ARK - РЕШЕНИЕ ЗА ПОСТОЯНЕН ИДЕНТИФИКАТОР ЗА ОНЛАЙН ИЗДАНИЕТО НА СПИСАНИЕ "БЪЛГАРСКА НЕВРОЛОГИЯ"

К. Камбуров

Софийски университет "Св. Климент Охридски"
Катедра "Библиотекознание, научна информация и културна политика"

ARK - A PERSISTENT IDENTIFIER SOLUTION FOR ONLINE EDITION OF THE "BULGARIAN NEUROLOGY" JOURNAL

C. Kamburov

Sofia University "St. Kliment Ohridski"
Department of Library and Information Studies and Cultural Policy

ABSTRACT

ARK (Archival Resource Key) persistent identifiers are a type of Uniform Resource Locator (URL) designed to provide long-term access to digital objects. This article explores the structure, functionality, and applications of ARK identifiers, highlighting their role in digital data storage and management. The advantages and challenges associated with ARK identifiers are discussed, providing an overview of their importance. The text reviews and analyzes implementation and practical application of ARK for the needs of the online edition of the journal "Bulgarian Neurology".

KEY WORDS: ARK, persistent identifiers

РЕЗЮМЕ

ARK (Archival Resource Key) постоянните идентификатори са вид Uniform Resource Locator (URL), предназначен да осигури дългосрочен достъп до цифрови обекти. Статията изследва структурата, функционалността и приложенията на ARK идентификаторите, подчертавайки тяхната роля в цифровото съхранение и управление на данни. Обсъждат се предимствата и предизвикателствата, свързани с ARK идентификаторите, предоставяйки преглед на тяхното значение. Показано е внедряването и практическото приложение на ARK за нуждите на онлайн изданието на списание "Българска неврология".

КЛЮЧОВИ ДУМИ: ARK, постоянни идентификатори

Въведение

В дигиталната ера необходимостта от постоянни идентификатори (Persistent Identifiers, PID) става все по-важна, за да се осигури дългосрочна достъпност и цитируемост на цифровите ресурси. Идентификаторите ARK, разработени от Калифорнийската цифрова библиотека (California Digital Library, CDL), са едно такова решение. ARK са проектирани да осигурят стабилна и постоянна връзка с цифрови обекти, дори когато основните системи за съхранение и софтуер и технологии за онлайн публикуване се развиват и съответно променят.

Archival Resource Key (ARK, ключ за архивен ресурс) е многоцелеви URL адрес, подходящ да бъде постоянен идентификатор за информационни обекти от всякакъв тип. Той се използва широко от издатели на научни издания, библиотеки, центрове за данни, архиви и музеи. Служи за предоставяне на надеждни препратки към съществуващи, електронно публикувани и достъпни в Интернет цифрови обекти (статии, файлове, набори от данни и др.). ARK представлява и схема за единен идентификатор на ресурси (URI, Universal Resource Identifier).

URL адрес, съдържащ Archival Resource Key, се отличава с етикета "ark:" след името на хоста на URL адреса. Подразбиращо се в дизайна на ARK е, че постоянството е въпрос на услуга, а не свойство на синтаксис за именуване.

Постоянните идентификатори и ARK

В края на миналия век са разработени стандарти за постоянни идентификатори за уеб ресурси, включително URN (Uniform Resource Name), PURL (Persistent Uniform Resource Locator), Handle и DOI (Digital Object Identifier). В тези стандарти косвените идентификатори на определен електронен ресурс (текстов файл, изображение, статия, книга и т.н.) се преобразуват в URL адреси, които се променят с времето.

Схващането е, че този допълнителен идентификатор (различен от текущия URL адрес на електронния ресурс) ще допринесе за устойчивостта във времето на електронно публикуваните документи в уеб мрежата. Научни данни и резултати публикувани в електронна форма, които променят местонахождението си и точката си за достъп биха били налични на ново място (друг URL) и постоянния идентификатор ще бъде насочен към новото място в мрежата.

Например:

DOI номерът *10.3389/fneur.2024.1420696* (зададен като обръщение от вида *https://doi.org/10.3389/fneur.2024.1420696*) пренасочва браузъра към URL:
https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2024.1420696/full, където е публикувана електронно статията:

Stoyanova-Piroth G, Milanov I, Stambolieva K. Association between pain threshold and manifested pain assessed using a PD-specific pain scale in Parkinson's disease. Front Neurol. 2024 Jul 25;15:1420696.

Тази статия има и допълнителни постоянни идентификатори в мрежата на PubMed - *PMID: 39131046; PMCID: PMC11312374.*

При смяна на домейна от страна на издателя или смяна на системата за управление на съдържанието, дори закриване на конкретния сайт този URL няма да бъде валиден.

Предполага се, че ако публикацията бъде достъпна на друго място в мрежата, то създаденият кратък път в базата данни на doi.org, а именно *10.3389/fneur.2024.1420696* ще бъде пренасочен към новото точно местонахождение с различен URL.

