

# DiVA-forum

2023-11-16

DiVA 

# DiVA-forum den 16 november 2023

## Agenda

- Information från förvaltning (Eva)
- Information från projektet (Linda)
  - Cora-utvecklingen (Olov)
  - Nya metadatamodellen m.m. (Michael, Petra och Sara)
  - Nya gränssnitt (Egil)
  - Drift av Cora och nya DiVA (Andreas och Palle)
- Frivillig frågestund med supporten (Michael, Petra och Sara)



# Information från förvaltning

Eva Agius



# DiVA Classic

- Ingen utveckling – huvudfokus på projektet.
- Medarbetarna i förvaltningen arbetar merparten av arbetstiden i projektet.
- Hittills i år två releaser:  
2023-04-19 version 2.39.0  
2023-10-25 version 2.40.0

[Release notes](#)



# Drift och systemunderhåll

- Som jag berättade på DiVA-forum i maj så har driften arbetat med att säkerhet och tillgänglighet förbättrats.
- Sett över våra servrar och nätverk, för att kunna dirigera om datamängd vid höga belastningar m.m.
- Vi har utökat övervakningen av serverna för att mäta datatrafik och belastningar.
- Innan sommaren skedde även en stor förberedelse av Andreas och Palle på driften inför underhållsarbetet den 1 juli.

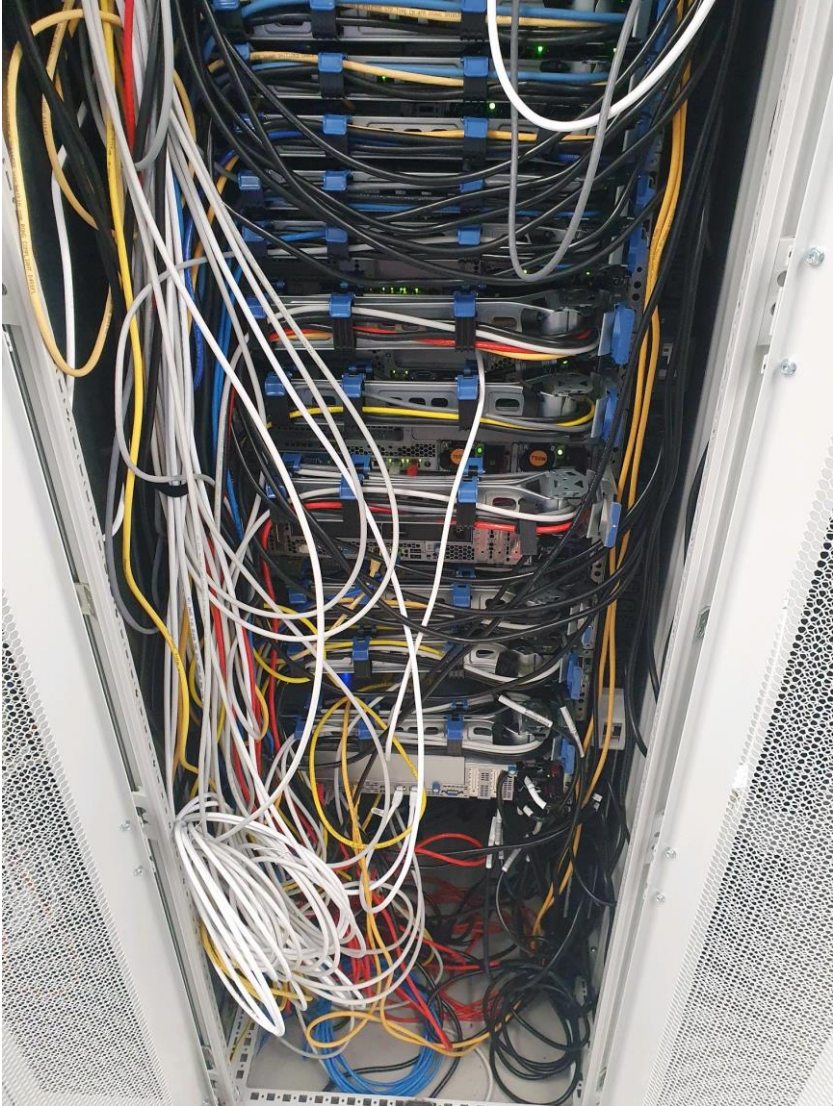


# Drift och systemunderhållet den 1 juli

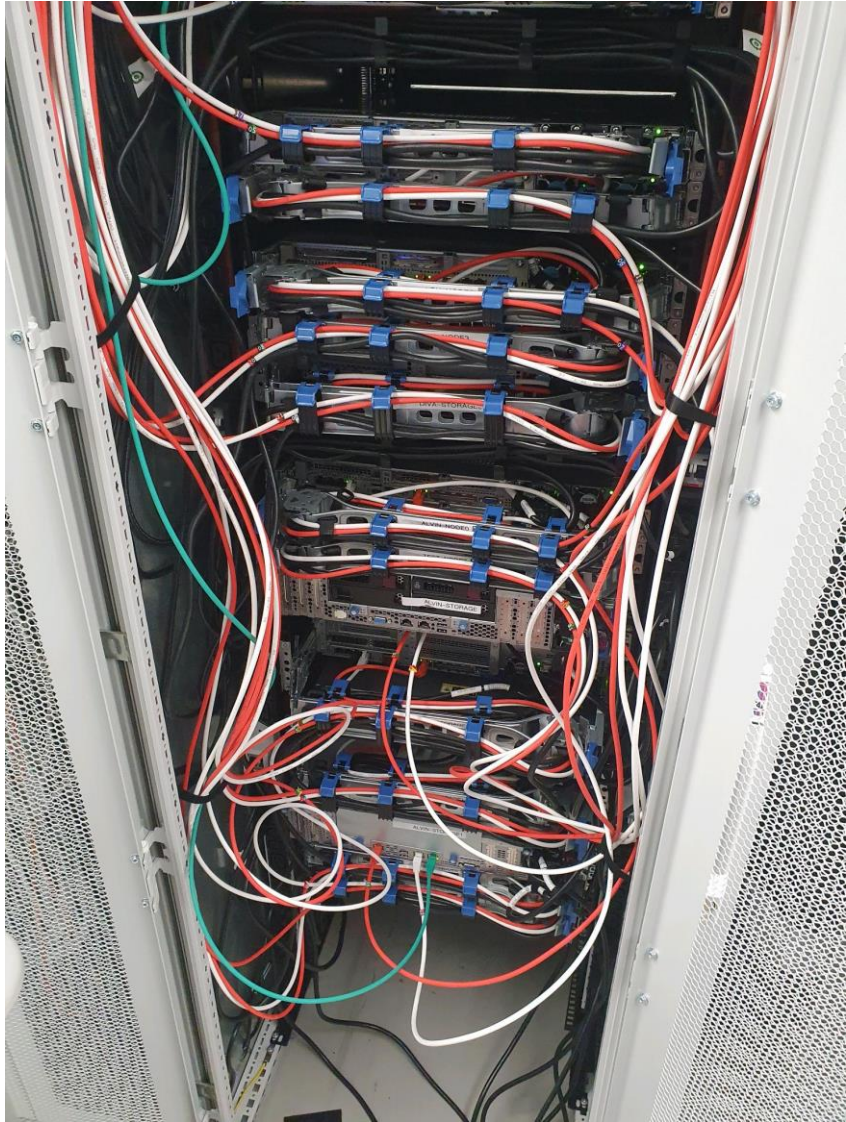
- Ett arbete som varit eftersatt länge för att vi inte haft bemanningen på drift att lägga den tid som behövs.
- Det har krävts en noggrann planering innan eftersom det var viktigt att allt skulle gå rätt till.
- Arbetet innebar en städning av serverrackets:
  - bytt ut alla nätkablar
  - färgkodat alla kablar och märkt upp dem
  - bättre kabelhantering
- Mycket lättare för driften att arbeta i racket nu när vi t.ex. ska installera/avinstallera servrar eller göra modifikationer på nätverket.
- Driften har fått bättre överblick om hur DiVA:s backend är strukturerad och har dessutom fått testa på att ta isär allt och bygga ihop det igen.



Före



Efter



# Praxisgruppen

- Praxisgruppen fortsätter att träffas ett par gånger per termin.
- Allt färre praxisfrågor som inkommer. Kan tyda på att dokumentationen är bra?

Mejla gärna frågor till  
[diva-support@ub.uu.se](mailto:diva-support@ub.uu.se)

## **Praxisgruppen:**

Alexandar Bobic, Stockholms universitet  
Marcus Israelsson, Luleå tekniska universitet  
Elisavet Koutzamani, Linköpings universitet  
Greta Linder, Södertörns högskola  
David Scheutz, Konstfack  
Michael Andersson, DiVA-support  
Eva Agius, DiVA-konsortiet





# DiVA:s publika Wiki och Jira försvinner

- Den typ av licens som vi har just nu (serverlicens) kommer att fasas ut av leverantören och licenserna kan inte förnyas från mars nästa år och leverantören utför då inte fler säkerhetsuppdateringar.
- Man kommer bara kunna köpa deras datacenterlicens framöver för installationer on-premis och dessa är en bra bit dyrare än serverlicenserna, ca 350 000 kr/år.
- En billigare lösning är att ha programmen som molntjänster, men på grund av informationssäkerhetsrisken är det ingen bra lösning.
- Förvaltnings- och utvecklargruppen kommer fortsätta använda programmen i vårt arbete och får köpa in sig på 15 Jira-licenser hos Uppsala universitet.
- **Förvaltningen och projektet kommer istället kommunicera med användare genom att använda hemsidan och e-post.**
- För att kunna följa ärenden i Jira funderar vi på hur vi kan ordna det på ett bra sätt.



# Ny hemsida

- Ny hemsida för DiVA är på gång eftersom Uppsala universitet byter webbplattform till SiteVision. ( [Start - DiVA \(diva-portal.org\)](http://diva-portal.org) )
- Flyttarbete kommer påbörjas under december.
- Nya webbplatsen beräknas vara klar i mars och då även innehålla informationen som finns i dagens wiki.
- Ett informationstillfälle kommer ges via Zoom i mars där vi visar hur ni hittar på webben. Mer info kommer.



Frågor?



