

**ОБРОБЛЕННЯ НАУКОВИХ ДАНИХ  
В УМОВАХ ІНФОРМАЦІЙНОГО «БУМУ»**

**М.З. ЗГУРОВСЬКИЙ, А.І. ПЕТРЕНКО**

Сучасна наука (е-наука) базується на обробленні потенційно величезних обсягів інформації. Розглянуто тенденції та перспективи розвитку е-науки в умовах інформаційного «буму», які охоплюють питання забезпечення якості та сумісності даних, використання метаданих і семантики даних, довгострокового їх збереження, інтелектуального оброблення та пошуку даних в існуючих джерелах, впливу даних на вибір платформи й її сервісно-орієнтованої архітектури, перспективи об'єднання Грід і хмарних обчислень, а також європейські ініціативи з цієї тематики.

**ВСТУП**

На початку ХХІ ст. постало питання: «Як багато інформації є у світі?» і було запропоновано декілька методик для оцінювання обсягу інформації [1]. Із 2000 р. та 2003 р. група дослідників у Берклі, оцінюючи кількість створеної інформації, дійшла висновку, що «більшістю від загального обсягу нової інформації є голосовий телефонний трафік зі своїм унікальним контентом (97 %)». У 2007–2008 рр. IDC (International Data Corporation — Міжнародна корпорація даних) підрахувала, що в 2007 р. «обсяг усього можливого для використання простору жорстких дисків, касет, CD, DVD і пам'яті (енергозалежної та енергонезалежної) склав 264 екзабайтів (1 екзабайт =  $10^{18}$  байт), а в 2010 р. — майже 2 зетабайта (1 зетабайт =  $10^{21}$  байт)». У 2008 р. дослідження зосередились на інформації, що споживається людьми при взаємодії з медіа пристроями, при цьому комп'ютерні ігри та фільми складають 99,2 % від цієї інформації. Упровадження широкопasmового Інтернету призвело до безперервного прискорення телекомунікаційного зв'язку (середньорічний темп зростання в 6 % для 1986–1993 рр., 23 % для 1993–2000 рр. і 60 % за 2000–2010 рр.). Людина хоче отримувати інформацію скрізь — на роботі, вдома, в авто, під час прогулянки тощо.

Бачення компанії IBM на зростання обсягів інформації у світі, яке враховує можливості Інтернет і жорстких дисків, відображено на рис. 1.

До оцінки зростання технологічного потенціалу оброблення інформації можна застосувати закон Мура, згідно з яким потужність комп'ютерів для оброблення інформації подвоюється приблизно в два рази кожні 14 місяців

протягом останніх десятиліть, у той час як для суперкомп'ютерів потужність подвоюється кожні 18 місяців. Зараз людство може зберігати кілька оптимально стиснутих зетабайтів даних на різних пристроях (рис. 2) і виконувати  $7 \cdot 10^{19}$  операцій за секунду на обчислювальних пристроях різного типу (рис. 3).

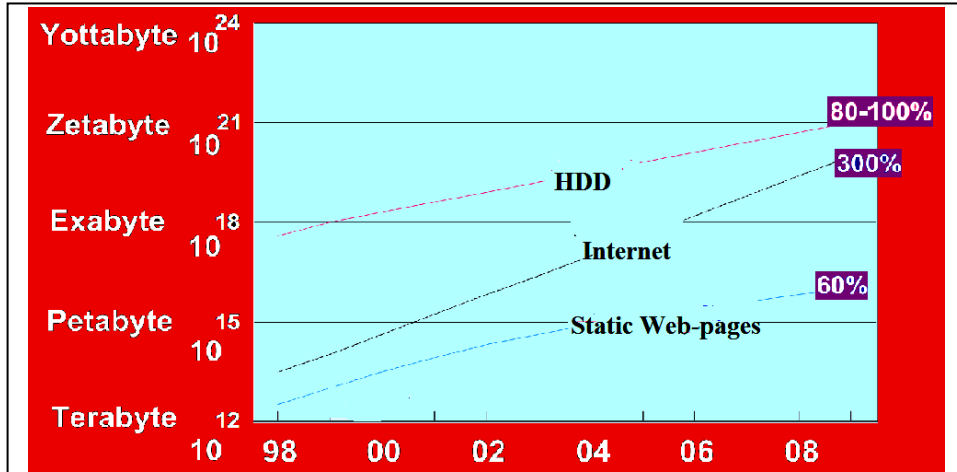


Рис. 1. Експоненційне зростання інформації у світі за даними ІВМ

Щорічно їхній обсяг зростає на 58 %, двонаправлена передача даних на 28 %, кількість збереженої інформації на 23 % [1].

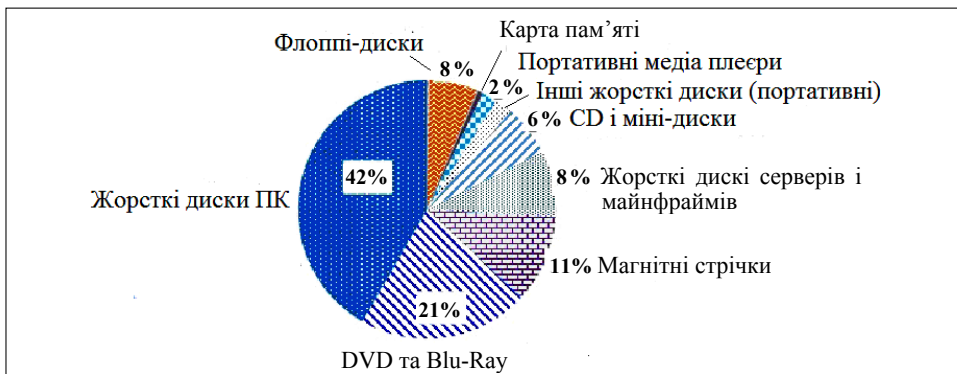


Рис. 2. Розподіл частки збереження даних різними засобами у світі

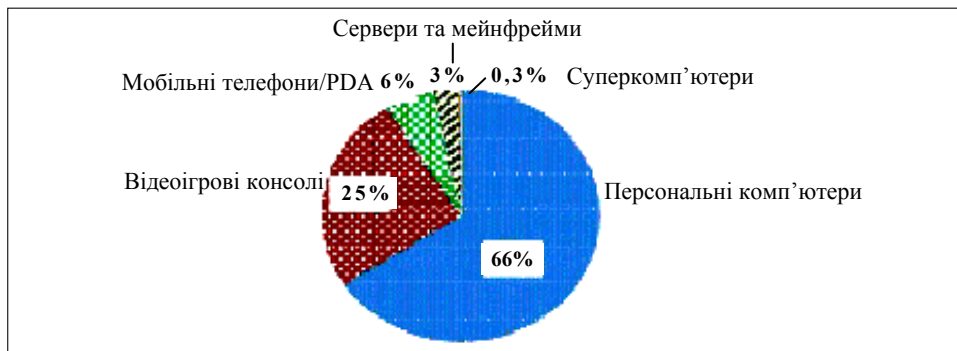


Рис. 3. Розподіл частки обчислення інформації різними засобами у світі

Варто зазначити, що ще з 2007 р. технологічні можливості людства виявилися недостатніми для збереження всієї інформації, яка генерується. Тому постає важливе питання відбору і тривалого збереження тієї частини інформації, яка забезпечує функціонування суспільства та визначає його ідентичність та перспективи подальшого розвитку. І таку ситуацію не може змінити поява таких велетенських сховищ, як Spectra T-Finity обсягом 3,6 екзабайтів даних.

