

MONA Referensgruppsmöte 2024-06-12

Närvarande:

Martin Brandhagen, Vetenskapsrådet
Susanne Dahllöf, MONA-supporten, SCB, ordförande
Olof Elfwering, MONA-supporten, SCB
Marie Eriksson Osborne; Vetenskapsrådet, gäst
Lisa Freed, MONA-supporten, SCB
Jonas Holm, MONA-supporten, SCB
Oscar Erixson, Uppsala universitet
Lucas Fischer Madsen, Göteborgs universitet
Martin Hällsten, Stockholms universitet
Marie Kåberg, uppdragsansvarig, SCB
Jenny Lindberg, enhetschef Uppdrag och extern service SCB
Helena Nilsson, Jönköpings universitet
Roujman Shahbazian, Stockholms universitet
Oscar Stålnacke, Umeå universitet
Johan Holmberg, Umeå universitet

Frånvarande:

Susanna Bylin, Vetenskapsrådet
Torun Österberg, Göteborgs universitet

Inledning

Susanne hälsade alla välkomna. Mötet inleddes med en presentationsrunda. Oscar Stålnacke och Johan Holmberg är nya i referensgruppen. Susanne söker representanter från Uppsala respektive Lunds universitet.

1. På gång på SCB

Jenny Lindberg, enhetschef, SCB, presenterade nedanstående fyra bilder:



Kombinerad insamling i AKU

- Kombinerad insamling har införts i Arbetskraftsundersökningen (AKU).
- Första insamlingsmånad med kombinerad insamling är april 2024.
- Insamlingen genomförs med webbenkät eller intervju.
- Webbenkät erbjuds inledningsvis i några utvald grupper på arbetsmarknaden.



Fyra prioriterade målbilder



Arbetsmarknadsstatistik baseras mer på registerbaserade data som varje månad inkommer till SCB och som sammanställs i Befolkningens arbetsmarknadsstatus, BAS. SCB har tidigare fokuserat på sin kommunikation med allmänheten för att göra SCB mer känd, men nu ligger fokus på expertanvändare. Alltmer digitalisering i samhället innebär också möjligheter till nya datakällor för statistik. Machine learning och AI undersöks om och hur det kan användas i verksamheten.

Finansieringsbehov inom tre områden

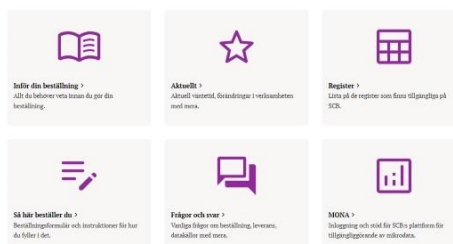
1. EU-initiativ som innebär tvingande krav och anpassningar av SCB:s verksamhet.
2. Långsiktiga satsningar för att hantera och använda data som en strategisk resurs
3. Åtgärder för att säkerställa kunskapen kring integration, kompetens och arbetsmarknad.



Mikrodata

Här finns information om hur du beställer mikrodata ur SCB:s register för forsknings- och statistikändamål.

Uppgifter om enskilda personer och företag som finns i SCB:s register är skyddade av allmän sekretess. Trots det finns undantag som tillåter SCB att lämna ut uppgifter till forsknings- eller statistikändamål. Utlämnande av uppgifter kan också ske till ett specifikt erbjuds- eller samtyckesregister. Utlämnande av uppgifter som personnummer och organisationsnummer till läkare. Inom SCB kan lämna ut uppgifterna givt i en prövning för utlämnande, om så kallad sekretessprövning.



Mikrodata-sidorna har uppdaterats för att förenkla beställningsprocessen och för att beställare enklare ska kunna hitta information. [Mikrodata \(scb.se\)](https://www.scb.se/mikrodata)

En fråga ställdes på mötet: Delar av en beställning till SCB går ibland till andra enheter än mikrodatasektionen, där väntetiden kan vara längre (gäller främst fastighetsdata), varför är det så?

