

Данни, данни - Големи данни! И какво?
Алианс за изследователски данни (RDA)

Проф. дфзн Ани Пройкиова,
съ-координатор на българския възел на Алианса за изследователски данни
Председател на Европейската стратегическа група за данни, изчисления и
цифрови инфраструктури

Алиансът за изследователски данни (Research Data Alliance, RDA) беше създаден през 2013 година като общност на съмишленици, които бяха убедени, че откритото споделяне на резултатите от научните изследвания във формата на правилно запазени данни дава широка перспектива за развитие на всички дялове на науката. Основание за тази убеденост беше, че достъпът до данните е възможност за всички учени, администратори, ученици и техните родители да направят неочаквани открития, неоткрити с рутинните анализи. Динамиката в постигането на знанията е сложен процес, чиито закони все още не познаваме добре, а любопитството често ни води до интересно познание (curiosity driven science). Интерпретацията на данните в реално време от физическия свят най-вероятно ще доведе и до въвеждането на различни нови бизнес услуги, които могат да дадат значителни икономически и социални ползи.

Идеята за отворен достъп е прекрасна и в хармония с новата парадигма – Отворена наука (Open science), но реализирането ѝ е изключително скъпо. За радост, няколко големи финансиращи организации (Европейската комисия, Националната научна фондация на правителството на Съединените щати, Националният институт за стандарти и технологии, САЩ и Министерството на иновациите на правителството на Австралия) споделиха разходите за изграждането на социалната и техническата инфраструктура, които да осигурят откритото споделяне и повторното използване на данни.

Установени бяха ясни изисквания към данните, за да се впишат в Отворената наука. Данните трябва да са намираеми (Findable), достъпни (Accessible), оперативно съвместими (Interoperable), повторно използваеми (Re-usable). Съкратено от английски: FAIR data, което звучи добре и за ухото, защото fair е честен и добър.

Алиансът RDA има приобщаващ подход, обхващащ всички етапи на жизнения цикъл на данните, ангажирайки *производителите* на данни, *потребителите* и *администраторите* (stewards), които осигуряват съхранение, обработка и обмен на данните. Алиансът успя да създаде неутралната социална платформа, където международните експерти в областта на научните изследвания се срещат, за да обменят мнения и да постигнат съгласие по различни теми, включително социалните препятствия при споделянето на данни, предизвикателствата в образованието и обучението, плановете за управление на данните и сертифициране на хранилищата им, дисциплинарна и интердисциплинарна оперативна съвместимост, както и технологични аспекти.

Алиансът е представен и управляван от Фондацията RDA, регистрирана във Великобритания като независимо юридическо лице, работеща в обществена полза. След Brexit, през април 2020 г. в Белгия беше създадена Фондация RDA

ASBL (RDAF BE) (Предприятие № 0745917132) със седалище Boulevard Louis Schmidt 24, 1040 Etterbeek. RDAF ASBL е организация с нестопанска цел, чиято мисия е да подкрепя общността на RDA в Европа. Генералният секретар на RDA, Хилари Ханахо, е главна изпълнителна директорка, а членовете на Европейския съвет Сандра Колинс, Ингрид Дило и Редакт Херцог са попечители заедно с Хилари Ханахоу.

Визията на RDA: Изследователите и иноваторите открито споделят данни в различни технологии, дисциплини и страни, за да посрещнат големите предизвикателства на обществото.

Мисията на RDA: RDA изгражда социалните и техническите мостове, които позволяват открито споделяне и повторно използване на данни.

Ръководни принципи на RDA

Отвореност - Членството е отворено за всички заинтересовани лица, които споделят Ръководните принципи на RDA. Срещите и процесите на общността на RDA са открити, а резултатите от работните групи по RDA са публично разпространени.

Консенсус (съгласие) - RDA се развива чрез постигане на консенсус сред членовете си. Процесите и процедурите на RDA включват подходящи механизми за разрешаване на конфликти.

Баланс - RDA се стреми да насърчи балансираното представителство на членовете си и заинтересованите общности.

Хармонизация - RDA работи за постигане на хармонизация между стандартите за данни, политиките, технологиите, инфраструктурата и общностите.