Нас цікать, перш за все, стан і перспективи використання наукових даних, тому що вони визначають прогрес у розвитку земної цивілізації (табл. 1).

**Таблиця 1.** Різні категорії наукових даних

Категорії даних	Фізика/хімія	Гео-/астро-науки	Біологічні науки
Дані, які можуть бути виміряні повторно	Більшість даних	Геологічні структури, зафіксовані зірки	Більшість даних
Дані, які можуть бути виміряні лише один раз	Більшість даних	Спалахи сонця, виверження вулканів, нові зірки	Рідкісні зразки скам'янілості
Незалежні від розташування	Більшість даних	Тектоніка мінералів	Більшість даних
Залежні для розташування	Більшість даних	Скелі, метеорологічні дані, астрономічні дані, скам'янілості	Рідкісні зразки скам'янілості
Первинні отримані під час спостереження або експериментальні дані	Кристаліграфічні дані, оптичні спектри	Сейсмографічні, погодні дані та їх записи	Фізіологічні дані (наприклад, кров'яний тиск), біохімічні дані (наприклад, композиція тканин і органів)
Комбінації первинних даних за допомогою теоретичної моделі	Фундаментальні константи кристалічних структур	Розподіл температури на поверхні сонця	Генетичні коди, модель м'язів і судин
Дані, які отримано теоретичним обчисленням	Властивості молекул	Сонячні затемнення	Передбачення поведінки залежно від генотипу
Визначені (детермістичні) дані	Більшість макроскопічних даних	Елементи планетарних орбіт	Розташування гена в хромосомах
Стохастичні дані	Дані полімерів, чутливі до їх структури	Склад ґрунтів і кам'яних структур, частота сонячних спалахів за одиничний інтервал	Більшість даних
Кількісні дані	Більшість даних	Метеорологічні та сейсмічні дані	Фізіологічні та біохімічні дані
Напівкількісні дані	Масштаб твердості матеріалів	Масштаб сили вітру	Фізіологічні та біологічні дані
Якісні дані	Формули хімічних структур, властивості нуклідів	Класифікація кам'яних структур, класифікації зоряних спектрів	Таксономічна класифікація організмів, послідовності амінокислот
Дані, які подані числами	Значення параметрів і констант	Метеорологічні дані	Фізіологічні та біохімічні дані
Дані, які подані графами або моделями	Фазові діаграми, стереоскопічні молекулярні діаграми, моделі молекул	Геологічні карти, погодні карти, карти неба на специфічній радіочастоті (наприклад 21 см)	Електроенцефалограми, електрокардіограми
Символічні дані	Математичні моделі та формули	Математичні моделі	Математичні моделі

Наукові інструменти та комп'ютерне імітаційне моделювання породжують величезні обсяги даних, для аналізу й організації яких потрібні нові наукові методи. Щороку обсяги наукових даних майже подвоюються, зараз йдеться про обробку петабайтних наборів даних (1 петабайт=10<sup>15</sup> байт). За оцінками вчених у наступні п'ять років буде вироблено наукових даних більше, ніж за всю історію людства [2]. Оскільки нові наукові інструменти володіють винятковою точністю, так само швидко поліпшується якість даних. Для аналізу цих даних з метою знаходження «тонких» ефектів, які не були враховані в попередніх дослідженнях, потрібні алгоритми, які одночасно можуть працювати з величезними наборами даних і виявляти при цьому дуже «тонкі» ефекти, які залишилися невиявленими за попередніми вимірюваннями. Обсяги інформації, що створюється, величезні і вони впливають не тільки на розвиток природознавчих, але й гуманітарних та соціальних наук, а також на науки про життя та здоров'я.

Це лавиноподібне зростання обсягів даних, або інформаційний «бум» призводить до нових підходів їх оброблення, коли дослідники вживають методи та засоби, керовані даними. Великі обсяги даних сприяють можливості нових наукових відкриттів, але при цьому створюють труднощі, пов'язані зі збереженням, пошуком і доступом до даних з їх аналізом.

Переконливим прикладом джерела петабайтних наборів даних є Великий Андронний Коллайдер, який працює в CERN (European Particle Physics Laboratory — Європейський центр фізики високих енергій) і який вироблятиме приблизно 10 петабайтів необроблених даних у галузі фізики високих енергій за рік і додатково створюватиме 100 петабайт оброблених даних за рік, що записані на дисках для наступного аналізу. Потрібна величезна обчислювальна потужність для оброблення такої кількості даних, які за допомогою Грид-інфраструктури WLCG (Worldwide LHC Computing Grid — світовий обчислювальний Грид Великого Андронного Коллайдера) [3] розподіляються до комп'ютерних центрів у 34 різних країнах на шести континентах. Крім того, керівництво CERN запрошує волонтерів, які згодні надати свої персональні комп'ютери для того, щоб допомагати обробляти дані коллайдера. За інформацією Геннадія Зінов'єва (завідувача відділом фізики високої щільності енергії Інституту теоретичної фізики ім.Боголюбова НАН України, представника України в CERN) ця ініціатива отримала широку підтримку у світі і з'явилися навіть цілі сім'ї, готові обробляти інформацію.

Іншим прикладом подібного джерела даних може слугувати новітній телескоп з міжнародного проекту SDSS (Sloan Digital Sky Survey — Слоунівський цифровий огляд неба), за допомогою якого досліджується спектр зірок на хвилях за межами видимого спектра, і за допомогою якого очікується отримання вже в перший рік його експлуатації 1,28 петабайтів інформації, що більше, ніж дав будь-який інший телескоп за всю історію. Кількість даних, доступних астрономам, постійно зростає зі зростанням кількості спостережень, які зроблені наземними чи космічними телескопами у всьому електромагнітному спектрі. Петабайтні набори даних притаманні проекту GEOSS (Global Earth Observation System of Systems — система систем глобального обстеження Землі), який базується на використанні супутникових даних для багатьох галузей господарства: прогнозування погоди та можли-

вих врожаїв, спостереження змін клімату й екологічного стану, розповсюдження стихійних лих (поводів, пожерів, засухи) та багато ін. [4]. А також проекту EISCAT (European Incoherent Scatter Scientific Association — Європейська наукова Асоціація некогерентного опромінювання), який за допомогою 100000 елементів радарних антен, розташованих у Норвегії, Фінляндії та Швеції, здійснює радарне дослідження від верхньої стратосфери до магнітосфери і далі з мікросекундною щільністю. Ця система генерує декілька терабайтів даних за день, її теперішня архітектура архівації і розподілення даних не дозволяє впоратися з таким обсягом даних, тому проектом передбачається використання нових технологій в розповсюдженні даних у відкритому доступі [5]. Перелік подібних проектів можна було б значно розширити. Слід відмітити вагомий роль міжнародної організації CODATA (Committee on Data for Science and Technology — Комітет з даних для науки і техніки) й її ініціатив ADMIRE (Advanced Data Methods and Information technologies for Research and Education — передові методи даних та інформаційні технології для досліджень і освіти), SD3 (The Scientific Data across the Digital Divide — наукові дані через цифровий вододіл) та GICSI (Global Information Commons for Science Initiative — глобальні інформаційні спільноти для наукових ініціатив) з організації збору, розміщення та використання даних, що відносяться до різних предметних галузей [6].