- Jenny svarade: Mikrodatasektionen hanterar merparten av data, det är nog (snart) endast fastighetsdata som hanteras av andra. Bedömningen är att antalet beställningar som inkluderar dessa data är så pass få att det inte är värt att bygga upp kompetens vid sektionen. I takt med att efterfrågan ökar kanske kompetensen hos mikrodatasektionen bör utökas för att förkorta väntetider vid beställning av dessa data.

2. Offentliga mikrodata

Susanne och Olof sammanfattade den utredning som genomförts av SCB på uppdrag av Vetenskapsrådet om offentliga mikrodata.

Offentliga mikrodata (Publicuse files)

Utredningen är klar.

Förslag på metod:

1. Urval av variabler
2. Omkodningar grupperingar och klassindelningar
3. Exkludering av individer med unika kombinationer av variabelvärden
4. Urvalsdragning



Omfattande informationsförlust.

Det föreslagna framställningssättet på bilden ovan innebär inbyggd bias då extremvärden och ovanliga kombinationer av värden inte kan komma med i urvalet.

SCB rekommenderar i nuläget inte syntetisering då det är oklart hur röjanderisken kan utvärderas.

Offentliga mikrodata (Publicuse files)

Inkomna frågor:

- Får forskare med tillgång till registerdata själva framställa offentliga mikrodata?
- Är pseudonymisering= anonymisering? General Court of the European Union T-557/20
- Behöver man lära sig att "tvätta" och hämta data från flera register?

Ska dataset framställas enligt förslagen? Justeringar?



Referensgruppen hade före mötet fått tillgång till rapporten och en fråga från Vetenskapsrådet: Ska dataset framställas enligt förslagen? Med anledning av den frågan fick SCB de motfrågor som framgår på bilden ovan.

- På frågan om pseudonymisering är detsamma som anonymisering svarade Jenny att när det gäller mikrodata från SCB så är pseudonymisering nästan aldrig samma sak som anonymisering. Även om inga direkt identifierande uppgifter ingår, så ingår ofta väldigt många uppgifter som gör

bakvägsidentifiering möjlig. När material endast innehåller någon enstaka variabel kan materialet dock ses som anonymiserat efter pseudonymisering.

- Får forskare framställa offentliga mikrodata?
 - SCB svarade att universiteten får undersöka om det finns stöd för det.
- Inget svar inkom på frågan om det finns behov av att träna på att tvätta data och själv göra sambearbetningar.

På mötet ställdes också följande frågor:

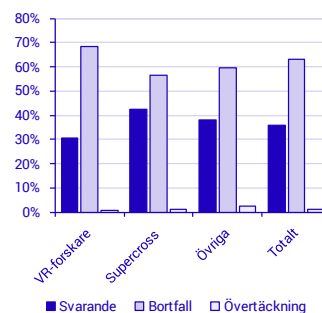
- Vad kommer kostnaden vara?
 - SCB har svårt att bedöma kostnaden då det är mycket som påverkar, exempelvis om förslagna variabler ska vara med eller om de ska bytas ut mot andra.
- Ska man kunna följa en individ?
 - Ja, i förslagen avses mikrodata där man kan följa samma personer över tid.

3. Användarundersökningen

Susanne visade några resultat från användarundersökningen där 528 användare av MONA svarade på en webbenkät i januari. Svaren delas i tre olika grupper av användare, VR-forskare, vilka är forskare vars användning av MONA finansieras av Vetenskapsrådet, användare som fått data från SCB för statistikändamål och som nästan enbart använder programvaran Supercross för enkla tabelluttag och gruppen övriga som främst har fått data från SCB för statistikändamål men analyserar data på liknande sätt som forskare gör.

