

**Техническо задание за създаване на  
автоматизирана система и национален портал  
(включително стандарти и спецификации) за  
отворен достъп до научна информация**

## Съдържание

1. Визия .....	3
2. Основни изисквания към автоматизирана система и национален портал за отворен достъп до научна информация .....	4
3. Архитектура на процесите в системата .....	8
4. Системни изисквания .....	9
5. Възможни приложения на системата .....	10
6. План етапи за разработване и внедряване .....	10
7. Досегашен опит при изграждане на подобни системи.....	10

## 1. Визия

Отвореният достъп до научна информация е от съществено значение за разпространението на научното знание в Европа<sup>1</sup>. Той има отражение върху науката, индустрията и обществото и е важен елемент за по-ефективното използване на вече съществуващи знания с научна стойност. Резултатите от научните изследвания - както публикациите, така и събраните данни, трябва да бъдат широко и бързо разпространявани чрез използването на електронни медии, платформи и бази данни. Това решение бе взето от Съвета на Европейския съюз още през 2016 година, като до края на 2018 година всяка държава-членка следва да разработи конкретен план и да докладва докъде е стигнала в изпълнението на тази задача. Отворената наука и в частност Европейският облак за отворена наука бяха водещ приоритет на Българското председателство, като основните принципи, по които бе постигната договореност, са обединяването на национални инфраструктури (репозиториуми, библиотеки и архиви) и възможност за създаване на единни стандарти за качване и ползване на данните.

През 2014 г. Министерството на образованието и науката разработи концепция за отворена наука<sup>2</sup> с цел прилагането на принципа на отворен достъп до научната информация, финансиран с публични средства и препоръки за постигане на набелязаната цел, между които:

- националната политика за отворен достъп да бъде формулирана въз основа на хибридният модел, при който качеството се осигурява от научните издания;
- да се обмисли договарянето на възнаграждението на авторите при „златния“ отворен достъп<sup>3</sup>, както и прозрачност при заплащането на услугите;
- да се създаде услуга за дълготрайно съхранение, която ще гарантира, че дигиталните издания могат да бъдат прочетени и използвани за дълъг период от време;
- да се планира как данните могат да бъдат архивирани и да се гарантира достъп до тях.

Най-важните стъпки в този план за действие са стартирането на технологична платформа за обмен на знания и популяризирането на научната продукция на българската

---

<sup>1</sup> <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access>

<sup>2</sup> [https://www.mon.bg/upload/4017/conception\\_OpenAccess\\_2rev.pdf](https://www.mon.bg/upload/4017/conception_OpenAccess_2rev.pdf)

<sup>3</sup> <https://www.springer.com/gp/authors-editors/authorandreviewertutorials/open-access/what-is-open-access/10286522>

научноизследователска общност, чрез създаване на електронно хранилище за търсене със свободен достъп и попълване на отворено съдържание за изследователски проекти с българско участие, техни научни резултати, публикации и експериментални данни.

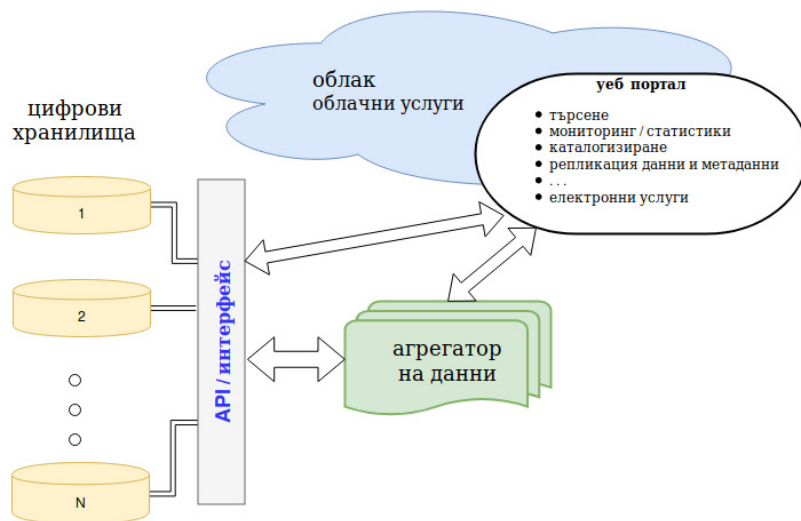
Това налага разработването на единни стандарти, които да осигуряват оперативна съвместимост на бъдещия национален портал с аналогичния Европейски портал OpenAIRE<sup>4</sup> за отворен достъп до съдържание на проекти по Рамковата програма на ЕС за научни изследвания и иновации, както и съвместимост между портала и базата от данни с аналитичните инструменти от националния абонамент за научни бази от данни на МОН.

