

Excellensutvärdering genom bibliometrisk analys

Forskning finansierad av Mistra, 2006-2023

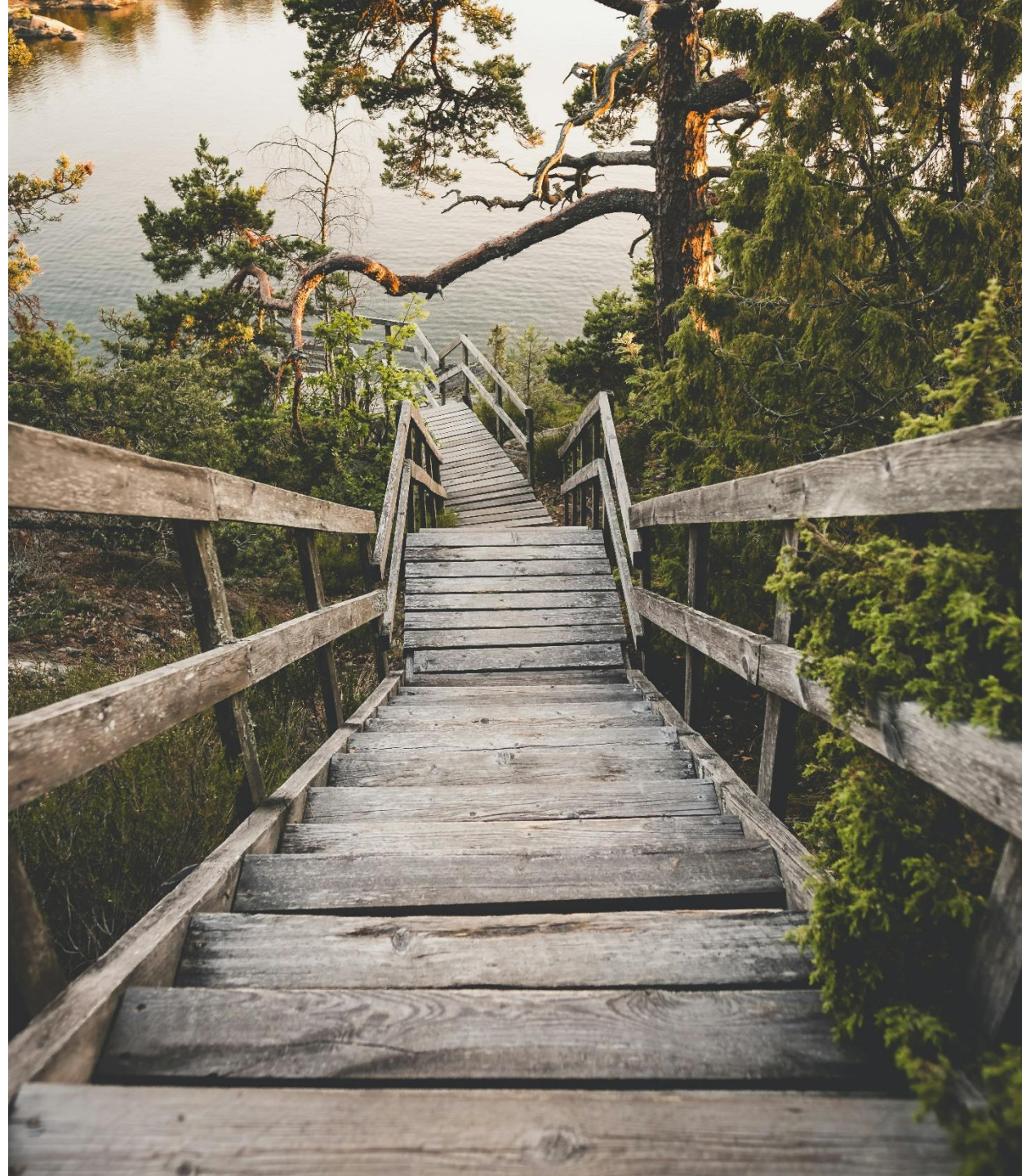
2024-06-03

Introduktion

Denna delutvärdering genomfördes av ADC från november 2023 till april 2024 och syftar till att mäta och analysera den vetenskapliga kvaliteten i Mistras forskningssatsningar genom bibliometrisk analys. Genom analysen identifieras bibliometriska resultat kopplat till styrka, specialisering och kvalitet från den forskning som bedrivs inom Mistras program. Resultaten kopplar till tilltänkta effekter från effektlogiken genom att identifiera vetenskaplig excellens, internationell samverkan samt prioriterade forskningsområden.

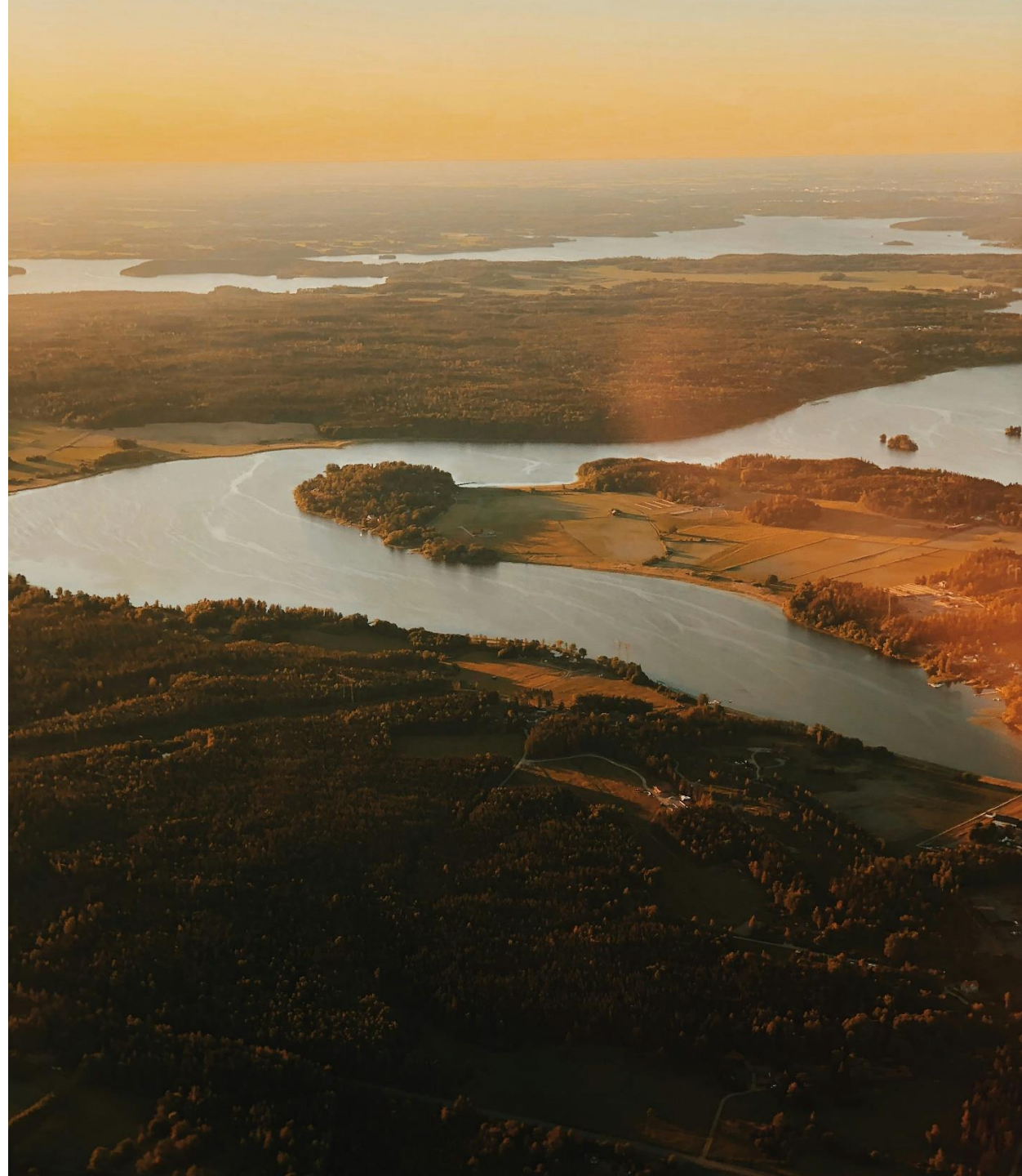
Arbetet genomförs på bas av den bibliometriska data som samlats in från de 29 forskningsprogram och tre centrubildningar som finansierats och påbörjats under perioden 2006-2020. ADC använder sig av publikationsdatabasen Scopus för att analysera publikationerna.

Den bibliometriska analysen utgår från etablerade metoder, och i möjligaste mån förhåller vi oss till de definitioner som Vetenskapsrådet använder i sina kartläggningar av det svenska forskningslandskapet.



Innehållsförteckning

1	Underlag.....	5
2	Bakgrund.....	8
3	Forskningsfält.....	13
4	Sampublicering.....	20
5	Excellens.....	24
	Appendix.....	32



Publikationer från Mistras satsningar håller hög vetenskaplig kvalitet, även om den relativa genomslagskraften har minskat över tid

ADC har på uppdrag av Mistra undersökt huruvida Mistras finansiering av forskningsprogram och centrumbildningar har resulterat i excellent forskning av högsta internationella klass. Arbetet baseras på den bibliometriska data som samlats in från de 29 forskningsprogram och tre centrumbildningar som finansierats och påbörjats under perioden 2006-2020. Publikationsdatabasen Scopus ligger till grund för analysen som enbart görs på tidskriftsartiklar och konferenspaper.

Excellens

Forskningen i form av publikationer som producerats inom ramen för Mistras satsningar håller hög vetenskaplig kvalitet. Mistrafinansierade publikationer har en högre vetenskaplig genomslagskraft än andra jämförbara publikationer under hela den studerade perioden. Den relativa genomslagskraften har dock minskat över tid.

Publikationerna blir i hög utsträckning publicerade i topprankade tidskrifter samtidigt som de även tillhör de mest högciterade inom sina forskningsfält. Stockholm Resilience Centre (SRC) utmärker sig i positiv bemärkelse gällande alla excellensmått, men även de andra satsningarna håller hög vetenskaplig kvalitet utifrån ett bibliometriskt perspektiv.