Завдання на сьогодні полягає в забезпеченні можливості зберігання постійно зростаючих обсягів даних і їх доступності для спільного використання, у наданні вченим ефективних засобів оброблення даних.

## **ЯКІСТЬ І СУМІСНІСТЬ**

«Сирі» дані, які отримані від інструментів або в результаті моделювання, надалі оброблятимуться. За термінологією NASA, «сирі» дані рівня 0 вивіряються та виправляються до наборів даних рівня 1, які комбінуються з іншими даними, породжуючи набори даних рівня 2. При цьому велика частина аналітичної роботи проводиться саме з наборами даних рівня 2, а в разі виявлення аномалій виконується «спуск» до даних рівня 1. Величезну кількість даних доведеться задалегідь обробляти для зменшення їх обсягу перед подальшим аналізом, але кожна стадія попереднього оброблення зменшує цю гнучкість.

Більшість дослідників витрачають багато часу на створення даних і значно менше часу на їх підготовку до розповсюдження (curating). Коли розглядається стаття для публікації, дані, що лежать в основі її висновків, рідко додаються та перевіряються. Проте це ключове положення для побудови вірогідної бази знань, що використовується і може бути доступна для майбутнього дослідження. Якість наукових даних визначається:

- властивою їм науковою суттю;
- форматом, в якому запам'ятовуються дані;
- документацією щодо наукових даних, або метаданих (даних про дату створення, джерело походження, ім'я власника даних) і контекстуальною інформацією.

У той час, як дослідники, можливо, підтримують ідею формалізації оцінки якості даних, виникає сумнів про те, чи це працюватиме на практиці. Перепони, що виникають, включають труднощі пошуку рецензентів, які мають достатню кваліфікацію, щоб зрозуміти й оцінити дані, і які згодні попрацювати, а також брак часу і грошей, потрібних для здійснення формального процесу рецензування даних. Хоча відкритий доступ до даних стає більш визнаним, він ще не є дійсністю для всіх наукових галузей, тому дослідники часто не мають достатнього стимулу, щоб робити свої дані доступними для використання іншими користувачами. Наукова спільнота традиційно визнає наукову публікацію, а не дані чи програмне забезпечення, які сприяли появі публікації. Але відношення до оприлюднення даних почало змінюватися. Нещодавно в Нідерландах фірмою DANS було винайдено метод для оцінки якості наукових даних, який називається «Data Seal of Approval» (DSA — Схвалення даних) [7]. Створений для суспільних наук DSA може також бути застосований до природознавчих наук, при цьому він гарантує, що наукові дані, які досліджуються, матимуть необхідну якість і можуть бути знайдені, визнані та використані надійно.

Варто вирішити питання про єдиний формат даних замість численних стандартів, по суті, по одному для кожної наукової дисципліни. Тому обмін даними поза кожної групи науковців стає проблематичним. Вимагається новий підхід із використанням розумних індексів і методів організації даних для скорочення обсягу пошуку; паралельної обробки та доступу до даних під час виконання пошуку у величезних наборах; потужних засобів аналізу, які можна було б застосувати до піднабору аналізованих даних.

Щоб сумісність була успішною, необхідні стандарти. Наприклад, використання метаданих може сприяти сумісності баз даних, оскільки дані можуть вживатися, інтерпретуватися і тлумачитися однаково. У суспільних науках центри даних регулярно використовують стандарт Data Documentation Initiative Metadata (метадані ініціативи документації даних). Наукові дисципліни також починають розвивати стандартизовану систему метаданих таким чином, щоб набори даних для різних наукових дисциплін були сумісними. Через Virtual Observatory (віртуальна обсерваторія) астрономами демонструють хороший приклад сумісності на практиці.

Прототипами можуть стати формати HDF (Hierarchical Data Format — ієрархічний формат даних), NetCDF (Network Common Data Form — мережева загальна форма даних) та FITS (Flexible Image Transport System — гнучка система передачі зображення), які забезпечують незалежний від платформи спосіб читання масивів, а також формати XML (eXtensible Markup Language — розширювана мова розмітки) та RDF (Resource Description Framework — модель представлення даних), що використані для побудови веб-сервісів, які забезпечують доступ через Інтернет до функцій будь-якої програми.

Рішення проблем сумісності в управлінні (менеджменті) даними є ключовим, щоб гарантувати доступність і можливість використання даних в інших наукових галузях.

## ДАНІ ПРО ДАНІ (МЕТАДАНІ)

Сьогодні для аналізу наукових даних використовуються такі технічні можливості:

- метадані та стандарти метаданих, які забезпечують розуміння даних одночасно людьми і програмами та дозволяють відстежувати походження даних;
- інструментальні засоби аналізу, які полегшують ученим процедури формування запитів і розуміння відповідей на них;
- паралельний доступ до даних, який підтримується новими індексними схемами та новими алгоритмами, і який дозволяє в інтерактивному режимі досліджувати петабайтні набори даних.

Для забезпечення простого доступу до даних, їх взаємообміну й інтеграції необхідно перейти до використання метаданих, або інформації про дані, наприклад, вимірювані атрибути, їх імена, одиниці виміру, точність, формат даних тощо. Найважливішим є те, що метадані включають інформацію про походження даних — як дані вимірювалися, генерувалися або обчислювалися. Бажано, щоб більшість із цих метаданих мала б генеруватися та використовуватися автоматично, при цьому знижувати інтелектуальне навантаження на вченого. Наявність добротних метаданих стає основною умовою спільного використання даних у різних наукових дисциплінах і для різних засобів аналізу та візуалізації даних.

Майже в кожній дисципліні нині ведуться роботи зі створення *онтологій* (званою також керованим словником). Це значно полегшує побудову інструментальних засобів, оскільки загальна згода щодо базових понять допомагає під час розробки засобів аналізу. Онтології, що використовуються комп'ютерами, допомагають побудувати семантичний веб, в якому додатки є сумісними на семантичному рівні, а не тільки синтаксично сумісними, як це реалізовано сьогодні під час використання поточного покоління веб-сервісів на основі інтерфейсів, що типізуються. Для реалізації елементарної семантичної алгебри може використовуватися SQL (Structured Query Language — мова структурованих запитів). Використання XML в сучасних системах управління базами даних (СУБД) відкриває двері для існуючих стандартів RDF та OWL (Web Ontology Language — мова опису онтологій для семантичного веб) [8].

Семантичні веб-технології можуть бути використані для:

- *інтеграції* даних, за допомогою якої дані з різних сховищ і різними форматами можуть бути застосовані в одному цілісному додатку;
- *відкриття* ресурсів і їх класифікації, щоб забезпечити кращу роботу систем пошуку даних;
- *каталогізації* опису контенту та змістових зв'язків даних, доступних на веб-сайті, сторінці або цифровій бібліотеці;
- *підтримки* інтелектуальних агентів програмного забезпечення з метою полегшення спільного використання й обміну знаннями;
- *оцінки* контенту;
- *опису* колекцій сторінок, які являють єдиний логічний «документ»;

- *опису* інтелектуальних прав власності веб-сторінок.

