

# ePIC

Persistent Identifiers for eResearch

<https://www.pidconsortium.eu>

# Úvod

Konsorciem ePIC cílí na vytvoření a udržování jednotného systému pro registraci a ukládání perzistentních identifikátorů pro vědeckou komunitu. Pro automatizaci registrace vzniklo API, které umožňuje přidávat například kontrolní součty nebo ukazatele na metadata.

## Persistentní identifikátory

S rostoucím množstvím vědeckých dat a jejich vzájemnou provázaností roste i potřeba zdroje jednoznačně identifikovat. Pro vědecké výzkumy je velice důležité, aby byly ověřitelné, což bez použitých dat může být u většiny výzkumů až nemožné. Z tohoto důvodu je důležité mít možnost se odkázat na data pomocí neměnného identifikátoru, který se odkazuje právě na tato data a na žádná jiná.

Vhodným přírovnáním jsou ISBN (International Standard Book Number); persistentní identifikátor pro knihy, jež umožňuje jednoznačně se odkázat (případně citovat) na konkrétní knihu.

Důležitými vlastnostmi persistentních identifikátorů jsou:

- trvanlivost
- actionable – je možné pomocí nich dohledat odkazovaný objekt (= provést rezoluci)
- jednoznačnost
- nepodléhají takzvanému „link rot“ (proces, při kterém stránka, na kterou odkaz ukazuje, „zemře“)

## Stav ePIC

ePIC bylo založeno roku 2009 konsorciem evropských partnerů. Aktuálními členy jsou [1]:

- CLARIN: European Research Infrastructure for Language Resources and Technology
- CNIC: Computer Network Information Center, Chinese Academy of Sciences, China
- CSC: IT Center for Science
- CSCS: Swiss National Supercomputing Centre

David Dobrovolný (456297)  
PV070 Digitální knihovny  
25. 11. 2018

- DKRZ: Deutsches Klimarechenzentrum
- GRNET: Greek Research and Technology Network
- GWDG: Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen  
PDC KTH: Center for High-Performance Computing at the KTH
- SURFsara: SURFsara is part of SURF, the collaborative ICT organization for higher education and research in the Netherlands

Rada konsorcia je složena ze zástupců plnohodnotných členů. Předseda rady je volen každoročně. Aktuálně jím je Ulrich Schwardmann z GWDG.

## Handle System

Handle System je systém pro přidělování a rezoluci persistentních identifikátorů. Byl implementován roku 1994 a spravován CNRI (Corporation for National Research Initiatives) až do roku 2015. Nyní je spravován DONA Foundation.

Identifikátor se skládá z prefixu, který identifikuje „naming authority“ a sufixu, který identifikuje samotný objekt. Prefix se skládá z jedné nebo více částí oddělených tečkami zaznamenávající hierarchii autorit. Sufix, nebo také lokální jméno, může být složený z libovolných znaků Unicode USC-2.

Příkladem identifikátoru může být: 20.1000/100.

- 20 odpovídá prefixu pro CNRI
- 1000 odpovídá podřízené autoritě v rámci prefixu 20
- 100 je lokální jméno objektu

Další příklady prefixů:

- 10 – Digital Object Identifier System
- 11 – identifikátory přiřazované ITU (International Telecommunication Union)
- 21 – přiřazované GWDG (ePIC identifikátory)

Je přiřazeno okolo 30 prefixů a 30 milionů identifikátorů [3, slide 7].

## Přidělování identifikátorů

Přidělování ePIC identifikátorů je rozděleno na přidělování prefixů a sufixů. Prefixy jsou přiřazovány institucím nebo velkým projektům samotným konsorciem ePIC. Tyto instituce poté přidělují sufixy. Zároveň zajišťují unikátnost, persistenci a obecnou správu přiděleného prefixu.

## Rezoluce identifikátorů

Pro rezoluci Handle System identifikátorů (tzn. i ePIC identifikátorů) je možné použít proxy <http://hdl.handle.net> s identifikátorem.

Například rezoluce pro identifikátor 21.012/xyz-123 by vypadala následovně:  
<http://hdl.handle.net/21.012/xyz-123>.

