

**Денис Слипецький,**

провідний інженер ІнБПМ НАН України

**Володимир Горбунов,**

провідний інженер ІнБПМ НАН України

**Олександра Сергєєва,**

провідний інженер ІнБПМ НАН України

## Створення електронного репозиторію Інституту біології південних морів НАН України на основі вільно поширюваного програмного забезпечення

У статті розглядається досвід створення відкритого електронного архіву (репозиторію) в Інституті біології південних морів Національної академії наук України (ІнБПМ НАНУ) на основі вільно поширюваного програмного забезпечення з відкритим кодом (DSpace).

К л ю ч о в і с л о в а: електронні архіви, відкритий доступ, OAI-PMH, OceanDocs.

### Вступ

У продовж практично всієї історії існування (починаючи від Біологічної станції) ІнБПМ мав свої видання («Труды Севастопольской биологической станции», «Биология моря», «Экология моря», «Морський екологічний журнал»). Традиційно в цих виданнях публікувалася значна кількість наукових праць співробітників ІнБПМ (і не тільки). Але оскільки тираж видань ІнБПМ був завжди невеликий, то значна частина наукових праць співробітників Інституту залишалася невідомою широкому науковому загалу, навіть вітчизняним ученим.

Використання сучасних інформаційних технологій дає змогу зробити наукові праці співробітників Інституту максимально доступними світовому науковому співтовариству, що, безсумнівно, сприятиме підвищенню авторитету (й інтересу до наукової діяльності) не тільки окремих співробітників, але й Інституту в цілому. Досягти цього можна, використовуючи так звані «електронні репозиторії»<sup>1</sup> (EP).

Інститутський репозиторій – це публічно доступний архів наукової організації, призначений для публікації результатів досліджень. Точніше буде сказати, що це система, що містить у собі не тільки програмне й апаратне забезпечення, але й політику розміщення інформації і документів, процедури та правила документообігу.

<sup>1</sup> Слова репозитарій / репозиторій не зафіксовані у словниках української мови. Оскільки слово запозичено з англійської мови (від repository), то у статті використано написання «репозиторій».

### Цілі і завдання

З огляду на потреби Інституту при створенні репозиторію були сформульовані такі завдання:

- зберігання публікацій Інституту в електронному вигляді;
- забезпечення доступу до них через мережу Інтернет;
- зберігання докладного опису публікацій (метаданих) відповідно до міжнародних стандартів;
- підтримка протоколу OAI-PMH<sup>2</sup> для індексування спеціалізованими системами (harvesters);
- забезпечення збереження публікацій Інституту (архівування);
- забезпечення інтеграції репозиторію Інституту в його інформаційну систему.