## 2. Основни изисквания към автоматизирана система и национален портал за отворен достъп до научна информация

Системата да бъде изградена със стандартизиран програмен интерфейс, който да позволява изграждането на мрежа от цифрови хранилища за различните организации, университети, институции, свързани през API интерфейси. Това ще осигурява възможност за агрегиране на данни и изграждане на облачни услуги, уеб портал с множество електронни услуги, работещи с всички цифрови хранилища. Чрез наличие на глобална система за индексирание, търсене и извличане на данни с една заявка от потребител се обработват множество хранилища, както и други услуги за съхранение на големи данни. Обработката, статистиките, виртуализацията, данните и метаданните да не са с фиксирано физическо местоположение и да се намират физически на различни места – фигура 1.

---

<sup>4</sup> <https://www.openaire.eu/>



**Фиг. 1. Схема на системата**

Системата да се състои от база от данни, система за управление на съдържание и достъп, уеб-интерфейс. Архитектурата да включва определен модел на данните с адаптивни схеми за метаданни, за работните процеси и функции за преглед/търсене.

Всяка организация поддържа, обновява и администрира своя архив върху компютри в самата организация. Националният портал се поддържа и администрира от организацията, отговаряща за цялата система.

Автоматизираната система като софтуерна платформа трябва да осигури на всяка една организация следните **базови функционалности**:

- архивиране на данни без ограничения за файлов формат;
- създаване и поддържане на база от метаданни;
- управление на достъпа, отчетност, статистики;
- организиране и управление на съдържанието според структурата на организацията по различните звена, секции, отдели, лаборатории, проекти и пр.;
- навигация и търсене в данни и метаданни;
- потребителски интерфейс (уеб базиран);
- програмен/машинен интерфейс за обработка или обмен на голям обем от данни;
- интерфейс за оперативна съвместимост с OpenAIRE<sup>5</sup> и EOSC<sup>6</sup>.

<sup>5</sup> <https://www.openaire.eu/>

<sup>6</sup> <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

Автоматизираната система като софтуерна платформа трябва да осигури на всяка една организация следните **допълнителни функционалности**:

- **вградени работни схеми**: първоначално разработени за цифрови библиотеки, вграден модел на данните и работни процеси за одобрение стандартно прилагани при библиотеки и цифрови архиви;
- **платформа за търсене**: позволява филтриране и разглеждане на всички публикувани цифрови обекти. Освен търсене по метаданни да има възможност за търсене и в текстово съдържание на общоприетите текстови файлови формати;
- **фасетъчно търсене**: позволява навигиране в многомерното информационно пространство чрез комбинирането на текстово търсене с постепенното стесняване на избора в всяко измерение;
- **без ограничения за файловете типове**: системата да съхранява файлове от всякакъв тип. Освен това тя да може автоматично да разпознава файлове в най-разпространените формати, например DOC, PDF, XLS, PPT, JPEG, MPEG, TIFF и др.;
- **метаданни**: по подразбиране да се използва схема на метаданните на базата на Dublin Core (QDC)<sup>7</sup>. Да предоставя възможности за разширение на тази основна схема или добавяне на потребителски QDC-подобни схеми. Системата да има възможност за импорт или експорт на метаданни от други основни схеми за метаданни, като MARC<sup>8</sup> или MODS<sup>9</sup>;
- **инструменти/добавки**: в архива да се предлагат набори от инструменти (пакетна обработка, за експорт, за редактиране на метаданни и т.н.);
- **сигурност**: достъп чрез собствена вградена система за автентификация /оторизация, но също така да може да се интегрира и с други системи, като LDAP<sup>10</sup> или Shibboleth<sup>11</sup>;
- **права и нива на достъп**: системата да позволява управление на правата за четене и промяна на ниво сайт, общност, колекции, публикуван обект и файл. Също така да има възможност да се делегират административни права поотделно за всяка общност или колекция. Контрол на достъпа по IP/интернет адреси и/или мрежи;

