

## Rekommendation för öppna data

# Fem stjärnor för öppna data



Digisam rekommenderar alla som arbetar med digitalt kulturarv att sikta mot den stjärnmodell som formulerats av Tim Berners-Lee, mannen bakom Internet. Modellen visar tydligt vägen mot den semantiska framtiden. Det är enkelt att ta de första stegen och man kan inte hoppa över något steg, utan varje steg bygger vidare på de tidigare.

★ En stjärna för öppna data: Gör din information tillgänglig på webben, i vilket format som helst, men under öppen licens.

★★ Två stjärnor för strukturerad/maskinläsbar data: Gör din information tillgänglig som strukturerad data (t.ex. i Excel istället för en inskannad bild av en tabell).

★★★ Tre stjärnor för öppen mjukvara: Använd icke-proprietära format (t.ex. CSV istället för Excel).

★★★★ Fyra stjärnor för länkbar data: Använd URI:er för att identifiera ting och öppna standarder från W3C (RDF och SPARQL) för att uttrycka påståenden om dem.

★★★★★ Fem stjärnor för länkad data: Skapa sammanhang genom att länka dina data till andras data.

## ★ En stjärna - data tillgängligt digitalt

Den som tillgängliggör sina data digitalt och med öppna licenser får en stjärna. Om man t.ex. har information tillgänglig på vanliga webbsidor och kompletterar med tydlig information om hur datat får vidareanvändas så är första stjärnan säkrad.

Det är bra för institutionen eftersom:

- Det är enkelt att publicera.
- Du slipper förklara för andra hela tiden att de är fria att använda dina data.

För användaren är det ett stort steg framåt att kunna använda informationen direkt i stället för att behöva ta kontakt, reda ut tillstånden och explicit beställa data från din organisation.

Användaren kan nu:

- Titta på datat.
- Skriva ut det.
- Ladda hem det och spara på sin egen hårddisk eller USB-sticka.
- Importera datat till sina egna system, ändra datat efter sina egna behov, dela datat med vem hen vill.

## ★★ Två stjärnor - bättre datakvalitet

Den som tillgängliggör data i maskinläsbara format säkrar informationsinnehåll som annars kan riskera att gå förlorat vid migrering/extrahering. Det är fortfarande enkelt att publicera, men ger ökad tillförlitlighet och säkrar två stjärnor.

Det är bra för institutionen eftersom:

- Det fortfarande är enkelt att publicera.

Användaren kan göra allt som gäller för ★-stjärnig data och dessutom:

- Direkt processa datat med proprietär<sup>1</sup> programvara, aggregera, kalkylera, visualisera det etc. vilket gör det enklare och mer attraktivt att använda.
- Exportera det till ett annat (strukturerat) format.

## ★★★ Tre stjärnor - öppna data

Den som siktar på tre stjärnor undanröjer behovet av investeringar i proprietär teknologi hos dem som ska vidareutnyttja informationen. Genom att arbeta med öppna format, som är

---

<sup>1</sup> <https://sv.wikipedia.org/wiki/Propriet%C3%A4r>

antingen väldigt enkla (t.ex. CSV) eller väldigt väl dokumenterade, skapar man förutsättningar för mer långsiktigt hållbar data. Dessutom minskar risken för felaktig bearbetning av information då proprietära format hanteras av tredje parts mjukvara (särskilt om fullständig dokumentation om formatet saknas).

Det är bra för institutionen eftersom:

- Det fortfarande är ganska enkelt att publicera.
- Men du kan eventuellt behöva konverteringar eller plug-ins för att exportera data från ett proprietärt format.

Användaren kan nu göra allt som gäller för ★★-stjärnig data, och dessutom:

- Manipulera datat hur som helst, utan att hindras av restriktioner i en viss programvara.

## ★★★★ Fyra stjärnor - enhetligt informationsuttryck och tydlig semantik

Den fjärde stjärnan ger en rad fördelar. Genom att dela upp datamängderna i poster (ting) med globalt unika identifierare (URI:er) möjliggör man för andra att referera till egna data på ett sätt som är standardiserat och effektivt.

Dataintegration med andra parter blir enklare då data uttrycks på ett sätt som inte är bundet till det egna datalagret. Istället beskriver man sina data med hjälp av redan existerande termer. Ofta kan man bygga vidare på andras arbete, och anpassa eller kombinera en eller ett par existerande informationsmodeller efter egna behov. Det innebär att andra parter har lättare att förstå informationsmodellen när de ska vidarenyttja datat, och att de i sin tur kan bygga vidare på din organisations insatser.

Då data uttrycks i RDF, ett standardiserat språk för hantering av information, finns redan många tjänster och verktyg som kan användas för att skapa, validera, lagra, maskinellt bearbeta, kombinera, redigera och utforska data med redan existerande frågespråk.

