources for a Bulgarian-centric knowledge graph. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 9, 65–74. <https://doi.org/10.55630/dipp.2019.9.5>
- Simov, K., Paev, N., Osenova, P. (2025). LLM-based Models for Transforming of Diachronic Bulgarian Spelling to Contemporary Bulgarian. In *Proceedings of CLARIN Annual Conference 2025 (CLARIN2025)*, Vienna, 2025. <https://www.>

clarin.eu/sites/default/files/CLARIN2025_ConferenceProceedings.pdf
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. <https://doi.org/10.1038/sdata.2016.18>

INSTITUTIONAL COOPERATION AS A FACTOR IN THE DIGITAL TRANSFORMATION OF THE IVAN VAZOV NATIONAL LIBRARY IN PLOVDIV

Abstract: *The present text examines the development of the digital resources and services of the Ivan Vazov National Library in Plovdiv (IVNL) in the context of contemporary technological and infrastructural changes in the library sector. Particular emphasis is placed on the creation and continuous improvement of the institution's Digital Library, based on the CultIS platform, as well as on the pursuit of greater accessibility, connectivity, and more efficient management of digitized cultural heritage. The discussion addresses solutions such as full-text search, the processing of OCR-generated texts and the optimization of administrative workflows.*

Special attention is given to the library's participation in the CLaDA-BG project, which creates conditions for cooperation with technological and academic partners in the fields of language technologies and artificial intelligence. The text analyzes the possibilities for improving the recognition and normalization of historical texts, semantically enriching content, and expanding user-oriented services. In this context, it also outlines the significance of the library's involvement in the PressMint project, through which Bulgarian-language collections of historical periodicals can be integrated into a pan-European research environment. In this way, digitization emerges not only as a means of preservation, but also as a strategic instrument for broader visibility, scholarly use, and the sustainable presence of Bulgarian cultural heritage in the digital age.

Keywords: digitization; digital library; cultural heritage; artificial intelligence; historical periodicals

Ivan Kratchanov, PhD Student

ORCID: 0000-0002-0430-7953

Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences Mathematical Linguistics
Sofia, Bulgaria

Head of Digitization Centre
Ivan Vazov National Library – Plovdiv
Plovdiv, Bulgaria
E-mail: digitization@libplovdiv.com

Dimitar Minev, Director

Ivan Vazov National Library – Plovdiv
Plovdiv, Bulgaria
E-mail: dimin@libplovdiv.com