

Metadane prezentacji

Projekt Open Data Support jest finansowany przez Komisję Europejską w ramach umowy SMART 2012/0107 „Część 2: Świadczenie usług na rzecz publikacji, udostępniania i wykorzystywania otwartych danych publicznych w Unii Europejskiej poprzez istniejące portale otwartych danych” (nr umowy 30-CE-0530965/00-17).

© Komisja Europejska, 2014



**OPEN
DATA
SUPPORT**

Moduł szkoleniowy 2.3

Stałe identyfikatory URI – tworzenie i zarządzanie

Cele szkolenia

Po zakończeniu modułu szkoleniowego powinieneś wiedzieć:

- czym jest identyfikator URI.
- dlaczego ważna jest stałość identyfikatora URI.
- jak tworzyć stałe identyfikatory URI zasobów danych i zarządzać nimi.

Treść

Niniejszy moduł zawiera...

- wprowadzenie do identyfikatorów URI.
- zasady tworzenia stałych identyfikatorów URI.
- wymogi serwisu dedykowanego stałych identyfikatorów URI.

Identyfikatory URI

jako wspólne identyfikatory elementów, np. osób, budynków, miejsc, zasobów informacji...

Czym jest identyfikator URI?

URI to

**“związły łańcuch znaków
identyfikujący abstrakcyjny lub
materialny zasób”**

[TBL et al, 2005].

- **“związły”**, tj. łańcuch nie może składać się ze spacji, tabulatorów i innych niepotrzebnych znaków;
- **“abstrakcyjny lub materialny”**, tj. URI może odwoływać się do elementu świata rzeczywistego, np. osoby, budynku, ale także elementów abstrakcyjnych, np. usługi czy dokumentu sieci Web.

Przykład



Kraj, np. Belgia

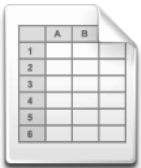
- <http://publications.europa.eu/resource/authority/country/BEL>



Publications Office

Organizacja, np. Urząd Publikacji UE

- <http://publications.europa.eu/resource/authority/corporate-body/PUBL>



Zbiór danych, np. lista organów krajowych wg kraju

- <http://publications.europa.eu/resource/authority/country/>

Najważniejsze właściwości

- **Stażość**, tj. identyfikator URI jest przypisany na stałe do danego zasobu. Jest stabilny, nie zmienia się ani nie znika.
- **Możliwość odwołania**, tj. klient użytkownika może odwołać się do identyfikatora URI w internecie i otrzymać treściwą informację zwrotną.
 - Jeżeli klient użytkownika jest *przeglądarką internetową*, wówczas informacja powinna wrócić w formie dokumentu HTML odczytywanego przez człowieka.
 - Jeśli klient jest *klientem RDF*, wówczas RDF powinien zostać zwrócony z *tego samego identyfikatora URI*.
- **Jednoznaczność**, tj. identyfikatory dokumentów sieci Web nie powinny mylić się z identyfikatorami innych zasobów.
 - Powinny istnieć osobne identyfikatory URI odwołujące się do autora strony internetowej i samej strony internetowej.

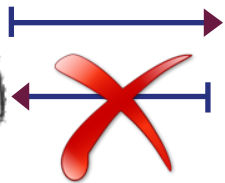
Najważniejsze założenia

- Aby tworzyć i zarządzać identyfikatorami URI, należy być **właścicielem** danej domeny i posiadać uprawnienia administratora.
- Domeny rządowe są zazwyczaj **zarządzane przez agendę rządową**. Przed rozpoczęciem skonsultuj się z kolegami.
- Stałe identyfikatory URI musi **wspierać zaufana infrastruktura internetowa**, zapewniona przez Twoją firmę lub przez inną organizację, np. w formie zasobu współdzielonego. Przed rozpoczęciem skonsultuj się z kolegami z działu IT.

Gdy identyfikator URI jest niestały lub nie można się do niego odwołać

Wyobraźmy sobie następującą sytuację...