Låg grad av sampublicering med näringslivet och varierande grad av internationell sampublicering

De studerade publikationerna uppvisar en relativt låg grad av sampublicering med näringslivsförfattare. Endast fyra procent av studerade publikationerna har sampublicerats med en författare affilierad med ett företag. Samtidigt har sampublicering med internationella författare skett i olika hög grad för de olika finansieringsformerna. SRC ligger något över graden av internationellt samarbete jämfört med ett svenskt benchmark medan andelen för forskningsprogrammen är klart under benchmark.

Sjunkande publikationsvolym per satsning över tid

Den totala publikationsvolymen för samtliga forskningsprogram och centrumbildningar uppgår till 2 844. 53 procent av dessa har producerats inom ramen för de 29 forskningsprogrammen, 35 procent inom ramen för SRC och de resterande 12 procenten inom de två olika centrumbildningarna Misum och Mistra Urban Futures.

Publikationsvolymen är relativt jämn under den studerade perioden trots att antal pågående satsningar har ökat över tid. Trenden visar således att det årligen publiceras färre artiklar och konferenspaper per satsning. Det går inte att förklara denna trend baserat på urval av program, programfas eller bristande datakvalitet.

Under den studerade perioden har Mistra infört två krav gällande publicering: 1) publicering ska ske med open access och 2) "Mistra" ska nämnas i publikationernas titel, abstract eller bland nyckelorden. Efter att dessa krav infördes har andelen open access-publicationer ökat till att vara i linje med ett svenskt benchmark och "Mistra" omnämns i betydligt högre utsträckning publikationerna. Kraven har således gett önskad effekt.

Hälften av publikationerna görs inom miljövetenskap

Hälften av publikationerna har gjorts i tidskrifter som klassificeras tillhöra *Environmental Sciences*. Andra vanliga områden att publicera inom är *Social Sciences*, *Agricultural and Biological Sciences* och *Business, Management and Accounting*. Det är framför allt de tre centrumbildningarna som bidrar till den stora publikationsvolymen inom *Social Sciences*. Drygt fyra procent av publikationerna är publicerades i tidskrifter som klassificeras som *Multidisciplinary*.



1. Underlag



Insamling av data

Urvalet av forskningssatsningarna till denna bibliometriska excellensutvärdering har gjorts i syfte att bygga vidare på Sweco EuroFutures AB:s utvärdering av Mistras genomförda program mellan 1996-2011. Totalt inkluderas 29 forskningsprogram och tre centrumbildningar i denna utvärdering (se lista över forskningsprogram och centrumbildningar på nästa sida).

De tre centrumbildningarna Misum, Mistra Urban Futures och Stockholm Resilience Centre (SRC) redovisas genomgående som två enheter där Misum och Mistra Urban Futures har lagts ihop medan SRC särredovisas. Detta val har gjorts i samråd med Mistra samt M30-kommittén mot bakgrund av SRC:s relativa storlek i förhållande till den samlade publikationsvolymen.

För att möjliggöra denna bibliometriska genomgång har publikationslistor från samtliga satsningar samlats in från respektive program/centrumansvarig. En del av satsningarna hade vid starten av denna delutvärdering redan rapporterat in publikationslistor i samband med sin slutrapportering. Dessa publikationslistor har skickats till program/centrumansvarig för bekräftelse innan de tagits med i analysunderlaget. ADC har tillsammans med Mistra kontaktat de program/centrumbildningar där publikationslistor saknats och bitt dessa inkomma med publikationslistor enligt en av ADC framtagen Excel-mall. Underlaget av publikationslistor är således självrapporterat av program/centrumansvarig.

Notera slutligen att SRC under den studerade perioden haft flertalet ytterligare finansiärer vid sidan av Mistra. Det har dock inte varit möjligt för SRC att markera exakt vilka publikationer Mistrafinansieringen har bidragit till. Således inkluderar underlaget från SRC samtliga publikationer som producerats från SRC under perioden 2007-2018.



Följande 29 forskningsprogram och tre centrubildningar ingår i den bibliometriska analysen

- **Forskningsprogram (genomförda):**

- E4 Mistra (2006-2014)
- PlantComMistra (2006-2013)
- SIRP (2006-2012)
- ENTWINED (2007-2013)
- MistraPharma (2008-2015)
- Mistra-SWECIA (2008-2015)
- Mistra Future Forests (2009-2016)
- Mistra Arctic Futures (2011-2013)
- Mistra Future Fashion (2011-2019)
- Mistra Biotech (2012-2019)
- Mistra Indigo (2012-2015)
- AquaAgri (2014-2017)
- Mistra Arctic Sustainable Development (2014-2018)
- Mistra Environmental Nanosafety (2014-2023)
- Mistra REES (2015-2023)
- The Seed Box (2015-2022)
- Mistra Financial Systems (2016-2020)

- **Forskningsprogram (pågående):**

- Mistra SAMS (2016-2024)
- Mistra STEPS (2016-2024)
- Mistra Carbon Exit (2017-2025)
- Mistra Geopolitics (2017-2024)
- Mistra Sustainable Consumption (2017-2025)
- Mistra TerraClean (2017-2025)
- Mistra Digital Forest (2018-2027)
- Mistra InfraMaint (2018-2028)
- Mistra Environmental Communication (2019-2027)
- Mistra Food Futures (2020-2024)
- Mistra SafeChem (2020-2024)
- Mistra Sport & Outdoors (2020-2024)

- **Centrumbildningar (genomförda):**

- Misum (2015-2025)
- Mistra Urban Futures (2010-2019)

- **Stockholm Resilience Centre (genomförd) (2007-2018)**



2. Bakgrund

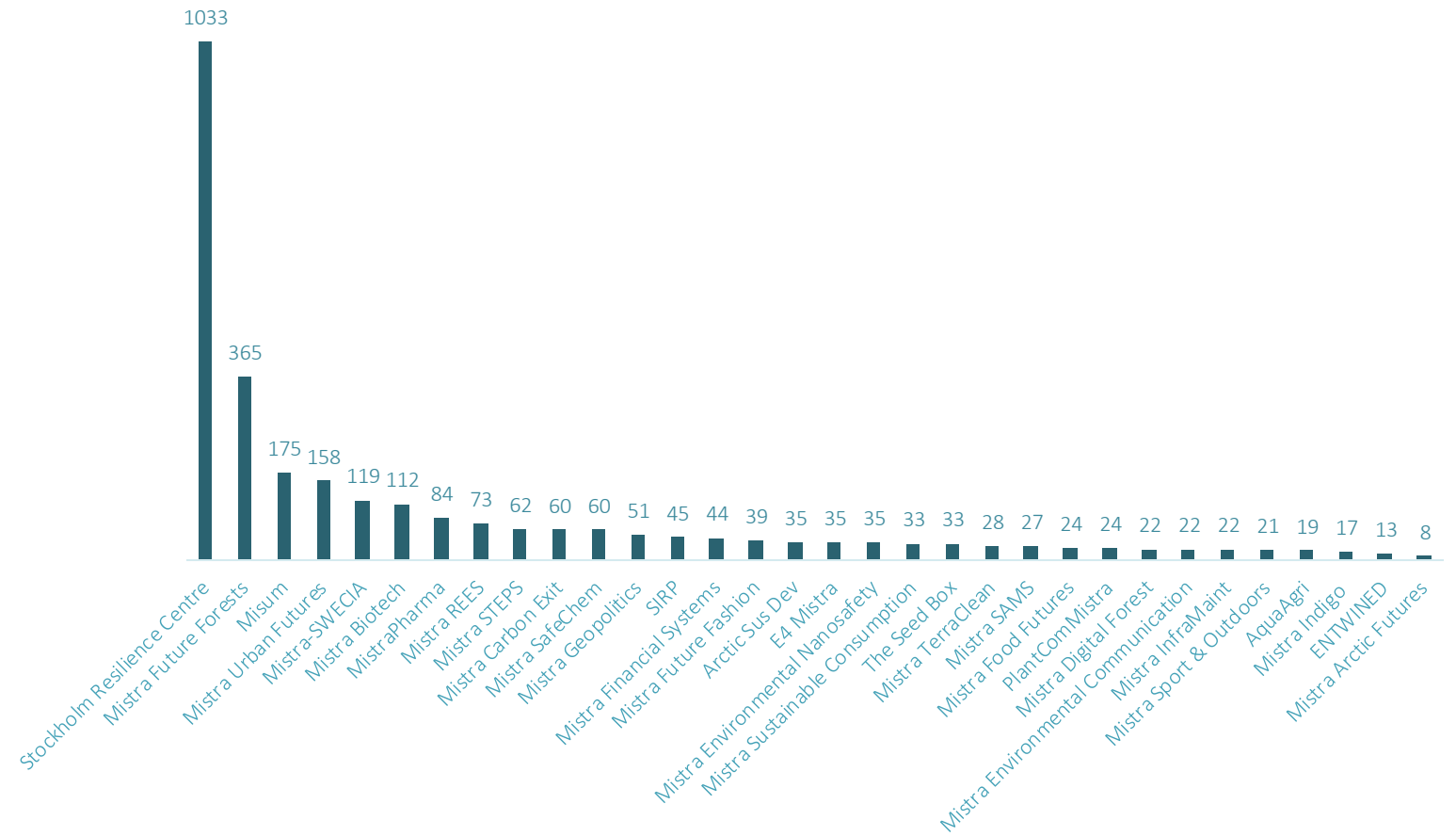


Stor variation sett till publikationsvolym

Det råder stor variation sett till antal publikationer som publicerats inom ramen för de olika satsningarna.

I de båda extremerna hittar vi Stockholm Resilience Centre och Mistra Arctic Futures. Totalt 1 033 publikationer har publicerats inom ramen för SRC, att jämföra med 8 inom ramen för Mistra Arctic Futures.

Näst flest publikationer har tillkommit inom ramen för Mistra Future Forests – totalt 365 publikationer. Andra satsningar med relativt hög publikationsvolym är Misum, Mistra Urban Futures och Mistra-SWECIA.