# Nuläge projektet Nya Alvin och Diva

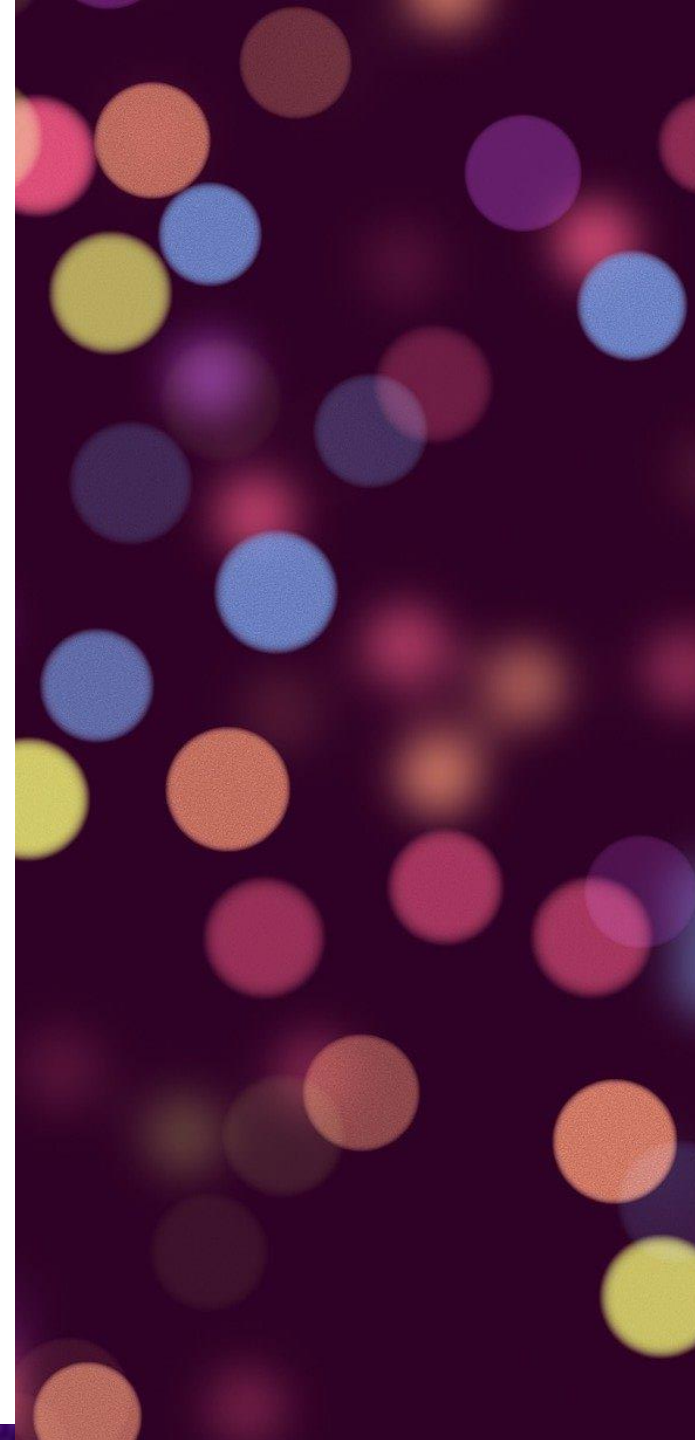
DiVA-forum 2023-11-16

Linda Jernstedt

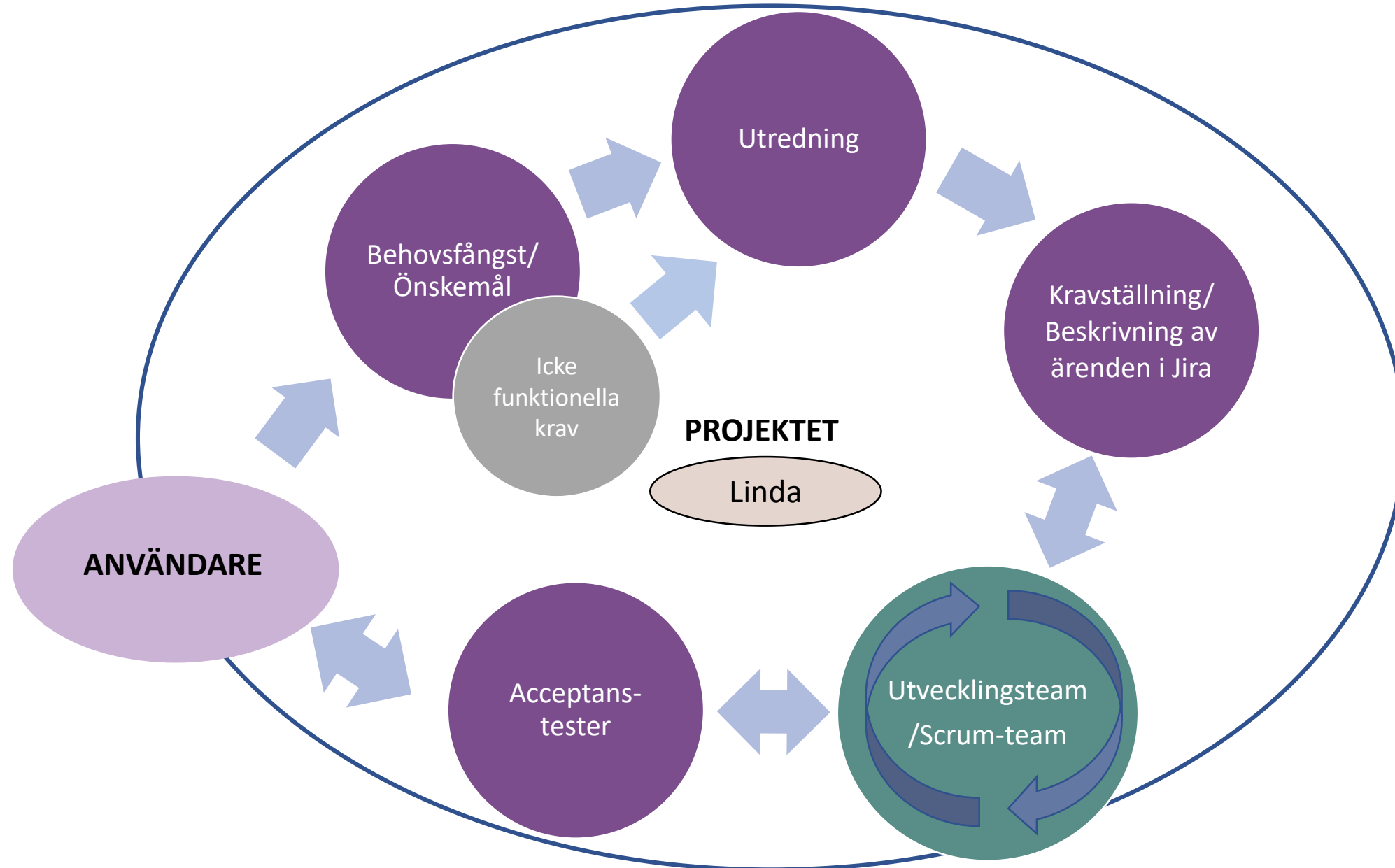


# Projektet Nya Alvin och DiVA

- Ny teknisk plattform för DiVA och Alvin. Grunden är gemensam, förgrenas till två system.
- Projektledare: Linda Jernstedt. 50 %
- Produktägare/förvaltningsledare DiVA: Eva Agius. 50 %
- Produktägare/förvaltningsledare Alvin, senior rådgivare DiVA: Stefan Andersson. 50 %
- Projektgrupp 14 personer, som också jobbar i förvaltningen av DiVA Classic.



# Vår projektorganisation



## Resurser i lila cirklar

**DiVA:** Michael, Sara, Petra, Eva + DiVA-specialister.

Stefan senior rådgivare.

**Alvin:** Kent-Inge och Stefan

**Drift:** Palle och Andreas

## Utvecklingsteamet

**Produktägare Alvin:** Stefan

**Produktägare DiVA:** Eva

**Scrum master:** Kent-Inge

**Cora-utveckling:** Olov, Pere och Marcus

**Front-end:** Egil och Erik

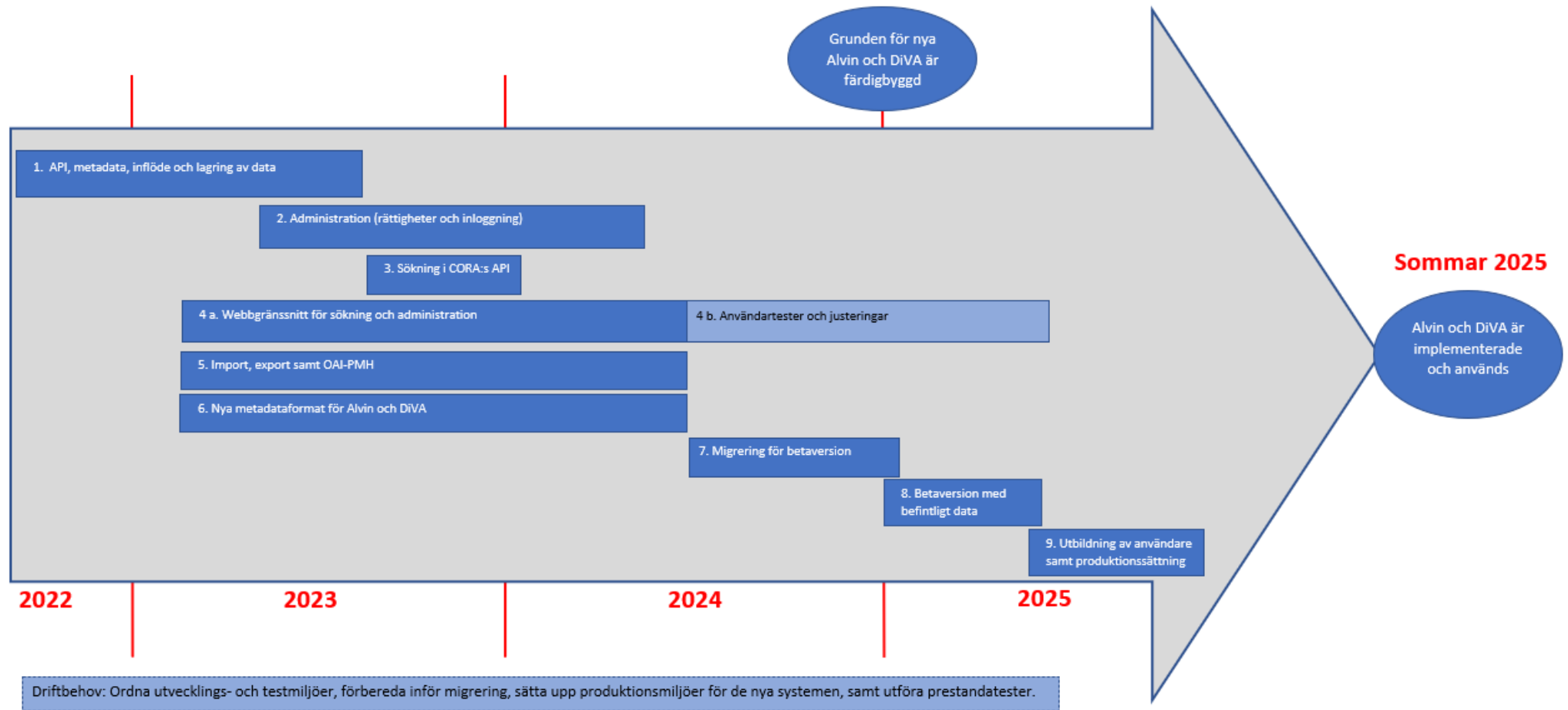
**Drift:** Palle och Andreas

# Nuläge

- Vad jag informerade projektets styrgrupp om i oktober:
  - Arbetet i projektet.
  - Ändringar i organisation och arbetssätt
  - Att vi testar att använda API:erna, redan nu underlättar det viss del av arbetet med att bygga systemet. Dock inte möjligt för medlemmar att testa än.
  - Försening och insatser runt försening.
  - Min rekommendation framåt.
- Hur har det gått sedan dess?
  - Det börjar lossna, hela utvecklingsteamet har nu tillräcklig förståelse av Cora och har kommit igång med att leverera resultat.
  - Ytterligare försening har dock tillkommit.



