

## Checklista: Metadatatstatus

### Vad är metadata?

Med ökad digitalisering krävs också ökad uppmärksamhet på metadata. Data om data alltså, eller strukturerad information som beskriver, förklarar, lokaliserar eller på annat sätt gör det lättare att återfinna, använda och förvalta informationsresurser.

Grundläggande registerdata om t.ex. titel, objektsnummer eller upphovsman ligger typiskt sett (i prioriterade/obligatoriska fasta fält) i katalogen eller samlingshanteringssystemet, och är ibland öppna för allmänheten. Men till metadata hör också förvaltningsinformation om sådant som konservering och in-/utlån, samt proveniens och andra mjukare uppgifter (ofta i fritext).

### Vad är det för fel?

Vanliga brister i metadata kan vara sådant som felstavning och andra skrivfel, ofullständiga uppgifter, inkonsekvent användning av datering, nyckelord och interpunktion eller dubbelregistrering och andra nedärvda brister från äldre tiders kataloger, rutiner och standarder. Det sägs ofta att kulturarvsinstitutionernas metadata är ”dålig”, men vad betyder det egentligen? Och hur vet man det?

Viktigt att framhålla är att vad som är kvalitativ metadata naturligtvis beror på syftet med katalogen, att detta kan ha skiftat över tid, och att kriterierna efter interna och externa behov kan skilja sig åt mellan olika myndigheter och institutioner. Men en enda tappad bokstav i ett namn, ”Stna Aronson”, kan göra det omöjligt att hitta posten för den som gör en korrekt sökning på författaren ”Stina Aronson”. Bristfällig metadata kan göra digitala material svåra eller omöjliga att hitta och därigenom få långtgående negativa konsekvenser för både forskning, förvaltning, lärande och andra intressen både internt och externt. För att kunna hålla ordning på sina digitala objekt eller filer behöver myndigheter och institutioner prioritera professionell förvaltning av metadata, vilket bl.a. förutsätter löpande förädling, komplettering och ”datatvätt”.

### Ingen är perfekt!

Manuellt arbete med kvalificerad katalogisering ger aldrig 100% korrekt resultat. Det är viktigt att komma ihåg att den helt perfekta katalogen inte finns, men också att alla kan bli bättre. Helt centralt är att metadata rutinmässigt kvalitetskontrolleras så tidigt som möjligt i reguljära arbetsflöden. De som arbetar operativt med katalogisering kan också behöva stöd för att hålla jämna steg med den digitala utvecklingen och kanske kompetensutveckling i metadata och innehållsstandarder. Nya möjligheter till välkommen hjälp från användare med taggning och annan crowdsourcing ställer också ökade krav på strukturerad kvalitetskontroll, så nu är den viktigare än någonsin.

## Vad kan man göra?

- **Omvärldsbevaka.** Börja med att lägga lite tid på omvärldsspaning och försök hitta goda exempel från andra arkiv, bibliotek och museer. Det finns mycket information att hämta på nätet, både via officiella webbplatser och mer informellt via bloggar och sociala medier.
- **Skapa tydliga rutiner för katalogisering.** Utarbeta tydliga rutiner för katalogisering, i synnerhet om flera personer, kanske med olikartad kompetens, kommer att vara med och skapa metadata. Formalisera riktlinjer för arbetet och se till att de är lätt tillgängliga (digitalt eller analogt) för dem som katalogiserar. Se också till att admin-gränssnittet är användarvänligt utformat, med en tydlig uppsättning obligatoriska fält. Överväg drop-down-menyer för att styra beskrivande innehåll mot större enhetlighet, alternativt lägg in länkar till relevanta index, thesaurier och auktoritetslistor.
- **Kompetensutveckla.** Prioritera resurser för att kunna erbjuda befintlig personal den kompetensutveckling de behöver. Då det ofta är svårt att hitta rätt kompetenser för nyanställning, är sådana insatser helt avgörande för en organisation som vill upprätthålla hög kvalitet i metadataförvaltning.
- **Användartesta.** Gör användartester med organisationens vanligaste användare - kollegor, allmänhet eller bådadera. Enkla online-enkäter kan ge mycket information om vilka krav användare har på databaserna, och peka på inkonsekvenser eller andra brister. För interna användare i synnerhet, t.ex. kollegor som är beroende av institutionens metadata för eget arbete och forskning, kan man överväga att bygga enkla tjänster för felrapportering.
- **Gör rutinmässiga kvalitetskontroller.** Skapa rutiner för regelbunden och konsekvent utformad kvalitetskontroll för att bedöma metadatastatus. Sådan uppföljning är särskilt viktig om organisationen nyligen implementerat förändrade rutiner eller arbetsflöden, då den gör det möjligt att spåra kvalitativa genomslag i katalogen.