Användarundersökningen 2024

	VR-forskare	Supercross-användare	Övriga	Totalt
Svarande	219	218	91	528
Bortfall	488	292	143	923
Övertäckning	5	5	6	16
Totalt	712	515	240	1467



Användarundersökningen 2024

Antal inloggade dagar aprildec 2023, andel per användargrupp

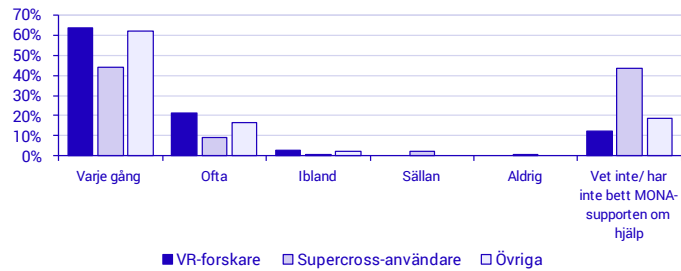
Antal inloggade dagar			
	VR-forskare	Supercross	Övriga
<10 dagar	27%	58%	25%
11-50 dagar	28%	26%	40%
51-149 dagar	32%	8%	24%
150+ dagar	13%	9%	12%
Totalt	100%	100%	100%

SCB

Det är vanligare att en VR-forskare har använt MONA fler dagar än en användare i de andra grupperna.

Användarundersökningen 2024

Hur ofta anser du att MONA-supporten hjälper dig tillräckligt snabbt?

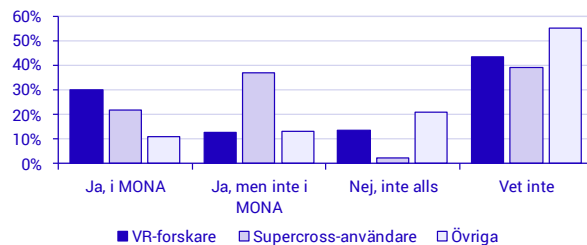


SCB

De svarande är i hög grad nöjda med MONA-supportens stöd.

Användarundersökningen 2024

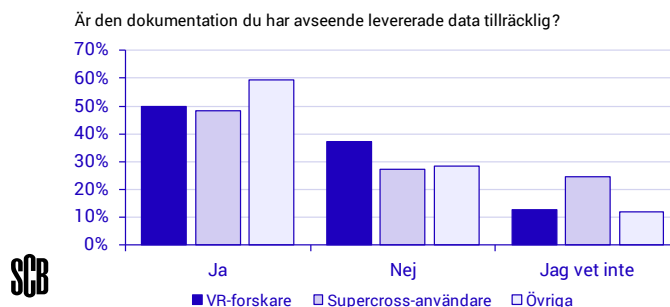
Har du planer på att i framtiden använda AI eller maskininläring när du analyserar data?



SCB

Ca 30 % av de forskare som svarat har planer på att använda AI/ML i MONA i framtiden.

Användarundersökningen 2024



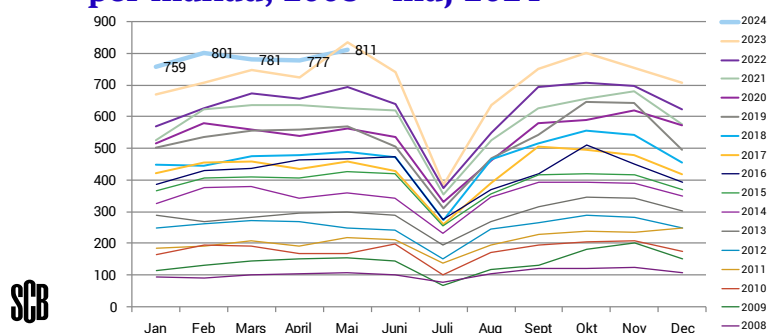
Ca 50% av forskare som svarat är missnöjda med dokumentationen av levererade mikrodata, framför allt saknar man variabelbeskrivningar.

Jonas demonstrerade den uppdaterade varianten av MONA-Metadatas som kommer att publiceras på mikrometadatas.scb.se. Verktöget blir ett komplement till nuvarande dokumentationsverktyg där man hierarkiskt söker efter information.

[Dokumentation av mikrodata](#). På mötet visades uppskattning av att det förekommer markering för "känslig personuppgift". SCB kunde dock meddela att den informationen tyvärr inte är heltäckande.