---

<sup>7</sup> <http://dublincore.org/>

<sup>8</sup> <http://www.loc.gov/standards/marcxml/>

<sup>9</sup> <http://www.loc.gov/standards/mods/>

<sup>10</sup> <https://searchmobilecomputing.techtarget.com/definition/LDAP>

<sup>11</sup> <https://www.shibboleth.net/>

- **възстановяване:** системата да позволява експортиране на съдържанието - резервни копия от файлове в AIP формат за възстановяване на цялото съдържание или на отделни общности, колекции и елементи;
- **стандарт:** системата да отговаря на стандартните протоколи и указания за най-добри практики за отворен достъп, оперативна съвместимост, импорт и експорт. OAI-PMH<sup>12</sup>/SWORD<sup>13</sup>/ OpenAIRE<sup>14</sup>;
- **интерфейси:** системата да предоставя интерфейси API в съответствие със съвременните уеб стандарти;
- **статистика:** системата да разполага с модули за статистика и анализ как се използват данните от цифровото хранилище. Няма да се използват лични данни на посетители и/или потребители в системата;
- конфигурируема **база от данни:** организациите да могат да избират PostgreSQL<sup>15</sup> или Oracle<sup>16</sup> за база от данни;
- **конфигурируемо хранилище за файлове:** в системата да могат да се съхраняват файлове както с помощта на локалната файлова система (по подразбиране), така и в облачни решения, такива като Amazon S3<sup>17</sup>;
- **цялост на данните:** в системата да има механизъм за изчисляване и мониторинг на цифрови отпечатьци на файловете данни с цел периодична проверка на целостта на данните, ако има нарушение - при дефект в данните да могат да се възстановят в оригиналната си версия;
- системата да е с **многоезикова поддръжка**, използва се Unicode/UTF-8<sup>18</sup> ;
- поддръжка на постоянни уникални **идентификатори** за всеки публикуван ресурс (опционално).