# Roadmap

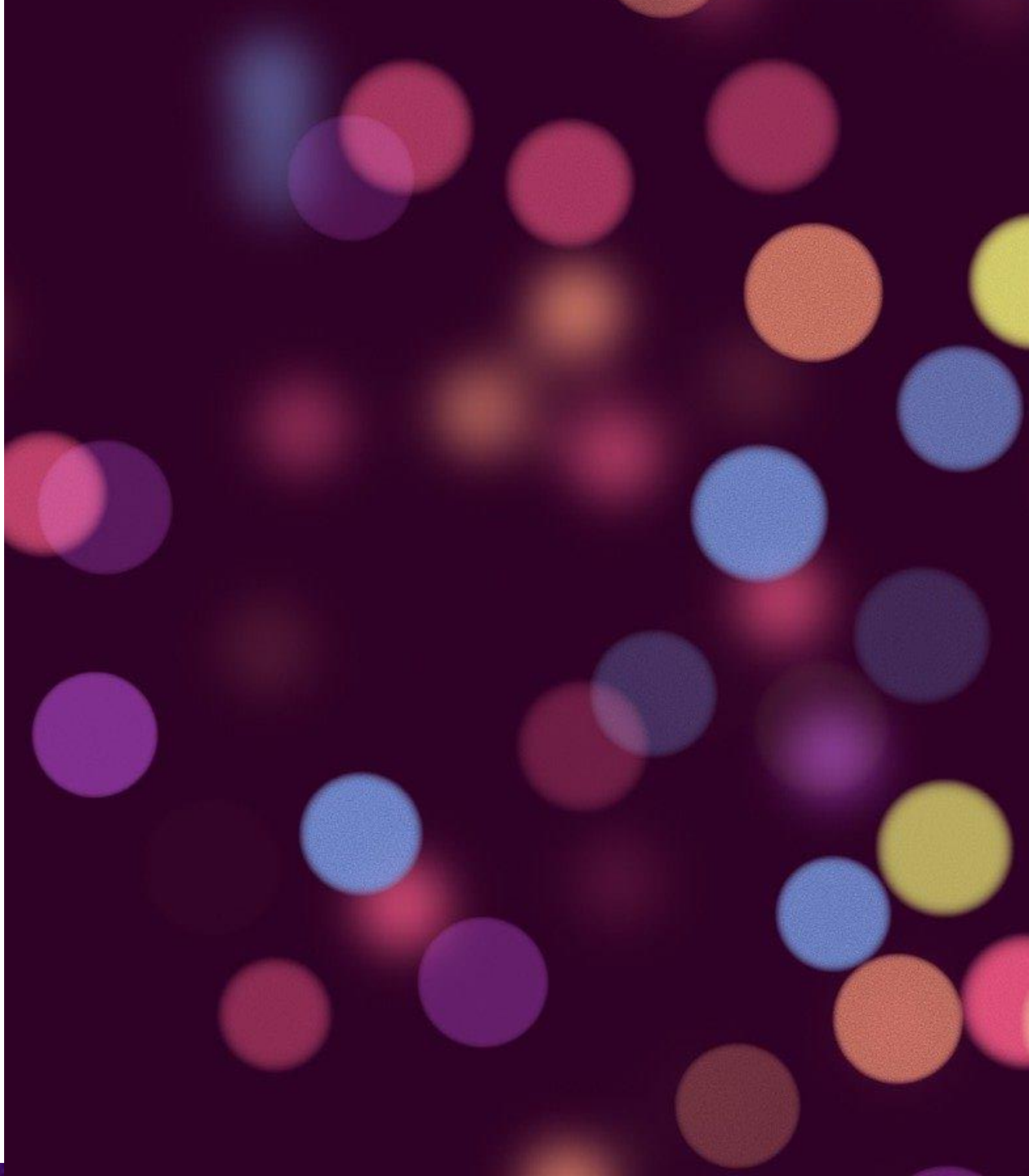


[Länk: Roadmap med tillhörande projektaktiviteter](#)



# På gång i projektet för er specialister

- Enkät om systemberoenden hos medlemmar skickas ut inom kort.
- Referensgruppen för DiVA:s nya användargränssnitt möts i december.



Frågor?



**DiVA powered by Cora**  
**2023-11-16**

Olov McKie

# Utveckling från förra mötet

- **Färdigställt valideringstyp**
- **Arbete pågår för binära filer**
- **Inloggning och rättigheter näst på tur**

# Övrig utveckling

- **Default label för variable**
- **Default rubrik för grupper**
- **MinNumberToShow default=1**
  
- **(JSClient) Klickbar info för alla variabeltyper och grupper**

# Annat som görs

- **Upplärning av ny personal**
- **Nya versioner**
  - **Postgres 16**
  - **Solr 9**
  - **Etc.**

# Målsättningar

- **REST-API**
  - Uppdaterat maskinläsbart format
  - Åtkomst till data i samma format som det är lagrat
  - Alla systemets funktioner åtkomliga via API:et

# Ny metadatamodell för DiVA

Michael Andersson, DiVA support





# Innehåll

- Arbete sen förra DiVA-forum
- Utgångspunkter och mål
- Skillnader mellan nuvarande och kommande system
- Några nyheter i modellen
- Vidare arbete



# Arbete sen förra DiVA-forum

- Arbete med grundläggande metadatamodell för output
- Utredning av mappning mellan befintliga och nya modellen
- Utredning av ändringar samt tillägg för export till andra system
  - Swepubmods, Openaire, Datacite



# Utgångspunkter och mål

- Nya modellen ska:
  - kunna ta hand om befintligt data
  - kompletteras och struktureras om för att möta krav/önskemål
- Målen för modellen är:
  - enklare underhåll och uppdateringar
  - ökad harmonisering med andra metadatamodeller



# Skillnader mellan nuvarande och kommande system

- Regelstyrt på alla nivåer
- Varje typ av post har en mall som måste följas
- All output delar en gemensam grundtyp som innehåller all möjlig metadata
- Varje typ av output - exempelvis samlingsverk – har en mall som definierar vilket metadata som ingår i den
- Vid varje skapande eller uppdatering av en post valideras innehållet mot mallen



# Skillnader mellan nuvarande och kommande system

- Upphov som exempel
- En övergripande grupp samlar alla typer av upphov som ska kunna registreras
  - Författare, redaktörer, annan upphovsperson, gruppförfattare och utgivande organisation
- För en enskild typ av output så skapas en mall där gruppen upphov är obligatorisk
- I mallen specificeras vilka typer av upphov som ska kunna registreras i outputtypen



# Skillnader mellan nuvarande och kommande system

- contributor

- author

- divaPerson
- givenName
- familyName
- birthYear
- deathYear
- email
- localId
- ORCID
- organisation
- otherOrganisations
  - otherOrganisation
  - rorId
- researchGroup
- otherResearchGroup

- editor

- divaPerson
- givenName
- familyName
- birthYear
- deathYear
- email
- localId
- ORCID
- organisation
- otherOrganisations
  - otherOrganisation
  - rorId
- researchGroup
- otherResearchGroup

- responsibleOrganisation

- organisation

- groupAuthor

- name
- organisation
- otherOrganisations
  - otherOrganisation
  - rorId
- url



# Några nyheter i modellen

- Harmonisering med Swepubs outputtyper
- Lägga till licensinformation
- Lokala kontrollerade listor för Forskargrupp
- Lokala kontrollerade listor för sökning
- Möjlighet att lägga till Gruppförfattare som upphov
- Möjlighet att lägga till ROR-id vid registrering av externa organisationer



# Vidare arbete

- Färdigställa modellen inför konvertering
- Återkoppling från er (mer information kommer)





Frågor?



# Bland gränssnitt och metadata

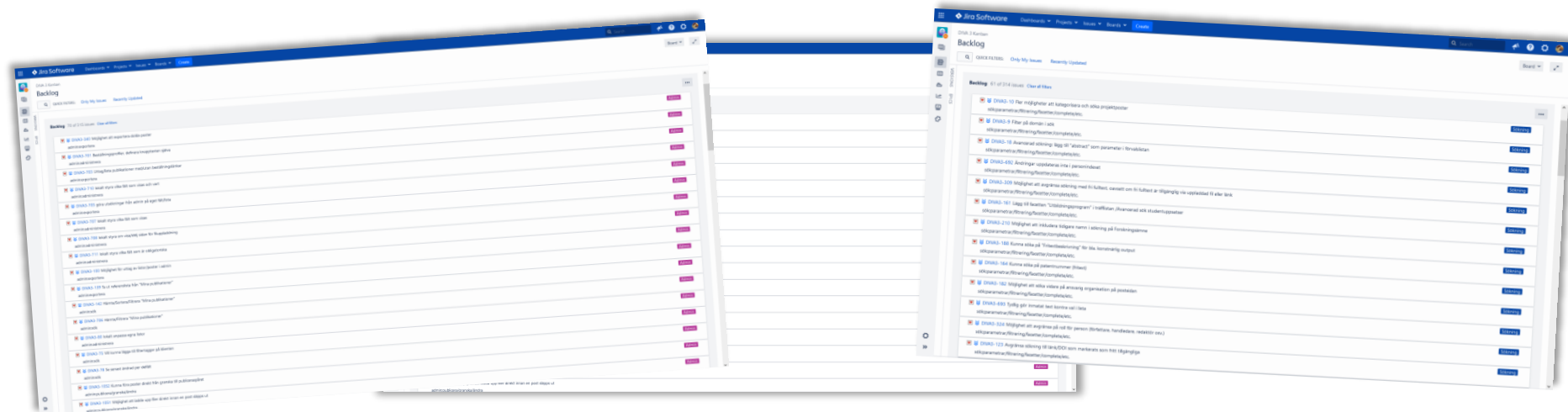
Dåtid, nutid och framtid

Sara Tobiasson och Petra Thorsson



# DiVA-forum i maj 2023

- Ta fram och skriva grundläggande funktionella krav för systemet
- Utgångspunkt från de önskemål som inkommit under åren
- Över 300 ärenden! Vi kommer inte kunna göra allt
- Sortering, gruppering och ställningstagande
- Överblick över det fortsatta arbetet



# Våren – publikationsflödet

- En användare ska kunna välja och registrera en publikationstyp
  - Vad finns idag?
  - Vilka önskemål finns?
  - Vad behöver vi?



# User Stories

En kort beskrivning av vad en användare vill uppnå i grundformen  
*som [typ av användare] vill jag [målet] för att [anledningen]*

Som kund vill jag kunna ta ut kontanter.