Семантичний веб-портал призначається для користувачів і забезпечує авторизований доступ певної групи користувачів до специфічних інформаційних ресурсів: баз даних, електронних архівів, програм тощо. Важливою функцією порталу є також організація взаємодії користувачів, наприклад, з іншими користувачами в складі проектної команди або віртуальної організації. Важлива також функція візуалізації, успіх якої визначається можливістю відображення запиту, сформульованого в концептуальному середовищі онтології предметної області, в запит(и) сервера аналізу (мета) даних.

Завдання на сьогодні: створити додаткові важливі інформаційні сервіси та надати можливість користувачеві об'єднувати дані більше, ніж з одного джерела інформації, поєднувати функціональності різних програмних інтерфейсів.

## ДОВГОСТРОКОВЕ ЗБЕРЕЖЕННЯ ДАНИХ

Для того, щоб дослідники в майбутньому могли використовувати дані, створені зараз, необхідне тривале їх збереження. Потрібний не лише менеджмент нових даних, але і дані, які створені в минулому, потребують також відповідного упорядкування та збереження. Перш за все, необхідні метадані, щоб надати дослідникам можливість зрозуміти дані, використовувати їх багато разів і повторювати, якщо потрібно, роботу колег.

Наприклад, тривале збереження даних дуже важливе в галузі електронної медицини. Ніхто не знає під час формування записів, скільки разів і коли вони знадобляться, тому спеціальні пошукові системи мають бути впроваджені для забезпечення майбутньої доступності та сумісності даних, що архівуються, з новими даними. Дуже важливо забезпечити безпеку цих записів. У разі електронних медичних записів лікарі, лікарні, страхові компанії і самі пацієнти, можливо, вимагатимуть доступу до них протягом десятиліть у майбутньому. Крім того, вони, можливо, будуть використані представниками різних країн з різними правилами, які можуть змінюватися в часі.

Еволюція технологій така, що програми, які були використані для створення інформації, імовірно, не будуть доступні в майбутньому для тих, хто побажає вивчити дані, що були архівовані. Можливе рішення полягає в збереженні даних в інваріантному для часу способі, наприклад, за допомогою мови XML, яка є набором правил для шифрування документів електронним засобом.

У галузі збереження даних бібліотеки та репозитарії можуть відігравати ключову роль. Бібліотекарі мають довгу історію підтримки записів, і могли б використовувати ці навички, щоб допомагати розповсюджувати і запам'ятовувати дані. Надаючи цю відповідальність централізованому обслуговуванню, дослідники мають бути впевнені, що дані запам'ятовуються, доступні та використовуються надійним сталим засобом для багатьох різних дисциплін. В Європі, наприклад, ЕС забезпечує доступ до інформації, що зберігається європейськими бібліотеками, архівами, музеями й аудіовізуальними архівами через Digital Libraries Initiative (Ініціативу цифрових бібліотек) [9].



Відпрацьована та поширена сьогодні реляційна база даних нікуди не зникне, але скоро з'являться інші види баз даних, наприклад, потокові. Це означатиме істотний відхід від того, що ІТ відділи та ділові користувачі використовували протягом десятиліть.

Потрібне впровадження нового покоління об'єктно-реляційних систем баз даних, які сприймають будь-який тип даних (чи то звичайне число, масив, рядок символів або складений об'єкт, такий як XML або HTML-документ) як інкапсульований тип, значення якого можуть зберігатися в полі запису. Такі системи баз даних забезпечують потужний асоціативний пошук (пошук за значенням, а не за місцем розташування), а також автоматичний паралельний доступ і виконання, що істотно для аналізу петабайтних даних. Розглянуті системи приховують поняття файла (він є всього лише контейнером даних) і працюють із колекціями даних. Вони забезпечують непроцедурний і паралельний пошук даних для швидкого знаходження піднаборів даних і можуть утворювати федеральні об'єднання багатьох різних джерел даних, дозволяючи програмам бачити їх як єдину колекцію даних. Вони також дозволяють програмам спиратися на будь-які атрибути даних.

Можна розглядати *NetCDF*, *HDF*, *FITS* та *Google Map-Reduce* (Map & Reduce — модель розподілених обчислень компанії Google) як приклади таких систем баз даних, що зароджуються: у них є мова для визначення метаданих, декілька стратегій індексування та проста мова маніпулювання даними, набір непроцедурного і паралельного програмування, набір засобів для створення, доступу, пошуку та візуалізації даних.

## ЗНАННЯ З ДАНИХ

Нині наукова громадськість орієнтується на створення потужних інтелектуальних засобів оброблення даних вільного використання у вигляді веб-сервісів, що полегшують фіксацію, організацію, аналіз, візуалізацію та публікацію даних. Ці засоби мають здійснювати інтелектуальний аналіз даних (Data mining) і навчати користувачів на основі цих даних (machine learning — машинне навчання). При цьому налаштування їх на аналіз конкретних даних здійснювалося б простою зміною скриптів, що описують потоки робіт з аналізу даних. Це нова технологія інтелектуального аналізу даних з метою виявлення прихованих закономірностей у вигляді вагомих особливостей, кореляцій, тенденцій та шаблонів. Сучасні системи вилучення знань із «сирих» даних, що використовують, засновані на методах штучного інтелекту, засобів уявлення й інтерпретації, це полегшує пошук розчиненої в петабайтних сховищах цінної інформації. В основу технології Data Mining встановлена концепція шаблонів (pattern), що відображають фрагменти багатоаспектних взаємостосунків у даних. Цими шаблонами є закономірності, властиві підрозділам даних, які можуть бути компактно виражені у формі, яка зрозуміла людині. Пошук шаблонів проводиться методами, не обмеженими апріорними припущеннями про структуру вибірки і вид розподілів значень аналізованих показників.

Для автоматизації цих засобів важливо мати якісні метадані. Збереження та поповнення цих даних в процесі обробки (походження даних) складуть основну перевагу інструментальних засобів наступного покоління.

ІТ фахівцям потрібно тісніше працювати з науковцями, щоб визначити, де аналітичні дослідження можуть бути фактично посилені, і де передбачається належне поєднання послуг, щоб оптимізувати можливості аналітики в підприємстві в цілому.