Publikationsvolymens utveckling över tid

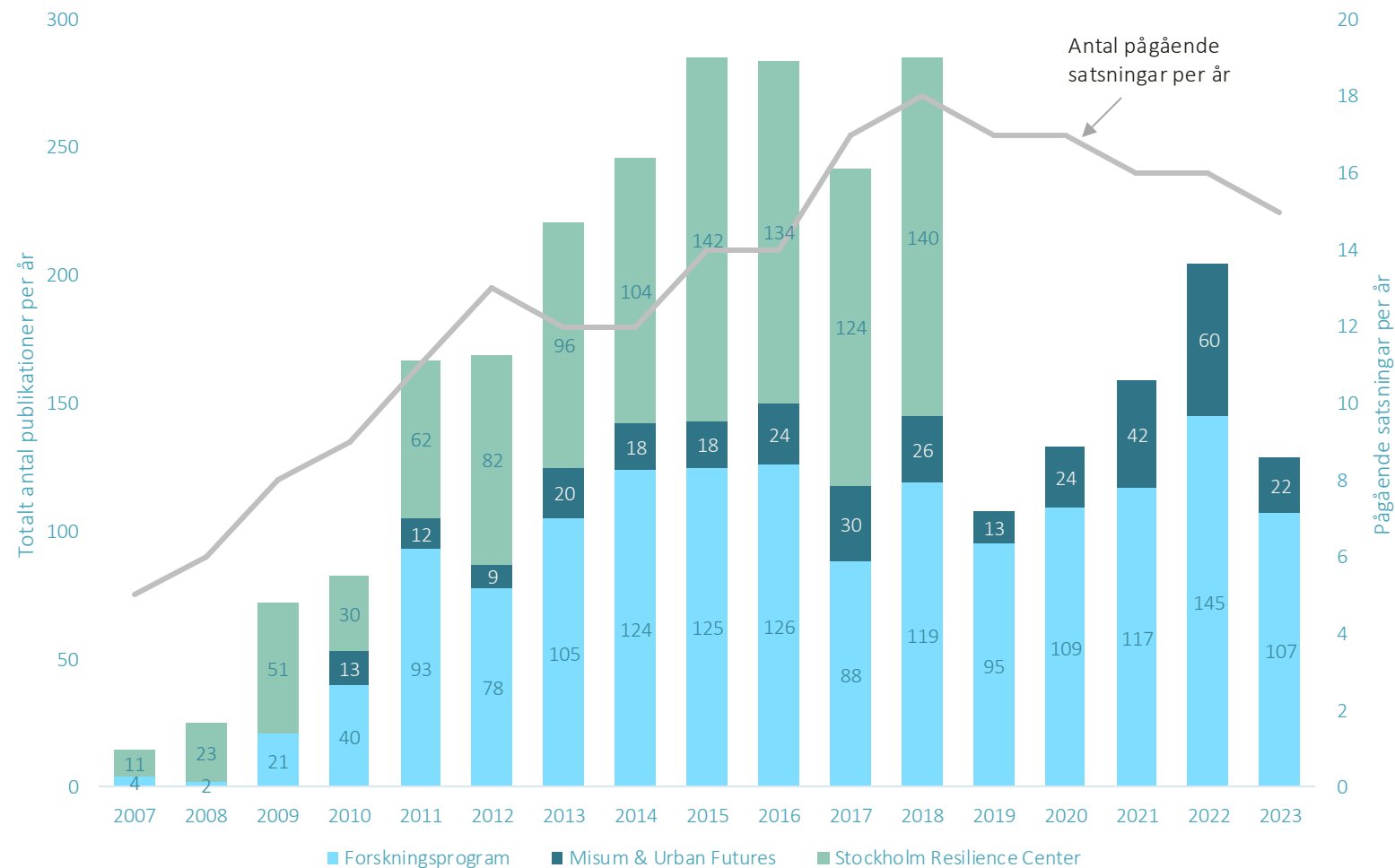
Staplarna i figuren visar hur publikationsvolymen för de undersökta satsningarna har utvecklats över tid mellan år 2007 och 2023. Linjen i figuren illustrerar antal pågående satsningar per år. Flest samtidigt pågående satsningar noteras under år 2018 då 18 olika satsningar var i gång.

Den låga publikationsvolymen under de inledande åren är en naturlig konsekvens av att inget av satsningarna i denna utvärdering existerade innan 2006. Det pågick fler program under denna period – men de har inkluderats i M20-rapporten varför de exkluderas från denna utvärdering. Detsamma gäller för slutet av perioden. Program som nyligen startats har exkluderats från denna utvärdering.

Den samlade publikationsvolymen ökar relativt stadigt fram till år 2018, varav Stockholm Resilience Centre (SRC) står för en betydande del. Nedgången i total publikationsvolym efter 2018 kan således härledas till SRC:s slut.

Trots att antal årligen pågående forskningsprogram ökar från nio till 15 mellan 2011 och 2018 syns ingen större ökning i publikationsvolym. 2011 publicerades 93 publikationer från forskningsprogram – att jämföra med 119 år 2018.

Underlaget visar inga tecken på att denna produktivitetsminskning kan härledas till utvärderingens urval av program, vilken programfas programmen befinner sig i eller varierande datakvalitet över tid.

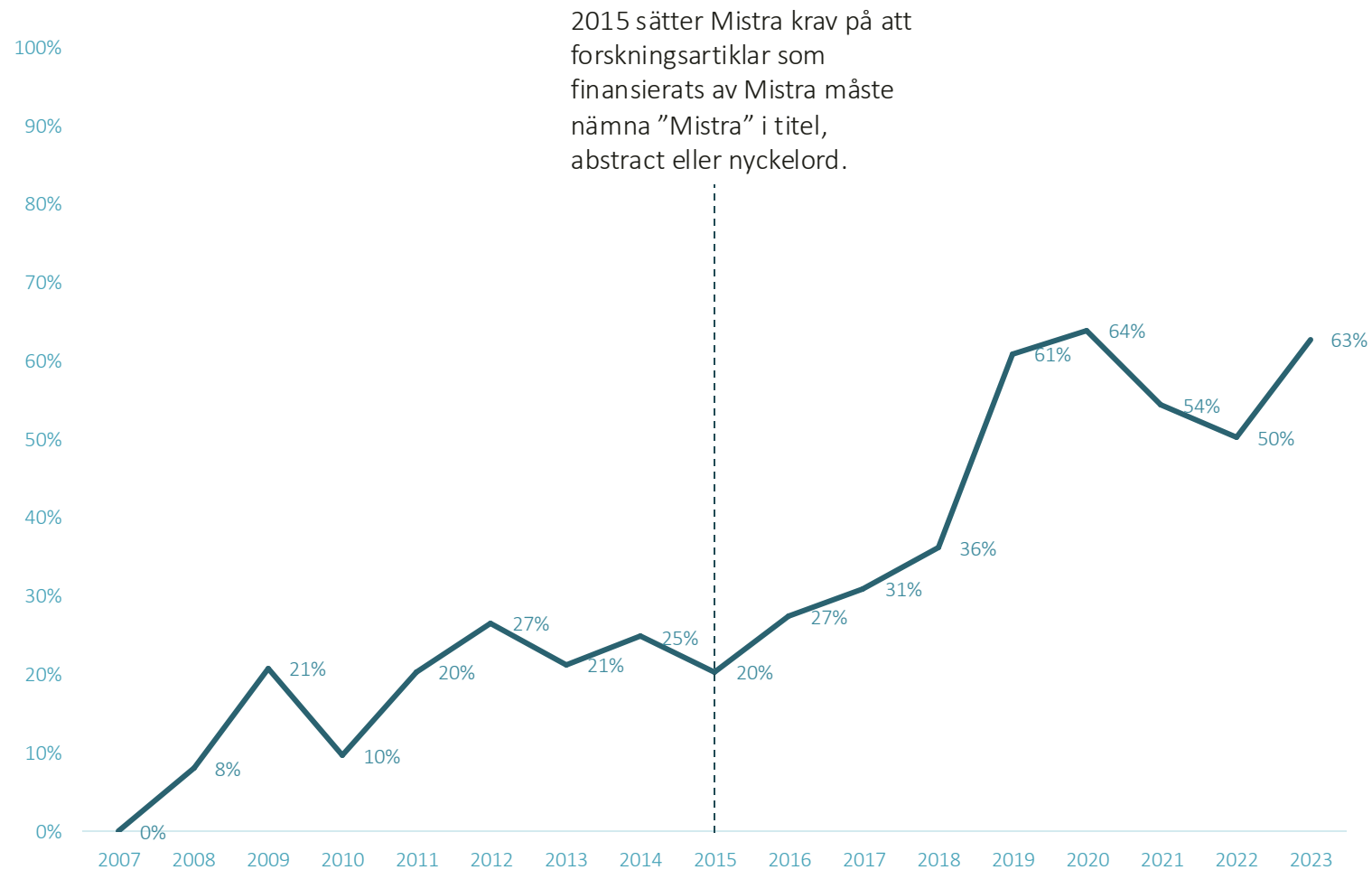


Andel artiklar som nämner Mistra i titel, abstract eller keywords

Grafen visar hur stor andel av de Mistrafinansierade artiklarna som nämner "Mistra" i titel, abstract eller bland nyckelorden.

Över tid syns en uppåtgående trend där andelen Mistrafinansierade artiklar som nämner "Mistra" i titel, abstract eller bland nyckelorden ökar över tid. En trolig anledning till den tydliga ökningen från och med 2015 är det krav som upprättas.

Trots den relativt kraftiga ökningen var det 2023 fortfarande bara 63 procent av artiklarna som nämnde Mistra.