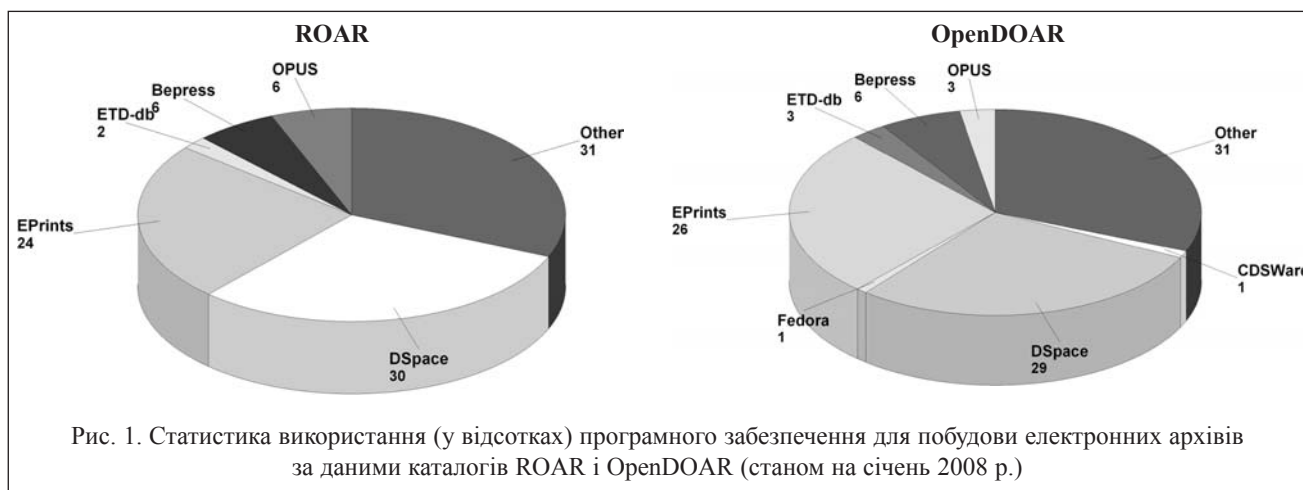
### Реалізація репозиторію на практиці

Сьогодні існує багато вільно поширюваних систем побудови електронних архівів, які підтримують протокол обміну метаданими OAI-PMH. Найбільш популярні серед них, за даними каталогів Registry of Open Access Repositories (ROAR)<sup>3</sup> і Directory of Open Access Repositories (OpenDOAR)<sup>4</sup>, DSpace, EPrints, Vepress (рис. 1).

<sup>2</sup> OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) – протокол обміну метаданими. <http://www.openarchives.org/pmh>

<sup>3</sup> Registry of Open Access Repositories (ROAR) <http://roar.eprints.org/>

<sup>4</sup> Directory of Open Access Repositories (OpenDOAR) <http://www.opendoar.org>



Групою розроблювачів, беручи до уваги популярність програмних продуктів і рекомендації із створення EP<sup>5</sup>, були проаналізовані системи DSpace, EPrints (як найбільш популярні) і Fedora (як одна з кращих із погляду архітектури) за такими критеріями<sup>6</sup>:

1. розширюваність;
2. легкість роботи з вихідним кодом;
3. налагоджені механізми безпеки (аутентифікація, авторизація, безпека публікації тощо);
4. сумісність (можливість інтеграції в інші репозиторії, підтримка протоколу OAI-PMH);
5. легкість установки і настроювання. Можливість підтримки множини інсталяцій на одній платформі;
6. легкість адміністрування (настроювання під різних користувачів);
7. локалізація (підтримування мов);
8. умови поширення;
9. якість і настроюваність документообігу в системі;
10. спільнота розроблювачів.

У результаті аналізу цих систем було прийнято рішення за основу електронного репозиторію Інституту взяти програмний продукт DSpace, точніше його версію, модифіковану в рамках проекту IOC/IODE UNESCO (Intergovernmental Oceanographic Commission/International Oceanographic Data and Information Exchange) OceanDocs<sup>7</sup>.

Причини вибору на користь системи DSpace (OceanDocs) стають зрозумілими при аналізі пере-

ваг і недоліків (із погляду групи розроблювачів ІмБПМ) її конкурентів.

Система EPrints виконана скриптовою мовою<sup>8</sup> Perl, що значно ускладнює подальшу інтеграцію побудованого на її основі репозиторію в інформаційну систему (ІС) Інституту, що розвивається на базі software frameworks<sup>9</sup> (Microsoft\_. Net<sup>10</sup> і Java Platform Enterprise Edition<sup>11</sup>), а також подальшу модифікацію і підтримку відповідно до планів розвитку інформаційної системи ІмБПМ.

Система Fedora виступає, по суті, «движком» електронного репозиторію, тобто реалізує програмний інтерфейс зберігання і доступу до об'єктів репозиторію, а також його адміністрування. Проте система, сама по собі, не має веб-інтерфейсу, який має бути реалізований окремо. Існуючі доступні реалізації такого інтерфейсу виконані скриптовою мовою PHP, що накладає, по суті, такі самі обмеження на подальшу їхню інтеграцію в ІС ІмБПМ, як і у випадку з EPrints. Створення ж веб-інтерфейсу власними силами на підходящій платформі «із нуля» забрало би значний час на шкоду розвитку ІС ІмБПМ.

Система DSpace (OceanDocs) побудована на платформі Java, що має всі необхідні засоби для подальшої інтеграції електронного репозиторію Інституту в загальноінститутську ІС. Крім того, DSpace має готовий, достатньо гнучкий у плані настроювань, веб-інтерфейс.

Найважливіші функціональні можливості DSpace:

- побудова ієрархічної моделі колекцій документів;

<sup>5</sup> Sourced from Open Society Institute – A Guide to Institutional Repository Software (3rd Edition, August 2004).