Общностно управление - RDA е обществен орган, ръководен от общността, състояща се от доброволчески организации, подкрепени от секретариата на RDA.

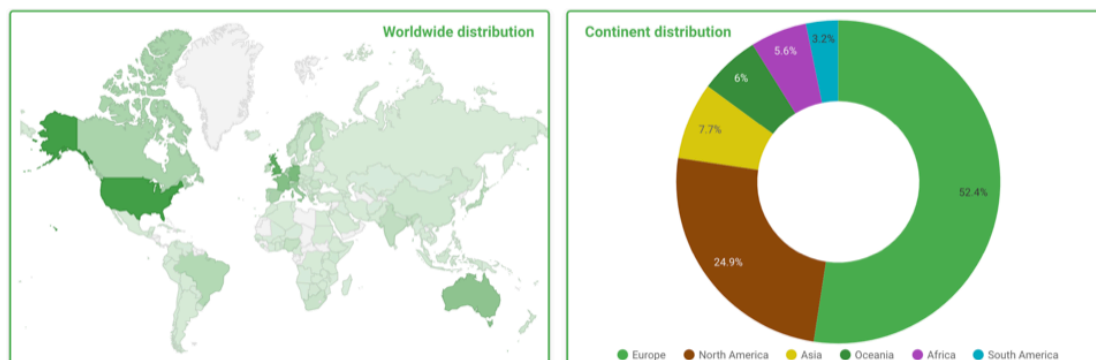
Нестопанска цел - RDA не популяризира, не одобрява или продава търговски продукти, технологии или услуги.

RDA Европа накратко

Европейският Plug-in¹ към глобалния алианс за изследователски данни RDA-получи финансиране от програмата за изследвания и иновации на Европейския съюз “Хоризонт 2020“, съгласно споразумението за безвъзмездна помощ № 777388. С почти 11 000 членове от 145 страни (август 2020 г.), RDA предоставя неутрално пространство, където членовете ѝ могат да се съберат чрез фокусирани глобални работни групи и групи за интереси, за да разработят и приемат инфраструктура, която насърчава обмена на данни и ръководените от данни изследвания и ускорява растеж на сплотена общност от данни, която интегрира участници в различни области на научните изследвания и географски граници. RDA Europe, европейският плъгин към Research Data Alliance, три пъти разшири мрежата си от национални възли и през 2019 г. съдържа вече 22 държави - Австрия, България, Великобритания, Германия, Гърция, Дания, Естония,

¹ Plug-in: Приставка (плъгин) представлява софтуерен компонент, който се инсталира в допълнение към съществуващо софтуерно приложение за да предостави на потребителя допълнителна функционалност, която не е част от базовото приложение.

Ирландия, Испания, Италия, Литва, Норвегия, Португалия, Румъния, Словения, Унгария, Финландия, Франция, Холандия, Хърватия, Чехия, Швеция. Целта на националните възли е възприемането на резултатите от RDA в съответния регион. За да подкрепи Европейската общност на експерти и практикуващи с данни от научни изследвания, RDA Europe предлага безвъзмездни средства за националните възли, изследователите в началото на кариерата им, експерти, посланици, разясняващи принципите на RDA.



Националните възли в България, Естония и Румъния се включиха при третото разширение на RDA Europe. Домакин на българския възел е Софийският университет "Св. Климент Охридски", Факултет по математика и информатика, който има опит в изграждане и поддържане на хранилища на данни и координира национални и европейски проекти в областта на големите данни (Big data). Най-големи са проектът УНИТе (Университети за Наука, Информатика и Технологии в е-обществото, 2018-2023)², финансиран от Оперативната Програма за наука и образование за интелигентен растеж и проектът ГЕЙТ (2018-2023), финансиран от програмата Туининг на Хоризонт 2020.



Основните цели на Българския възел на RDA са: включване в разработването на националната програма за цифрови технологии в страната (2020-2030), подпомагане на съвместните дейности на съществуващите национално значими проекти в областта на данните и принос към стратегията на ЕС за отворена наука.