- Vem?
- Vad är målet? Vad behöver vi bygga?
- Vad är anledningen? Vilket är syftet?
  - Anledningen är inte alltid nödvändig, men kan tydliggöra skälet till att bygga funktionen



# Avvägningar

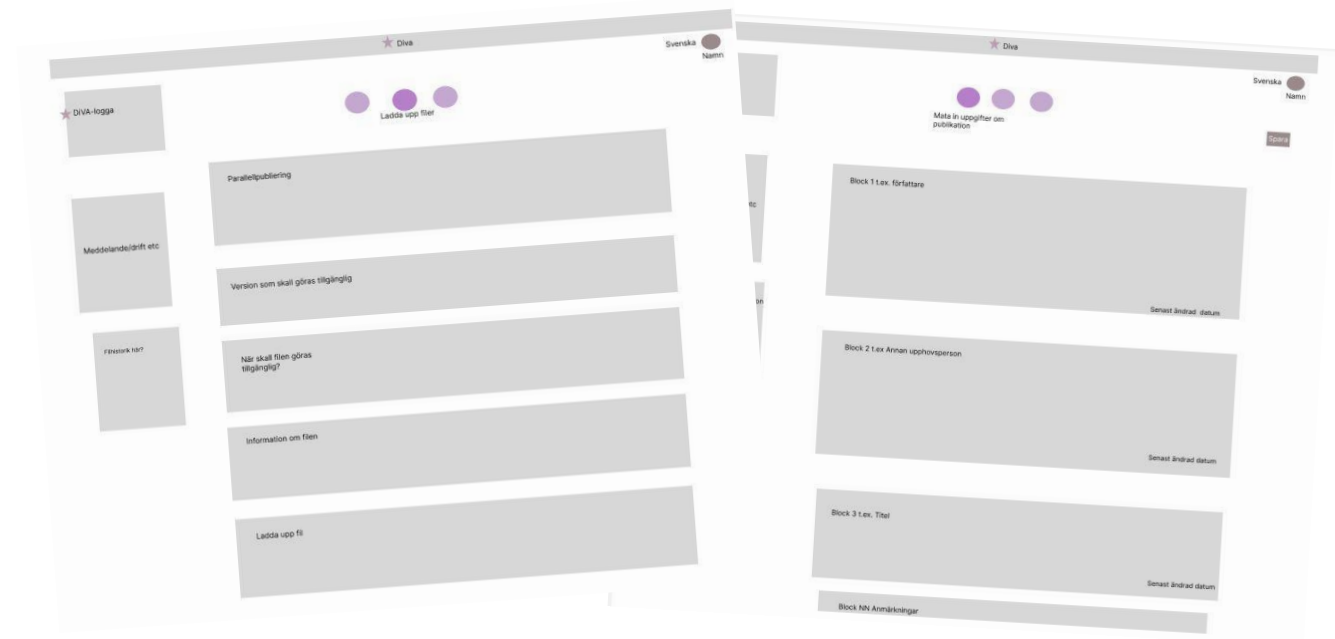
- Ta hand om befintligt data
- Önskemål och framtida behov
- Exporter till olika system



# Hösten – administrationsgränssnittet

- Vad behöver en administratör kunna göra?
- Vad behöver en domänadministratör kunna göra?

- Vad finns idag?
- Vilka önskemål finns?
- Vad behöver vi?



# Det fortsatt arbetet

- Fortsätta identifiera och definiera vad som behöver tas med i det fortsatta arbetet
- Arbeta fram och sätta metadatamodellen





Tack!  
Frågor?



# Nya gränssnittet

Egil Swenning Leyser



# Utmaningar

- Skisserna var inte jätteutförliga, men vi har tagit inspiration från dem.
- La våren på att bygga grafiska komponenter som vi använder oss av nu.
- Jobbet sen senaste DiVA-forumet har varit att koppla ihop dem med metadatan vi får från Cora.



# Utmaningar

- Utforskat olika vägval och i och med det så har vi valt att koppla den hårdare mot Cora, något vi inte gjorde tidigare.
- Nu ser vi att det kommer att fungera och vara möjligt att jobba vidare med.
- I och med detta så har vi haft intern utbildningstid nu under hösten.



# Lösningar

- Läsa så mycket vi kan från Cora, både metadata och presentation.
- Detta gör att förvaltningen i fortsättningen kommer att vara enklare.



# Presentation

Författare 1 ?

Koppla personpost +

Hämta sparade personuppgifter +

Förnamn \*

Du behöver fylla i efternamnet

Efternamn

Lokal användaridentitet ?

ORCID-identitet ?

Skapa ORCID-ID →

Födelseår

Institution, avdelning eller program

Forskargrupp ?

E-postadress

Annan organisation ?

Lägg till författare +

Lägg till annan upphovsman +

^ Title ?

Language ?

Main title ?

Sub title ?

^ Alternative title ?

Language ?

Type ?

Main title ?

Sub title ?



Frågor?



# **Driftsmiljö DiVA Cora**

Palle Raabjerg & Andreas Mattsson

**DiVA Konsortiet**

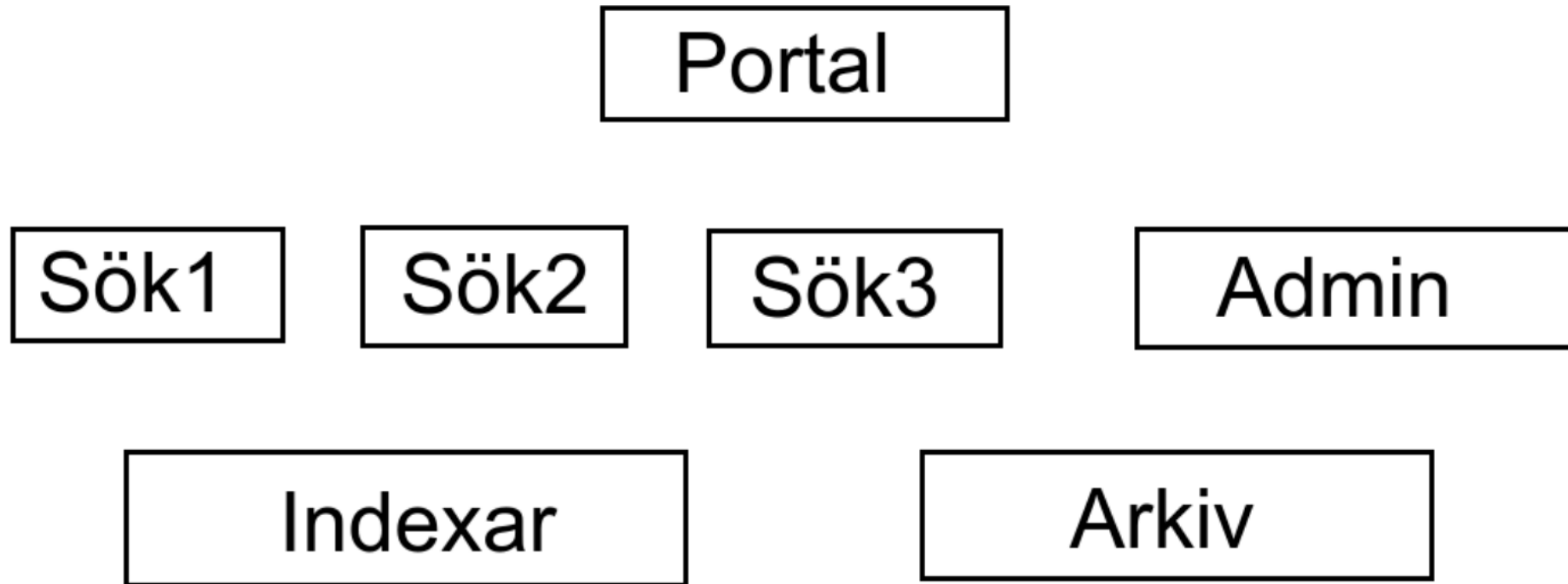


# Mål

- Snabbare svar
  - Varnish Cashe
- Högre tillgänglighet
  - Clustermiljö

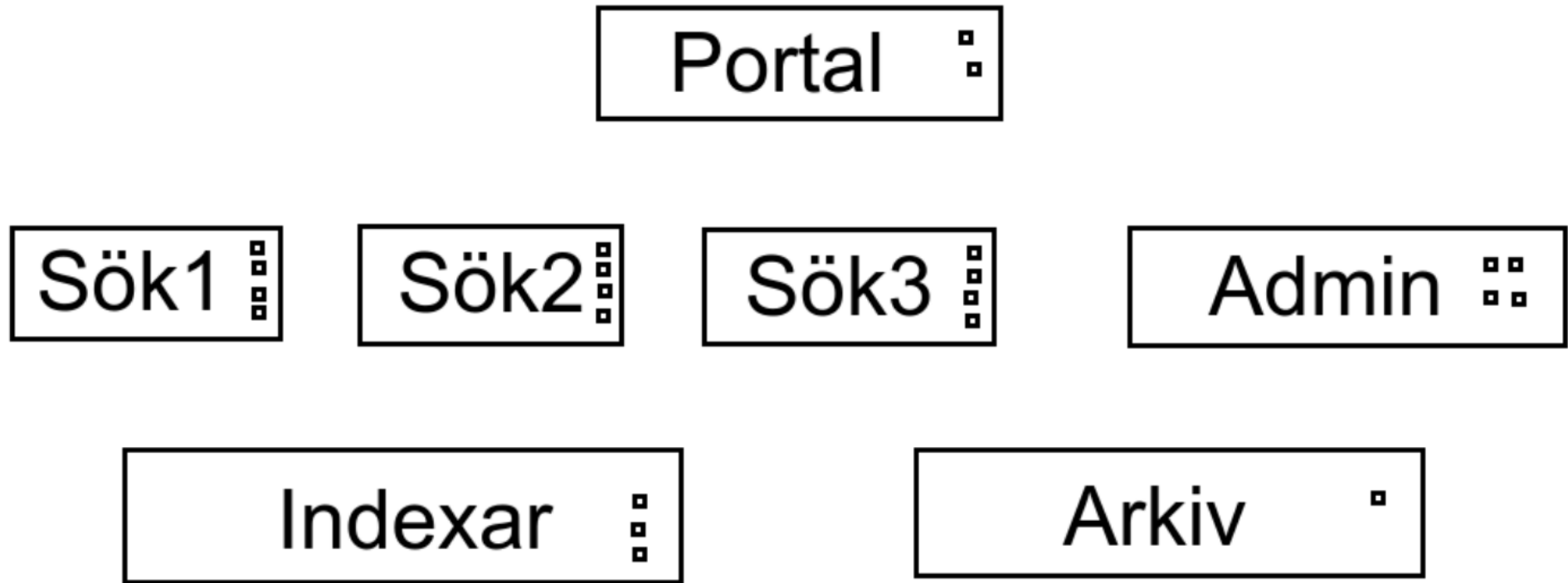
# Driftsarkitektur

Nuvarande driftsarkitektur: Ett antal statiskt uppsatta servrar med specifika roller.