<sup>6</sup> Technical Evaluation of selected Open Source Repository Solutions. <https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf>

<sup>7</sup> OceanDocs (E-repository for Ocean Publications) <http://www.oceandocs.net>

<sup>8</sup> Скриптова мова (англ. scripting language) – мова програмування, розроблена для запису «сценаріїв», послідовностей операцій, що їх користувач може виконувати на комп'ютері.

<sup>9</sup> <http://ru.wikipedia.org/wiki/Framework>

<sup>10</sup> [http://ru.wikipedia.org/wiki/Microsoft\\_.NET](http://ru.wikipedia.org/wiki/Microsoft_.NET)

<sup>11</sup> <http://ru.wikipedia.org/wiki/J2EE>

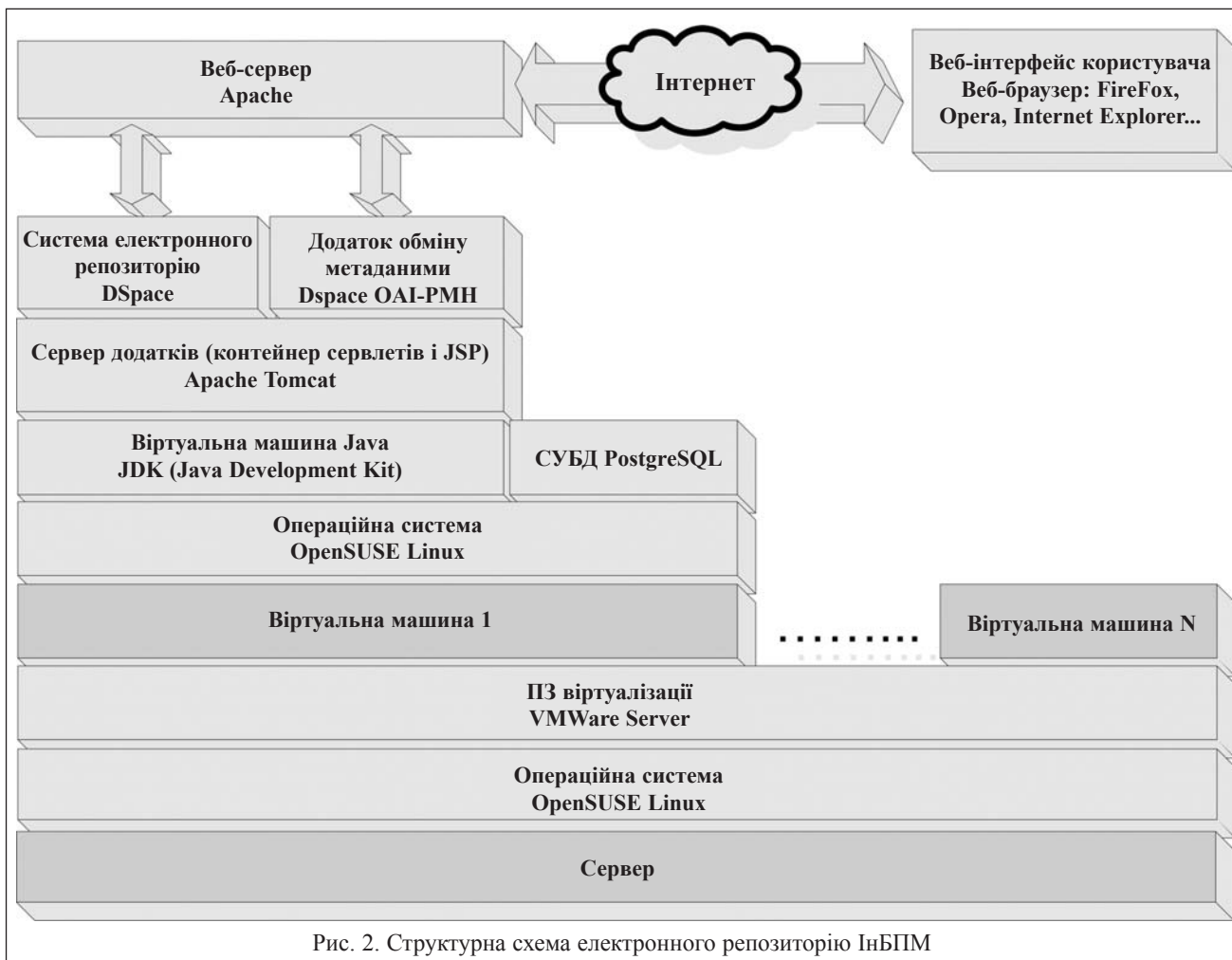


Рис. 2. Структурна схема електронного репозиторію ІнБПМ

- гнучке налаштування прав доступу до колекцій;
- інтерфейс, що піддається достатньо добре налаштуванню.