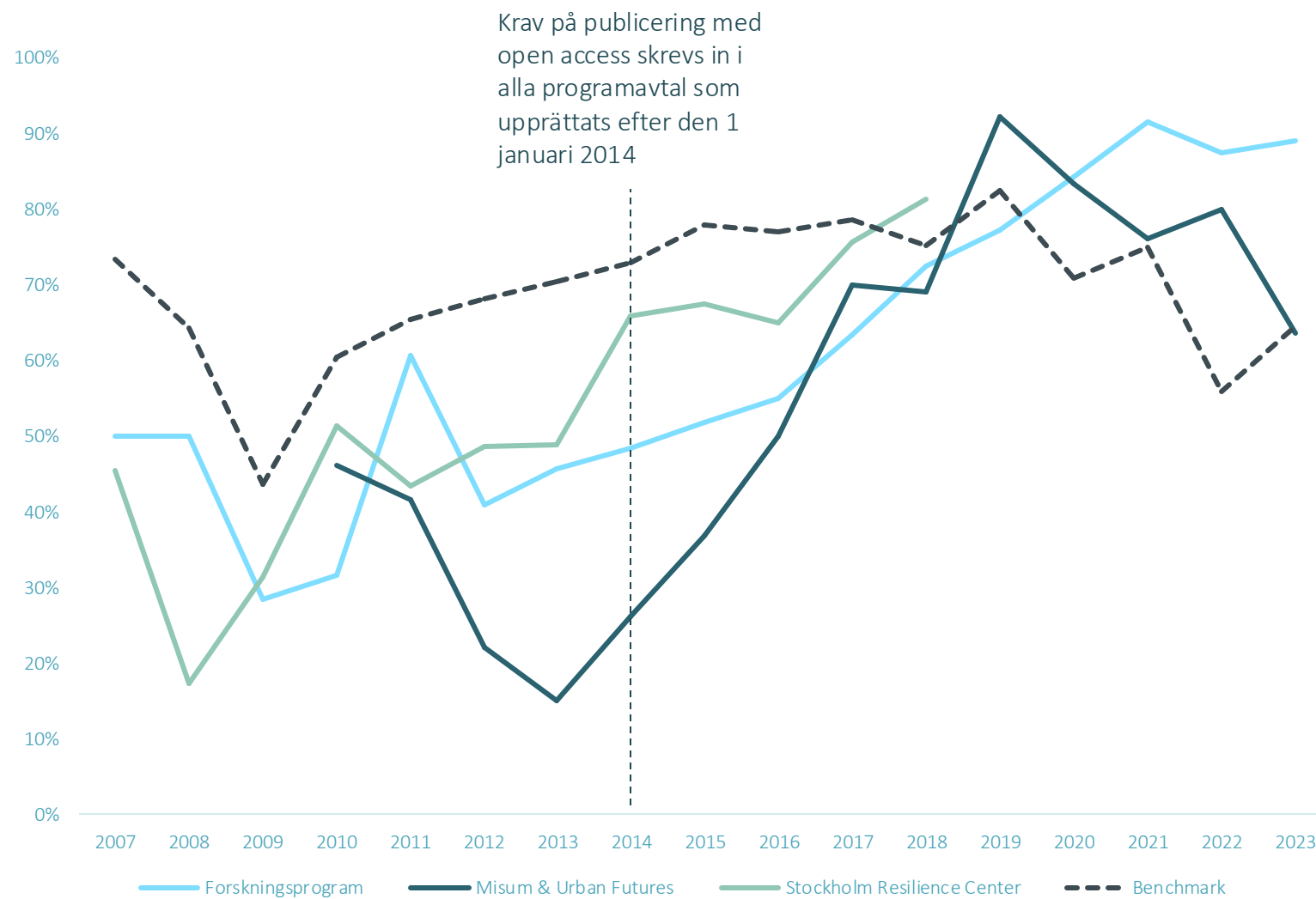


Tillgängligheten av publikationer finansierade av Mistra har ökat över tid

Utvecklingen för andelen publikationer som är publicerade i en tidskrift med öppen tillgång (open access) visas i grafen per finansieringsform. Det finns olika kategoriseringar för öppen tillgång* och i denna analys definieras alla publikationer med en typ av öppen tillgång som att den är öppen tillgång.

Andelen publikationer med öppen tillgång för forskningsprogram, Stockholm Resilience Centre, Misum och Mistra Urban Futures har ökat över de undersökta åren. Under de tidigare åren låg forskningsprogrammen och centrumbildningarna under jämförande benchmark**. Detta har förändrats och under de senare åren har forskningsprogrammen och de aktiva centrumbildningarna en högre andel open access än benchmark.

En tydlig trendförändring skedde för Misum och Urban Futures i samband med att Mistra införde krav på open access-publicering i nya programavtal. Misum och Urban Futures hade till en början en relativt låg andel open access-publicationer men sedan Mistras krav har andelen ökat till liknande nivåer som för forskningsprogrammen och Stockholm Resilience Centre.



*För definitioner av de olika typerna, se [Scopus](#).

** Benchmark: Alla publikationer publicerade i samma journaler som Mistrafinansierade publikationer.

Källa: ADC, Scopus

3. Forskningsfält



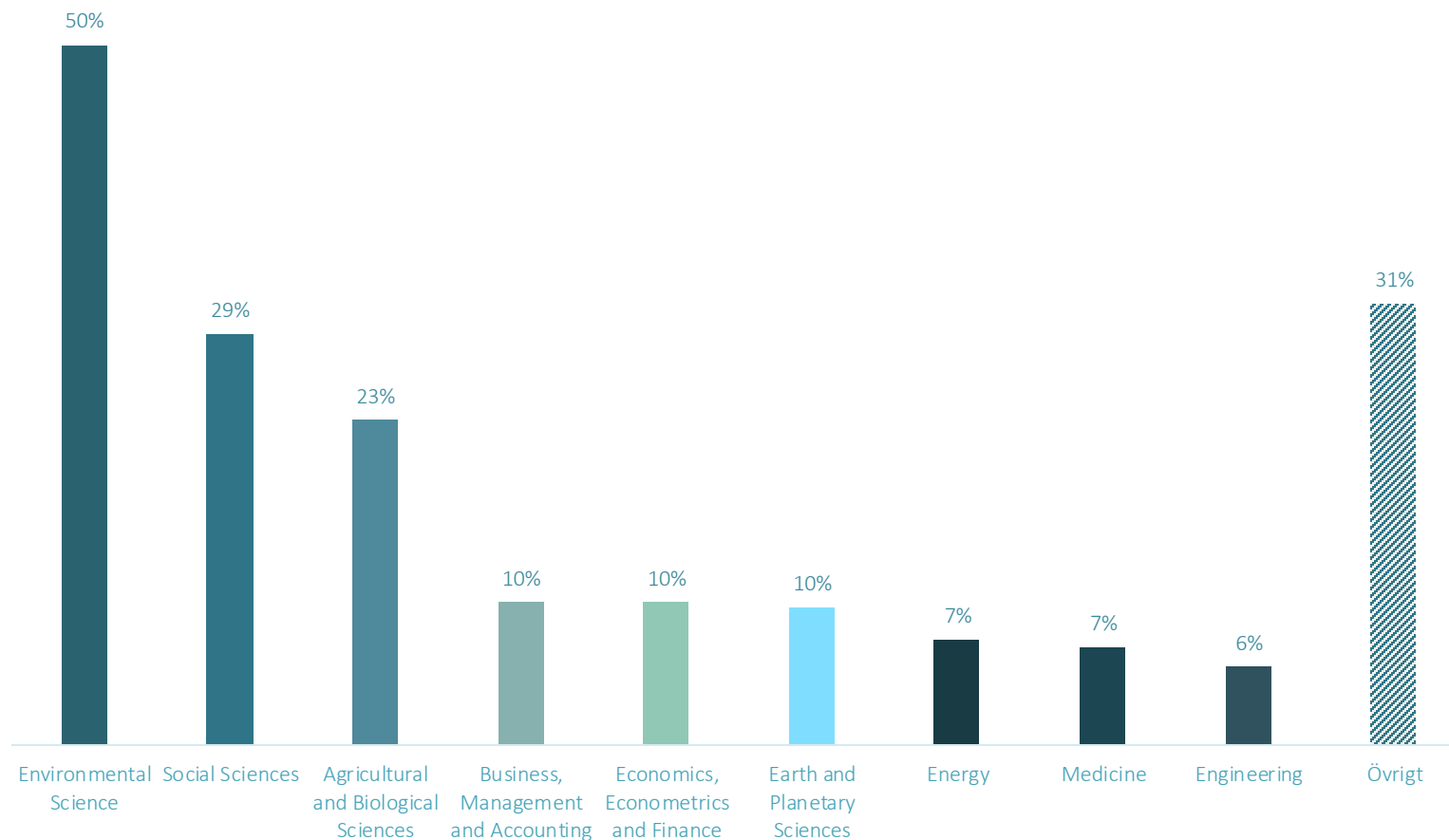
Hälften av publikationerna är publicerade i tidskrifter för miljövetenskap

De vetenskapliga artiklarna publiceras i tidskrifter som kategoriseras in i olika ämnesområden av Scopus. Värt att notera är att respektive publikation kan tillhöra fler än ett ämnesområde.

Figuren visar hur stor andel av den totala publikationsvolymen som har publicerats i tidskrifter med respektive ämnesområdesfokus.

Det största ämnesområdet är Environmental Science. 50 procent av alla studerade artiklar har publicerats i tidskrifter inom det ämnesområdet. Efter Environmental Sciences är tidskrifter inom Social Sciences (29 procent) och Agricultural and Biological Sciences (23 procent) de vanligaste ämnesområdena att publicera inom för de Mistrafinansierade publikationerna.

Inom Övrigt-kategorin finns ämnesområden som Biochemistry, Genetics and Molecular Biology, Multidisciplinary, Chemistry, Arts and Humanities, Decision Sciences, Computer Science och Psychology. Drygt fyra procent av publikationerna är publicerade i tidskrifter som klassificeras som Multidisciplinary.

