

Posttyper i DiVA på Cora

Bakgrund

Det finns ett antal ärenden i DiVA:s backlog¹ sedan 2015 som handlar om att anpassa DiVA till SwePubs nationella indelning av forskningsoutput. De flesta har aldrig genomförts men i och med förflyttningen till Cora blir det möjligt att utforma de nya posttyperna så att en fullständig mappning blir möjlig. SwePubs format har dessutom uppdaterats ett antal gånger sedan 2015. Den aktuella listan över forskningsoutput är från 2019² och är kopplad till SwePubs formatspecifikation 3.0.

Samtidigt skulle en lösning där man utformar varje SwePub-typ som egen posttyp enligt SwePubs platta modell vara klumpig och opraktisk för den som registrerar poster i DiVA. Huvudstrategin bör därför fortsatt vara som 2015 att de huvudtyper man väljer att registrera i huvudsak är de samma som nu men att det finns möjlighet att via undertyper specificera ytterligare. Detta gör också det mer flexibelt om nya outputtyper skulle uppstå som i sådana fall förhoppnings kan läggas som undertyper istället.

Dock behöver vissa anpassningar göras för Cora.

Gemensamt rotelement

Bibliografiska poster från Cora bör ha ett enhetligt rotelement när användaren hämtar poster i Cora-formatet via API:et eller export. Idag används elementet *publication* i Fedora vilket skulle passa mindre bra eftersom nu *publication* är en särskild typ av output. I Alvin har posterna roten *record* (oavsett resurstyp).

För att göra detta upprättas en abstrakt posttyp (motsvarar kolumnen Output i tabellen nedan) som de underliggande posttyperna ingår i.

Löpande id:n

I Alvin hänvisar också roten *record* till dess id:n som motsvarande heter *alvin-record:XXXX* för bibliografiska poster, dvs. en gemensam löpande svit för alla resurstyper där befintliga post-id:n förs över till Cora och sviten fortsätter efter migrering. I Alvin har *alvin-id:n* och *urn:nbn* samma löpnummer.

För DiVA behöver man fastställa dels om/hur man behåller dagens post-id:n (*diva2:XXXXX*) samt hur de olika *urn:nbn*-serierna skall upprätthållas. I DiVA har dessa en separat numrering för varje domän som inte är kopplad till DiVA:s eget post-id.

Posttyper i Cora

Posttyper i Cora skapas utifrån kolumnen 3 Typ nedan, dvs. i princip samma som i Classic. Genom att skapa en särskild posttyp för varje typ blir det möjligt att definiera vilka speciella fält och villkor som gäller för varje typ, dvs. artikel, bok etc. Denna typ är den användaren väljer vid manuell registrering i formulär.

För att kunna identifiera vilken outputtyp (kolumn 2) en typ tillhör läggs ett fast "final value" in i ett eget element som sätts automatiskt när man registrerar en typ. Likaså läggs ett fast värde in som typ

¹ <https://jira.epc.uu.se/browse/DIVA-2273>

² https://www.kb.se/download/18.d0e4d5b16cd18f600e1eca/1611326441553/SIFO_2019.pdf

när man valt att registrera t.ex. en bok. Dessa element definieras för varje posttyp och obligatoriska 1-1.

För vissa typer krävs dessutom att en undertyp anges (kolumn 4). Dessa undertyper kopplas till respektive typ och anges vid registreringen av varje. Det kan finnas defaultvärden. Om undertyper finns är de obligatoriska 1-1. De motsvarar i princip dagens underkategorier.

En bibliografisk post kommer alltså att ha:

1. Ett gemensamt abstrakt rotelement.
2. En obligatorisk outputtyp som sätts automatiskt vid manuell registrering. Värdet motsvarar SwePub.
3. En obligatorisk typ som sätts automatiskt vid manuell registrering efter val av typ. Värdet motsvarar SwePubs kategorier om inga undertyper finns, annars enligt "egen" lista.
4. En undertyp som är obligatorisk om den finns som sätts manuellt vid manuell registrering (defaultvärde kan finnas). Värdet motsvarar SwePubs kategorier.

Exempel på struktur

```
<root>
  <outputType>publication</outputType>
  <type>article</type>
  <subType>journal-article</subType>
</root>
```

Transformation till SwePub

```
<genre authority="kb.se" type="outputType">publication/book-chapter</genre>
```

skapas genom

```
<genre authority="kb.se" type="outputType">
  Värdet av outputType
  /
  Om subtype finns
    Värdet av subType
  Annars
    Värdet av type
</genre>
```

```
</genre>
```

Utgångspunkt för grundstrukturen i Cora baserat på dels nationell indelning dels dagens Classic kan se ut så här:

Output (abstrakt)	Outputtyp	Typ	Undertyp
	Publikation	Bok	-
		Samlingsverk	Samlingsverk
			Special-/temanummer av tidskrift
		Kapitel/bidrag	Samlingsverk
			För-/Efterord
			Rapport
			Encyklopedi
		Rapport	Rapport
			Working paper
		Artikel	Vetenskaplig tidskrift
			Forskningsöversikt
			Inledande text
			Recension
			Övrig tidskrift
			Dags-/nyhetstidning
		Doktorsavhandling	-
		Licentiatavhandling	-
		Textkritisk utgåva	-
		Preprint	-
		Övrig	-
	Konferens	Paper	Paper
			Poster
			Övriga
		Proceedings	-
	Konstnärlig	Konstnärligt arbete	-
		Dokumenterat konstnärligt forskningsprojekt (doktorsavhandling)	-
	Immaterialrättslig	Patent	-
		Övriga	-
	Övrig	Dataset	-
		Programvara	-

Kommentarer

- I Classic är strukturen för artikel lite ologiskt i förhållande till övriga typer. Undertyperna Forskningsöversikt och Recension är egna typer medan övriga undertyper återfinnes under Artikel i tidskrift. Om man vill behålla Forskningsöversikt och Recension som egna typer i DiVA skulle det naturligtvis gå.
- För doktors- och licentiatavhandling kommer man sannolikt i DiVA fortsatt vilja ha egna typer för monografier respektive sammanfattningar eftersom dessa har lite olika underelement.
- I DiVA behövs fortsatt typen Dissertation (äldre avhandling) som inte finns i SwePub.
- För DiVA behövs utöver forskningsoutput en typ för studentuppsatser.

Nya typer i DiVA på Cora

- Textkritisk utgåva är en ny typ som inte finns som typ i Classic men som inte heller passar som undertyp.
- Dokumenterat konstnärligt forskningsprojekt (doktorsavhandling) blir en egen typ under Konstnärlig output.
- Immaterialrättslig, övriga samt Övrig, programvara finns inte i DiVA men är troligen inte heller nödvändiga. Behöver utredas.

Mappningar från Classic

- Underkategori för Samlingsverk kan bara mappas mot Samlingsverk eftersom inga underkategorier finns i Classic.
- Underkategori för Kapitel kan bara mappas mot Samlingsverk eftersom inga underkategorier finns i Classic.
- Textkritisk utgåva. Mappning från Classic ej möjlig maskinellt.
- Två underkategorier till Artikel – Letter och Editorial – motvaras av endast en i SwePub: inledande text. Mappning möjlig men behovet av två olika behöver utredas.
- Underkategorierna meetingAbstract för artikel samt abstract för konferensbidrag i Classic skall enligt nationell praxis anges som "Övriga konferensbidrag". Denna mappning möjlig men behöver utredas.
- Mappning mot konstnärlig doktorsavhandling kan göras om konstnärligt arbete är ikryssat i Classic.

/Stefan 2022-10-10