Незважаючи на деякі мінуси порівняно з розглянутими вище системами (немає розвинуеного механізму побудови розподілених репозиторіїв, недостатні механізми оптимізації при збільшенні навантаження на систему) DSpace є найбільш оптимальним варіантом у нашому випадку. Крім того, зміни, внесені в DSpace у рамках проекту OceanDocs, значно підвищують зручність використання репозиторію для зберігання документів за тематикою робіт Інституту.

Одним з найбільш важливих доробків є можливість вибору типу документа, який поміщується в репозиторій. Якщо в оригінальному DSpace ця процедура має суто інформативний характер, то в OceanDocs це дає змогу сховати від користувача ті поля метаданих, що не призначені для опису даного типу документа. DSpace дозволяє зробити такі налаштування, але тільки для колекції у цілому, і для цього необхідно вручну редагувати спеціальний файл. OceanDocs не накладає подібних обмежень – у ту саму колекцію можуть додаватися різні типи документів, крім того, відповідає необхідність ручного редагування конфігурації системи.

Ще одна важлива зміна стосується поля «цитовання». Якщо в DSpace у цьому полі користувач сам вписує посилання на документ, за правилами цитування (що на практиці спричиняє численні помилки), то в OceanDocs додані спеціальні поля (номер сторінки, кількість сторінок, том, рік видання тощо), і посилання на документ формується системою автоматично за заданими правилами. Це значно знижує трудомісткість заповнення полів метаданих документа, а також усуває джерело можливих помилок при заповненні.

Звісно, є й інші важливі доповнення, але вищеписані найбільш помітні з погляду кінцевого користувача системи.

У результаті проведених досліджень і тестів для установлення електронного репозиторію ІнБПМ були обрані всі необхідні компоненти. Як-от:

- операційна система OpenSUSE 10.2;
- контейнер сервлетів Apache Tomcat 6;
- сервер реляційних баз даних PostgreSQL 8.