## **ПОШУК ДАНИХ**

Виявлення необхідної інформації ускладнюється з появою петабайтних наборів даних. Для того, щоб керувати своїми розподіленими наборами даних, багато наукових організацій адаптують для своїх потреб SRB (Storage Resource Broker — брокер ресурсів збереження), розроблений у суперкомп'ютерному центрі в Сан-Дієго [10]. SRB — це проміжне програмне забезпечення, завданням якого є об'єднання наборів розподілених даних та представлення цих наборів користувачу як цільних та логічно зв'язаних. Індексация та отримання даних, що збережені під керуванням SRB, виконується за допомогою каталогу метаданих MCAT, який надає користувачеві можливість отримувати доступ до даних із використанням атрибутів логічних імен, а не посилань на конкретні місця розміщення тощо. Хоча такий підхід забезпечує гнучкість, але він стикається з труднощами функціональної сумісності у випадках виконання пошукових операцій на розподіленій множині даних, різні частини якої використовують різні користувацькі розширення метаданих. Оскільки SRB використовується багатьма організаціями різного профілю, кожна із них сформувала власний набір мета даних, властивий конкретній галузі, а поточні механізми пошуку використовують певні ключові слова. Для вирішення проблеми сумісності необхідно мати детальні відомості щодо структури всіх метаданих та словників, які використано в системі. Нещодавно зроблено спробу покращити SRB через онтологічні описи для підтримки довільних схем метаданих та посилення таким чином пошукових можливостей. Здатність робити логічний висновок, використовуючи інформацію із онтологій щодо відносин, дозволяє семантичним пошуковим системам долати багато з недоліків, які пов'язані з існуючими методами пошуку. Можливість формального опису взаємовідносин між різними схемами метаданих в онтології дозволяє використовувати її як посередника для об'єднання гетерогенних мультидисциплінарних репозиторіїв даних.

Інтеграція семантичної складової у систему зберігання даних полегшує також її застосування разом із різноманітними Грід-сервісами. Сенс запропонованого рішення полягає у створенні розширення до SRB, що реалізує семантичний шар над існуючим інструментарієм. Інтерфейс системи дозволяє завантажити додаткові користувацькі онтології і це розширення є незалежним від основного ядра SRB, тому немає необхідності в перекомпіляції існуючих інсталяцій. Створений таким чином додаток потребує більше часу на свою роботу, що пов'язано із проведенням логічного висновку та аналі-

зом завантажених онтологій, при чому швидкодія пошуку залежить від вживаних механізмів логічних висновків.

В цілому, інтеграція семантичної складової дозволила покращити якість та обсяг пошукових результатів системи.

## **МЕРЕЖА ЦЕНТРІВ НАУКОВИХ ДАНИХ**

За наявності петабайтних наборів даних необхідно впроваджувати новий стиль роботи наукових центрів даних. Сьогодні у більшості випадків учений копіює файли на локальний сервер і оперує наборами даних із використанням своїх власних ресурсів. Проте все частіше набори даних є настільки великими, а прикладні програми настільки складними, що економніше перемістити програми до даних і передавати в наукові центри тільки запити користувачів та отримувати на них відповіді, а не переміщати початкові дані і додатки в локальну систему користувача. Розподіл даних і програм є штучним: ніхто не може побачити дані без використання програм, а більшість програм, у свою чергу, керується даними.

Передбачається реплікація даних центрами даних у різних географічних місцях, що забезпечить доступність даних і запобіжить їх втраті. Пошкоджені дані автоматично відновлюватимуться на основі запасних копій без переривання доступу до бази даних. Для роботи з петабайтними наборами даних вимагаються величезні масиви пам'яті і тисячі обчислювальних вузлів, що сьогодні найефективніше забезпечується Грід або хмарними обчислювальними інфраструктурами.

Варто забезпечити дієву співпрацю та сумісність даних Наукових центрів різної підлеглості: як системи 53-ьох Світових центрів даних (World Data Centers — WDC), підлеглих ICSU (International Council of Scientific and Technical Information — Міжнародний комітет із наукової та технічної інформації), так і центрів, об'єднаних за федеральним принципом: центрів SDSS, NCBI (National Center for Biotechnology Information — Національний центр біотехнологічної інформації), NSIDC (National Snow and Ice Data Center — Національний центр снігових і льодових даних), MMPDN (Materials Properties Data Network — мережа даних про властивості матеріалів), Euro VO Data Centre Alliance (Європейський альянс центрів даних про віртуальні організації), AIMS Data Centre (Азійський центр даних), Databases of the European Institute for Environment and Sustainability (база даних Європейського інституту зовнішнього середовища і сталого розвитку), Google's Palimpsest (бібліотека наукових даних компанії Google) та ін. Це дозволить організувати одночасний пошук даних у семантично об'єднаних базах цих різних центрів за допомогою Грід-сервісів для різних галузей науки і техніки (фізики, хімії, біології, геофізики).

## **ПЛАТФОРМИ ВИЗНАЧАЮТЬСЯ ДАНИМИ**

Настав час, коли архітектура платформи оброблення буде вибратися за умови впорання з величезними обсягами даних і складністю управління, а не

з умов здатності підтримувати конкретні додатки (застосування). Тобто, ІТ і наукові лідери починають змінювати свою орієнтацію на вибір платформ даних: від монолітних систем з центральними серверами для обслуговування до розподілених сервісно-орієнтованих систем, в яких додатки складаються з композиції окремих веб-сервісів, що можуть бути повторно використовані й за межами підприємства. Таким чином, прикладні послуги стають утилітами. При цьому передбачається еволюція також засобів, якими науковці ведуть дослідження, формулюючи нові виклики і можливості для ІТ. Стане ціннішою інформація для наукових структур, яку індивідууми розміщують у соціальних мережах, ніж традиційна й ізольована інформація, яку можна отримати від реєстрації індивідуумів на корпоративних веб-сайтах.

Центр тяжіння буде зміщуватися від звичних рішень для комп'ютерних систем до розвитку хмарних інфраструктур, які забезпечать збільшену функціональність і гнучкість, використовуючи змішування суспільних і приватних хмарних застосувань і платформи сервісів. Необхідно передбачити поступовий перехід до використання хмарних обчислювальних ресурсів, коли веб- і Грід-сервіси (а також Data Mining) перемістяться до SaaS (Software as a Service — програмне забезпечення як сервіс), засоби розподілення та підтримки наукових даних і ресурсів до PaaS (Platform as a Service — платформа як сервіс) і наукові центри безпосередньо до IaaS (Infrastructure as a Service — інфраструктура як сервіс).

## **ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ВИЩОГО РІВНЯ У ВИГЛЯДІ СЕРВІСІВ**

Завдяки програмним сервісам програмісти мають змогу зосередитись на тому, аби зробити їх функціональнішими із більшою ефективністю, гнучкістю, скоріше просувати їх до впровадження і ринку, також співпрацювати з предметними експертами. Далі ці сервіси стандартизуються і стають частиною застосування для вибраної прикладної платформи. Висока пропускна спроможність підключення до Інтернет і наявність стандартів дозволить програмним сервісам слугувати структурними блоками під час формування мережових застосувань вищого рівня.

## **ТЕНДЕНЦІЇ РОЗВИТКУ ІТ ПЛАТФОРМ**

Можна виділити декілька тенденцій, що характеризують сучасний прогрес у розвитку засобів ІТ:

1. Прогрес в інформаційних технологіях триває, збільшуючи темп порівняно з минулими роками.

Нові технології збереження даних з надзвичайно високими щільністю та швидкістю доступу, зростаюча пропускна спроможність оптичних ліній і сучасні нанотранзистори, забезпечують це зростання. Уже на сьогодні функціонують декілька суперкомп'ютерів із обчислювальною спроможністю, яка перевищує 10 петафлопс (1 Петафлопс =  $10^{15}$  операцій за секунду). Хоча в наступному десятилітті очікуються проявлення істотних бар'єрів для майбутнього прогресу в багатьох галузях, які пов'язані з фізичними основа-

ми сучасних технологій, але історія ІТ показала, що завжди знаходяться нові технології, які долають ці обмеження.

