

UTKAST

# Råd gällande beständiga länkar

Nationellt ramverk för öppna data

# Innehållsförteckning

|   |          |
|---|----------|
| <b>1. Nationellt ramverk för öppna data</b> .....                     | <b>2</b> |
| 1.1. Råd gällande beständiga länkar.....                              | 2        |
| 1.2. Vem vänder sig råden till?.....                                  | 2        |
| 1.3. Dokumentets struktur .....                                       | 2        |
| 1.4. Vilka delar av arbetet med öppna data tas upp i dokumentet?..... | 3        |
| 1.5. Vägledande principer .....                                       | 3        |
| <b>2. Vad är en URI?.....</b>   | <b>4</b> |
| <b>3. Syftet med beständiga länkar .....</b>                          | <b>6</b> |
| <b>4. Råd för beständiga länkar .....</b>                             | <b>7</b> |

# 1. Nationellt ramverk för öppna data

Dessa råd gällande beständiga länkar är utarbetade av Kommunförbundet Stockholms län, Peter Krantz AB, PwC och Sveriges Kommuner och landsting med stöd av finansiering från Vinnova. De är en del i ett nationellt ramverk för öppna data som innehåller vägledande principer, råd, mallar och verktyg som är framtagna för att erbjuda ett stöd till kommuner och landsting i deras arbete med öppna data. Övriga delar av det nationella ramverket finns publicerade på <http://www.skl.se/oppnadata>.

Råden är licensierade under en Creative Commons Erkännandelicens (CC-BY). Det innebär att materialet är fritt att använda, sprida, göra om, modifiera och bygga vidare på, även i kommersiella sammanhang så länge som uppgifter om källa och upphovsman anges enligt följande: Råd gällande beständiga länkar utarbetade av Kommunförbundet Stockholms län, Peter Krantz AB, PwC och Sveriges kommuner och landsting med stöd av Vinnova.



## 1.1. Råd gällande beständiga länkar

Detta dokument ger en introduktion till och en översikt av beständiga länkar eller URI:er och hur arbetet med dessa bör se ut för de data som kommunen/landstinget själv äger eller kontrollerar på annat sätt. Vid publicering av öppna data är det viktigt att det finns ett enhetligt och beständigt sätt att referera till datat och det är där beständiga länkar kommer in i bilden. Dokumentet presenterar en översikt av hur URI:er fungerar samt lyfter fram principer och riktlinjer för arbetet med beständiga länkar.

En URI är en identifierare som pekar ut en resurs på ett datornät. Vid publicering av öppna datakällor är det viktigt att dessa identifierare inte ändrar utseende eller plötsligt pekar ut någon annan resurs, de behöver alltså vara beständiga. Vid arbete med öppna data i andra länder har ett antal riktlinjer tagits fram som syftar till att öka beständigheten hos URI:erna.

De principer som presenteras nedan följer de riktlinjer<sup>1</sup> som tagits fram inom EU och är anpassade för att fungera i en svensk kontext. Det är även en vidareutveckling och utökning av de principer<sup>2</sup> som Domstolsverket tagit fram tidigare.

## 1.2. Vem vänder sig råden till?

Råden vänder sig till personer som har en mer teknisk roll och som ansvarar för eller deltar i utformningen av en kommuns eller landstings it-system. Beslut om att använda sig av beständiga länkar kan tas som ett övergripande policybeslut eller av respektive systemägare i kommunen/landstinget.

## 1.3. Dokumentets struktur

Inledningsvis beskrivs vad en URI är och vad syftet är med beständiga länkar. Därefter ges ett antal rekommendationer för hur beständiga länkar bör utformas.

---

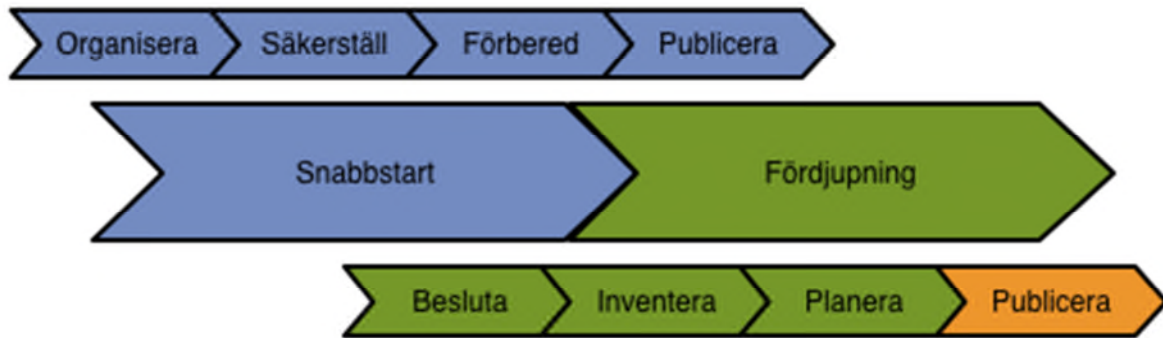
<sup>1</sup> Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC, Europeiska kommissionen (2012)