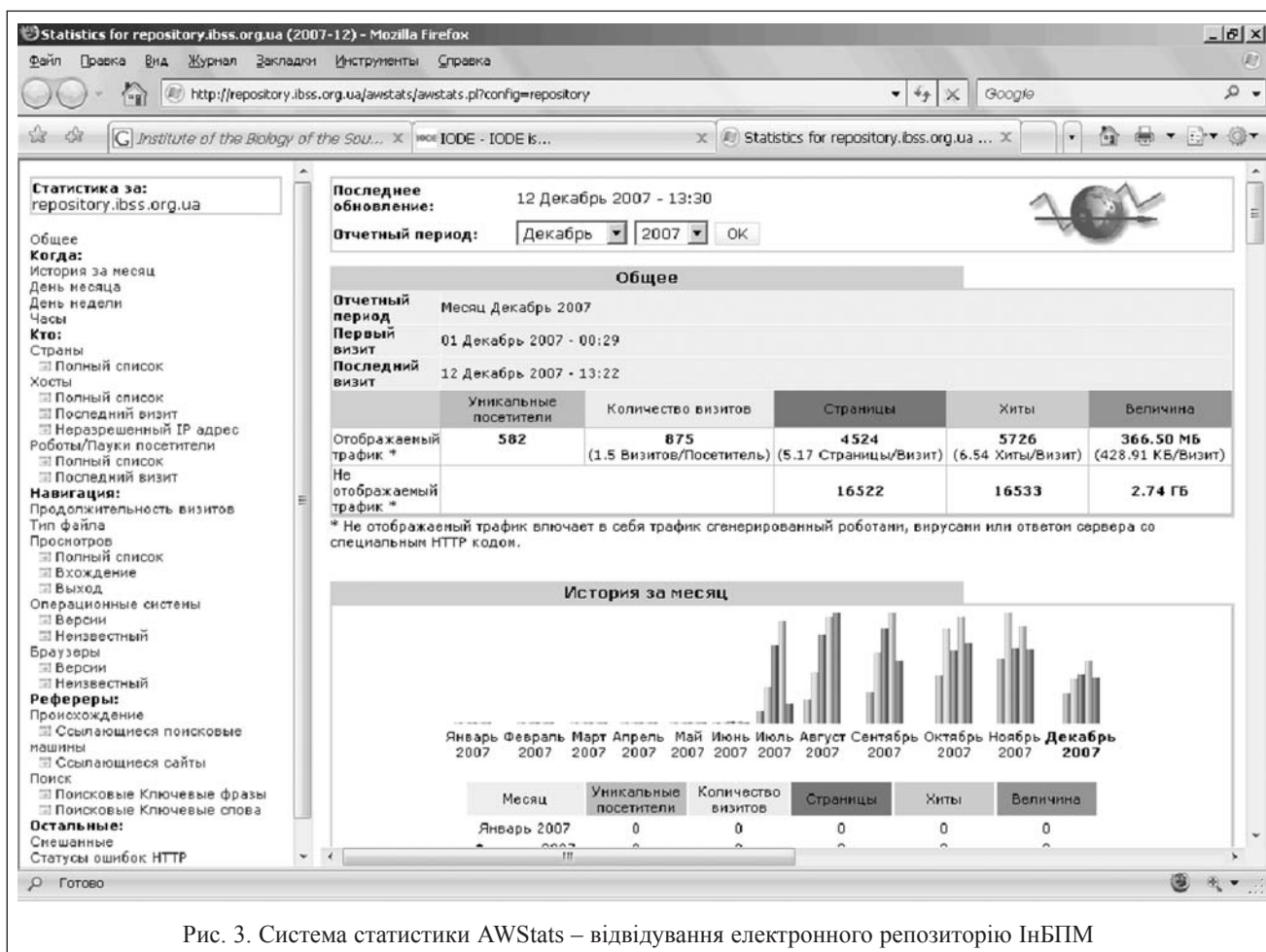


Рис. 3. Система статистики AWStats – відвідування електронного репозиторію ІнБПМ

Структурну схему ЕР ІнБПМ наведено на рис. 2.

Структуру колекцій документів було розроблено співробітниками бібліотеки ІнБПМ. Також співробітники бібліотеки заносять публікації у репозиторій і заповнюють необхідні поля метаданих.

Багато організацій довіряють роботу зі створення електронних репозиторіїв своїм інформаційним підрозділам. Проте це не цілком виправдано – технічні спеціалісти можуть настроїти роботу системи, але не можуть працювати з публікаціями на фаховому рівні<sup>12</sup>. Саме бібліотека повинна займатися підтримкою (наповнення, структурування тощо) електронного репозиторію установи, тому що в бібліотеці мають бути сконцентровані всі засоби й інструментарій для пошуку інформації з різноманітних джерел – чи то друковані, чи електронні матеріали.

Для забезпечення ефективного контролю за роботою ЕР ІнБПМ до нього було підключено систе-

му статистики й аналізу робочих журналів AWStats (рис. 3, доступна на сайті розроблювача <http://awstats.org/>), що дає змогу збирати достатньо докладну інформацію про роботу репозиторію: кількість унікальних відвідувачів, кількість візитів і їхня тривалість, пошукові ключові фрази і слова, посилання на репозиторій із зовнішніх джерел, найбільш популярні сторінки та інше.

Відкритий науковий архів, створений в ІнБПМ НАНУ, зареєстрований кількома збирачами метаданих (harvesters), що підтримують протокол OAI-PMH:

- Open Archives Initiative (<http://openarchives.org/>);
- Registry of Open Access Repositories (ROAR) (<http://roar.eprints.org/>);
- Avano (<http://www.ifremer.fr/avano/>);
- The Directory of Open Access Repositories – OpenDOAR (<http://www.opendoar.org/>);
- The University of Illinois OAI-PMH Data Provider Registry (<http://gita.grainger.uiuc.edu/registry/>);
- OAIster (<http://oaister.umdl.umich.edu>).

## Результати

У даний час ЕР ІнБПМ доступний у мережі Інтернет за адресою <http://repository.ibss.org.ua> (рис. 4).