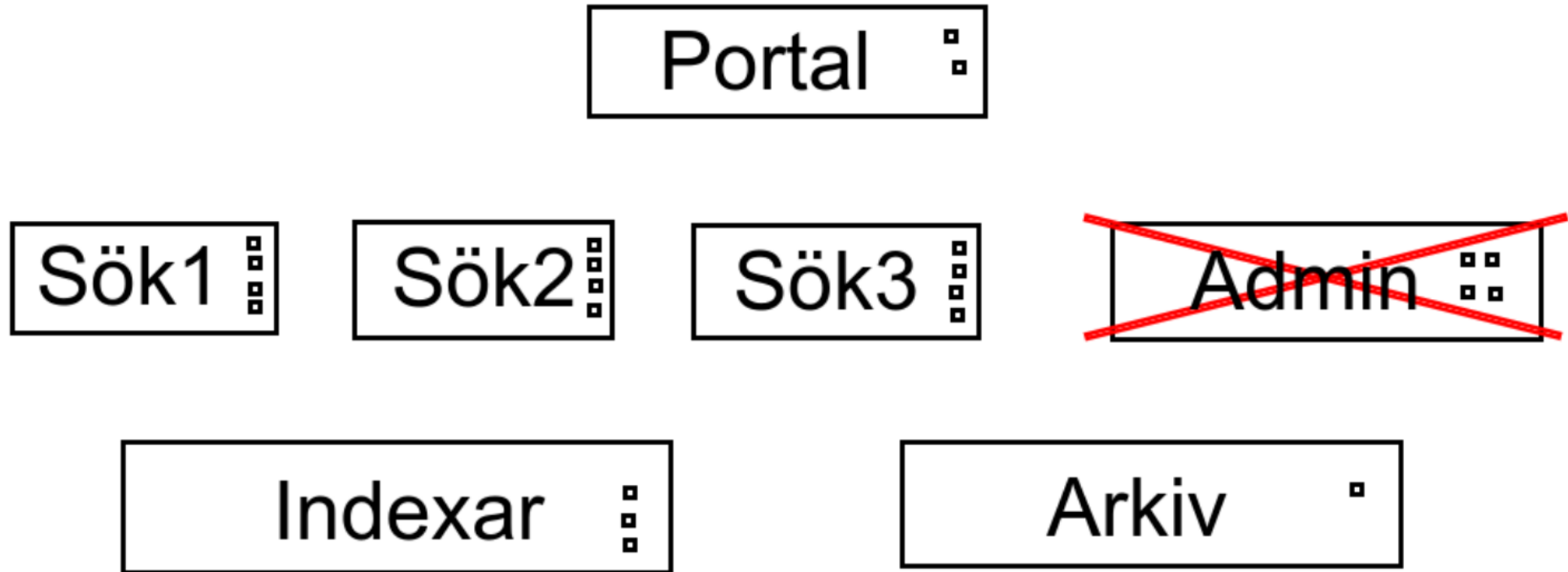
# Driftsarkitektur

Services är bundna till specifika servrar.



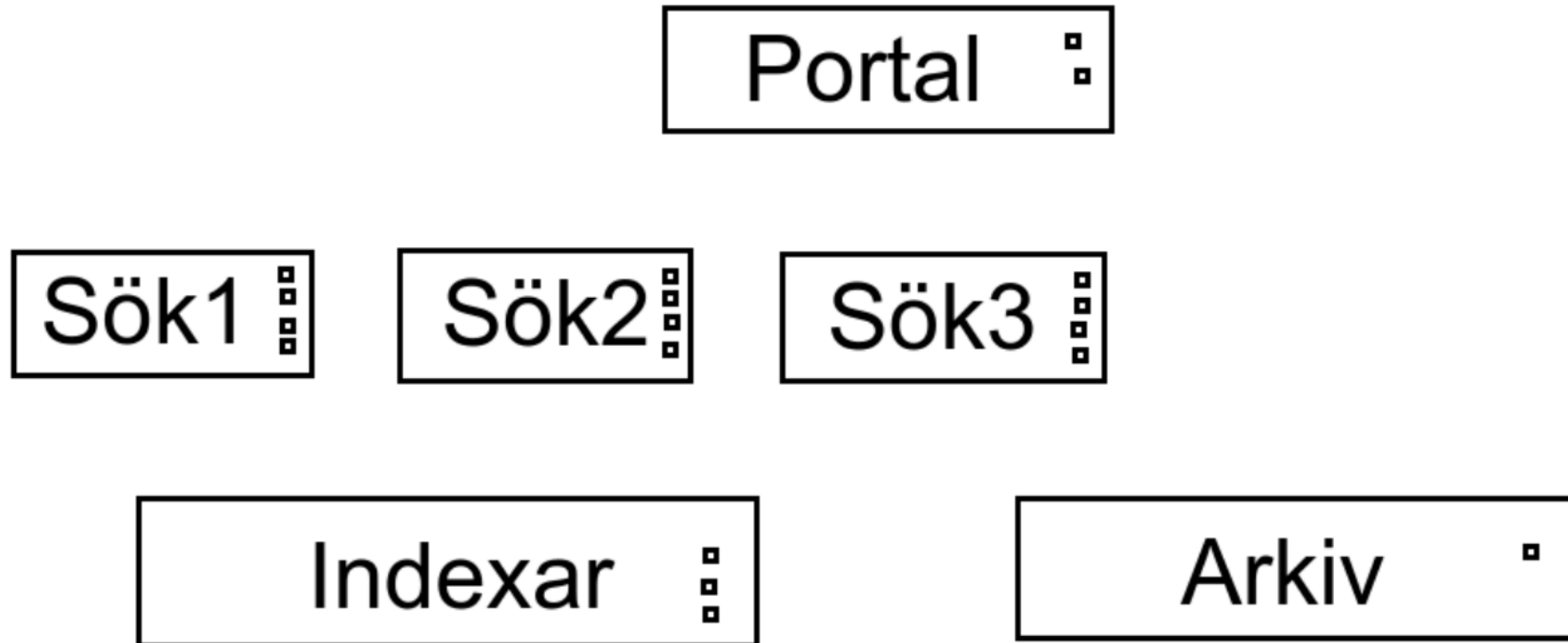
# Driftsarkitektur

Om en server av någon anledning dör eller tas ur drift...



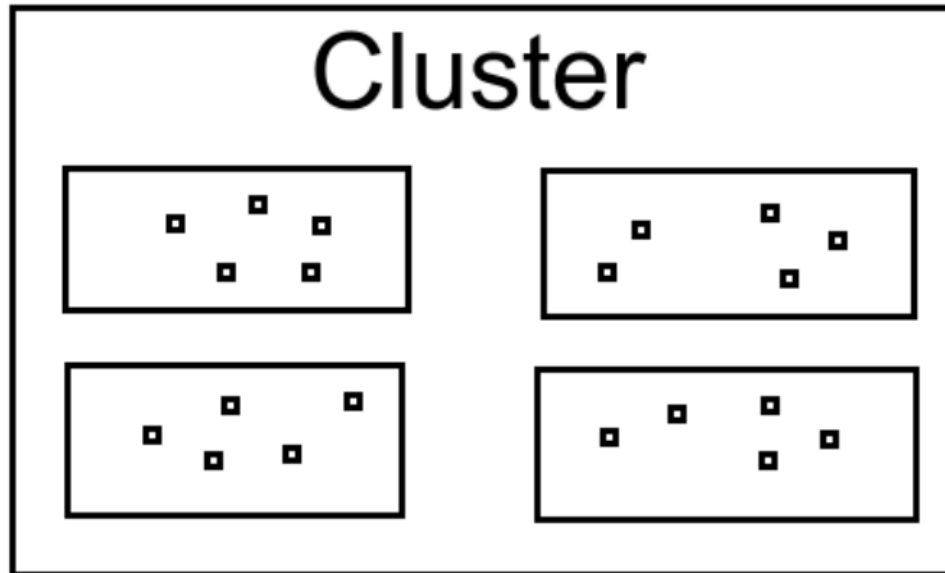
# Driftsarkitektur

...så försvinner just nu även de bundna services, vilket kan störa driften av systemet i helhet. Det är bland annat därför vi i dagsläget jobbar med "underhållsfönster" där vi stänger av det hela när vi behöver uppdatera eller uppgradera systemen.



# Driftsarkitektur

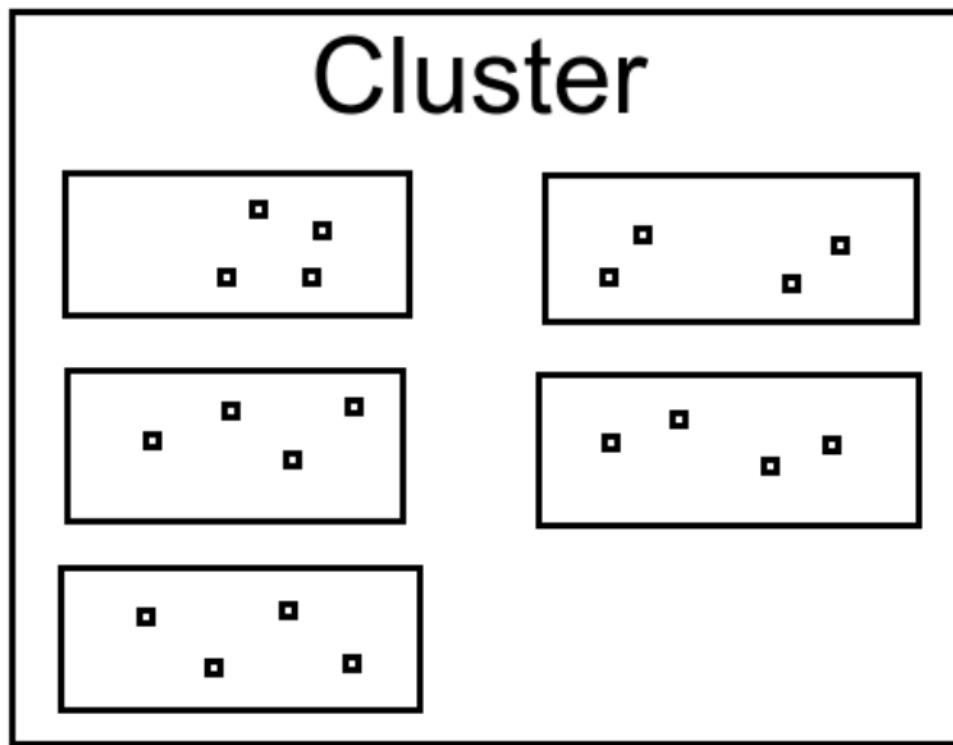
Med clusterhantering kan vi sätta upp ett villkorligt antal servrar och definiera en "applikationsstack" som bland annat specificerar hur de olika services behöver kommunicera med varandra och hur många instanser det ska finnas.



# Driftsarkitektur

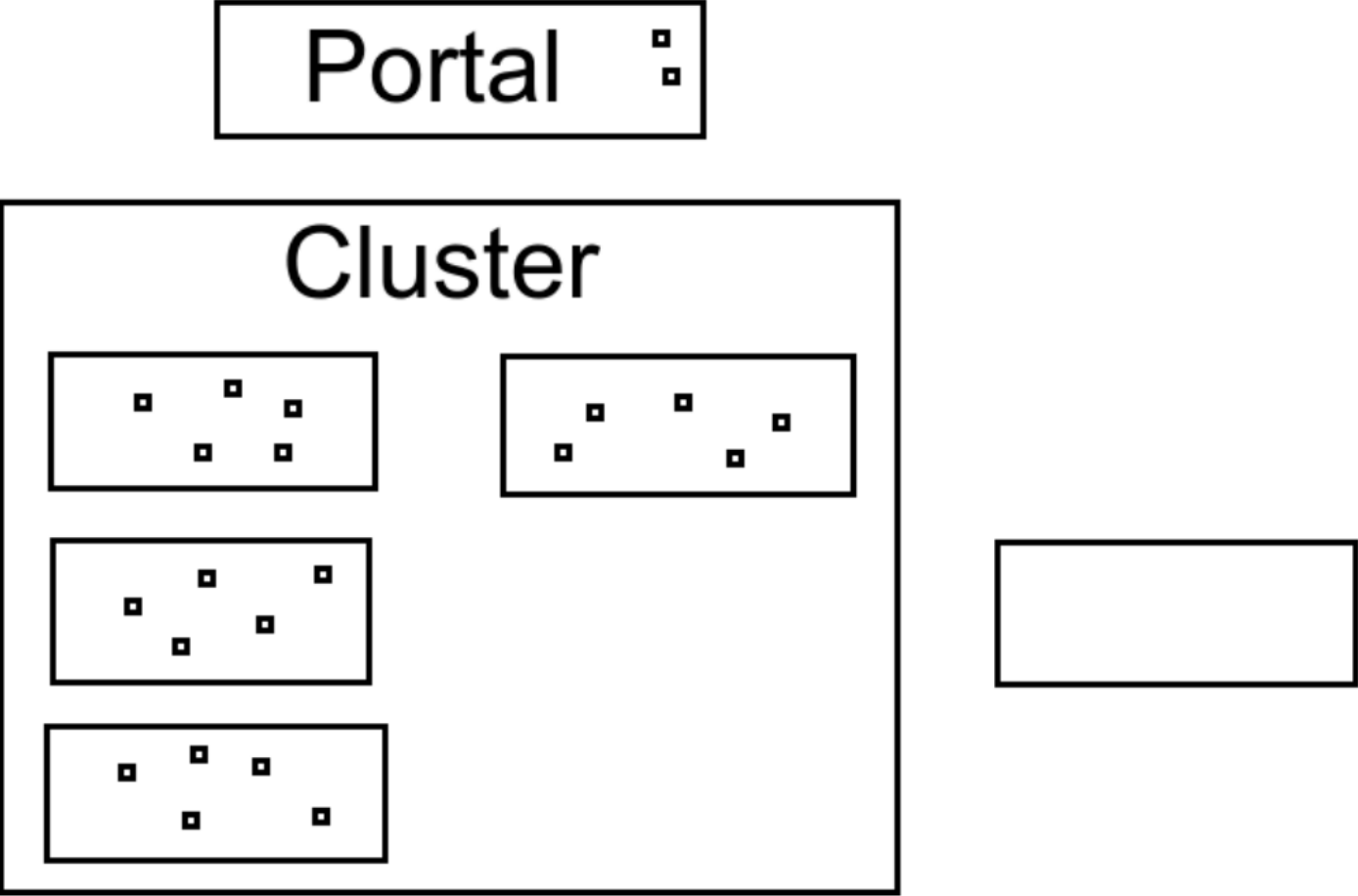
Då ser clusterhanteringen till att dela ut serviceinstanser på de olika serverarna.