Цікаві можливості відкриваються у зв'язку з винайденням плівки (електронного паперу), на який можна виводити зображення. Це призведе до того, що монітори почнуть нестримно поширюватися і станеться розрив між пристроєм, через який надходить інформація, умовно кажучи, системним блоком, і монітором. Зараз ведуться активні дослідження безпроводного інтерфейсу між пристроєм та монітором, оскільки передбачається, що такі монітори будуть скрізь і користувачі зможуть під'єднуватися до ближчого та зручнішого у використанні. Під'єднання може виконуватися за допомогою мережі LTE (Long Term Evolution — назва 4G технології мобільної передачі даних), яка на відміну від Wi-Fi може забезпечити швидкість передачі до 300 Мбіт/с у повітрі.

2. Розподілені обчислення скрізь змінюють взаємодію людей і об'єктів із цифровим світом, роблячи персональні пристрої домінуючим засобом інформаційного доступу.

Перша хвиля такої взаємодії відбулася в 1980-ті рр. і зв'язала людей академічного та бізнесового суспільств електронною поштою. Друга хвиля (1990-ті рр.) дозволила науковим структурам і зростаючій кількості людей взаємодіяти через електронну пошту, мережеві браузері і засоби співпраці, засновані на Інтернет. Третя хвиля взаємодії (2000-і рр.) поєднувала штучне інтелектуальне окілля з різних об'єктів, щоденних речей і безлічі задавачів, які підключені до цифрового світу й інтегруючих ІТ, з особистою і діловою активністю людини. Четверта хвиля, що тільки розпочинається, передбачає використання природних засобів взаємодії, притаманних людям, не тільки через звичайну мову, але і мову жестів (положення тіла, пильний погляд, ручні рухи), щоб виказати свою емоцію, настрій, відношення й увагу (Multimodal Human computer interaction — багатомодальна взаємодія людини з комп'ютером), а також використання інтерфейсу, що керується контентом.

3. Застосування ІТ у майбутньому буде динамічним, адаптивним і таким, що постійно оптимізується, а також залежать від потужної ділової аналітики й управління знаннями для виживання.

Обсяги даних та інформації збільшилися не лише завдяки науковому інструментарію й автоматизації моделювання об'єктів і процесів, але й завдяки електронній комерції та автоматизації бізнесових процесів (наприклад, ERP (Enterprise Resource Planning — планування ресурсів підприємства), SCM (Supply Chain Management — системи управління ланцюгами поставок), CRM (Customer Relationship Management — управління взаємовідносинами з клієнтами)). В умовах високої конкурентоздатності тільки організації із адаптивною структурою, які здатні опрацювати ці дані, зможуть вижити та розвиватися. Досконале управління й аналіз даних дозволять успішним підприємствам швидко та гнучко реагувати на ринкові події (розпізнавати їх і реагувати). Наприклад, ланцюги статичного постачання сьогодні будуть замінені динамічною торгівлею на електронних бізнесових ринках.

4. Майбутнє ІТ-технологій пов'язане з персональними хмарними обчисленнями.

Хмарні обчислення — модель зручного за вимогою мережевого доступу до розподіленої конфігуруємої множини обчислювальних ресурсів (мереж, серверів VM (Virtual Machine — віртуальна машина), сховищ, додатків і сервісів), які можуть швидко обиратися та змінюватися з мінімальними менеджерськими зусиллями, або з мінімальною взаємодією з постачальниками послуг (рис. 4).

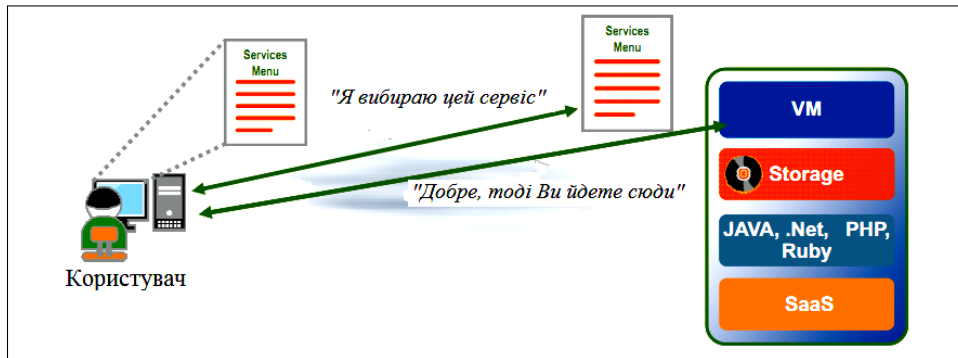


Рис. 4. Користувач у сервісно-орієнтованому середовищі

Хмарні обчислення характеризуються п'ятьма істотними властивостями (самообслуговуванням за бажанням, широкосмуговим мережевим доступом, пошуком і об'єднанням ресурсів, суттєвою гнучкістю, оцінюванням послуг); трьома згаданими вище сервісними моделями SaaS, PaaS, IaaS і чотирма моделями розгортання (приватна Хмара, громадська Хмара, соціальна Хмара, гібридна Хмара). Ключові базові технології включають: швидкі глобальні мережі, потужні, недорогі серверні комп'ютери, і високопродуктивну серверну віртуалізацію для технічного забезпечення.

Хмарна обчислювальна модель обіцяє значну економію витрат, які поєднані зі зростаючим застосуванням ІТ. Переконаливим є те, що уряди та промисловість починають користуватися цією технологією в умовах нинішніх економічних труднощів. При цьому хмарні обчислення підтримують багато традиційних підходів до організації наукових центрів даних і проектуванню програмних додатків для підприємств і менеджменту. Хмарні обчислення зараз починають використовуватися, проте досі невирішені повністю питання безпеки, сумісності та гнучкості, що гальмує ширше застосування.

## ХМАРНІ ТА ГРІД ОБЧИСЛЕННЯ ДЛЯ Е-НАУКИ

Хмарні та Грід обчислення розвиваються паралельно і використовуються в сучасній е-інфраструктурі суспільства. Дослідження зв'язків між ними і тенденціями їх розвитку дозволяють краще організувати розподілені обчислення в академічних і комерційних е-інфраструктурах, об'єднуючи можливості цих двох існуючих сьогодні важливих парадигм.

Грід є об'єднанням комп'ютерів, які зазвичай належать різним власникам і географічно розподілені, але користувачі можуть розділяти доступ до

цих об'єднаних ресурсів. Прикладами можуть бути е-інфраструктури EGEE (Enabling Grids for E-science — Грід-інфраструктура для наукових досліджень в Європі) в Європі та OSG (Open Science Grid — відкритий науковий Грід) у США.

Хмара є об'єднанням комп'ютерів, які належать одному власнику, але при цьому користувачі можуть орендувати доступ до цих ресурсів, що розділяються. Прикладами можуть бути Amazon's Elastic Compute Cloud (веб-сервіс, який надає обчислювальні потужності в Хмарі і належить компанії Amazon), Google's App. Engine (сервіс хостингу сайтів і web-додатків на серверах Google з безкоштовним ім'ям), IBM's Enterprise Data Centre (центр даних підприємства компанії ІБМ).