<sup>12</sup> Еременко Г. О. Стратегические направления развития электронных информационных ресурсов и сервисов (Научная электронная библиотека eLIBRARY.RU). [http://www.elibrary.ru/projects/conference/turkey2005/presentations/Science\\_online.doc](http://www.elibrary.ru/projects/conference/turkey2005/presentations/Science_online.doc)

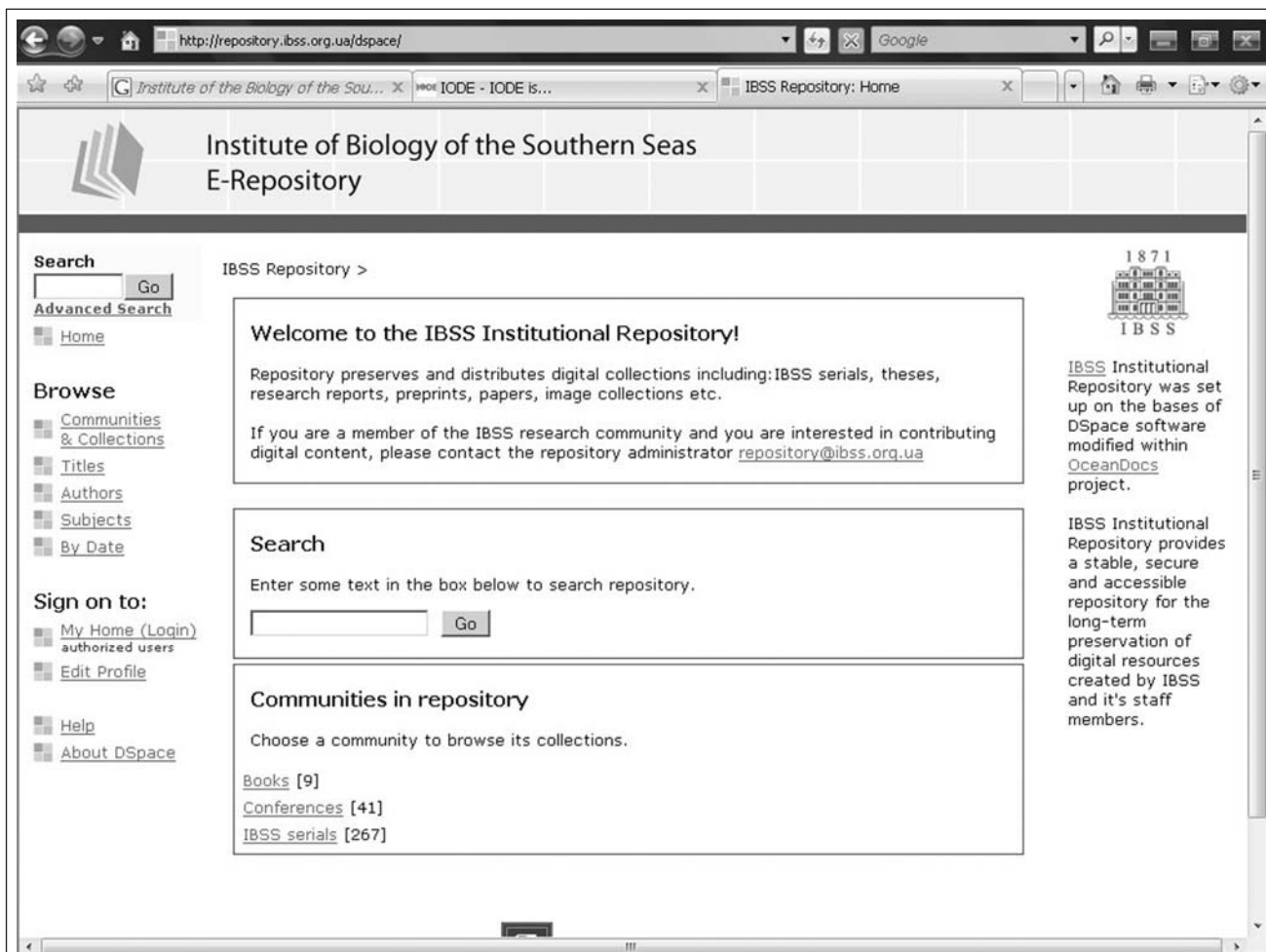


Рис. 4. Домашня сторінка електронного репозиторію ІнБПМ

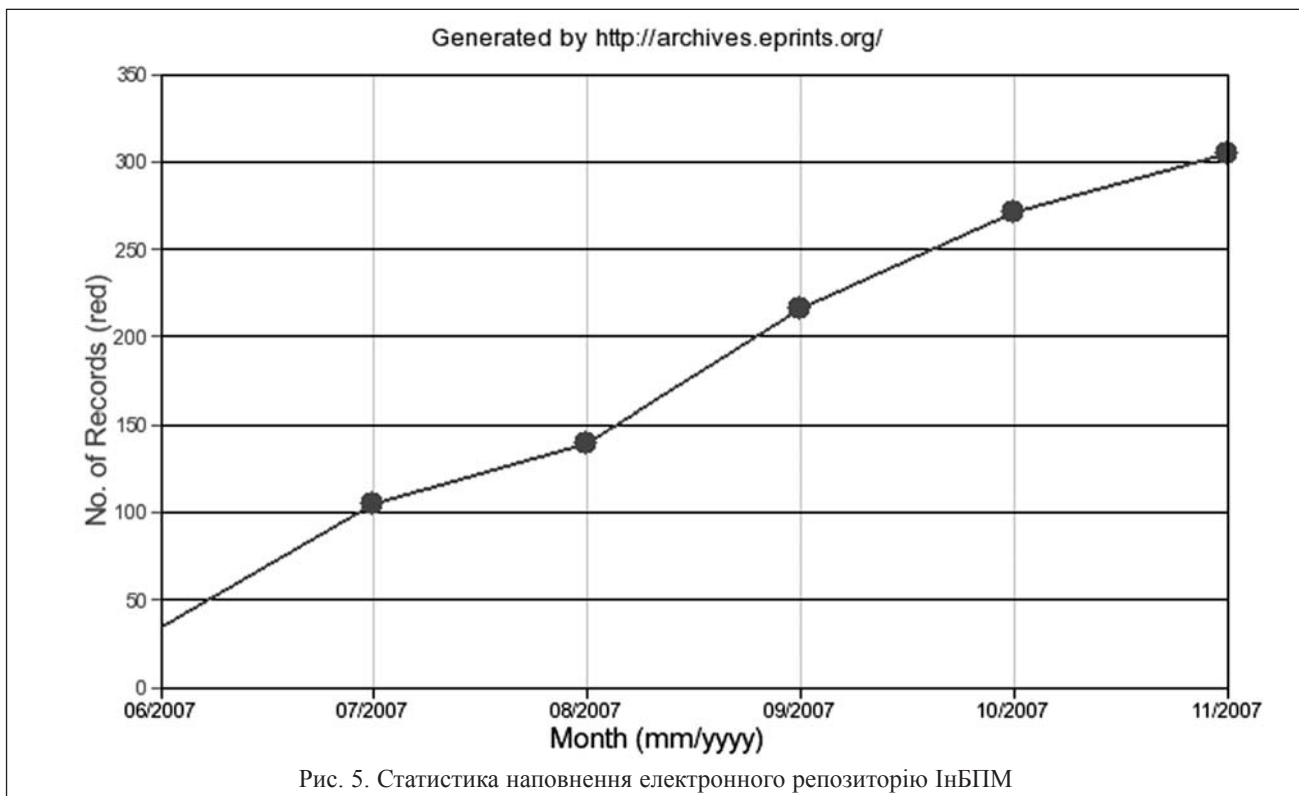


Рис. 5. Статистика наповнення електронного репозиторію ІнБПМ

За перші 5 місяців роботи в репозиторій було поміщено понад 300 публікацій (рис. 5). Це статті з часописів «Морський екологічний журнал» і «Екологія моря», книги і монографії, видані ІнБПМ НАНУ. Також репозиторій містить матеріали конференцій, проведених Інститутом (тези і презентації) і декілька постерів, представлених на міжнародних конференціях співробітниками Інституту.