Om vi köper in en ny server kan vi lägga till den i clusterhanteringen, och så kommer clustret kunna fördela ut serviceinstanser på den.



# Driftsarkitektur

Det betyder även att vi kan "dräna" en server för service-instanser, ta den ur drift, uppgradera den...

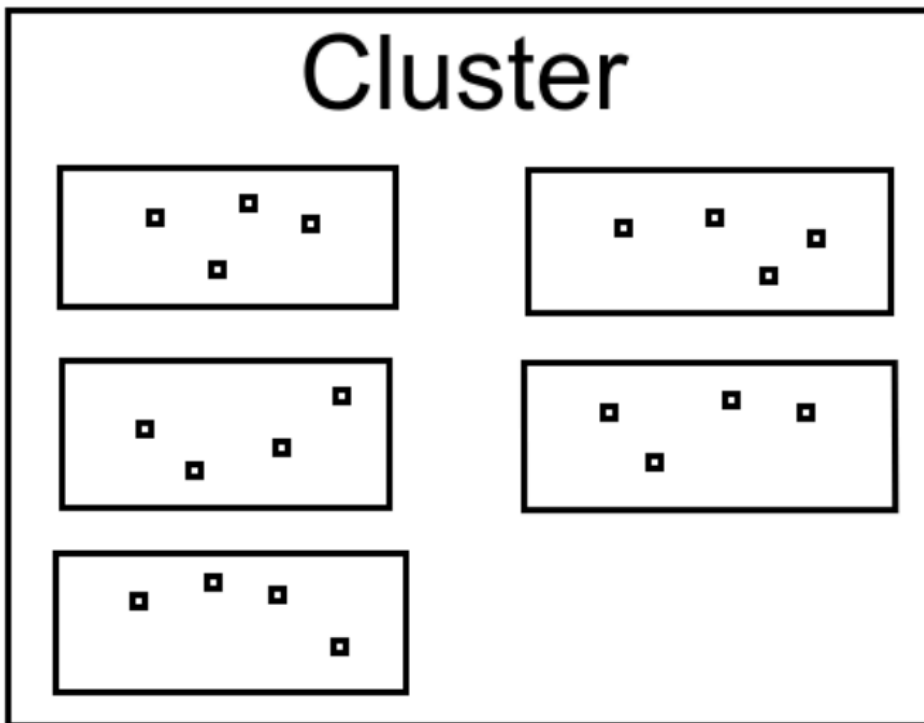




# Driftsarkitektur

...och driftsätta den igen utan störa driften av hela systemet.

Detta möjliggörs i stor grad också av att vi jobbar med REST-api. Det betyder att det inte finns någon "state" som kan förstöras när en service-instans ska flyttas. För att "flytta" en instans innebär i själva verket att den startas upp på en ny maskin och stängas ner på den gamla.

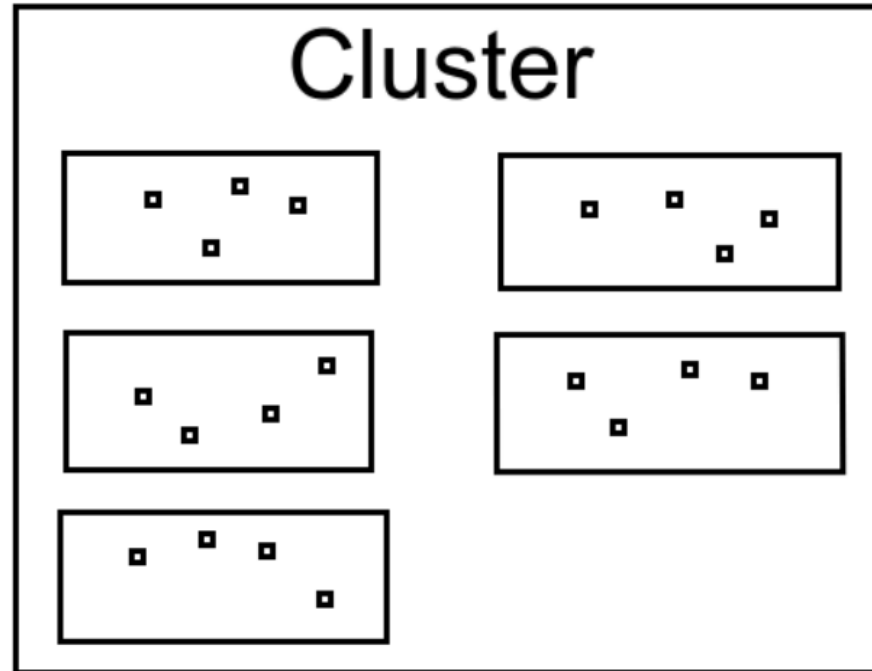


# Driftsarkitektur

Ett cluster innebär dock också en viss utmaning med att få till lagringen på ett bra sätt, då denna ska kunna tillgås från alla cluster-servrarna.

Vissa services som ska skriva till den får då troligen bara finnas som en enda instans i clustret.

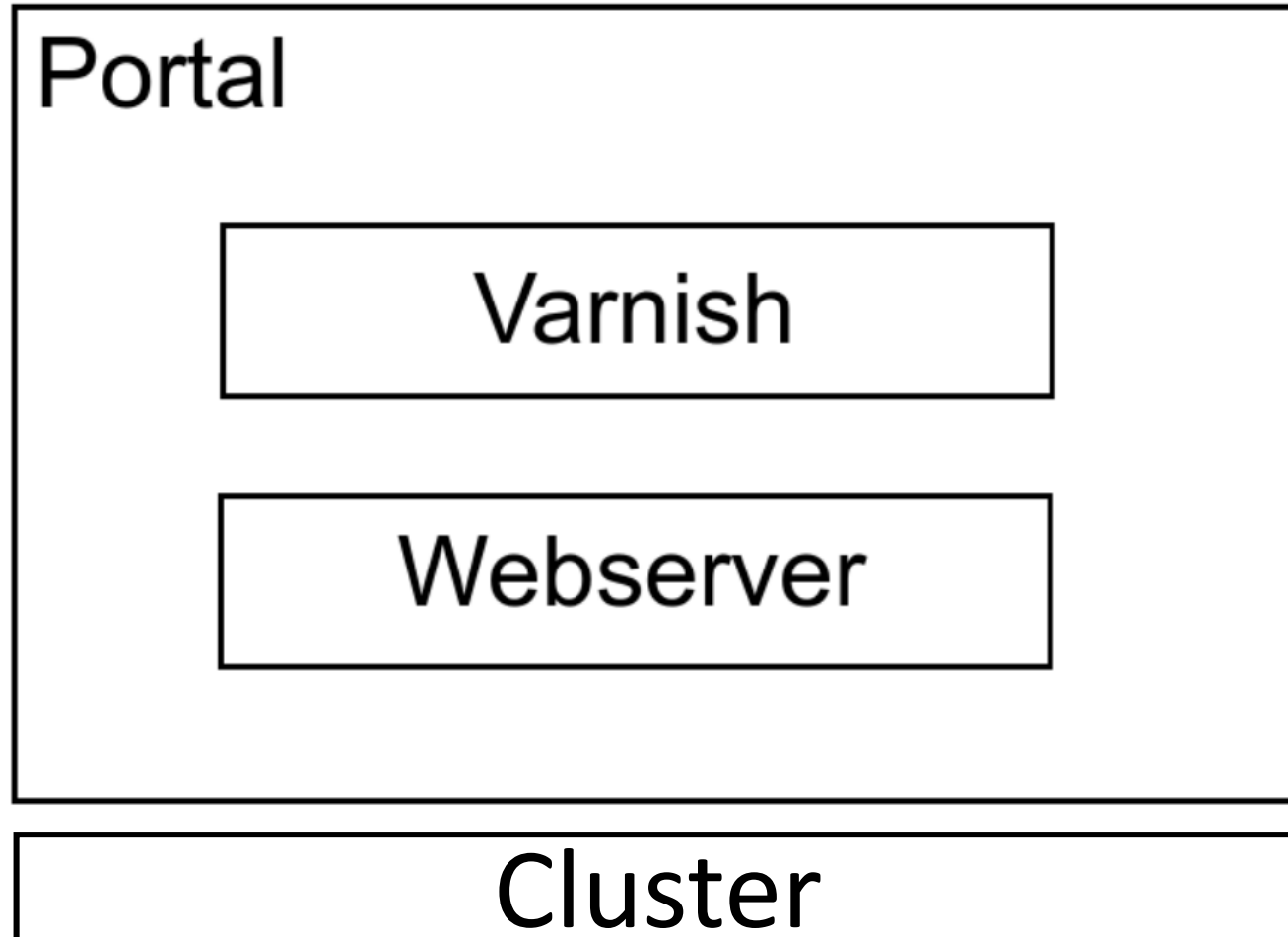
Portal 



Gemensam lagring?