Wyszukajmy opis „Irlandii” na liście kodów państw.



http://foo.org/concept_tid

Nie znaleziono



Tworzenie stałych identyfikatorów URI dla zbiorów danych

Zalecenia i przeciwwskazania

Używaj standardowego formatu URI

`http://{domain}/{type}/{concept}/{reference}`

- **{domain}** to połączenie hosta i odpowiedniego sektora.
- **{type}** jedna z nielicznych możliwych wartości określających rodzaj identyfikowanego zasobu. Typowe przykłady:
 - 'id' lub 'item' w przypadku elementów świata rzeczywistego;
 - 'doc' w przypadku dokumentów opisujących te elementy;
 - 'def' w przypadku pojęć;
 - 'set' w przypadku zbiorów danych;
 - ciąg zależny od kontekstu.
- **{concept}** może być zbiorem, rodzajem identyfikowanego elementu świata rzeczywistego lub nazwą modelu pojęciowego;
- **{reference}** konkretny element, termin lub pojęcie.

Twórz URI wykorzystując dostępne identyfikatory

- Istniejące identyfikatory zasobów, np. klucze bazy danych, powinny zostać uwzględnione w identyfikatorze URI.
 - Wykorzystuj identyfikatory stałe.
 - Wykorzystuj identyfikatory standardowe zamiast wewnętrznych kodów funkcjonujących w jednym systemie.
- Przykładowo, jeśli firma posiada w krajowym rejestrze gospodarczym identyfikator AB123456, identyfikator URI firmy może wyglądać następująco:

`http://businessdata.gov/id/company/AB123456`

Stosuj przekierowania 303 do zasobów świata rzeczywistego

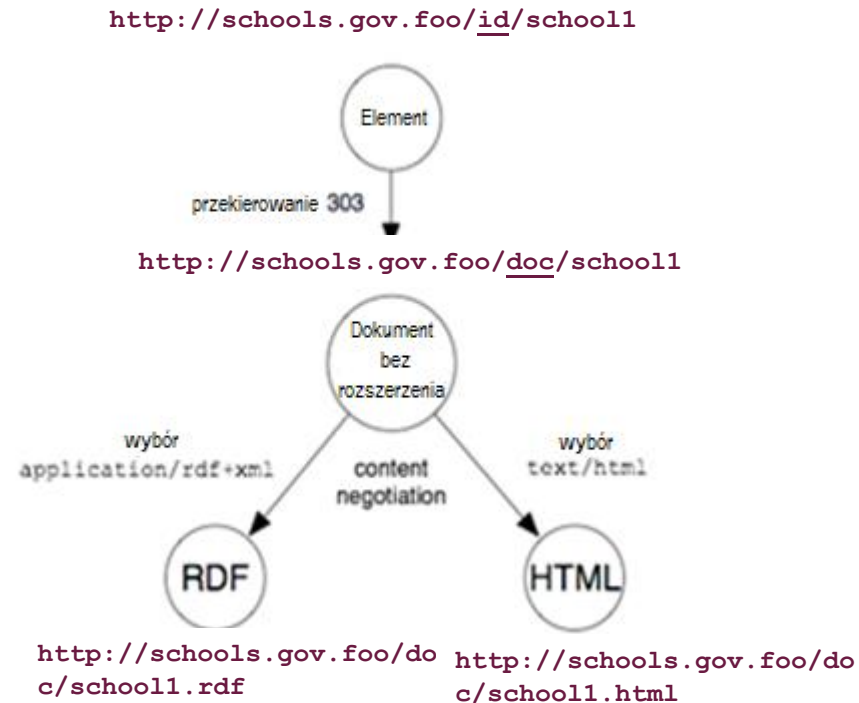
- Ponieważ **zasoby świata rzeczywistego** (tj. zasoby inne niż dokumenty, np. osoba, firma, miejsce...) **nie posiadają odpowiedniej reprezentacji**, przydatne jest przekierowanie do dokumentu sieci Web zawierającego informacje na temat danych zasobów..
- **Umożliwia rozróżnienie** między zasobem świata rzeczywistego a reprezentującym go dokumentem.
- Przykładowo, rząd tworzy następujące przekierowania reprezentujące szkoły podstawowe:
 - `http://schools.gov.foo/id/school1`
 - `http://schools.gov.foo/id/school2`