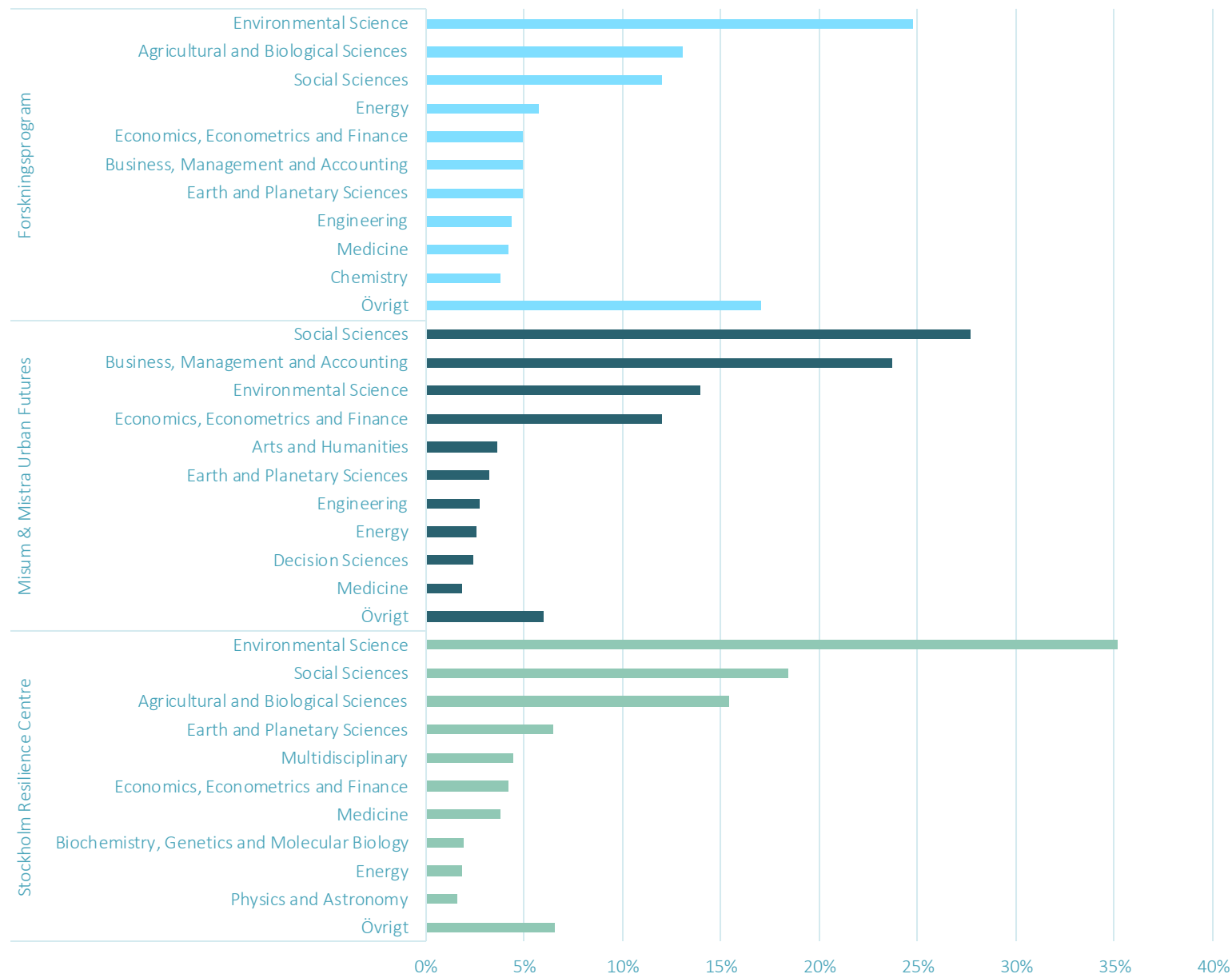


Fördelning per finansieringsform

Till höger visas hur stor andel av det totala antalet publikationer som har publicerats i tidskrifter med respektive ämnesområdesfokus per finansieringsform.

Det största ämnesområdet bland forskningsprogrammen är Environmental Science (25 procent), följt av Agricultural och Biological Sciences (13 procent) och Social Sciences (12 procent). Noterbart är att Övrigt-kategorin är den näst största bland forskningsprogrammen. Detta indikerar att forskningen är utspridd över många olika discipliner vilket i sin tur kan indikera att programmen inte är tillräckligt vetenskapligt fokuserade eller profilerade.

Fördelningen mellan ämnesområden för Stockholm Resilience Centre speglar i hög grad den fördelningen som gäller för samtliga finansieringsformer (som syns på föregående sida). Det största ämnesområdet är Environmental Science (35 procent), följt av Social Sciences (18 procent) och Agricultural och Biological Sciences (15 procent).



4. Sampublicering

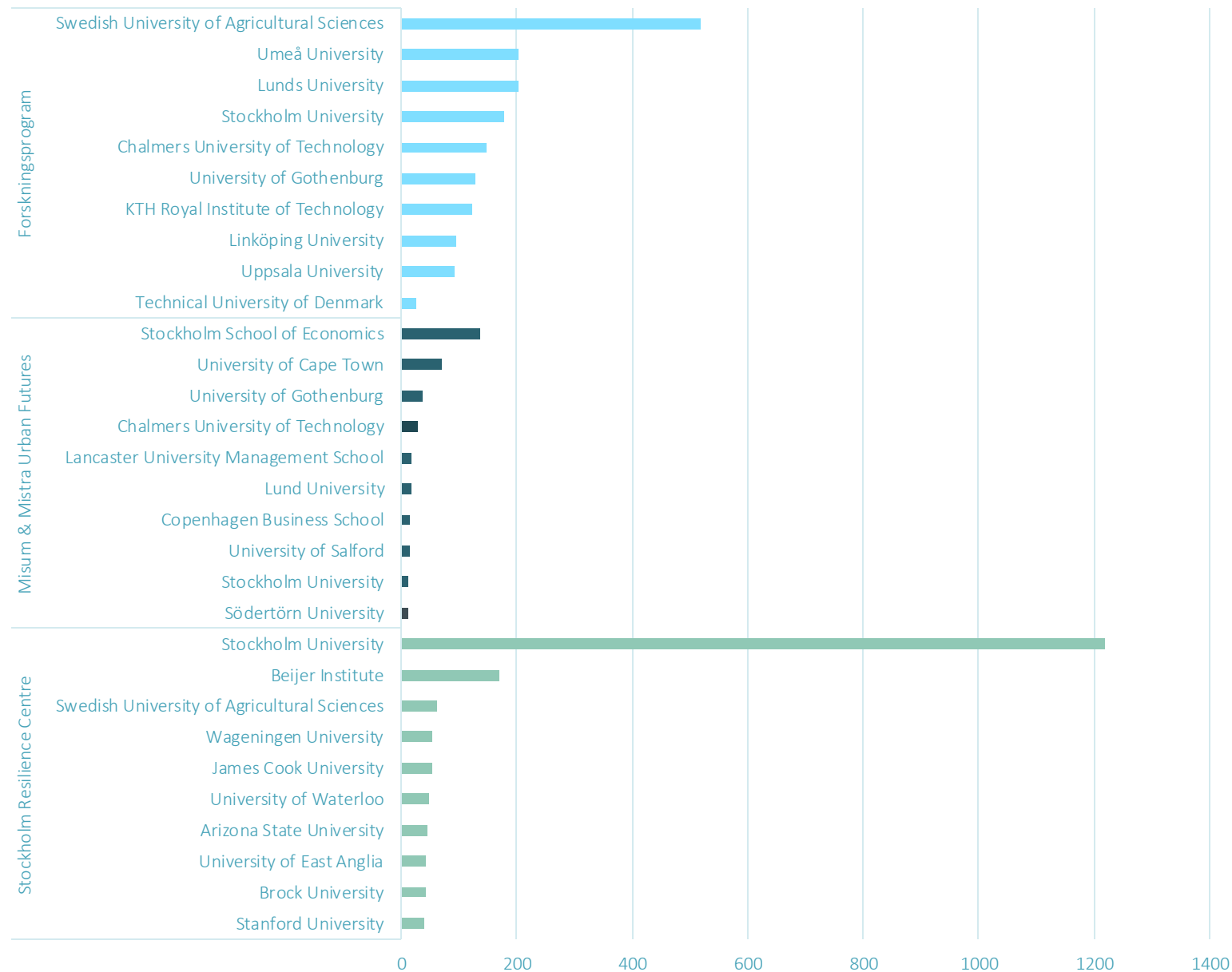


Forskare från Stockholms universitet publicerar mest

Figuren illustrerar de tio mest förekommande affilieringarna för författare av de Mistrafinansierade publikationerna mellan 2007 och 2023, per finansieringsform.*

Över samtliga finansieringsformer är den vanligast förekommande affilieringen Stockholms universitet med 1398 publikationer, vilket är mer än dubbelt så många som för den näst mest förekommande affilieringen, Sveriges lantbruksuniversitet (SLU).

Skillnaden mellan de olika finansieringsformerna är naturlig och beror på program/centrumvärd samt dessas forskningssamarbeten.



*En publikation kan ha fler än en affiliering.

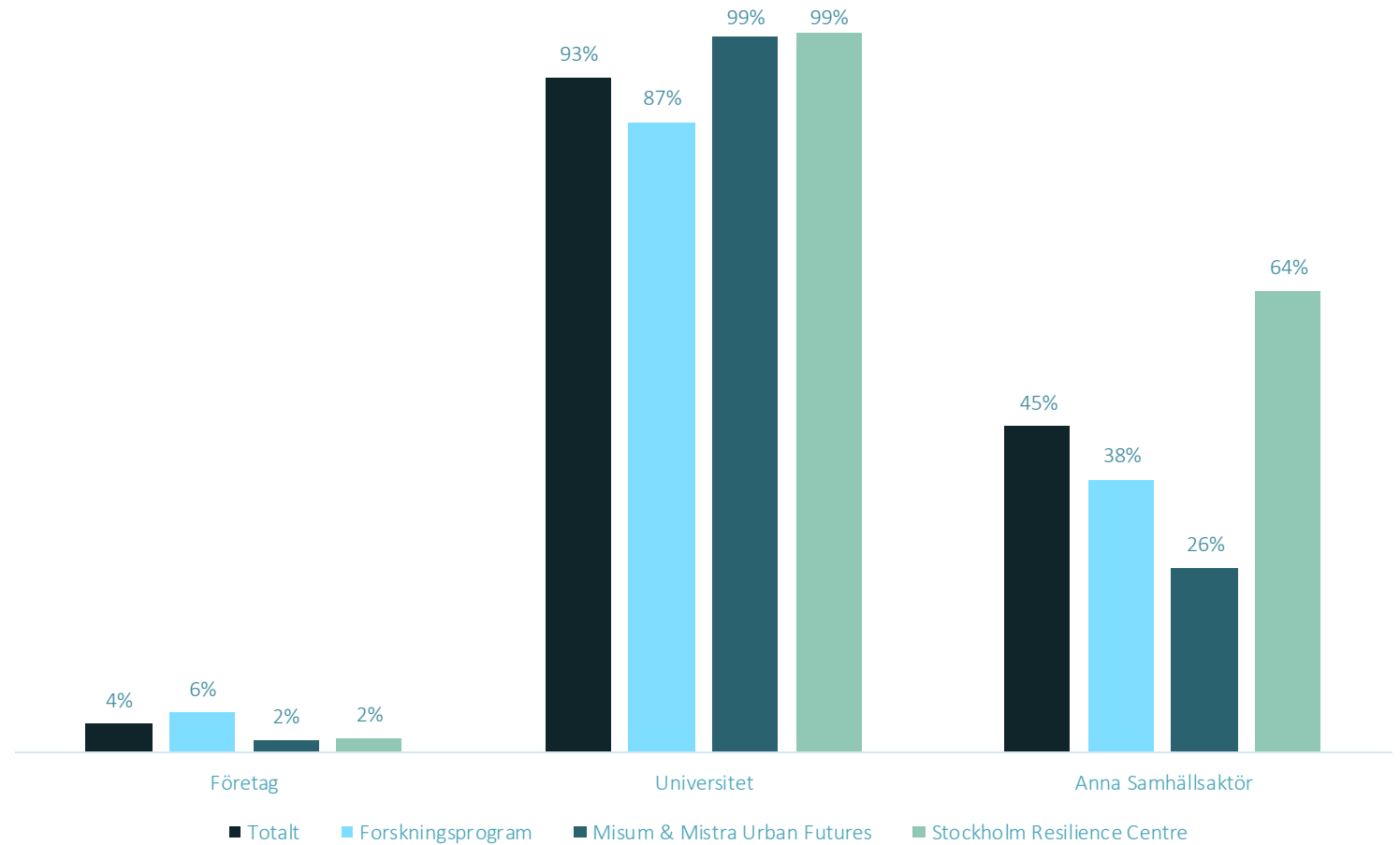
Låg grad av sampublicering med näringslivet

I figuren visas andelen av publikationerna som är affilierade med lärosäten, företag eller andra samhällsaktörer*.