Изискване към всеки възел е да привлича национални членове чрез ангажиране в дискуссионни форуми (кръгли маси) и разпространение на обсъдените теми, да се свърже с финансиращи организации и министерства, за да лобира за управление и споделяне на данни и да насърчава приемането на резултатите, предоставяйки поне една история или статия за списанието CODATA. Разширяването на мрежата с възли от Източна Европа е постижение за Европейския съюз, тъй като създава устойчива среда за включване на следващото поколение изследователи.

² Елисавета Гурова, УНИТе, Наука (2020)

Българският възел <https://rda.bg/> включва партньори от университетите, изследователските организации, НАЦИД, дирекция Наука – Министерство на образованието и науката.

Резултатите от работата на Алианса RDA са решенията за техническа и социална инфраструктура, разработени от работните групи или групи по интереси на RDA, които позволяват обмен на данни, обмен и оперативна съвместимост.

Тези резултати имат важно въздействие в две области: решаване на проблеми и включване и / или приемане в инфраструктурни среди от лица, проекти и организации. Като организация, целта на RDA е да разшири осведомеността и приемането на тези резултати, а оттам и тяхното въздействие, във всички региони на света. Резултатите от RDA са продукти на съответната работна група или група по интереси и трябва да бъдат демонстративно разработени и одобрени от групата. Всеки продукт трябва да има съответната работна група или група по интереси, посочена като автор, когато е подходящо.

Резултатите от RDA се класифицират като *препоръки* от Алианса за изследователски данни (официални, одобрени резултати от RDA групи), *подкрепящи резултати* (полезни решения от работните групи и групите по интереси на Алианса) или *други резултати* - повече информация можете да намерите на <https://www.rd-alliance.org/recommendations-outputs>.

Пример от естествените науки е работната група за идентификация на инструментите, която се стреми да намери решение за форматиране на данните на базата на глобално уникално идентифициране на измервателните уреди, използвани в естествените науки.

Измервателните прибори, като сензори, използвани в науката за околната среда, ДНК секвенсори, използвани в науките за живота или лабораторни инструменти, използвани за медицински области, са широко разпространени в повечето области на приложните науки. Способността за свързване на активен инструмент (инстанция) с типа инструмент и с по-широкия контекст, в който инструментът работи (включително генерирани данни, други инструменти и платформи, хора и производители и т.н.) е от изключително значение, особено за автоматизирана обработка на такава контекстуална информация и за интерпретация на генерираните данни.

Идентифицирането и описанието на модели и частни случаи на инструменти набира скорост. Няколко дисциплини използват установени контролирани речници (стандартизирани термини) за идентифициране на устройства. Напредъкът в веб технологиите на Semantic Sensor (кодиране на сензорни описания, които са машинно четими и оперативно съвместими) доведе до нови схеми за метаданни на инструмента. Някои постоянни идентификатори (PID), като универсално уникални идентификатори (UUID) или цифрови обектни идентификатори (DOI), вече се използват.

Важна активност на RDA е убеждаването на големи потребители на данни (издателски къщи, информационни агенции, университети) да възприемат предложените стандарти за събиране, съхранение и обновяване на данните, за да

бъде обменът им възможен и полезен. Пример е издателството УИЛИ (Wiley), което възприе препоръките на *Работната група за регистрация на FAIR данни: свързване на стандарти и стандарти за метаданни, хранилища и политики*. В свое интервю, Крис Граф - директор, Изследователска почтеност и издателска етика в Wiley, казва:

“Wiley е издател на изследователски книги и ние обичаме това, което правим. Защото обичаме любопитството на изследователите, което разпалва въпроси. Тези въпроси водят до открития. Водещи до решения на предизвикателства, пред които всички сме изправени. Ние от Wiley свързваме тази работа, която изследователите правят с изобретатели и новатори. Ние помагаме да превърнем новите знания, които изследователите създават в прогрес, за да променят света. RDA помага на Wiley да направи откритото публикуване на научните изследвания по-лесно за изследователите като резултатите на *Работната група на RDA за обмен на връзки с ученици, Работната група за регистрация на FAIR данни* и *Групата по интереси за стандартизация и прилагане на политиката за данни*”.

Всеки заинтересован може да получи повече информация за групата FAIR данни от уеб-страницата ѝ: <https://rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html>

Българският възел на RDA Еуропа организира семинари, конференции и целеви срещи със своите членове.