---

<sup>12</sup> <https://www.openarchives.org/pmh/>

<sup>13</sup> <https://api.archives-ouvertes.fr/docs/sword>

<sup>14</sup> <https://www.openaire.eu/>

<sup>15</sup> <https://www.postgresql.org/>

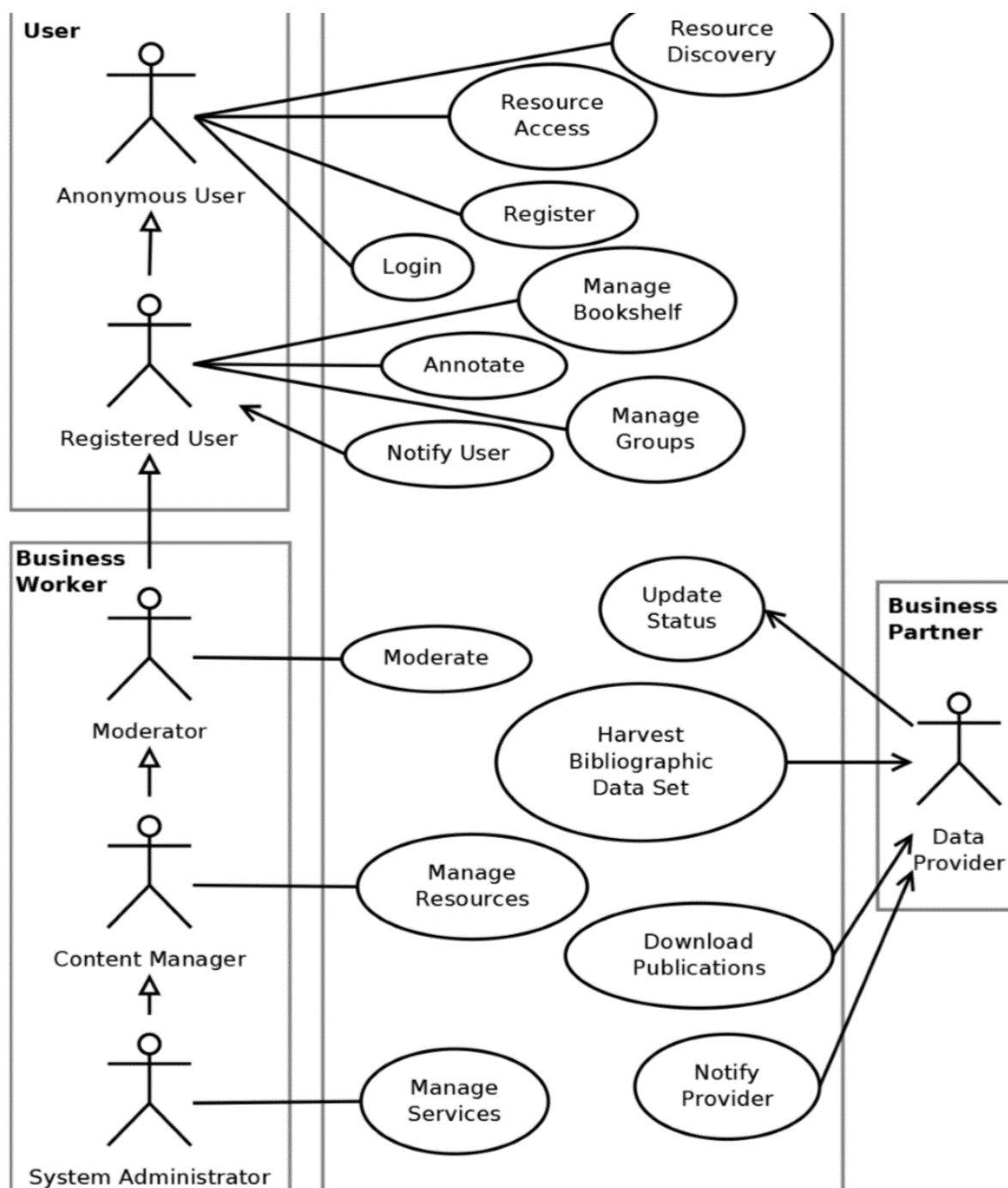
<sup>16</sup> <https://www.oracle.com/database/>

<sup>17</sup> <https://aws.amazon.com/s3/>

<sup>18</sup> <https://www.utf8-chartable.de/unicode-utf8-table.pl>

### 3. Архитектура на процесите в системата

Общата архитектура на процесите на системата трябва да е като дадената на фигура 2 схема, заимствана от EUDML<sup>19</sup>.



Фиг. 2. Архитектура на системата

<sup>19</sup> <https://eudml.org/>



## 4. Системни изисквания

### Системни изисквания за софтуер:

- UNIX-подобна операционна система (Linux, Ubuntu, CentOS, HP/UX);
- Java JDK 7 или 8+ (OpenJDK<sup>20</sup> или Oracle JDK<sup>21</sup>);
- Apache Maven 3.0.5<sup>22</sup> или по-нова версия (3.3.9 +) (Java build tool);
- Apache Ant 1.8 или по-нова версия (Java build tool);
- релационна база от данни: PostgreSQL 9.4<sup>23</sup> или по-нова версия (с pgcrypto) или Oracle;
- Servlet Engine<sup>24</sup> (Apache Tomcat 8 или по-нова, Jetty, Caucho Resin или еквивалент);
- валиден SSL сертификат<sup>25</sup>;
- наличен пощенски адрес на потребител, който ще бъде използван от системата.

### Системни изисквания за хардуер (препоръчителни изисквания за минимален и среден клас конфигурация):

Съвременен сървър не по-стар от 5 години със следните характеристики:

- ~ 4 – 8 GB памет (RAM)
- ~ 4 – 6 GB за Tomcat
- ~ 2 – 4 GB за база данни (PostgreSQL<sup>26</sup> или Oracle<sup>27</sup>)
- 500 GB - 1 TB за цифровото хранилище (или достатъчно място за съхранение на всички данни, които дадена организация иска да съхранява в системата в зависимост от очаквания обем и тип на данните).