Грід і Хмари мають такі спільні ознаки: вони забезпечують доступ до відділених комп'ютерних ресурсів і забезпечують сервіси для користувачів.

Грід на сьогодні є досить поширеною формою організації розподілених обчислень, яка виникла з ініціативи наукової спільноти фізиків і стала звичною для інших галузей е-науки (концепція сформована у 1997 р., а Грід-система EGEE побудована в 2004 р.). Навпаки, Хмари знаходяться зараз на експериментальному етапі розвитку (виникли в 2007 р.) і їх послуги пропонуються лише декількома провідними ІТ-компаніями. Послуги Грід як форми співпраці науковців часто можуть бути безкоштовними, у той час, як Хмари надають лише комерційні послуги. Крім того, Грід концентрується на забезпеченні доступу до різних ресурсів *багатьох сайтів*, а Хмара розрахована на надання ресурсів із обчислень і пам'яті *на замовлення*.

Хмарні обчислення довели свою перевагу в ефективності та спрощенні обслуговування у випадках, коли користувачу потрібен доступ до сконцентрованих однорідних ресурсів. Але специфічні ІТ вимоги наукової спільноти (перш за все, з можливості співпраці) виправдовують подальше існування Грід-інфраструктур, тому що існуючі зараз комерційні Хмари ще не в змозі підтримувати складні сценарії спільних досліджень, які потребують науковці. Схожі і відмінні ознаки Грід і Хмари подані в табл. 2.

Звичайно, слід очікувати подальший розвиток хмарних послуг. Але Грід, що розвивається колективно науковою громадою, у своїх послугах, здається, буде завжди випереджати послуги Amazon та Google та ін., тому що науковці добре розуміють, що їм потрібно, і концентрують свої зусилля на оперативному задоволенні своїх потреб в інтересах розвитку науки.

Природно постає питання про доцільність і можливість об'єднання Грід- і хмарних сервісів, і про можливі здобутки від цього для різних галузей науки, починаючи з ядерної фізики та закінчуючи генною інженерією. Очікуються, що такі здобутки будуть корисними як для користувачів (зменшення коштовності та тривалості обчислень), так і для провайдерів (зменшення операційних витрат на підтримку функціонування Грід-сайтів).

Хмари можуть успішно використовуватися як локальні ресурси, а Грід об'єднувати ці ресурси в національні е-інфраструктури. Тому доцільно дослідити наслідки використання хмарних технологій (наприклад віртуалізації) в існуючих Грід-інфраструктурах, з одного боку, і можливості побудови Грід-сервісів поверх віртуальних інфраструктур, з іншого.

Таблиця 2. Схожість і розбіжність Грід і Хмари

Ознака	Грід (наприклад EGEE)	Хмара (наприклад Amazon)
Призначення	Забезпечує доступ до обчислювальних ресурсів і пам'яті, що спільно використовуються з комп'ютера користувача	Забезпечує доступ до обчислювальних ресурсів і пам'яті, що орендуються з комп'ютера користувача
Провайдери	Спільнота дослідницьких інститутів і університетів по всьому світі	Великі індустріальні компанії
Користувачі	<ul style="list-style-type: none"> <li>• Спільнота вчених</li> <li>• Віртуальні організації, складені вченими, які розміщені по всьому світу</li> </ul>	<ul style="list-style-type: none"> <li>• Малі та середні комерційні фірми</li> <li>• Учені, яким потрібні великі обчислювальні потужності</li> </ul>
Платники за послуги	Державні провайдери та наукові організації, які отримують громадські гранти	Провайдери Хмари сплачують за комп'ютерні ресурси, а користувачі — за їх оренду
Розташування	В обчислювальних центрах, розподілених за різними сайтами, країнами і континентами	У приватних центрах даних провайдерів Хмар, які часто централізовані
Функціонування	Грід є відкритою технологією. Користувач і провайдер можуть приймати участь у менеджменті Грід-системи	Хмара є приватною технологією. Тільки провайдер ресурсів знає точно, як в його Хмарі здійснюється менеджмент даних, обслуговування черг, організовано захист даних тощо
Причини використання	<ul style="list-style-type: none"> <li>• Немає потреби створювати та підтримувати свій власний комп'ютерний центр.</li> <li>• Можна виконати більший обсяг робіт і вирішити складніші задачі.</li> <li>• Можна обмінюватися даними з членами вашої розподіленої команди</li> </ul>	<ul style="list-style-type: none"> <li>• Немає потреби створювати і підтримувати свій власний комп'ютерний центр.</li> <li>• Можна швидко отримати додаткові ресурси під час роботи</li> </ul>
Корисність	Грід-системи було впроваджено для вирішення безлічі задач із обмеженим часом виконання, які потребують або виробляють великі обсяги даних	Хмари краще підтримають довгострокові сервіси і задачі зі значним часом виконання
Переваги	<p><b>Співпраця:</b> Грід надає платформу для розподіленої співпраці вчених.</p> <p><b>Власність:</b> провайдери ресурсів зберігають власність на ресурси, які вони внесли в Грід.</p> <p><b>Прозорість:</b> Грід-технології є відкритими, що посилює довіру та робить процеси прозорими.</p> <p><b>Пружність:</b> Грід-систему розміщено на багатьох сайтах, що зменшує ризик у випадку відмови одного з сайтів</p>	<p><b>Гучність:</b> користувач може швидко збільшити ресурси, потрібні йому для забезпечення піку активності, без довгого погодження й очікування.</p> <p><b>Надійність:</b> провайдер ресурсів бере на себе фінансові зобов'язання із забезпечення якості послуги, що надається (наприклад, Amazon повертає користувачу частково кошти, якщо обсяг послуги знизиться на 99,9 %).</p> <p><b>Простота використання:</b> користувач може порівняно просто та швидко почати вирішувати свої задачі</p>
Недоліки	<p><b>Надійність:</b> Грід базується на множині розподілених сервісів, які підтримуються розподіленим персоналом, що може призводити до їх неузгодженості.</p> <p><b>Складність:</b> будувати й експлуатувати Грід складно, тому користувач повинен мати певний рівень досвіду</p>	<p><b>Загальність:</b> Хмари не пропонують багатьох високорівневих сервісів, які притаманні Грід.</p> <p><b>Безпека:</b> користувач з цінними даними може не довіряти їх зовнішньому провайдеру.</p> <p><b>Непрозорість:</b> технології, що використовуються в Хмарах для гарантії надійності та безпеки, не є публічними.</p> <p><b>Жорсткість:</b> Хмара зазвичай розміщується на одному сайті, що збільшує ризик виходу з ладу всієї системи</p>



## ЄВРОПЕЙСЬКІ ІНІЦІАТИВИ

Для науки про інтенсивні дані (Data Intensive Sciences — науки з інтенсивним використанням даних), на жаль, поки що немає жодних Європейських проектів, подібно тим, що мають місце в галузях мережевих технологій, високопродуктивних обчислень і Грід та хмарних технологій. Тому e-IRG (e-Infrastructure Reflection Group — група супроводження e-інфраструктур) вирішила разом з ESFRI (European Strategy Forum on Research Infrastructures — Європейський форум стратегічного розвитку наукових інфраструктур) створити групу спеціального призначення, щоб узагальнити численні Європейські ініціативи, які пов'язані з менеджментом наукових даних, і сприяти визначенню загальних принципів і практик у цій галузі. Звіти e-IRG [13] підкреслюють важливість метаданих у допомозі полегшити довгострокове збереження даних і рекомендують постійно (незмінне у часі) сприяття сховищам збережених даних. Програми PARADE (Partnership for Advanced Data in Europe — співпраця для важливих даних у Європі) та GRL2020 (Global Research Library 2020 — бібліотека глобальних досліджень 2020) також підтримують такі заклики [14]. Біла книга PARADE [15] призвала до широкої співпраці з менеджменту даними для допомоги в їх збереженні й обробленні. Програма GRL2020 прийшла до подібного висновку, запропонувавши Європейським державам працювати разом, щоб розвивати інструменти та методи для довгострокового збереження даних.