Zobacz również:

Cool URIs for the Semantic Web. <http://www.w3.org/TR/cooluris>

Przekierowanie 303 i content-negotiation

- Identyfikatory URI, do których ma miejsce odwołanie, powinny zwrócić **kod przekierowania 303** do dokumentu opisującego dany element.
- Serwer sieci Web musi zostać skonfigurowany, aby **przekierować**:
 - z `http://schools.gov.foo/id/school1`
 - do `http://schools.gov.foo/doc/school1`
- Obowiązuje **zasada modyfikacji identyfikatora URI**, zazwyczaj poprzez zastąpienie wartości „id” przez „doc”.
- Możliwe są **różne wersje**, np. RDF, XML, HTML...



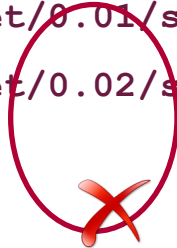
Zobacz również:

Cool URIs for the Semantic Web.

<http://www.w3.org/TR/cooluris>

Nie dodawaj numeru wersji do identyfikatora URI

- Zbiory danych, modele pojęciowe, ontologie, taksonomie i słowniki są publikowane w kolejnych wersjach będących rezultatem cyklicznych zmian/aktualizacji.
- Identyfikator URI powinien być **stały niezależnie od wersji**.
 - Do identyfikatora URI **nie powinno się dołączać** numerów wersji ani informacji dot. statusu.
- Przykładowo, zbiór danych szkół zostaje zaktualizowany z wersji 0.01 do 0.02. Jeśli numer wersji znajduje się w identyfikatorze URI, przy każdej aktualizacji musi on być zmieniany.
 - `http://schools.gov.foo/set/0.01/schools`
 - `http://schools.gov.foo/set/0.02/schools`



Unikaj autoinkrementacji tworząc nowe identyfikatory URI

- Zwykła inkrementacja licznika podczas tworzenia URI dużego zbioru danych jest łatwa, jednak może powodować poważne problemy.
 - Co dzieje się w przypadku aktualizacji zbioru danych i konieczności ponownego przypisania identyfikatora URI? Skąd pewność, że sekwencja znaków będzie identyczna?

Czy to oznacza, że nigdy nie powinno się tego robić?

Autoinkrementacja URI jest dopuszczalna, jeśli:

- nigdy nie zostanie powtórzona;
- zostaje powtórzona, gdy powstaje identyczny identyfikator URI dla tych samych danych wejściowych, a nowy URI zostaje stworzony wyłącznie dla nowych elementów.

Nie korzystaj z ciągów zapytań

- Ciąg zapytań (np. „?param=value”) to tekst dołączany na końcu adresu URL, zawierający dane dla aplikacji internetowych, np. parametry wyszukiwania pozwalające na znalezienie terminów w bazie danych.
 - Ciągi zapytań są niestale, gdyż **zależą od konkretnej implementacji**. Dlatego też nie należy ich umieszczać w URI.
- Przykład: firma znajdująca się w rejestrze NBR byłaby zidentyfikowana jako