Při rezoluci si proxy server vyžádá z Global Handle Registries (GHR) adresu Local Handle Server (LHS) pro daný prefix. Local Handle Server spravuje všechna lokální jména náležící jednomu prefixu.

## Kompatibilita s DOI

Aby se zabránilo problémům způsobeným různou interpretací znaků v různých kódováních, jsou ePIC PID sufixy omezeny na tisknutelné znaky ASCII (znaky s hodnotou 33-126). To zajišťuje kompatibilitu s Digital Object Identifier (DOI).

Některé znaky by mohly zamezit rezoluci identifikátorů v URL. Například nejsou v URL povoleny nebo mají vedlejší účinek (mezi takové znaky patří například ; @ ? = { } [ ]).

## Persistence

Všechny ePIC identifikátory, kromě těch s prefixem ve tvaru 21.Txyz (kde xyz je libovolný validní řetězec), jsou perzistentní a není možné, aby byly smazány. Identifikátory s prefixem ve tvaru 21.Txyz jsou používány pro testovací účely a je možné, a někdy i nutné, je smazat. Oproti ostatním identifikátorům jsou odlišeny písmenem „T“ (test).

Identifikátor nesmí být smazán ani v případě, že objekt, na který odkazuje, již není dostupný (i z důvodu jeho smazání). Takové identifikátory stále nesou metadata již nedostupného objektu s případným důvodem jeho nedostupnosti.

## API

Pro usnadnění práce s persistentními identifikátory ePIC bylo vytvořeno API. API je dostupné na stránce GitHub (<https://github.com/pidconsortium/ePIC-API-v2>). Protože si uživatelé částečně definují svůj sufix identifikátoru, API nedefinuje celou strukturu. Je ale dostatečně obecné, aby splnilo potřeby různých vědeckých komunit. Bližší specifikace, zdrojový kód a dokumentace jsou dostupné na GitHubu.

## Vlastní hodnocení

Můj pohled na ePIC je velice pozitivní. Myslím si, že je velká potřeba jednoznačně identifikovat nejen knihy a vědecké práce, ale i samotná data využitá v těchto pracích. Důležité jsou i vztahy mezi těmito daty a reprezentace těchto vztahů.

S rostoucím množstvím dat (v roce 2017 CERN překročil hranici 200 petabytů uložených dat [2]) roste i potřeba efektivní organizace, skladování a vyhledávání v těchto datech. Tyto potřeby ePIC API dobře pokrývá.

Další výhodou identifikátorů ePIC je jejich kompatibilita s DOI a persistence i v případě nedostupnosti odkazovaného objektu.

## Zdroje

- [1]: [https://www.pidconsortium.eu/?page\\_id=74](https://www.pidconsortium.eu/?page_id=74)
- [2]: <https://home.cern/news/news/computing/cern-data-centre-passes-200-petabyte-milestone>
- [3]: [https://www.rd-alliance.org/sites/default/files/attachment/20160901-RDA\\_EU\\_View\\_on\\_PID\\_Systems\\_Garching-Ulrich\\_Schwardmann-Building\\_and\\_Maintaining\\_a\\_Registry\\_for\\_PID\\_Info\\_Types.pdf](https://www.rd-alliance.org/sites/default/files/attachment/20160901-RDA_EU_View_on_PID_Systems_Garching-Ulrich_Schwardmann-Building_and_Maintaining_a_Registry_for_PID_Info_Types.pdf)
- <https://www.pidconsortium.eu>

David Dobrovolný (456297)  
PV070 Digitální knihovny  
25. 11. 2018

- [https://en.wikipedia.org/wiki/Persistent\\_identifier](https://en.wikipedia.org/wiki/Persistent_identifier)

## Metadata

<dc:creator>David Dobrovolný</dc:creator>

<dc:title>ePIC</dc:title>

<dc:date>2018-11-25</dc:date>

<dc:subject>Konsorcium pro správu persistentních identifikátorů</dc:subject>

<dc:description>Esej o konsorciu, které zajišťuje registraci a správu persistentních identifikátorů (PID) pro vědecká data.</dc:description>

<dc:type>text</dc:type>

<dc:language>cz</dc:language>