4. Information från SCB om MONA

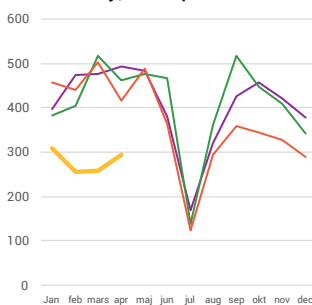
Antal unika inloggade användare, per månad, 2008 – maj 2024



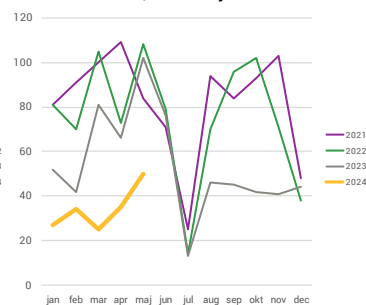
Susanne visade diagram över den ökande användningen av MONA. Nu finns fler än 2 100 användarkonton, Ca 1475 användarkonton har använts under 2023. Antalet unika användare per månad har överstigit 800.

Aktuellt 2024

Besvarade mejl, 2021- april 2024



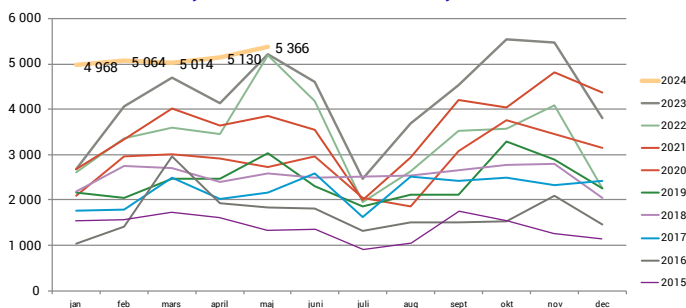
Besvarade samtal, 2021- maj 2024



SCB

Det nya inloggningssättet som infördes under våren 2023 har inneburit att färre användare behöver kontakta MONA-supporten för hjälp.

Antal batchjobb, 2015 – maj 2024

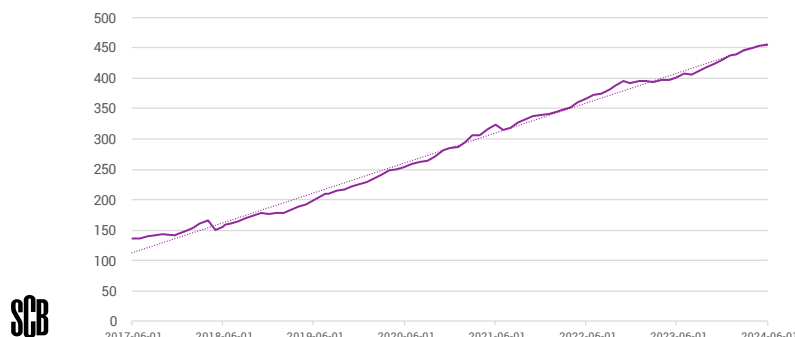


SCB

Fortsatt hög användning av Batch-miljön under 2024 framgår av diagrammet som visades. Flera i referensgruppen vittnade om hög belastning i batch-miljön.

Susanne berättade om ett test där man med manuella rutiner kan få projektdata överförda till BIANCA-klustret på Uppmax vid Uppsala universitet, där det finns betydligt bättre prestanda. Om någon i referensgruppen vill testa, kan man anmäla det. Referensgruppen var mycket positiv till möjligheten att använda resurser utanför MONA.

Lagring i TB, juni 2017- maj 2024



Diagrammet över lagring i projektmappar visar en fortsatt ökad lagring. Det är viktigt att var och en löpande raderar data man inte behöver, det är inte bara en fråga om kostnader utan också att mängden data påverkar återställningstiden vid ett haveri.

Aktuellt 2024

- **Införa rollen projektadministratör**
 - En eller flera, per projekt
 - Registrera nya och ta bort befintliga användares behörighet från projekt
 - Ta hand om material från tidigare användares MyDocuments
 - Registrera slutdatum på projekt och användare
 - Följa lagringstillväxten och annan användning
- **Utveckla en nyMONAWeb**
 - Tillgänglighetskrav
 - SCB:s grafiska profil
 - På sv och eng
 - Mobilanpassad

MONA-supporten är äntligen igång med arbetet med den nya MONAWebben, förutom en allmän uppdatering av design och tillgänglighet (t.ex. anpassning för mobil och personer med funktionsvariationer) tillkommer ny funktionalitet för användare, handläggare, MONA-supporten samt för den nya rollen projektadministratörer. Olof visade och berättade.