[http://businessdata.gov/NBR/id/company?id="AB123456"](http://businessdata.gov/NBR/id/company?id=) ❌

zamiast

<http://businessdata.gov/NBR/id/company/AB123456>

Nie podawaj informacji o właścicielu

- Szablon stałego identyfikatora URI **nie powinien zawierać nazwy organizacji bądź projektu**, który utworzył URI.
- Przykładowo, firma AB123456 zarejestrowana w rejestrze National Business Register (NBR) figurowałaby jako
`http://businessdata.gov/NBR/id/company/AB123456` ❌
- Po kilku latach NBR zostaje przemianowany na rejestr National Company Register (NCR). Dlatego też wszystkie identyfikatory URI musiałyby zostać zaktualizowane.
- W tym przypadku stały identyfikator URI powinien wyglądać tak:
`http://businessdata.gov/id/company/AB123456`

Nie używaj rozszerzeń plików

- Rozszerzenia stanowią informację o rodzaju pliku określonego dokumentu.
- Nie należy podawać rozszerzeń plików w stałych identyfikatorach URI.
- Przykładowo, identyfikator URI zbioru danych szkół w państwie członkowskim powinien wyglądać tak:
 - `http://data.gov.foo/set/schools`a nie tak:
 - `http://data.gov.foo/set/schools.csv`
- Rozszerzenie pliku może stanowić część metadanych dokumentu.
 - np. `dcat:mediaType` w słowniku DCAT W3C do opisu zbiorów danych.

Obsługa stałych identyfikatorów URI zasobów danych

Korzystaj z serwera dedykowanego

- Należy uruchomić zaufany serwer dedykowany, **niezależny od twórcy danych**.
- **Łatwość przenoszenia** i obsługi przez inną osobę w razie potrzeby.
 - Dublin Core korzysta z purl.org
 - data.gov.uk i publications.europa.eu również są niezależne od jakichkolwiek instytucji rządowych
- Brak konieczności tworzenia jednego serwera dla wielu dostawców danych.
 - Większe ryzyko, ponieważ stwarzałoby to punkt podatności na awarię, ale
 - Łatwiejsze zarządzanie i większa wydajność w stosunku do kosztów.

Podsumowanie

Identyfikator URI to “zwięzły łańcuch znaków identyfikujący abstrakcyjny lub materialny zasób”.



Korzystaj z szablonu

`http://{{domain}}/{type}/{concept}/{reference}`

Wykorzystuj istniejące identyfikatory

`http://education.data.gov.uk/id/school/123456`

Udostępnij wersje w różnych formatach

`http://data.example.org/doc/foo/bar.html`

`http://data.example.org/doc/foo/bar.pdf`

Używaj przekierowania 303 elementów rzeczywistych

`http://www.example.com/id/alice_brown`

Używaj serwisów dedykowanych

niezależnych od twórcy danych

10 zasad stałych URI



Informacje o właścicielu

`http://education.data.gov.uk/ministryeducation/id/school/123456`

Informacje o numerze wersji

`http://education.data.gov.uk/doc/school/v1/123456`

Autoinkrementacja

`http://education.data.gov.uk/id/school1/123456`

`http://education.data.gov.uk/id/school1/123457`

Ciąg zapytań

`http://education.data.gov.uk/doc/school?id=123456`

Rozszerzenia plików

`http://education.data.gov.uk/doc/schools/123456.csv`

Zobacz również:

10 zasad stałych identyfikatorów URI. <https://joinup.ec.europa.eu/node/53858>

Pytania dla grup



<http://www.visualpharm.com>

Czy w Waszym państwie istnieją krajowe wytyczne dotyczące URI? Jeśli tak, jakie są podstawowe zasady?



<http://www.visualpharm.com>

Czy w Waszym państwie istnieje serwer dedykowany stałym identyfikatorów URI? Jeśli tak, jak nazywa się instytucja zarządzająca serwerem? Jeśli nie, dlaczego?

Rozwiążcie test online!

Dziękujemy!
...i czekamy na pytania.

Prezentacja została przygotowana przez Open Data Support

Zastrzeżenia

1. Poglądy wyrażone w niniejszej prezentacji odzwierciedlają jedynie opinie autorów i w żadnym wypadku nie stanowią oficjalnego stanowiska Komisji Europejskiej.