Розміщення наукових публікацій, статей і іншої інформації щодо наукової діяльності у мережі Інтернет є ефективним і корисним, тільки якщо ця інформація доступна для пошуку. Пошукові машини є своєрідними «воротами в Інтернет», оскільки 70 % користувачів починають роботу в Інтернеті з відкриття сторінки пошукової системи. Але, давши запит пошуковій системі на пошук, зазвичай у відповідь користувач одержує у кращому разі сотні сторінок, а найімовірніше – тисячі або десятки тисяч<sup>13</sup>. Причому в переважній більшості – так зване пошукове «сміття», у котрому лише зернинки цінної і корисної інформації.

Усі сучасні пошукові системи не просто індексують зміст різноманітних інтернет-ресурсів, а застосовують т. зв. «Індекс цитування», що цілком повторює однойменний репозиторій, використовуваний у науковому світі. Розміщення публікації у репозиторії Інституту, що підтримує протокол обміну метаданими OAI-PMH, дає змогу анонсувати її на множині включених у ІВА (Ініціатива відкритих архівів)<sup>14</sup> систем, що підвищує її ранг для пошукових сервісів. Причому багато з таких систем існують уже не один рік і підтримуються найавторитетнішими співтовариствами: The University of Illinois, The National Science Foundation, Massachusetts Institute of Technology. Це і дає так звану «якість» посилань.

З огляду на той факт, що вже до кінця 2006 року більше 60 % загального числа входів із пошукових систем на сайти відбувалося з Google<sup>15</sup>, можна

стверджувати, що рейтинг цієї пошукової системи є дуже показовим для будь-якого електронного ресурсу.

Слід зазначити, що за увесь час існування ЕР ІнБПМ (із 14 червня 2007 р.) серед посилань на першій сторінці результатів пошуку Google за ключовими фразами з іменами співробітників Інституту, назвою Інституту і т.д. незмінно присутнє посилання на ЕР ІнБПМ (а також веб-сайт ІнБПМ <http://www.ibss.org.ua/>). Можна дійти висновку, що застосування сучасних систем управління контентом і ЕР дає змогу значно підвищити «індекс цитованості» Інституту в мережі Інтернет.

### Висновки

Для створення наукових архівів необхідно використовувати програмне забезпечення, що підтримує протокол обміну метаданими OAI-PMH. Це дасть змогу забезпечити сумісність з іншими подібними архівами або спеціалізованими пошуковими системами.

Для організацій, що працюють у галузі наук про Землю, оптимальним рішенням при створенні ЕР є використання системи DSpace/OceanDocs.

Науковий архів необхідно реєструвати в максимально можливій кількості каталогів, таких як ROAR, OpenDOAR, Avano, OAIster тощо. Це не тільки дає змогу скористатися всіма перевагами протоколу OAI-PMH, але і значно підвищити рейтинг публікацій із наукового архіву в пошукових системах.

Науково-дослідним установам для максимізації оприлюднення наукових досягнень і результатів необхідно взяти на себе самоархівування<sup>16</sup> статей і матеріалів із результатами досліджень, аж до обов'язкового архівування певних типів документів.

\* \* \*

Автори статті висловлюють подяку IOC/IODE UNESCO за надання матеріалів за проектом OceanDocs; IOC Project Office for IODE – за організацію стажувань співробітників ІнБПМ; науковій бібліотеці ІнБПМ – за ініціювання цієї роботи й активне поповнення репозиторію новими документами.

(Переклад з російської Н. Маслакової)

<sup>13</sup> «Издание массы книг и научных трудов становится бедствием, так как солидное, ценное и необходимое все чаще исчезает в огромном потоке ненужных изданий, и надвигается такая угроза, что все достойные внимания издания захлебнутся в этом потоке», – писав німецький дослідник Адольф фон Харнак ще у 1911 р. (Цит. за: *Фабиан Б.* Книги, библиотеки и гуманитарные исследования / Отв. ред. В. П. Леонов. – СПб., 1996. – С. 258).

<sup>14</sup> Open Archives Initiative <http://www.openarchives.org>

<sup>15</sup> [http://weblogs.hitwise.com/bill-tancer/2006/09/latest\\_search\\_numbers\\_google\\_c.html](http://weblogs.hitwise.com/bill-tancer/2006/09/latest_search_numbers_google_c.html)

<sup>16</sup> Самоархівування (self-archiving) – самостійне завантаження автором електронної версії своєї публікації у відкриті електронні архіви.