Конференцията на 24 юли 2020 г., беше посветена на управлението на данни и беше проведена електронно с участието на колеги от Албания, Германия, Испания и Франция. Българските участници от Института по информационни и комуникационни технологии (ИИКТ) – БАН, Института по математика и информатика (ИМИ) – БАН и софтуерната компания Ontotext със специалисти от отдел „Наука“ на Министерството на образованието и науката (МОН), Националният център за информация и документация, Софийският университет, изнесоха поканени доклади, които бяха обсъдени и много интересни предложения за бъдещи дейности бяха направени.

Г-жа Карина Ангелиева, заместник-министър на образованието и науката, представи националната политика в областта на отворената наука. Тя изтъкна, че първоначалните политики са били насочени към изискването за открито разпространение на резултатите от научните изследвания, постигнати от публично финансирани изследвания. Сега вече обхватът на тези политики е разширен и предвижда насърчаване на практиките на отворената наука във всички сегменти на научните изследвания – достъп до оригинални данни, както и достъп до изследователски инфраструктури.

Проф. Красен Стефанов, координатор на Националната научна програма по Информационни и Комуникационни Технологии в науката, образованието и сигурността (<https://npict.bg/>), обясни връзката между дейностите на българския възел на RDA Еуропа и целите на програмата. Възможностите на българските изследователи да участват в работни групи по RDA бяха представени от проф. Мария Нишева, координатор на RDA BG Node и заместник-декан на Факултета по математика и информатика (ФМИ). Доцент Димитър Василев, ФМИ, член на

работната група на RDA за препоръки за оперативна съвместимост на данните за пшеницата, говори за модел на интеграция на данни за хлебната пшеница. Взаимодействията между RDA Еуроге и Европейския облак за отворена наука (EOSC) бяха представени от проф. Ана Пройкиова, съ-координатор на българския възел на RDA, член на групата за взаимодействие между Европейския Стратегически Форум (ESFRI) и EOSC, председател на стратегическата работна група на ESFRI за данни, изчислителни и цифрови инфраструктури.

Проф. Тодор Гюров, ИИКТ – БАН, представи проекта „Национални инициативи за отворена наука“ (NI4OS, <https://ni4os.eu/>), финансиран от Хоризонт 2020, и възможните съвместни дейности с българския възел на RDA Еуроге.

Г-жа Християния Анчева представи българския портал за отворена наука – новосъздадена платформа, финансирана от МОН. Проф. Петър Станчев, ИМИ – БАН, говори за възможните бъдещи дейности на българския възел на RDA Еуроге и OpenAire-България (<https://www.openaire.eu/os-bulgaria>).

Д-р Олег Константинов предостави връзката към филма „RDA BG Node в действие“:

<https://www.youtube.com/watch?v=4HWsfZtQt7k&feature=youtu.be>

който той композира от интервюта, взети от членове на българския възел на RDA: проф. Ана Пройкиова, проф. Мария Нишева, проф. Кирил Симов и д-р Димитър Илиев, координатори на ClAda-BG (<https://clada-bg.eu/en/>), проф. Красен Стефанов, г-жа Ваня Грашкина, директор на НАЦИД, доц. Елиза Стефанова, заместник ректор на Софийския университет, г-жа Карина Ангелиева, заместник-министър със сектор „Наука“, МОН.

Българският възел на RDA Еуроге чрез своите членове участва в консултациите при изготвяне на националната стратегия за цифровизация. Достъп до пълния текст на стратегията е осигурен чрез <https://rda.bg/en/news>, <https://rda.bg/en/node/31> : Р Е Ш Е Н И Е № 493, от 21 юли 2020 година ЗА ПРИЕМАНЕ НА НАЦИОНАЛЕН СТРАТЕГИЧЕСКИ ДОКУМЕНТ „ЦИФРОВА ТРАНСФОРМАЦИЯ НА БЪЛГАРИЯ ЗА ПЕРИОДА 2020-2030 Г.“

Голямото предизвикателство пред всички партньори в българския възел на RDA Еуроге е консолидирането на работата им така, че нейните резултати да са полезни за всички участници в сложния процес на възприемане на Отворената наука като нов стандарт за образование и научни изследвания.

София
12.08.2020 г.