5. Synpunkter från referensgruppen

- Roujman Shahbazian, SU
 - Jättebra jobb. Metadata-sidan kommer bli jätteuppskattad

- Martin Hällsten, SU
 - Ingen input. Noterar att det är högt tryck i batch-miljön
- Oscar Erixson, UU
 - Ingen speciellt att rapportera. Föreslår att frågan om replikering för tidskrifter tas upp på nästa möte.
- Helena Nilsson, JU
 - Inga synpunkter. Det som presenterats idag ser jättebra ut
- Lucas Fischer Madsen, GU
 - Uppskattar införandet av Projektadministratörer
- Oscar Stålnacke, UmU
 - Instämmer med föregående, mycket spännande på gång. Finns det tankar i dagsläget att utöka kapaciteten?
 - I dagsläget kommer kapaciteten inte att utökas. SCB ligger efter i tidsplaneringen gällande nytt datacenter. När det är på plats kommer det finnas lite mer resurser.
 - Det är tung belastning, nu får man räkna med att det tar lite tid innan man kan köra något i batch-miljön. Främst den servern för långa jobb som kan bli trång.
 - SCB noterade önskemålet
- Johan Holmberg, UmU
 - Mycket spännande och bra saker på gång. Låter bra att kunna överföra till Uppmax.
 - Skulle gärna se Julia och/eller Matlab i MONA.
 - MONA:s batchsystem är användarvänligt i jämförelse med andra miljöer.
- Martin Brandhagen, VR
 - Intressant att se vad som händer.

6. Metadata på registerforskning.se - Marie Eriksson Osborne

Marie berättade om Vetenskapsrådets arbete med utveckling av en metadatakatalog som innehåller beskrivning av många, och inte bara SCB:s register. Marie behöver input för det fortsatta utvecklingsarbetet och söker kontaktpersoner. Vilka register och vilka data är intressanta att visa i verktyget? Vem kan Marie vända sig till med sådana frågor framöver? Man får gärna anmäla sitt intresse till Marie.

Marie visar en demo av RUT-verktyget. [Metadatakatalog - En tjänst från Vetenskapsrådet \(registerforskning.se\)](#).

VR tittar på anpassningar och översättningar till EHDS-initiativet.

På mötet ställdes frågan om hur lång är kön idag för att beställa data och om sådan information kan visas i realtid?

- Jenny svarade att på [Aktuellt \(scb.se\)](https://www.scb.se/aktuellt) finns information om aktuell kö på SCB.

7. Övriga frågor

- Under hösten kommer SQL-servern MQ02 att ersättas. Alla databaser kommer att migreras, men servernamn behöver uppdateras. Alla användare kommer att informeras.
- SCB kommer att utreda ny lösning som ersätter terminalserver
- Det finns ingen konkret tidsplan för det SCB:s nya datacenter ännu. Planen är att bytet ska ske utan att användare påverkas
- Ny hantering av R-paket.
- SPSS ser ut att kunna erbjudas även fortsättningsvis
- Varning för Windows recall. Ny funktion på Copilot-Plus-datorer som heter Recall som hjälper dig att komma ihåg vad du höll på med sist. Allt du gör spelas in och lagras lokalt. [Windows Recall – en mardröm för integriteten? | Computer Sweden](#)

Nästa möte

Nästa möte till hösten. Susanne skickar kallelse.

Förbered gärna för erfarenhetsutbyte om de krav vetenskapliga tidskrifter ställer på tillgång till kod och/eller data före publicering. Vad gäller? Erfarenheter och lösningar.

Om det finns utrymme på nästa möte kan en handläggare som tar emot beställningar av mikrodata bjudas in.

Susanne tackade alla för deras medverkan.