93 procent av de samlade publikationerna har minst en författare som är affilierad till ett universitet. 45 procent har affilieringar till andra samhällsaktörer vilket kan vara exempelvis forskningsinstitut. Endast 4 procent av publikationerna har minst en författare som är affilierad till ett företag.** Detta visar att företagen inom satsningarna inte engageras genom sampublicering utan snarare på andra sätt.

Samtliga finansieringsformer har låg grad av näringslivssamarbete sett utifrån sampubliceringar. Forskningsprogrammen har något högre grad än centrumbildningarna.

Andelen av publikationerna som är affilierade med andra samhällsaktörer varierar mellan finansieringsformerna. 38 procent av publikationerna genererade av forskningsprogrammen har minst en författare som är affilierad med andra samhällsaktörer. För Stockholm Resilience Centre ligger nivån på 64 procent.



*En publikation kan ha flera affilieringar

** Jämför exempelvis mot tre KK-miljöer där andelen samförfattade artiklar med näringslivet ligger mellan 9-18 procent ([länk](#)).

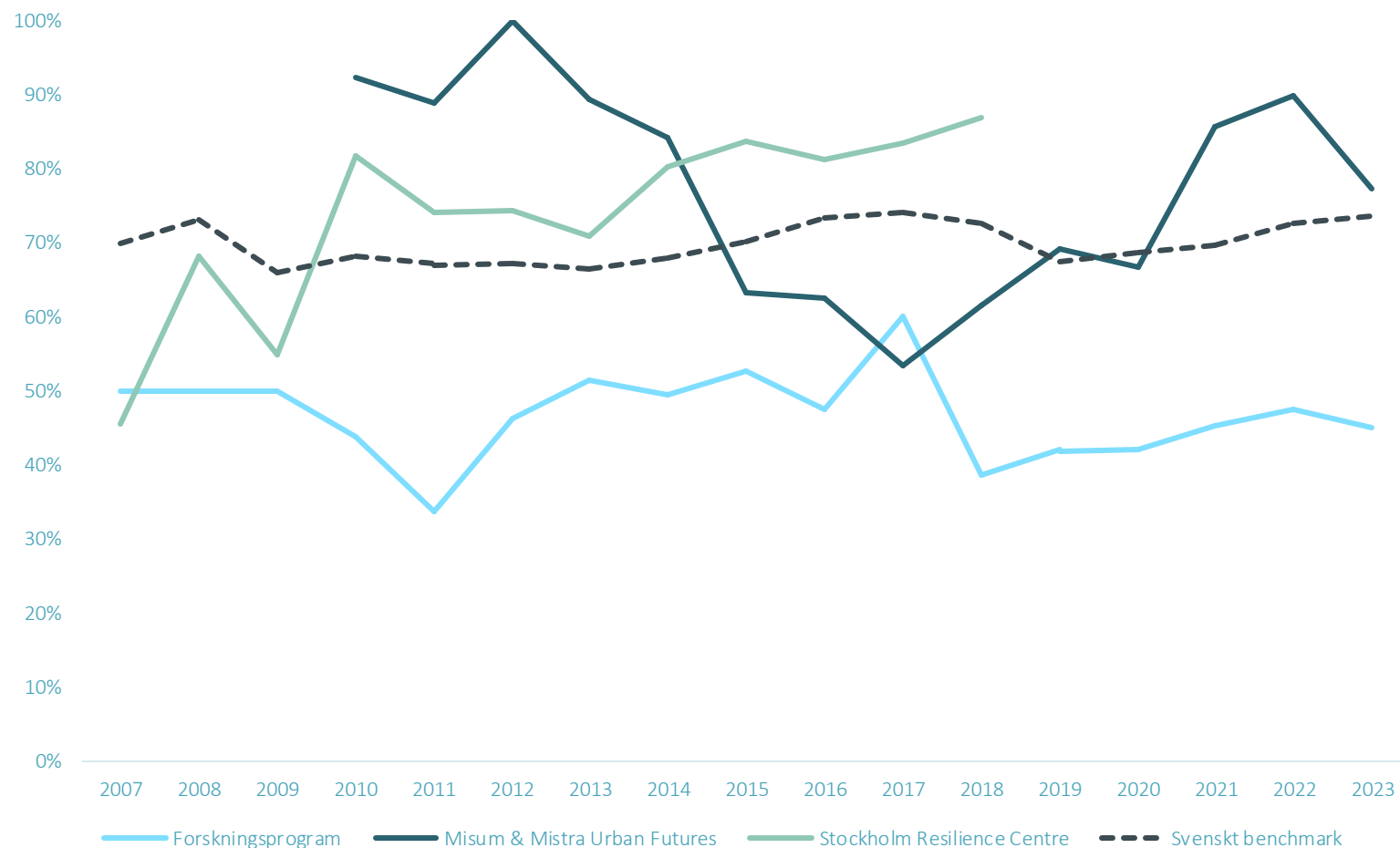
Källa: ADC, Scopus

Graden av internationell samverkan varierar över tid och skiljer sig mellan de olika finansieringsformer

Till höger visas andel sampublicerade publikationer med åtminstone en författare med internationell affiliering per finansieringsform över tid. Vidare visas även svenskt benchmark*.

Forskningsprogrammen har lägre grad av internationellt samarbete jämfört med benchmark under hela den studerade perioden. Nivån ökar mellan 2011-2017 och når högsta nivån på 60 procent för att sedan sjunka till 39 procent 2018. Totalt sett har flest internationella sampubliceringar skett med Danmarks Tekniske Universitet för forskningsprogrammen.

Stockholm Resilience Centre ligger något över snittet och håller en relativt jämn nivå över tid med undantag för de inledande åren. De högsta nivåerna nåddes i slutet av satsningen då 87 procent av publikationerna var sampublicerade med minst en författare med internationell affiliering. Wageningen University är den internationella aktör som SRC sampublicerat mest med.



*Benchmark: Alla publikationer publicerade i samma journaler som Mistrainfinansierade publikationer som har åtminstone en svensk affiliering.

Källa: ADC, Scopus

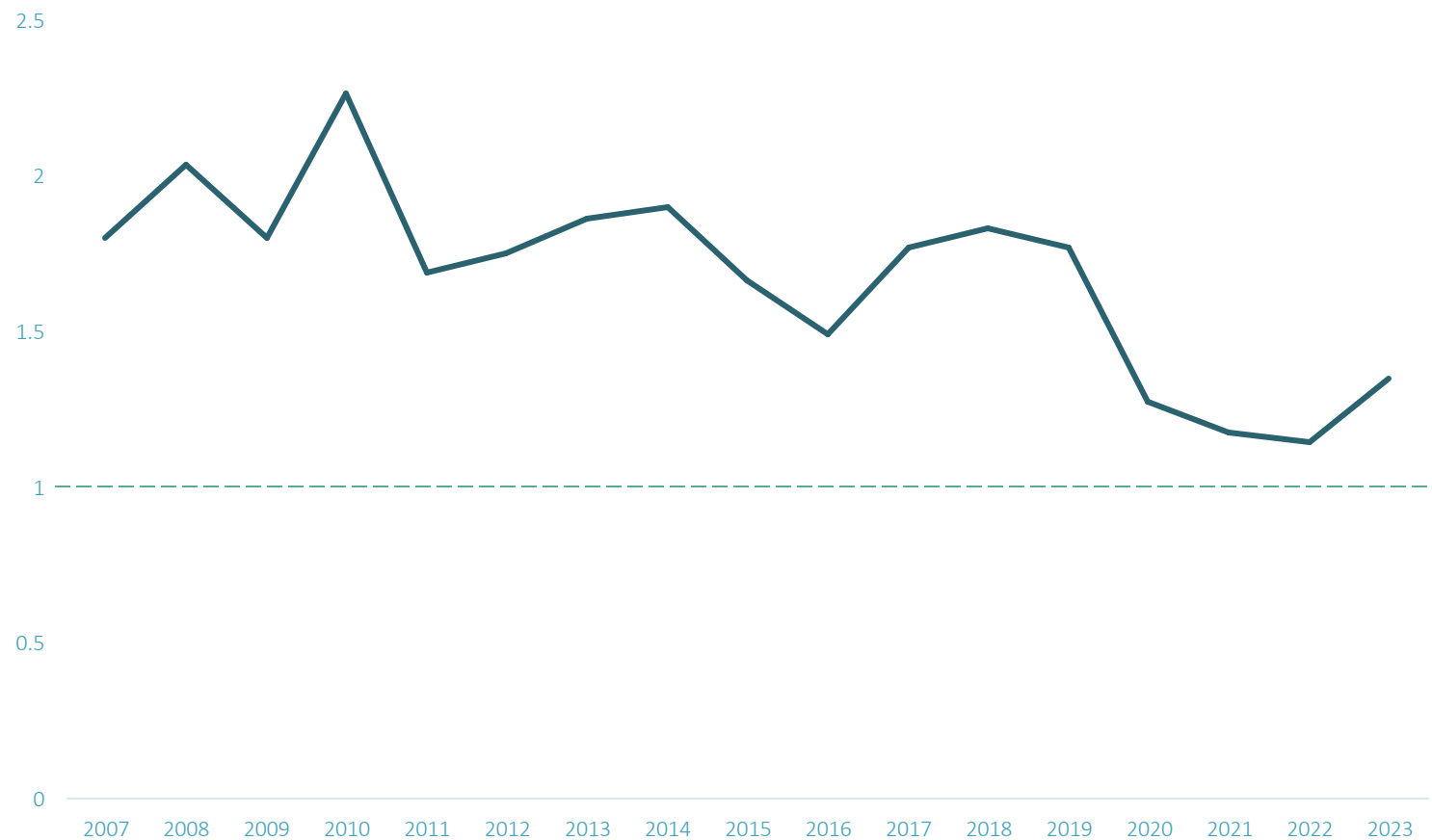
5. Excellens



Avtagande vetenskaplig genomslagskraft över tid men högre än jämförbara publikationer

Figuren visar den årliga förändringen av vetenskaplig genomslagskraft och ska tolkas som citeringsgrad i förhållande till publikationer från samma tidskrifter, under samma period. Ett värde över 1,0 innebär att MISTRAfinansierade publikationer blir citerade i högre grad än genomsnittet. Ett värde under 1,0 innebär att de MISTRAfinansierade publikationerna blir citerade i lägre grad än genomsnittet. Exempelvis innebär ett värde på 0,99 att publikationerna blir citerade 1 procent färre gånger än genomsnittet.