Komisja Europejska nie gwarantuje prawidłowości informacji zawartych w niniejszej prezentacji i nie ponosi odpowiedzialności za ich wykorzystywanie.

Komisja Europejska nie promuje, nie poleca ani nie popiera żadnych produktów, specyfikacji, procesów lub usług, o których wspomniano w prezentacji pod postacią nazwy handlowej, znaku towarowego, nazwy producenta bądź inną.

Autor uzyskał zgodę od właścicieli praw autorskich lub ich przedstawiciela na wykorzystanie wszelkich fragmentów dokumentów objętych ochroną własności intelektualnej, w tym ilustracji, map i wykresów.

2. Niniejsza prezentacja została starannie opracowana przez firmę PwC, która nie gwarantuje jednak, że zawarte w niej informacje są pełne i prawidłowe. Firma PwC nie ponosi odpowiedzialności za informacje ani wszelkie decyzje bądź następstwa decyzji podjętych na podstawie informacji zawartych w niniejszej prezentacji. PwC nie ponosi odpowiedzialności za ewentualne szkody powstałe w wyniku wykorzystania informacji zawartych w prezentacji. Przedstawione informacje mają charakter ogólny i powinny służyć jedynie jako wskazówki w kwestiach będących przedmiotem zainteresowania ogólnego. Niniejsza prezentacja nie może zastępować profesjonalnej porady w jakiegokolwiek kwestii. Nikt nie powinien podejmować decyzji na podstawie jakiegokolwiek kwestii poruszonej w niniejszej publikacji bez uprzedniego zasięgnięcia profesjonalnej porady.

Autorzy:

Nikolaos Loutas, Michiel De Keyzer and Stijn Goedertier

Odnosi

Slajd 6:

- T. Berners-Lee, R. Fielding and L. Masinter (2005) "Uniform Resource Identifier (URI): Generic Syntax". <http://tools.ietf.org/html/rfc3986>

Slajdy 11-22:

- UK Government, CTO Council, Designing URI sets of the UK Public Sector. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-uri-sets-uk-public-sector.pdf
- EC ISA Programme, Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

Slajdy 14-15:

- Cool URIs for the Semantic Web, <http://www.w3.org/TR/cooluris>

Literatura uzupełniająca (1/2)

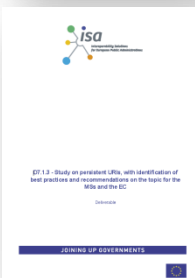


T. Berners-Lee, R. Fielding and L. Masinter (2005) "Uniform Resource Identifier (URI): Generic Syntax". <http://tools.ietf.org/html/rfc3986>



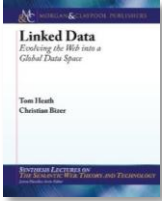
UK Government, CTO Council, Designing URI sets of the UK Public Sector.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-uri-sets-uk-public-sector.pdf



EC ISA Programme, Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

Literatura uzupełniająca (2/2)



Linked Data: Evolving the Web into a Global Data Space. Tom Heath and Christian Bizer.

<http://linkeddatabook.com/editions/1.0/>

Podobne projekty i inicjatywy



Projekt LOD2 FP7, <http://lod2.eu>



W3C Cool URIs for the Semantic Web

- <http://www.w3.org/TR/cooluris>
- <http://www.w3.org/wiki/GoodURIs>



URI Design Principles: Creating Unique URIs for Government Linked Data, <http://logd.tw.rpi.edu/instance-hub-uri-design>



Urząd Publikacji Komisji Europejskiej, <http://publications.europa.eu>

Data.gov.uk, <http://data.gov.uk/linked-data>



Dołącz do nas...

Odwiedź nas



Open Data Support

<http://www.slideshare.net/OpenDataSupport>



Open Data Support

<http://goo.gl/y9ZZI>

Śledź nas



@OpenDataSupport

Dołącz do nas



joinup

<http://www.opendatasupport.eu>

Kontakt

contact@opendatasupport.eu