## Gör en kvalitetskontroll

1. Utse ansvarig/-a för processen, en innehållskunnig informationsspecialist eller ett litet team.
2. Fatta beslut om i vilken utsträckning påträffade felaktigheter ska rättas direkt i den levande databasen.
3. Ladda regelbundet, två-fyra gånger om året, upp ett antal testposter till OpenRefine<sup>1</sup>.
4. Fassettera och klustra testposterna i OpenRefine, identifiera och dokumentera felaktigheter såsom felstavningar, inkonsekvent användning av versaler eller tomma fält.

---

<sup>1</sup> <http://openrefine.org>

5. Sammanställ dokumentationen så att eventuella kvalitetsförändringar kan iakttas över tid. Detta är särskilt viktigt om organisationen t.ex. implementerat nya rutiner för katalogisering.

## Använd OpenRefine

Vi rekommenderar OpenRefine, ett gratisverktyg för den som vill gå i närkamp med sina data i vad som på engelska kallas *data wrangling*. Verktöget har bl.a. använts i en rad LOD-projekt<sup>2</sup> inom kulturarvssektorn, men fungerar också bra för den som vill undersöka sub-set av metadata och underlättar identifiering och loggning av felstavning och andra brister i katalogen.

OpenRefine bygger på öppen källkod, utvecklades till en början av Metaweb som Freebase Gridworks, och blev GoogleRefine 2010. Sedan Google slutade utveckla verktöget 2012 lever det vidare som OpenRefine. Det är ett stationärt baserat program som kan laddas hem och användas lokalt. En delmängd metadata kan sedan exporteras från en institutions digitala katalog eller samlingshanteringssystem och laddas upp till OpenRefine som ett nytt projekt. Observera att verktöget inte är gjort för datatvätt. Rättelser kan alltså inte maskinellt laddas tillbaka till organisationens databas utan måste göras manuellt.

Syftet med tjänsterna i OpenRefine är att förbereda dataset för återanvändning, kanske genom att köra dem mot kontrollerade vokabulärer, extrahera namn-entiteter eller publicera dem som länkade öppna data (LOD). Särskilt värdefullt vid en kvalitetskontroll är att verktöget gör det enkelt att granska en datamängd och kvantifiera förekomster av felstavning, tomma fält, skiljetecken och andra brister.

## Fasettera

Gränssnittet i OpenRefine ser ut som ett kalkylblad, med rader och kolumner, men det fungerar mer som en databas. Fasettering kan användas på kolumnerna för att lista olika typer av cellvärden och antalet gånger som dessa värden uppträder. En fasett i kolumnen Nyckelord kan till exempel visa 20 poster för "Väder" och tre poster för "Vädr". Det gör det enkelt att identifiera de poster som är felskrivningar.

## Klustra

Om man använder klustrings-verktöget på en kolumn lägger det samman ord som är lexikalt lika, och det blir enklare att identifiera felstavning eller inkonsekvent användning av versaler. Särskilt användbart är att verktöget, precis som när det gäller fasettering, anger antalet förekomster av ett specifikt cellvärde (en stavning t.ex.) intill värdet. Det underlättar undersökning av de inkonsekventa posterna och gör det lättare att rätta felen.

---

<sup>2</sup> LOD står för Linked Open Data, eller Länkade Öppna Data. Se vidare "Vitbok länkade öppna data": <http://lankadedata.se/vitbok/introduktion.html>

## Lära mer!

På nätet finns många fria kurser och andra resurser för den som vill lära sig mera om OpenRefine. Ett bra ställe att börja på kan vara svenska MetaSolutions Länkade data<sup>3</sup>, eller belgiska Free your metadata<sup>4</sup> som ger bra vägledning till snabb nytta, även för nybörjare.

## Lästips

*Europeana Metadata Quality Task Force Report* (2015)<sup>5</sup>

Seth Van Hooland & Ruben Verborgh *Linked Data for Libraries, Archives, and Museums* (2014)<sup>6</sup>

Matthias Palmér & Hannes Ebner *Vitbok Länkade Öppna Data* (2014)<sup>7</sup>

Ruben Verborgh & Max De Wilde *Using OpenRefine* (2013)<sup>8</sup>

Checklistan har utarbetats med stöd av en guide från Digital Repository Ireland (DRI): McCarthy, Kate (2014), "Metadata Quality Control". Dublin: Royal Irish Academy. DOI: 10,3318 / DRI.2015.1 (CC:BY 4.0)

Version 1.0. Licens CC-0. En senare version kan finnas på [www.digisam.se](http://www.digisam.se)

Digisam är ett samordningssekretariat för digitalisering, digitalt bevarande och digitalt tillgängliggörande av kulturarvet. Samordningssekretariatet är en av regeringen beslutad verksamhet vid Riksarkivet.

---

<sup>3</sup> <http://lankadedata.se>

<sup>4</sup> <http://freeyourmetadata.org>

<sup>5</sup> <http://pro.europeana.eu/publication/metadata-quality-task-force-report>

<sup>6</sup> <http://book.freeyourmetadata.org>

<sup>7</sup> <http://lankadedata.se/vitbok>

<sup>8</sup> <https://www.packtpub.com/big-data-and-business-intelligence/using-openrefine>