В началото на нашия век Джон Кунзе (John Kunze) от Калифорнийския университет и Р. Роджърс (R. Rodgers) от Националната медицинска библиотека на САЩ публикуват първия си проект на „Схема за постоянен идентификатор на ARK". Тази методология е проектирана в отговор на нуждите на техните две организации. Обяснявайки мотивите им за създаването на нова система, Кунзе посочва, че за разлика от децентрализираната структура на мрежата, Handle и DOI са свързани с централизирани системи, които допълнително

имат и такса за включване. Той ги определя като противоречащи „на имплицитния принцип, че интернет стандартите не трябва да одобряват контрола от който и да е субект върху достъпа до мрежовите ресурси на друг субект“. Други решения, като URN са безплатни, но им липсват услуги като резолвер [начинът, по който номерът ще се конвертира в URL адрес].

За разлика от тези системи за постоянни идентификатори, схемата ARK предполага, че „постоянството е въпрос на услуга, ... нито е присъщо на даден обект, нито му е предоставено от определен синтаксис за именуване. Тогава, това което един идентификатор следва да направи, за да реши проблема с постоянството, е да посочи ангажимента на организацията издател.“ Съответно, в схемата ARK, идентификаторите ще се отнасят не само за уеб ресурс, но и за „обещание за управление“ и метаданни за ресурса. Ако уеб сървърът е запитан с ARK, той трябва да върне самия ресурс или някакъв негов заместител, като например „съдържание вместо цял документ“. Ако въпросителен знак бъде добавен към ARK обаче, той трябва да върне и описание - метаданни - кой, какво, кога и защо се отнасят до цифровия обекта. Схемата също така включва ръководство за електронни цитати на ресурси и прост формат за структуриране на тези метаданни. Ако бъдат добавени два въпросителни знака, сървърът следва да върне политиките на доставчика относно „постоянство на обекти, именуване на обекти, адресиране на обектни фрагменти...”.

Калифорнийската цифрова библиотека интегрира схемата ARK през 2002 г. и публикува софтуера Noid (Nice Opaque IDentifiers) за управление на ARK и други идентификатори през 2004 г. Други големи потребители на ARK са Portico, Internet Archive и Националната библиотека на Франция, Лувъра, Смитсониън и др.

През 2018 г. Калифорнийската дигитална библиотека и DuraSpace обявяват сътрудничество, наречено ARK Alliance, за изграждане на международна общност около ARKs и тяхното използване. Към 2024 г. над 1400 институции са се регистрирали да използват ARK като са създадени над 8 милиона постоянни идентификатора ARK.

Българското дружество по неврология е регистрирано като издател (53357) от 2023 г. а към 2024 г. са съставени над 160 ark идентификатора – за всички публикувани статии в списанието от 2019 г. до този брой, включително.

Структура на ARK идентификаторите и внедряване в онлайн изданието на списание “Българска неврология”

https://NMA/ark:/NAAN/Име[Квалификатор]

NAAN: Name Assigning Authority Number - задължителен уникален идентификатор на организацията, която първоначално е наименувала обекта

NMA: Орган за картографиране на имена - незадължително и заменяемо име на хост на организация, която в момента предоставя услуга за обекта

Квалификатор: незадължителен низ, който разширява основния ARK, за да поддържа достъп до отделни йерархични подкомпоненти на обект (например допълнителна информация, която може да бъде добавена към ARK за указване на определен аспект или версия на ресурса.)

Наример за статията от текущия брой на списание Българска неврология (том 25, бр.2)

Вълкова, М., & Гергелчева, В. (2024). Синдром на полиневропатия, органомегалия, едрокринопатия, м-протеин и кожни промени (POEMS) – литературен обзор. Българска Неврология, 25(2), 51–54. изтеглен от https://www.nevrologiabg.com/journal/index.php/neurology/article/view/172

е присвоен ARK

https://n2t.net/ark:/53357/NEU.v25i2.172

където:

https://n2t.net/ е организацията за картографиране на имена (може да бъде променяна, за разлика от други системи за постоянна цифрова идентификация)

ark:/53357/ е етикетът ark, последван от конкретния Name Assigning Authority Number. В случая, това е издателският номер на Българското дружество по неврология

NEU.v25i2.172 е уникалният електронен ресурс, т.е. конкретната статия с описание на абстрактната информация, авторската отговорност и метаданните на български и английски език.

При обръщение към етикета, браузърът отправя запитване към n2t.net, който пренасочва към сървъра на издателя - 53357. Системата на издателя връща отговор за URL адреса на обекта NEU.v25i2.172.

За да може браузърът чрез ARK да достигне до този ресурс, който се намира на URL:
https://www.nevrologiabg.com/journal/index.php/neurology/article/view/172 е необходим резолвър.