<sup>2</sup> URI-principer: Elektroniska identifierare i Rättsinformationssystemet, Domstolsverket (2013)

## 1.4. Vilka delar av arbetet med öppna data tas upp i dokumentet?

De vägledande principerna för öppna data är uppbyggda dels kring att vissa delar har en högre prioritet än andra och dels utifrån att det ska gå att snabbt komma igång med arbetet för att sedan fördjupa det. Råden i detta dokument omfattar åtgärder som är aktuella när kommunen/landstinget ska fördjupa sitt arbete med öppna data.

Nedan markeras de faser som behandlas i dokumentet med orange färg.

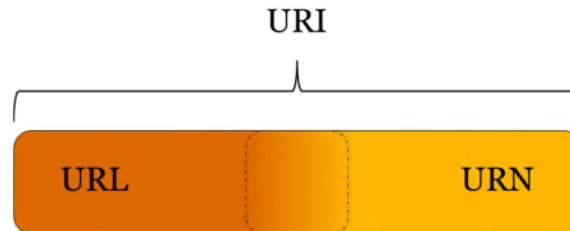


## 1.5. Vägledande principer

De vägledande principerna för öppna data rekommenderar att kommuner och landsting använder beständiga länkar när de tillgängliggör data för vidareutnyttjande.

## 2. Vad är en URI?

URI står för Uniform Resource Identifier och är en textsträng som används på Internet eller i ett datornätverk för att identifiera en abstrakt eller fysisk resurs. En URI kan klassificeras som en hänvisare eller som ett namn, eller som en kombination av båda. Om URI:n utöver att identifiera en resurs även hänvisar till platsen där resursen finns kallas den URL (Uniform Resource Locator), om URI:n används för att namnge en resurs kallas den för URN (Uniform Resource Name).



En typ av URI som de flesta är bekanta med är de adresser vi använder för att peka ut webbsidor på Internet med och som läses med hjälp av en webbläsare.

Exempel på URI:er:

```
http://www.w3.org/  
mailto:anders.andersson@exempel.se  
ftp://ftp.is.co.za/rfc/rfc1808.txt
```

Syftet med att publicera URI:er är inte att de bara ska vara identifierare, de bör även kunna avrefereras, det vill säga att den som efterfrågar URI:n bör få mer information om det som URI:n representerar. Om det är en webbläsare som gör förfrågan till en URI ska det som kommer tillbaka vara en webbsida som kan läsas av en människa. Om det är en RDF-klient som efterfrågar en URI ska svaret vara en RDF-fil, från samma URI.

### 2.1.1. Tekniken bakom

Allt som är uppkopplat mot Internet har en adress som gör att andra datorer eller enheter kan kontakta dem, dessa adresser kallas för IP-adresser och består av en serie siffror. Dessa nummer är inte nödvändigtvis statiska utan tilldelas när en enhet kopplar upp sig, och för att hålla koll på vilket nummer en dator har vid ett givet tillfälle används något som kallas DNS (Domain Name System). DNS är ett system som kopplar ihop IP-adresserna med textsträngar som dels är lättare för människor att komma ihåg och dels gör det möjligt att behålla samma adress även när innehållet flyttas till andra servrar. En URI är en sådan textsträng.

För att bättre förstå diskussionen om beständiga URI:er är det två aspekter av tekniken som behöver förklaras.

- **URI:er är enbart textsträngar.** De innehåller i sig ingen mening utan identifierar bara en resurs. Det finns inget i textsträngen som berättar vad det är för typ av resurs. Exempelvis säger inte URI:n `http://exempel.se/dokument.pdf` att det faktiskt finns ett PDF-dokument på adressen, det är fullt möjligt att det är något annat som skickas tillbaka.
- **Servrar är smarta och flexibla.** De kan konfigureras att göra betydligt mer än att bara returnera en statisk fil. URI:er kan avrefereras till olika format beroende på vem det är som efterfrågar resurser och nya varianter kan tillkomma senare.