### Примери за съхранение:

- 15 000 грт мрежови дискове в RAID достъпни чрез гигабитова връзка за съхранение на базата данни и индексите;

---

<sup>20</sup> <https://openjdk.java.net/>

<sup>21</sup> [https://docs.oracle.com/javase/8/docs/technotes/guides/install/install\\_overview.html](https://docs.oracle.com/javase/8/docs/technotes/guides/install/install_overview.html)

<sup>22</sup> [https://docs.oracle.com/javase/8/docs/technotes/guides/install/install\\_overview.html](https://docs.oracle.com/javase/8/docs/technotes/guides/install/install_overview.html)

<sup>23</sup> <https://www.postgresql.org/about/news/1592/>

<sup>24</sup> [https://docs.oracle.com/cd/A91202\\_01/901\\_doc/java.901/a90213/introduction.htm](https://docs.oracle.com/cd/A91202_01/901_doc/java.901/a90213/introduction.htm)

<sup>25</sup> <https://www.digicert.com/ssl->

[certificate/?msclkid=06d05f3fcfc91851957304c28dc720da&utm\\_source=bing&utm\\_medium=cpc&utm\\_campaign=SSL%20\(Generic\)&utm\\_term=ssl%20certificate&utm\\_content=SSLCertificate](https://www.digicert.com/ssl-certificate/?msclkid=06d05f3fcfc91851957304c28dc720da&utm_source=bing&utm_medium=cpc&utm_campaign=SSL%20(Generic)&utm_term=ssl%20certificate&utm_content=SSLCertificate)

<sup>26</sup> <https://www.postgresql.org/>

<sup>27</sup> <https://www.oracle.com/database/>

- 7,400 грт мрежови дискове в RAID, достъпни чрез гигабитова връзка за съхранение на данни, чийто размер може лесно да бъде разширен.

## 5. Възможни приложения на системата

- институционални хранилища
- електронно управление на записи – регистри
- цифровизация и дългосрочно съхранение на документи, данни, свободен достъп до научни резултати.

## 6. План етапи за разработване и внедряване

- създаване и конфигурация на необходимата програмна среда, операционна система, мрежови настройки, инсталиране на необходими допълнителни софтуерни пакети и програмни инструменти за една организация – 6 месеца;
- разработване и експериментално внедряване на прототипи на цифрово хранилище в организации избрани от МОН – 6 месеца;
- обучение на потребители и тестване – 6 месеца;
- създаване и внедряване на портал на системата - агрегиране и обмен на данни – 12 месеца;
- стартиране на системата в работен режим. Обучение на потребителите в системата – 6 месеца;
- администриране и поддръжка на системата по време на функционирането.

## 7. Досегашен опит при изграждане на подобни системи

ИМИ-БАН<sup>28</sup> има значителен многогодишен опит в организацията на дейности, популяризиращи отворения достъп до научни резултати, бази данни и хранилища. През последните години институтът е организиран девет национални информационни дни и една обучителна школа за отворен достъп до научни резултати. Учени от института имат значителен брой публикации (над 20) по темата и технологичен опит, ресурси и реализации (вкл.

---

<sup>28</sup> <http://math.bas.bg/>

експериментални внедрявания) на приложените софтуерни платформи. Те участват в редица работни комисии и срещи на национално и международно ниво по тематиката за България по европейските проекти OpenAire<sup>29</sup> и EuDML<sup>30</sup>. В ИМИ-БАН действа цифрово хранилище с отворен достъп в областта на математиката и информатиката. Практически в България действат реално и пълноценно три цифрови хранилища с отворен достъп до научни резултати: в ИМИ-БАН<sup>31</sup>, ФМИ на СУ<sup>32</sup> и НБУ<sup>33</sup>.

---

<sup>29</sup> <https://www.openaire.eu/>

<sup>30</sup> <https://eudml.org/>

<sup>31</sup> <http://math.bas.bg/>

<sup>32</sup> <https://www.fmi.uni-sofia.bg/>

<sup>33</sup> <https://www.nbu.bg/>