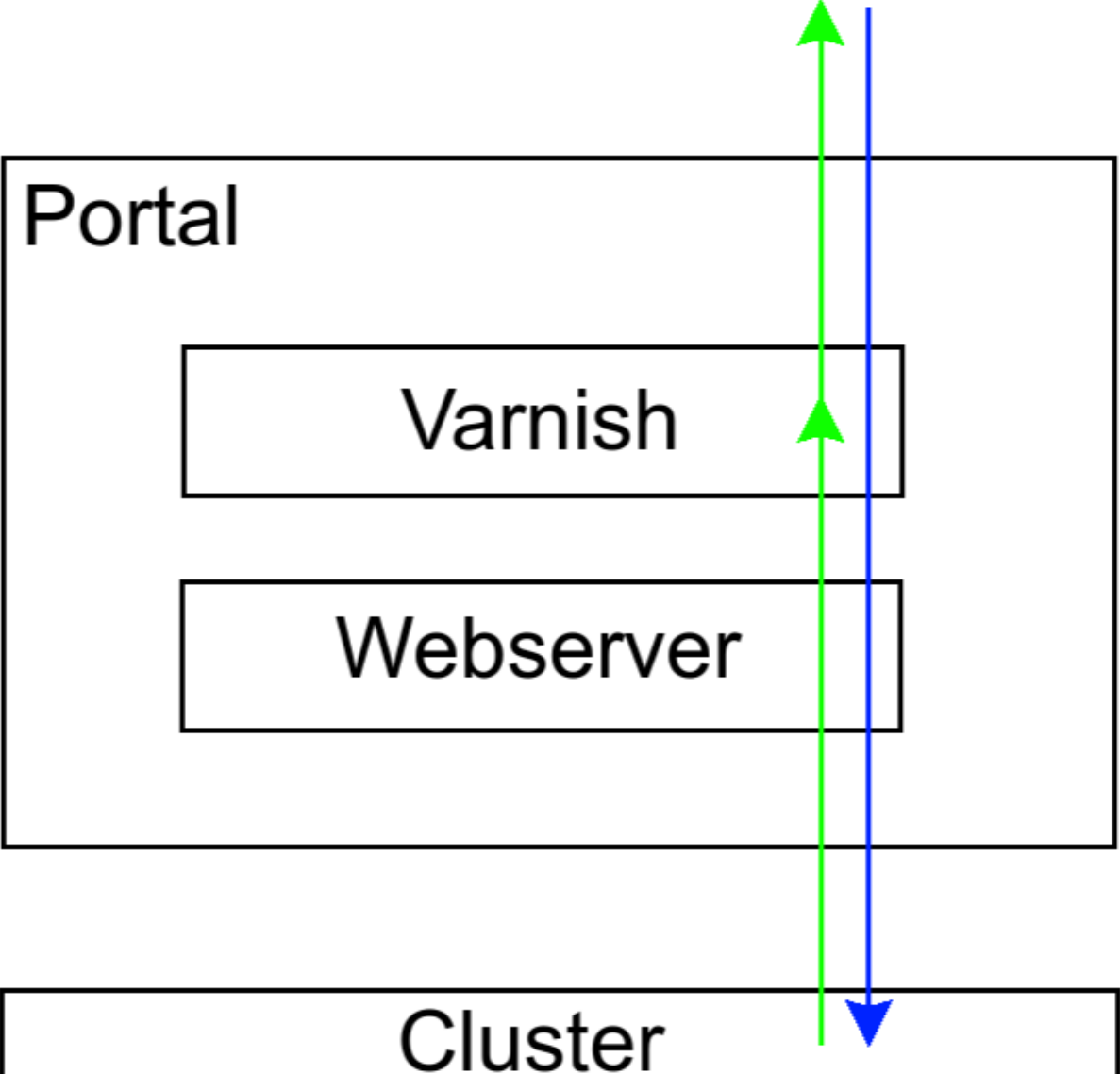
# Driftsarkitektur

Varnish är en http-cache, och kallas även ofta för en "webb-accelerator".  
Planen nu är att det ska sitta en sådan framför webservern i det nya systemet.



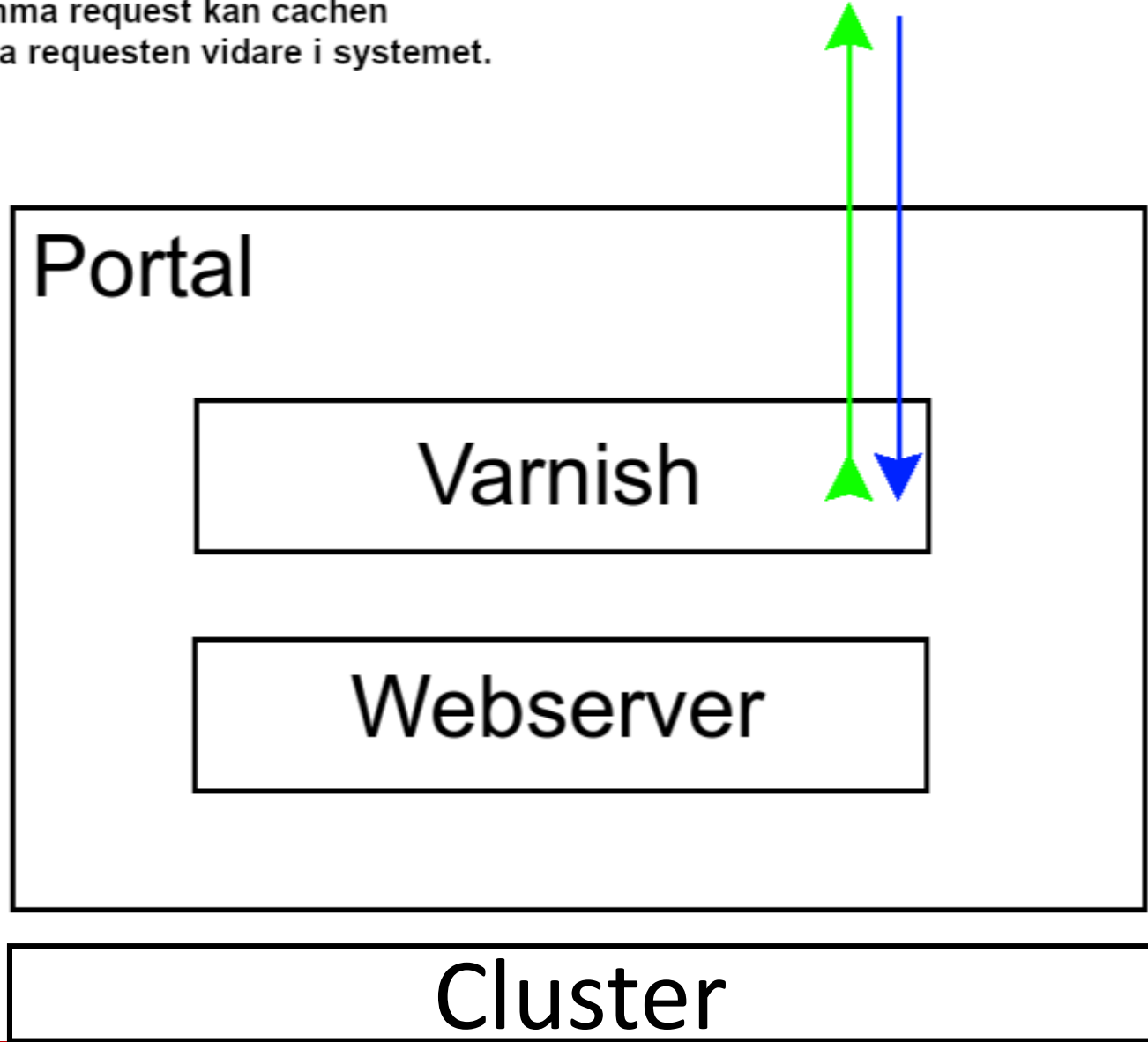
# Driftsarkitektur

När en request kommer in i systemet, så sparas responsen i cachen.



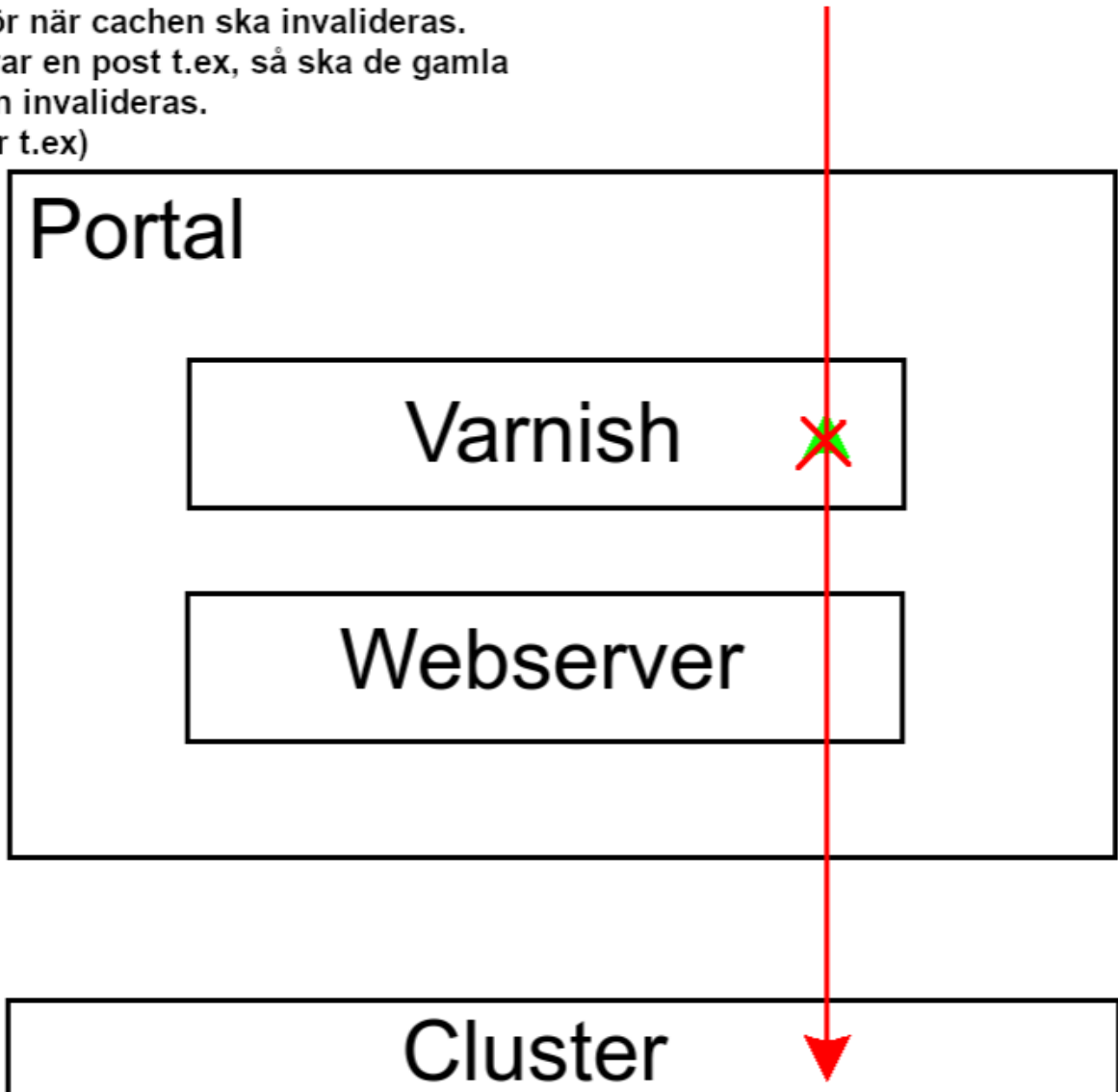
# Driftsarkitektur

Nästa gång någon skickar samma request kan cachen servera svaret igen utan skicka requesten vidare i systemet.