Резолвърът е уебсайт, който е специализиран в препращането на входящите идентификатори. За самото препращане се използва термина - резолюция. Пренасочването обикновено се реализира с програма и/или допълнение (плъгин) към внедрената система за редакция на съдържание, като за списание "Българска неврология", това е Open Journal System.

Добавката за ARK за OJS предлага възможности за асоциране на суфикси към различни цифрови ресурси.

Може да бъде зададен ark номер за изданието (контретен брой на списанието) като цяло.

Ползваният вариант за онлайн изданието на списание "Българска неврология" е идентификатор на ниво статия. Т.е. пренасочването е към самата статия, а не към конкретен вариант на типографското й представяне (PDF, HTML, EPUB и т.н.)

Шаблонът използван за обектите е от вида: NEU.volumeXXXissueXXX.OJSID. Съответно за примера с NEU.v25i2.172 абревиатурата се отнася както следва:

NEU е за "Българска Неврология" с ISSN 2815-2522 (Online); v25 - том 25; i2 - бр.2; 172 - ID в OJS за съответната публикувана статия.

Заключение

Предимствата на системата включват:

- Стабилност - осигурява се относително независима и сигурна връзка към цифрови ресурси;
- Оперативна съвместимост - може да се използва в различни платформи, подобрявайки интеграцията на ресурсите;
- Поддръжка на метаданни - връзките може да бъдат обогатени и с метаданни, подобрявайки откриването и управлението на тези ресурси;
- Мащабируемост - може да се използва за широк набор от цифрови обекти, от отделни файлове до големи набори от данни.

Предизвикателствата пред разпространението на схемата ARK включват участието на организации и институции, допълнителен квалифициран човешки ресурс, което може да бъде пречка за по-широкото използване.

Постоянните идентификатори на ARK са ценен инструмент за осигуряване на дългосрочна достъпност и цитируемост на цифровите ресурси. Тяхната стабилност, възможност за преобразуване и гъвкавост ги прави подходящи за широк спектър от приложения в цифрови библиотеки, управление на изследователски данни, архивно съхранение и публикуване. Въпреки че съществуват трудности по отношение на приемането на системата и управлението, ползите от идентификаторите на ARK ги правят критичен компонент от пейзажа за цифрово съхранение.

ЛИТЕРАТУРА

1. ARK Alliance. General identifier concepts and conventions. https://arks.org/about/identifier-concepts-and-conventions/
2. ARKs discussion group. https://groups.google.com/g/arks-forum?pli=1
3. California Digital Library, Archival Resource Key (ARK) Identifiers, 2018. http://n2t.net/e/ark_ids.html
4. Clark T. Using ARKs in AI for biomedical research. 2022. [video] https://youtu.be/7HYYR0tGGcw?t=542
5. Kunze J., Bermès E. ARK Identifier Scheme (v.39) for technical details. 2024. https://arks.org/assets/documents/2024/ark_spec_39.pdf
6. Kunze J., Rodgers. R. The ARK Identifier scheme, 2008. https://n2t.net/ark:/13030/c7cv4br18
7. Kunze J., Winston. D. Getting Started with ARK Persistent Identifiers. 2024. [video] https://youtu.be/-RkMGFCGRic
8. Kunze J. Towards Electronic Persistence Using ARK Identifiers, 2003. California Digital Library. University of California. https://n2t.net/e/Towards_Electronic_Persistence_Using_ARK_Identifiers.pdf
9. Meyerl, J. ARK Alliance: An Interview with John Kunze. bloggERS. Society of American Archivists Electronic Resources Section. 2021. https://saaers.wordpress.com/2021/09/14/ark-alliance-an-interview-with-john-kunze/
10. Resolving Names to Things. The N2T API and UI. 2021. https://n2t-dev.n2t.net/e/n2t_apidoc.html
11. Vera Y.P. ARK Plugin Guide for OJS. https://github.com/yasielpv/pkp-ark-pubid/files/8398101/ARK.plugin.guide.for.OJS.pdf
12. Vera Y.P. ARK plugin for the Open Journal System (OJS). https://github.com/yasielpv/pkp-ark-pubid
13. Wimalaratne SM, Juty N., Kunze J., Janée G., McMurry JA, Beard N., Jimenez R., Grethe JS, Hermjakob H., Martone ME, Clark T. Uniform resolution of compact identifiers for biomedical data. Sci Data. 2018 May 8;5:180029. doi: 10.1038/sdata.2018.29. PMID: 29737976; PMCID: PMC5944906.

Адрес за кореспонденция:
Кирил Камбуров
бул. "Цариградско шосе", № 125, блок 1
София, България
E-mail: k_kamburov@phls.uni-sofia.bg

Correspondence address:
Cyril Kamburov
125, Tsarigradsko Shose Blvd. Block 1
Sofia, Bulgaria
E-mail: k_kamburov@phls.uni-sofia.bg