En annan aspekt av serverkonfigurationer är så kallad innehållsförhandling (eng. Content Negotiation), vilket betyder att typen av svar som ges av servern beror på vilken agent det är som efterfrågar URI:n. En användaragent skickar med information om vem den är och får svaret baserat på detta. En webbläsare kan få ett HTML-dokument eller en PDF-fil, en RDF-läsare får en RDF-fil och om det dyker upp nya format på vilka informationen kan representeras kan dessa läggas till senare. Men det är fortfarande samma URI.

Sammanfattningen av detta blir att URI:erna kan vara väldigt enkla och strukturerade efter en verklig uppdelning och inte kopplat mot vilken teknik som används. Något som är viktigt när det gäller

beständiga URI:er då tekniken med största sannolikhet kommer att ändras flera gånger under URI:ns livstid.

### **3. Syftet med beständiga länkar**

Vi är vana vid att litterära verk, patent och legala dokument kan refereras till entydigt både nu och i framtiden, och många av våra strukturer bygger på dessa entydiga referenssystem. Syftet med beständiga länkar är detsamma, men för resursers som finns tillgängliga via Internet.

När offentliga förvaltningar utbyter data med varandra ställs höga krav på att alla involverade parter är överens om vad det är för data som utbyts, det får därför inte råda någon oenighet om vad referenserna faktiskt betyder. Detta är även något som blir mer intressant när det gäller samarbeten över gränserna, inom EU till exempel.

## 4. Råd för beständiga länkar

De första ansatserna till att ta fram riktlinjer för beständiga URI:er gjordes i slutet av 1990-talet. Dessa riktlinjer har sedan utvecklats och anpassats för olika situationer. De rekommendationer som presenteras här är de senaste vid utgången av 2012 och användes vid utvecklingen av data.gov.uk som är regeringens i Storbritannien initiativ kring öppna data.

### 4.1.1. Design och format

Det rekommenderade formatet för beständiga URI:er följer de riktlinjer som tagits fram inom EU<sup>3</sup> och har använts vid ett flertal lyckade scenarier. Formatet är uppbyggt kring ett generellt mönster, och dess betydelser beskrivs nedan:

`http://{domän}/{typ}/{koncept}/{referens}`

{domän}

Med *domän* avses en kombination av värd (ex. data.gov.uk, europa.eu, libris.kb.se) och relevant sektor (ex. utbildning, transport, kultur). Sektorn kan antingen anges som en subdomän eller vara första delen i sökvägen: sektor.värd.se/ eller värd.se/sektor/

{typ}

Med typ menas en av en liten uppsättning termer som definierar vilken typ av resurs som identifieras. Det är följande fyra typer som ska användas:

- **id** för fysiska objekt
- **dok** för dokument
- **def** för definition av ett koncept
- **set** för ett dataset

{koncept}

Konceptet kan vara en samling av något, det kan vara typen på det identifierade objektet (ex. skola, väg, stad) eller namnet på ett konceptschema (ex. språk).

#### Regler för val av koncept:

- Konceptet ska vara ett substantiv
- Konceptet ska stå i obestämd form singular
- Konceptet ska vara den vedertagna termen
- Ordet som väljs ska vara det som närmast beskriver konceptet.

{referens}

Referensen pekar ut det specifika objektet, termen eller konceptet: /skola/1212 eller /vaeg/e4 eller /stad/stockholm eller /spraak/swe.

Exempel

#### Storbritannien

`http://transport.data.gov.uk/doc/road/B3178`

sektor = transport

värd = data.gov.uk

domän = transport.data.gov.uk

typ = doc (anger att URI:n pekar ut ett dokument)

koncept = road (anger att det utpekade dokumentet handlar om en väg)

---

<sup>3</sup> Europeiska kommissionen (2013)



referens = B3178 (anger att den specifika vägen är väg B3178)

### Sverige, Kungliga Biblioteket

<http://libris.kb.se/bib/11284499>

sektor = libris

värd = kb.se

domän = libris.kb.se

typ = bib (anger att det är en bokreferens som utpekas)

referens = 11284499 (anger att den specifika boken är "ett urval dikter" av Karin Boye)

## 4.1.2. Teckenstandard och namngivning

### Svenska tecken

I URI:er ska endast gemener från a-z, siffror, bindestreck och understreck användas. Övriga skiljetecken, mellanslag, accenter eller specialtecken i redan befintliga identifierare skall tas bort eller ersättas av bindestreck eller understreck. Om en befintlig identifierare innehåller svenska tecken (ex. väg och språk) ska dessa översättas enligt följande:

| Bokstav | Översättning |
|---------|--------------|
| å       | aa           |
| ä       | ae           |
| ö       | oe           |

### Språkidentifiering

För att ange vilket språk en identifierad resurs följer ska ISO 639-3<sup>4</sup> för språkidentifiering användas (ex. swe för svenska och eng för engelska).