MISTRAfinansierade publikationer har en högre vetenskaplig genomslagskraft än andra jämförbara publikationer under hela den studerade perioden. Den relativa genomslagskraften har dock minskat över tid. Det högsta uppmätta värdet är för 2010 då MISTRAfinansierade publikationer citerades 127 procent oftare än jämförbara publikationer. Det lägsta uppmätta värdet är för 2022 då MISTRAfinansierade publikationer i genomsnitt citerades 15 procent mer än jämförbara publikationer.

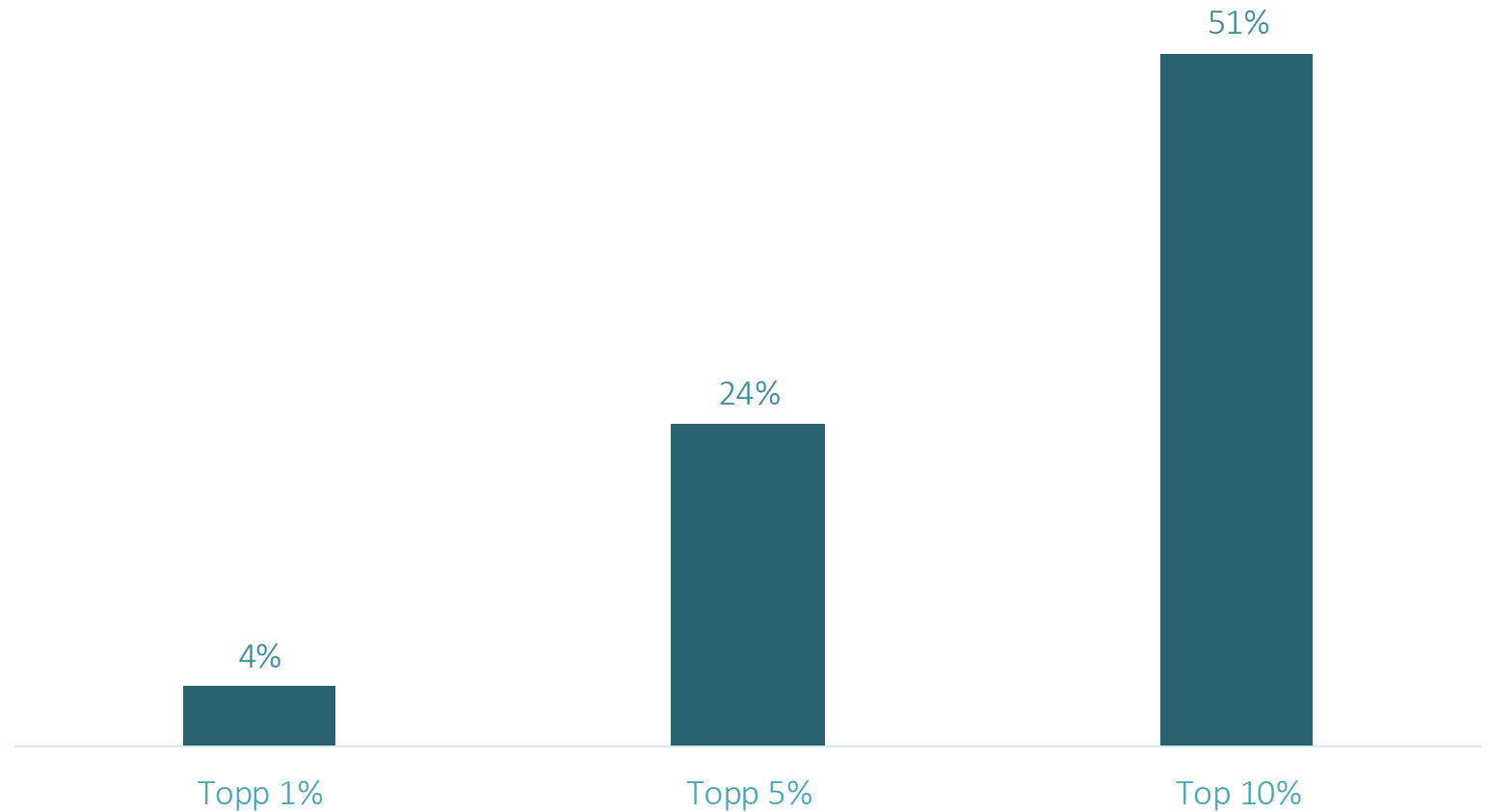


Mistrafinansierad forskning blir i hög utsträckning publicerad i topprankade tidskrifter

Figuren beskriver hur stor andel av de vetenskapliga publikationerna i denna utvärdering som är publicerad i topprankade tidskrifter inom respektive ämnesområde*.

De finansierade publikationerna blir i hög utsträckning publicerade i topprankade tidskrifter. 4 procent av publikationerna är publicerade i topp 1 procent rankade tidskrifter. Nästan en fjärdedel av publikationerna är publicerade i topp 5 procent rankade tidskrifter och drygt hälften av publikationerna är publicerade i topp 10 procent-rankade tidskrifter.

På nästa sida redovisas dessa mått per finansieringsform.



Källa: ADC, Scopus

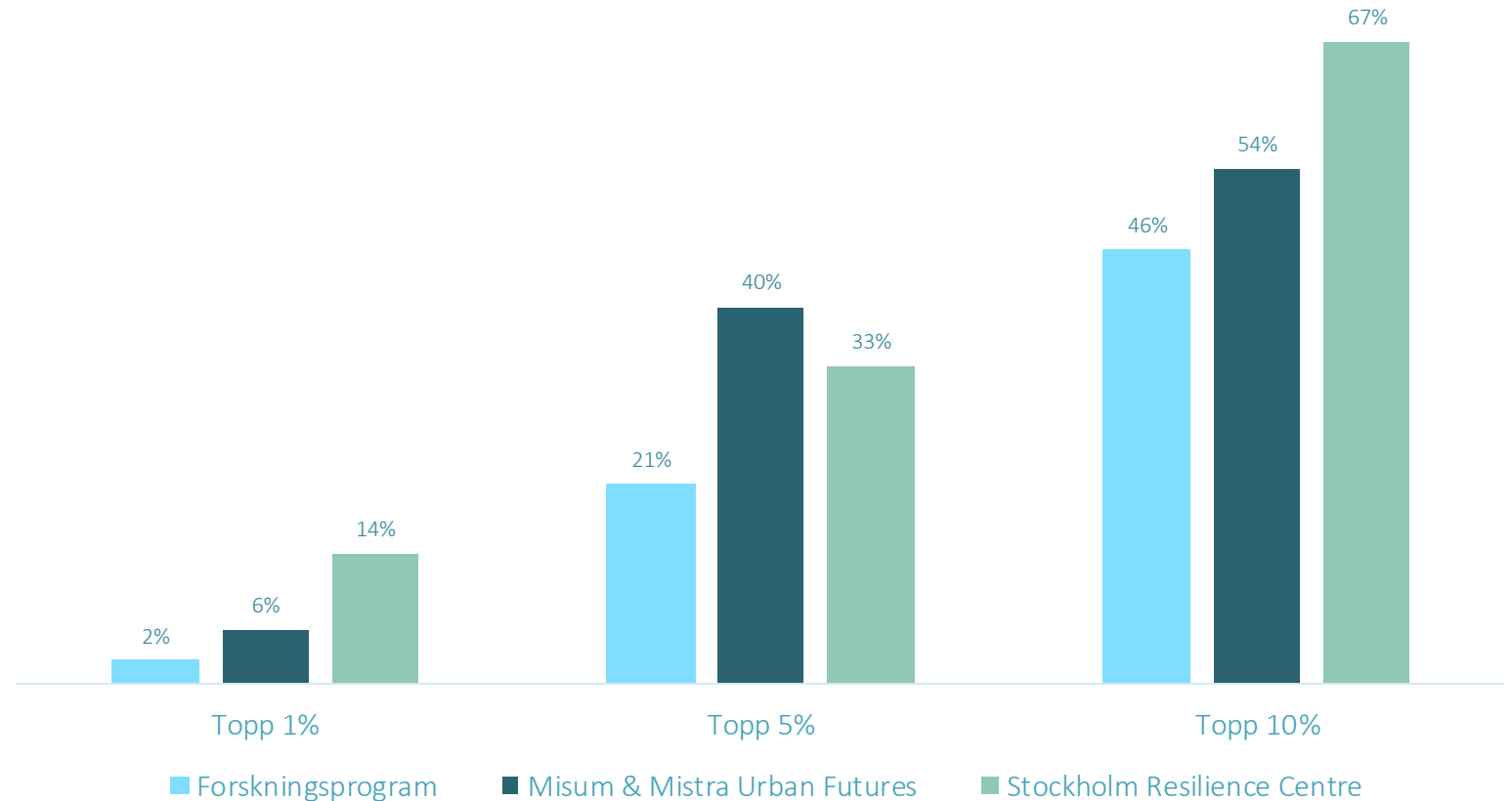


* En mer detaljerad beskrivning av topprankade tidskrifter finns i Appendix.