Європейська Комісія розпочала низку проектів: від METAFOR, який пов'язаний із даними зі зміни клімату, до IMPACT, що досліджує структури протеїнів, які охоплюють оброблення наукових даних в європейській e-інфраструктурі e-IRG. У кінці 2009 р. було опубліковано звіт e-IRG, що містить рекомендації щодо стану та майбутнього цієї інфраструктури, а саме:

- переконати дослідників приводити описи метаданих, їх потрібно створювати одночасно з даними і вони мають бути з доступними для постачальників ресурсів і послуг;
- поліпшити доступ до наборів даних, враховуючи зв'язки з оригінальними матеріалами дослідження та джерелами походження;
- заохочувати до створення міждисциплінарних і неспецифічних для вибраної дисципліни даних;
- підтримувати комунікацію та співпрацю між виробниками даних (для кращої сумісності і багатократного використання рішень) та інфраструктурою;
- скоординувати європейські міжнародні зусилля з отримання технічних, організаційних і політичних вимог для побудови екосистем глобальної інфраструктури наукових даних (GRDIs — Global Research Data Infrastructures), оскільки сьогодні швидко зростаюча кількість наукових даних архівуються в розподілених сховищах, при чому кожен із них має специфічний доступ і технології управління, використовує різні формати файлів та інструменти метаданих.

Європейська Комісія також приділяє велику увагу розвитку хмарних технологій в Європі, хоч поки в Європі (на відміну від США) немає постачальників хмарних обчислень [18].

## ВИСНОВКИ

Доступ до інформації взагалі та до наукових даних зокрема критичний до безперервного наукового та технологічного прогресу. Після обговорення різноманітних інформаційних вимог і тенденцій розвитку ІТ можна зробити такі висновки:

- сучасні інформаційні ресурси та механізми їх використання мають враховувати постійно наростаючий «інформаційний бум», завдяки чому щороку обсяги наукових даних майже подвоюються;

- зараз йдеться про обробку петабайтних наборів даних, що потребує розроблення інтелектуальних методів організації даних для скорочення обсягу пошуку, паралельної обробки та доступу до даних під час пошуку у величезних наборах;

- за наявності петабайтних наборів даних вимагається нова методологія роботи наукових центрів, яка передбачає переміщення прикладних програм до даних і передачу в наукові центри тільки запитів і отримання відповіді, а не переміщення початкових даних і додатків у локальну систему користувача;

- для роботи з петабайтними наборами даних вимагаються величезні масиви пам'яті та тисячі обчислювальних вузлів, що сьогодні найефективніше забезпечуються Грід- або хмарними обчислювальними інфраструктурами;

- для забезпечення простого доступу до даних, їх взаємообміну й інтеграції необхідно перейти до використання метаданих, або самоописів, а також онтологій, які забезпечують розуміння даних як інструментальними засобами, так і людьми;

- нині платформа оброблення даних визначається більше самими даними, а її архітектуру орієнтовано на сервіси, з яких процедурою композиції можна за бажанням користувача складати прикладні додатки оброблення даних. Серед головних принципів такої SOA (Service-Oriented Architecture — сервісно-орієнтована архітектура) виокремлюють такі: максимальне повторне використання, модульність, здатність до поєднання (композиції), функціональна сумісність, відповідність стандартам;

- бажано об'єднати можливості хмарних і Грід-обчислень для посилення засобів е-науки.

Важливо в наступні роки забезпечити можливість збереження постійно зростаючих обсягів даних, зробити їх ефективно керованими і доступними для спільного використання, надати вченим ефективні розподілені засоби інтелектуальної обробки даних.

## ЛІТЕРАТУРА

1. *Martin Hilbert, Priscila López. The World's Technological Capacity to Store, Communicate, and Compute Information*, 10 February, 2011. — [www.sciencexpress.org / 10 February 2011 / Page 5 / 10.1126/science.1200970](http://www.sciencexpress.org / 10 February 2011 / Page 5 / 10.1126/science.1200970).
2. *Accenture Technology Vision 2011. — The Technology Waves That Are Reshaping the Business Landscape.* — [www.accenture.com/us-en/technology/technology-labs/Pages/insight-accenture-technology-vision-2011.aspx](http://www.accenture.com/us-en/technology/technology-labs/Pages/insight-accenture-technology-vision-2011.aspx).

3. *WLCG*. — <http://lcg.web.cern.ch/lcg/>.
4. *GEOSS*. — [www.epa.gov/geoss/](http://www.epa.gov/geoss/).
5. *EISCAT*. — [www.eiscat.uit.no/index.html](http://www.eiscat.uit.no/index.html).
6. *CODATA*. — [www.codata.org/](http://www.codata.org/).
7. *DSA*. — [www.datasealofapproval.org/](http://www.datasealofapproval.org/).
8. *David R. Lide*. The impact of information technology on the access to science. — <http://archive.unu.edu/unupress/unupbooks/uu07ee/uu07ee05.htm>.
9. *Digital Libraries Initiative*. — [http://ec.europa.eu/information\\_society/activities/digital\\_libraries/index\\_en.htm](http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm).
10. *SRB* (Storage Resource Broker). — [www.sdsc.edu/srb/index.php/What\\_is\\_the\\_SRB](http://www.sdsc.edu/srb/index.php/What_is_the_SRB).
11. *Paul Horn*. The Future of Information Technology (ppt), University of Colorado, 14 September, 2000. — [www.cs.colorado.edu/events/lectures/horn/horn.pdf](http://www.cs.colorado.edu/events/lectures/horn/horn.pdf).
12. *Jason Hiner*. The future of IT jobs? It's in three types of roles, 26 July, 2011. — [www.techrepublic.com/blog/hiner/the-future-of-it-will-be-reduced-to-three-kinds-of-jobs/8717](http://www.techrepublic.com/blog/hiner/the-future-of-it-will-be-reduced-to-three-kinds-of-jobs/8717).
13. *e-IRG*. — [www.e-irg.eu](http://www.e-irg.eu).
14. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. — [www.fourthparadigm.org](http://www.fourthparadigm.org).
15. *PARADE*. — [www.csc.fi/english/pages/parade](http://www.csc.fi/english/pages/parade).
16. *GRDI2020*. — [www.grdi2020.eu](http://www.grdi2020.eu).
17. *ESFRI*. — [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri).
18. *European Commission*, The future of Cloud computing: opportunity for European
19. *Cloud computing beyond*. — 2010. — [www.cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf](http://www.cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf).

Надійшла 13.02.2012