# Driftsarkitektur

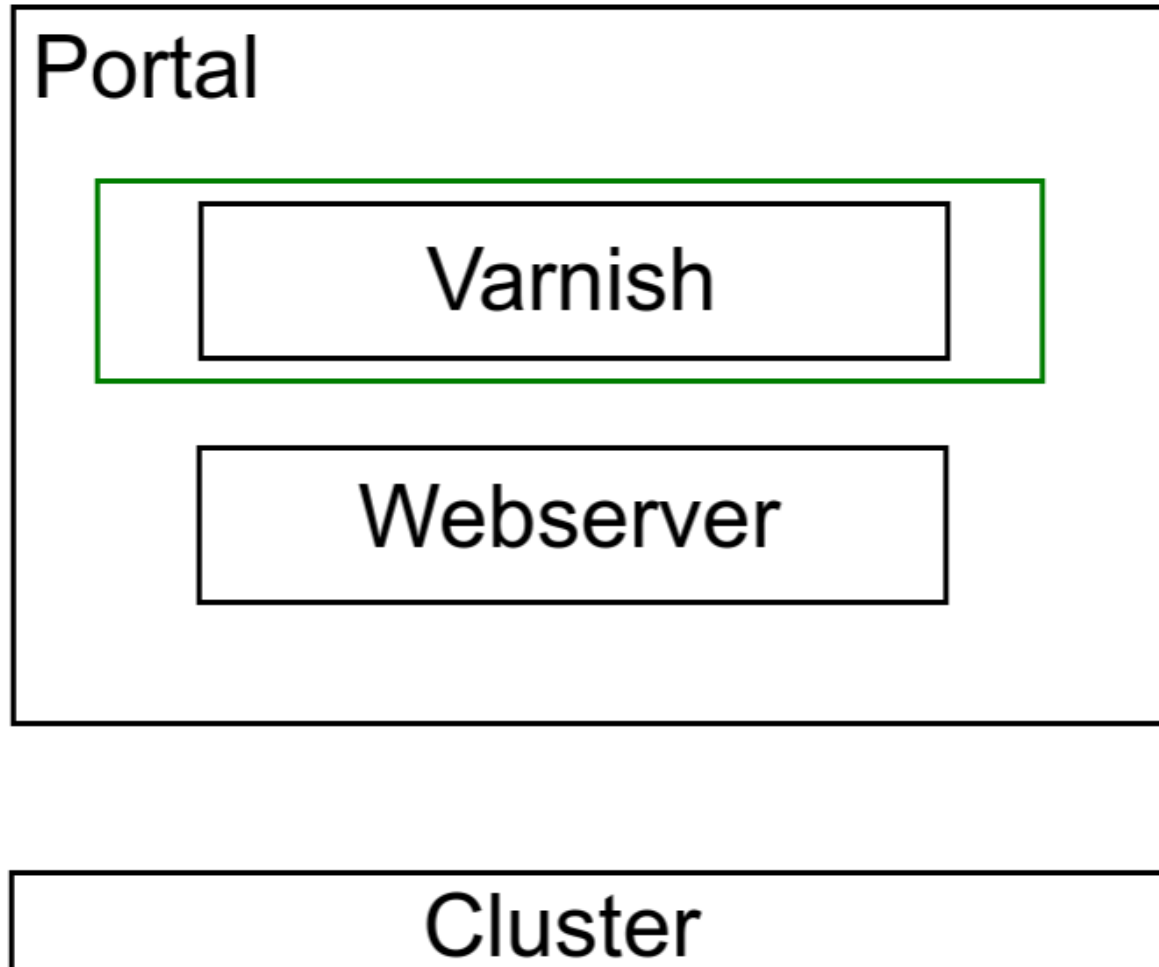
Men för att säkra att svaren vi skickar ut alltid är up-to-date så måste vi sätta upp regler för när cachen ska invalideras. När en administratör uppdaterar en post t.ex, så ska de gamla svaren för den posten i cachen invalideras. I många fall (svar på sökfrågor t.ex) där regelbaserad invalidering inte är möjligt måste vi troligen sätta upp lämpliga timeouts.



# Driftsarkitektur

Anledningen till att introducera detta är att en cache kan svara mycket snabbt. Särskild Varnish är känt i industrin för att vara optimerat till att kunna reagera på mikrosekunder...

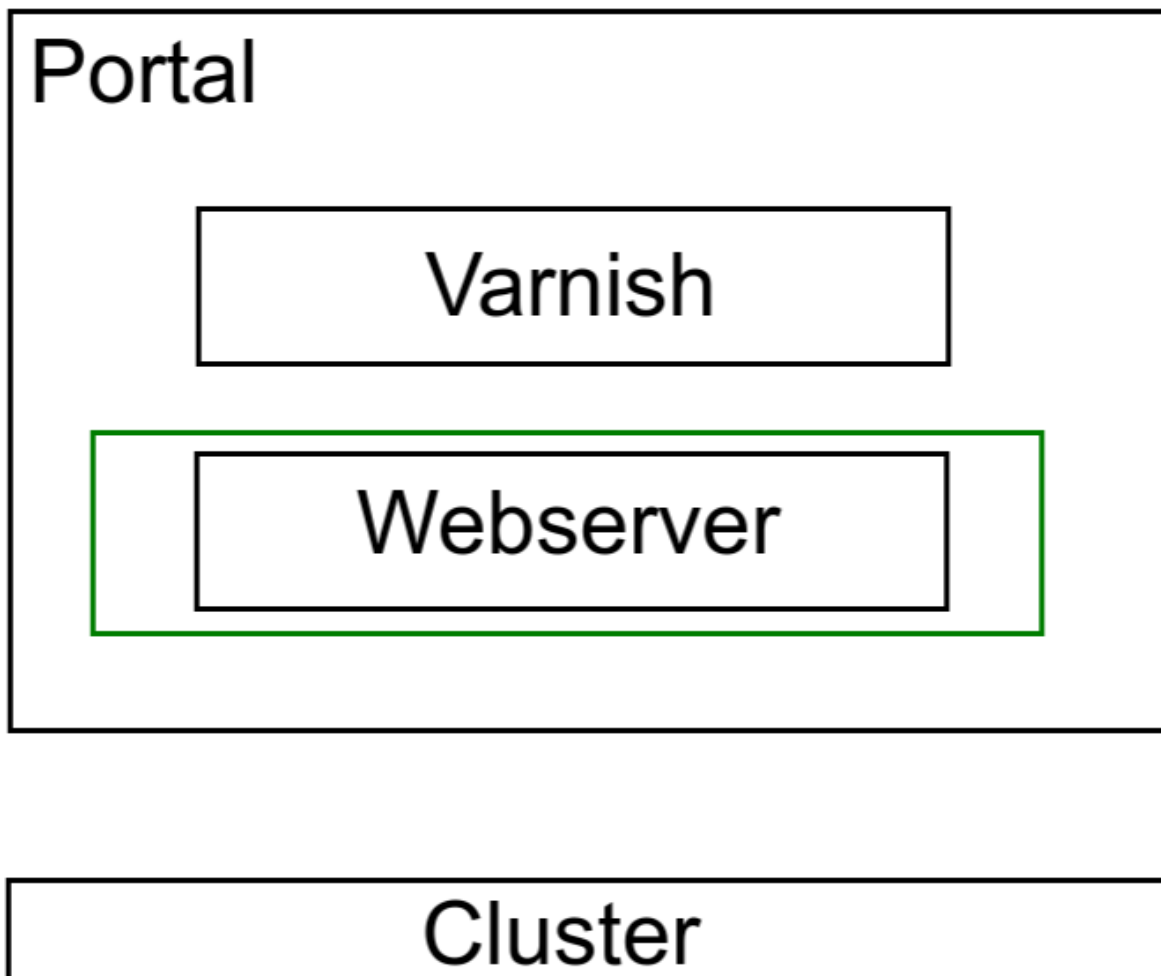
## Mikrosekunder



# Driftsarkitektur

...där de flesta webbservrar tar millisekunder bara för att servera statisk innehåll.

## Millisekunder





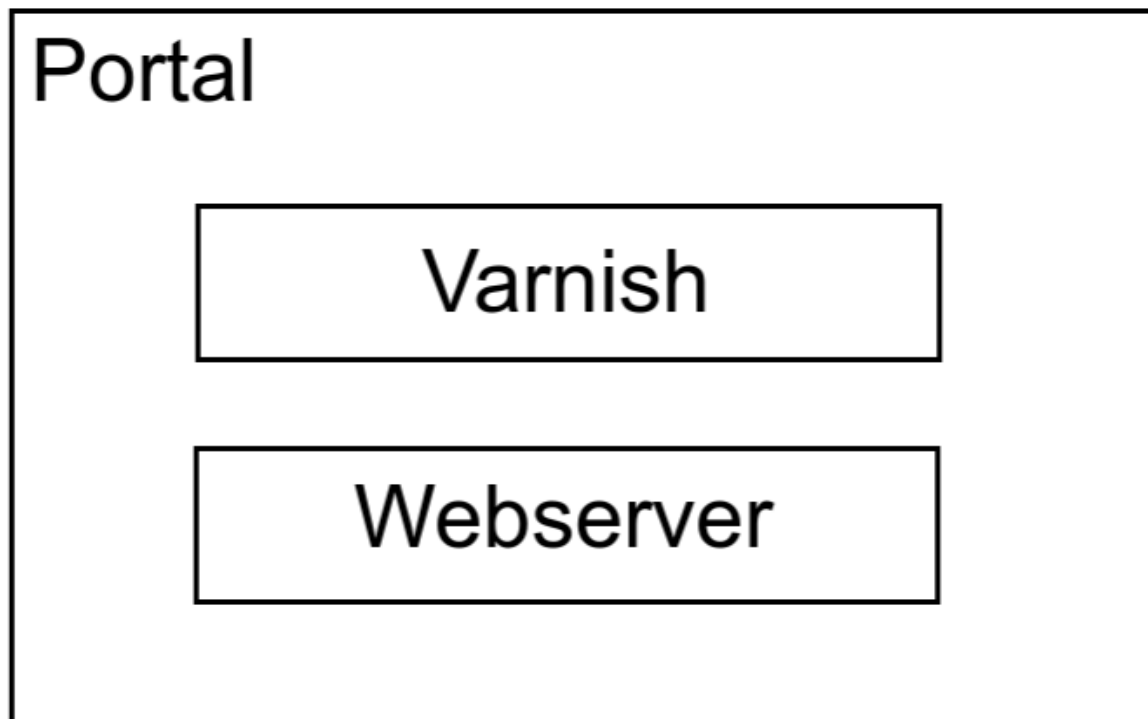
# Driftsarkitektur

Och beroende på requesten kan ett svar från en service nere i applikationsstacken potentiellt komma att räknas i sekunder.

Och det är inget fel på applikationen, det är bara realiteten i att logiken som ska svara på requesten blir mer komplicerat.

I slutändan, ju fler svar vi kan "outsourca" till Varnish, ju snabbare blir den genomsnittliga svarstiden och ju mindre last ska vi hantera i applikationsstacken i clustret.

## Sekund(er)?



# Önskade praktiska effekter

- Lägre last på services
- Snabbare svar
- Möjlighet för rullande uppdateringar
- Mindre behov av ”underhållsfönster”
- Mera redundans t.ex. vid hårdvarufel

# Pågående arbete

- Uppsättning av experimentell miljö
- Uppsättning av applikationsstack i clusterhanteringen
  
- Frågor?