## 4.1.3. Minimalt med information

En beständig URI ska i princip bara innehålla information som inte kommer att förändras. Mycket information kopplad till en resurs kommer potentiellt att ändras: status, författare, ägare, titel, ämne m.m. Eftersom offentliga enheter byter namn och ansvarsområden med jämna mellanrum får inte URI:erna vara beroende av en struktur där till exempel informationsägare finns med.

## 4.1.4. Senaste versionen

En klass av URI:er som kan nämnas extra är den som hänvisar till den senaste versionen av en resurs. URI:n måste vara stabil men innehållet kan ändras ofta. W3C har ett bra exempel på detta:

URI:n <http://www.w3.org/TR/vocab-org/> pekar alltid på den senaste versionen av organisationens ontologi. Den senaste versionen (vid skrivtidpunkten) publicerades den 23 oktober 2012 och har en egen fast URI på <http://www.w3.org/TR/2012/WD-vocaborg-20121023/>, men så fort en ny version publiceras är det den som kommer att finnas på <http://www.w3.org/TR/vocab-org/>. Detta kan jämföras med framsidan på en nyhetstidnings portal.

<sup>4</sup> <http://www-01.sil.org/iso639-3/default.asp>

### 4.1.5. Administration och struktur

För att skapa bra beständiga URI:er räcker det inte att bara följa ett mönster eller en struktur, det finns flera andra aspekter som är bra att ta hänsyn till. Nedan följer ett antal punkter som har visat sig vara viktiga för att skapa beständiga URI:er:

Följ mönstret `http://{domän}/{typ}/{koncept}/{referens}/`

Se punkt 4.1.1 *Design och format* för mer information.

#### Återanvänd identifierare

Om resurser redan är unikt refererade ska dessa identifierare återanvändas. Om det till exempel redan finns ett nummersystem för att identifiera skolor kan skolan 12345 refereras till som:

```
http://utbildning.exempel/id/skola/12345
```

Det är dock viktigt att komma ihåg att endast identifierare som själva är beständiga ska användas.

#### Använd 303-redirect för fysiska ting

Om en URI avrefereras till ett fysiskt objekt som inte kan skickas över ett datornät ska URI:n svara med http-svarskoden 303 och hänvisa vidare till ett dokument som beskriver objektet.

#### Länka multipla representationer

Olika representationer av samma resurs ska alla länka till varandra med lämplig metod och teknik. I HTML ska till exempel 'link'-elementet användas med 'rel'-värdet satt till 'alternate', och i RDF ska 'dcterms:hasFormat' användas.

#### Använd en särskild tjänst

I alla lyckade implementeringar av beständiga URI:er har en oberoende särskild tjänst som är fristående från dataägarna använts. Genom att lyfta ut hanteringen av URI:erna kan driften lättare flyttas runt om något skulle hända med organisationen.

#### Undvik att ange ägarskap

Namnet på den organisation eller det projekt som skapar en URI ska inte vara en del av URI:n då URI:n blir obsolet om det sker en organisationsförändring eller om projektet läggs ner eller byter namn.

#### Undvik versionsnummer

Även om konceptscheman, ontologier, taxonomier och vokabulärer ofta går igenom många iterationscykler ska versionsnummer och statusinformation inte vara en del av URI:n.

#### Undvik automatisk uppräknings

När det skapas URI:er för stora datamängder kommer processen att automatiseras. Ett sätt att göra detta på är att räkna upp ett tal som blir id-numret på resursen. Om detta förfarande används skulle URI:erna för två skolor kunna bli:

```
http://utbildning.exempel/id/skola/12345
```

```
http://utbildning.exempel/id/skola/12346
```

Enda tillfällena då automatisk uppräknings får användas är när processen aldrig kommer att upprepas eller om det kan garanteras att exakt samma URI:er skapas om processen måste upprepas.

#### Undvik frågesträngar

Frågesträngar av typen `?param=value` används ofta i URI:er för att skicka med uppslagsvärden till en databas. Detta sätt är dock inte stabilt då det ofta förutsätter att en specifik teknik används. Använd istället det format som beskrivits i punkt 4.1.1 *Design och format* och använd server-konfigurationen för att tolka URI:erna.

### **Undvik filändelser**

Filändelser är ofta beroende av den teknik som används för att representera informationen och detta är en lösning som inte är stabil då tekniken med stor sannolikhet kommer att förändras. Låt istället serverkonfigurationen avgöra vilken typ av fil som ska returneras och på vilket sätt.