Det är bra för institutionen eftersom:

- Du kan ha kontroll över datat på hög detaljeringsnivå och därigenom optimera tillgång, flöden, cachning etc.
- Andra dataförvaltare nu kan länka in mot dina data och hjälpa dig vidare mot 5 stjärnor!
- Men du måste ofta lägga en del tid på att städa och tvätta dina data.
- Du behöver ge dina dataposter unika identifierare (URI:er) och tänka igenom hur du vill presentera datat.
- Du behöver antingen använda befintliga designmönster<sup>2</sup> eller skapa egna.

---

<sup>2</sup> <https://sv.wikipedia.org/wiki/Designm%C3%B6nster>

Användaren kan nu göra allt hen kunde med ★★★-data, och dessutom:

- Länka till datat från en annan plats på webben, eller lokalt.
- Bokmärka datat.
- Återanvända delar av datat.
- Eventuellt återanvända befintliga verktyg och sammanställningar, även om hen bara förstår vissa delar av de designmönster som använts vid publicering.
- Att förstå strukturen i en datagraf (RDF) kan dock uppfattas som svårare än att förstå tabeller (Excel/CSV) eller datatråd.
- På ett säkert sätt kombinera datat med andra datakällor. URI:er igår i en global struktur, så om två ting refererar till samma URI är det antagligen avsiktligt, vilket betyder ett steg närmare ★★★★★-stjärnig data!

## ★★★★★ Fem stjärnor - länkade öppna data

Den femte stjärnan ger flera ytterligare fördelar. Den gör det lättare att kvalitetshöja dina data genom att länka till väletablerade och väl uttänkta termer och begrepp istället för att skapa egna eller skriva fritext. Du kan använda data, begrepp och termer från andra datakällor direkt när behov uppstår utan att först behöva fokusera på tekniska aspekter av dataintegration som import, konvertering och drift/underhåll.

Effektivt använd länkning kan leda till en fördjupad specialisering då du kan fokusera på att underhålla de delar av datat som är unika för din organisation och lägga mindre tid på information som redan finns i andra datakällor.

Länkar till andra datakällor ökar förtroendet för att dina data är genomtänkta, ungefär på samma sätt som referenser i artiklar visar på att informationen där är tillförlitlig och förankrad i ett större sammanhang. Länkar ut ökar också synligheten för dina data, då de blir en del av det större länkade datamolnet, vilket i ett längre perspektiv kan leda till återanvändning i nya sammanhang, och länkar in från andra förvaltare.

Det är bra för institutionen eftersom:

- Dina data blir lättare att upptäcka.
- Du ökar värdet av dina data.
- Din egen organisation kommer att dra nytta av länkarna på samma sätt som användarna.
- Men du behöver lägga resurser på att länka dina data till andra data på webben.
- Du kan behöva bevaka och laga brutna eller felaktiga länkar.

Användaren kan göra allt hen kunde med ★★★★★-stjärnig data och dessutom:

- Upptäcka mer (relaterad) data när hen använder datat.
- Direkt få överblick över data-schemat.

- Men hen måste också hantera brutna länkar på liknande sätt som med 404 felmeddelanden på webben.
- Att koppla data till en länk vilken som helst är lika vanskligt som att använda innehåll från en webbsida vilken som helst. Försiktighet, tillit och sunt förnuft behövs fortfarande.

Den som börjat arbeta med öppna data kan ladda hem stjärnor att märka dem med från Linked Data Research Centre Lab på Irland<sup>3</sup> eller få sina resurser certifierade genom Open Data Institute.<sup>4</sup> Man bör också registrera sina öppna data på den nationella portalen för Öppna data.<sup>5</sup>

För mer fördjupning i ämnet se vidare Meta Solutions Vitbok Öppna Data<sup>6</sup> eller (med en mer teknisk ingång) Linked Data Basics for Techies, som utarbetats av Cristopher Gutteridge på Southampton university i England.<sup>7</sup>

Denna rekommendation för arbetet med öppna data är baserad på underlag från MetaSolutions Vitbok Öppna Data (2014) som i sin tur utgår från Tim Berners-Lees femstjärniga modell, först publicerad 2010.<sup>8</sup> Den som vill veta mer om Berners-Lee och tankarna bakom stjärnorna kan ta del av hans TEDtalks.<sup>9</sup>

---

<sup>3</sup> Linked Open Data Star Badges <http://lab.linkeddata.deri.ie/2010/lod-badges>

<sup>4</sup> Certify you're open data <https://certificates.theodi.org>

<sup>5</sup> Öppna data <http://oppnadata.se>

<sup>6</sup> Vitbok Länkade Öppna Data <http://lankadedata.se/vitbok>

<sup>7</sup> Linked Data Basics for Techies [http://openorg.ecs.soton.ac.uk/wiki/Linked\\_Data\\_Basics\\_for\\_Techies](http://openorg.ecs.soton.ac.uk/wiki/Linked_Data_Basics_for_Techies)

<sup>8</sup> Linked Data <http://www.w3.org/DesignIssues/LinkedData.html>

<sup>9</sup> Tim Berners-Lee TEDtalks <http://www.ted.com/search?q=berners-lee>