14 procent av publikationerna från SRC publicerad i topp 1 procent rankade tidskrifter

14 procent av publikationerna som producerats inom Stockholm Resilience Centre är publicerade i topp 1 procent rankade tidskrifter. Vi ser även att 33 respektive 67 procent av deras publikationerna blir publicerade i topp 5 och 10 procent rankade tidskrifter.

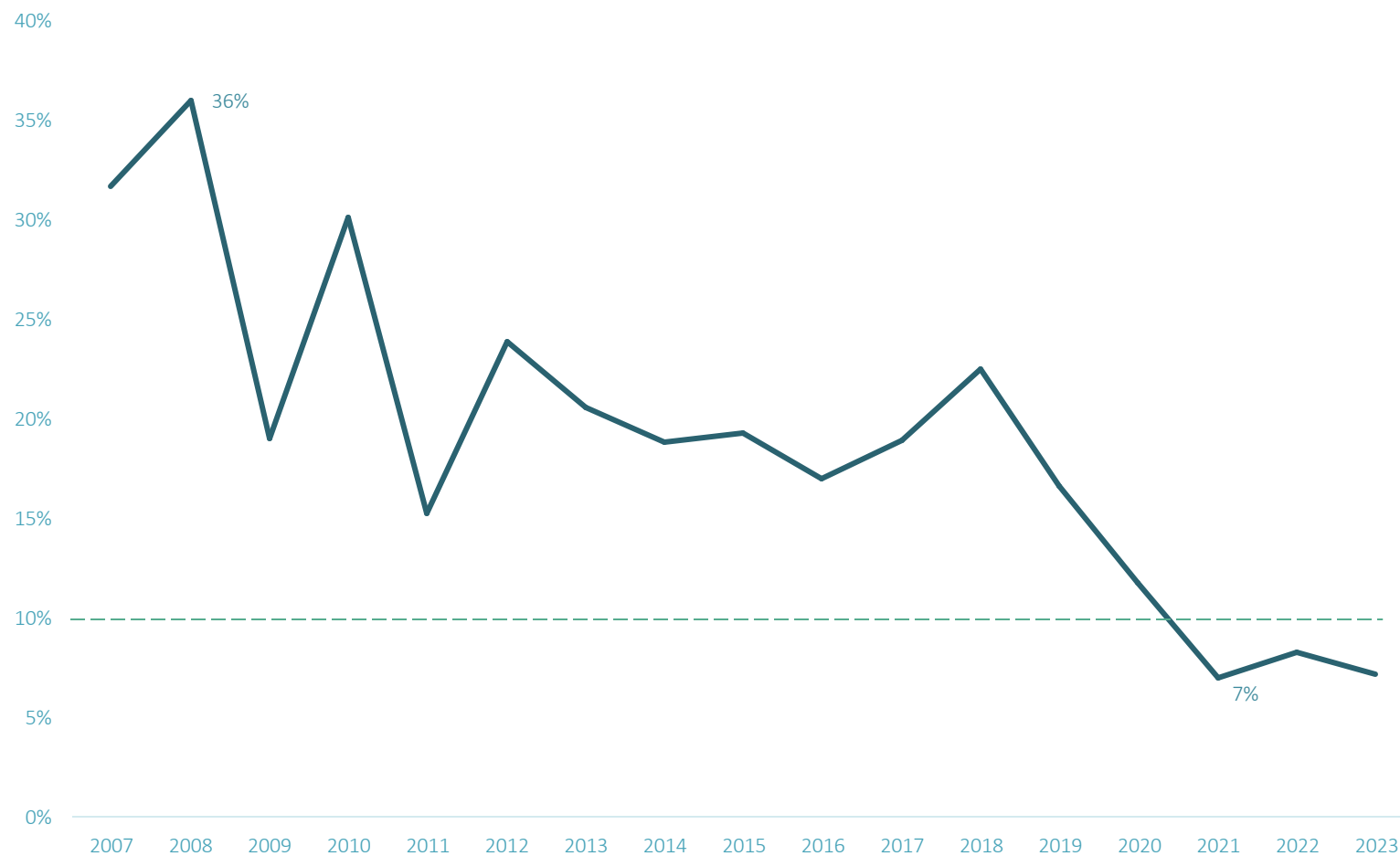
Publikationerna från forskningsprogrammen blir till något lägre grad jämfört med SRC och centrubildningarna publicerade i topprankade tidskrifter. Trots detta håller de ändå hög nivå. 2 procent av publikationerna är publicerade i topp 1 procent rankade tidskrifter. Vi ser även att 21 respektive 46 procent av deras publikationerna blir publicerade i topp 5 och 10 procent rankade tidskrifter.



Avtagande andel högciterade artiklar

Ett ytterligare mått som beskriver vetenskaplig kvalitet är andel toppciteringar. Till skillnad från måttet på föregående sida, andel publikationer publicerade i topprankade tidskrifter, tar detta mått inte hänsyn till vilken tidskrift som artikeln publicerats i utan endast till ämnesområde och publiceringsår.* Grafen visar andel Mistrafinansierade publikationer som tillhör de topp 10 procent mest citerade publikationerna inom respektive ämnesområde per publiceringsår.

Andelen Mistrafinansierade publikationer som tillhör de topp 10 procent mest citerade publikationerna har minskat under den studerade perioden. År 2008 var den uppmätta andelen högst – 36 procent av de Mistrafinansierade publikationerna tillhörde de topp 10 procent mest citerade inom sina respektive ämnesområden. Detta kan kontrasteras mot år 2021 när endast 7 procent av publikationerna tillhörde de topp 10 procent mest citerade. Figuren visar på en tydligt nedgående andel högciterade artiklar.



* En mer detaljerad beskrivning av andel toppciteringar finns i Appendix.

Källa: ADC, Scopus

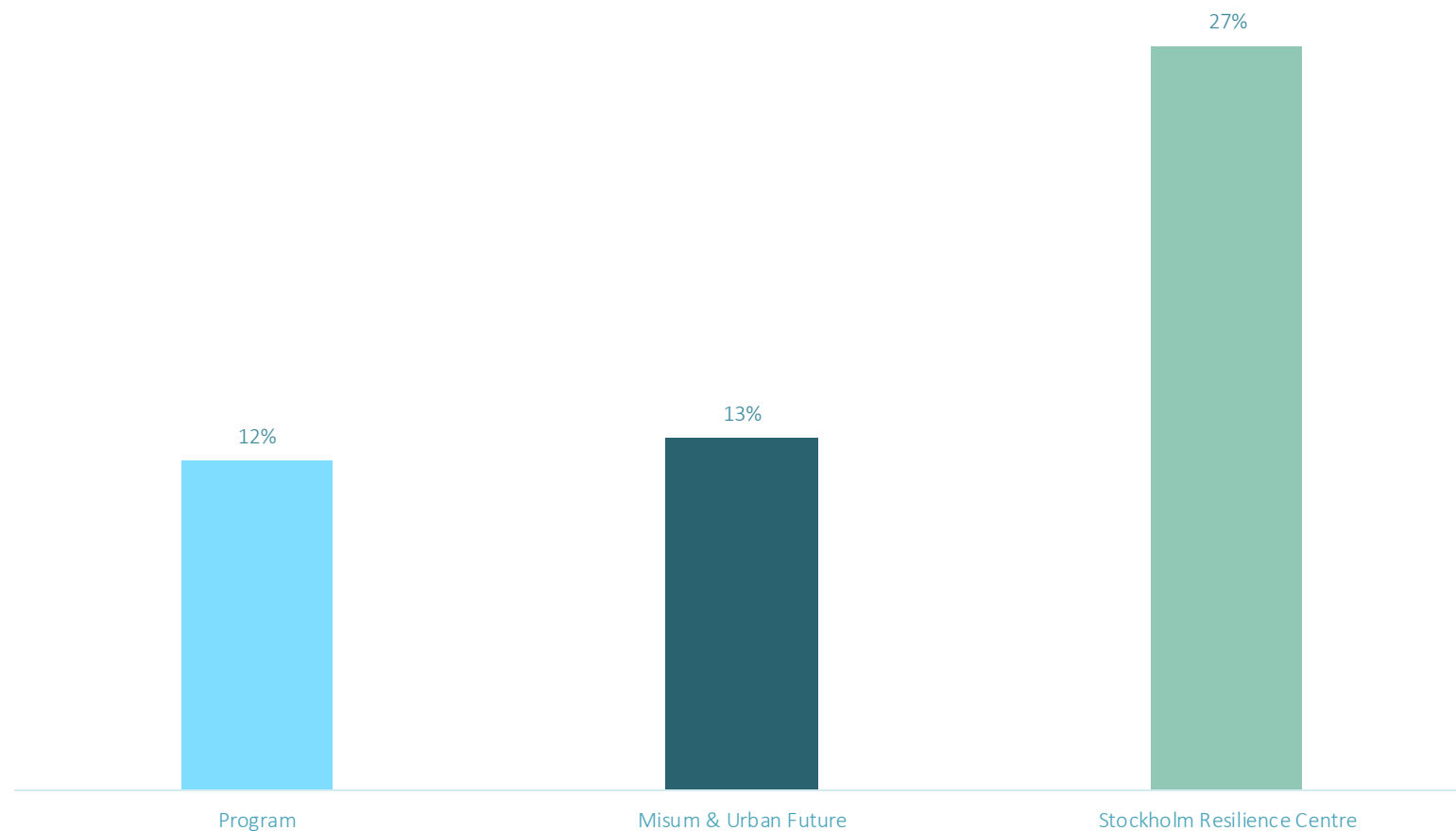
Total andel högciterade artiklar per finansieringsform

Publikationerna från samtliga typer av Mistrasatsningar blir i hög grad citerade. Figuren visar den totala andelen, per finansieringsform, av publikationerna som tillhör de 10 procent mest citerade publikationer inom respektive ämnesområde.

Andelen publikationer från forskningsprogrammen som tillhör de 10 procent mest citerade publikationerna är 12 procent.

Stockholm Resilience Centre utmärker sig genom att 27 procent av dess publikationerna tillhör de topp 10 procent mest citerade publikationerna.

På nästa sida visas utvecklingen över tid per finansieringsform.



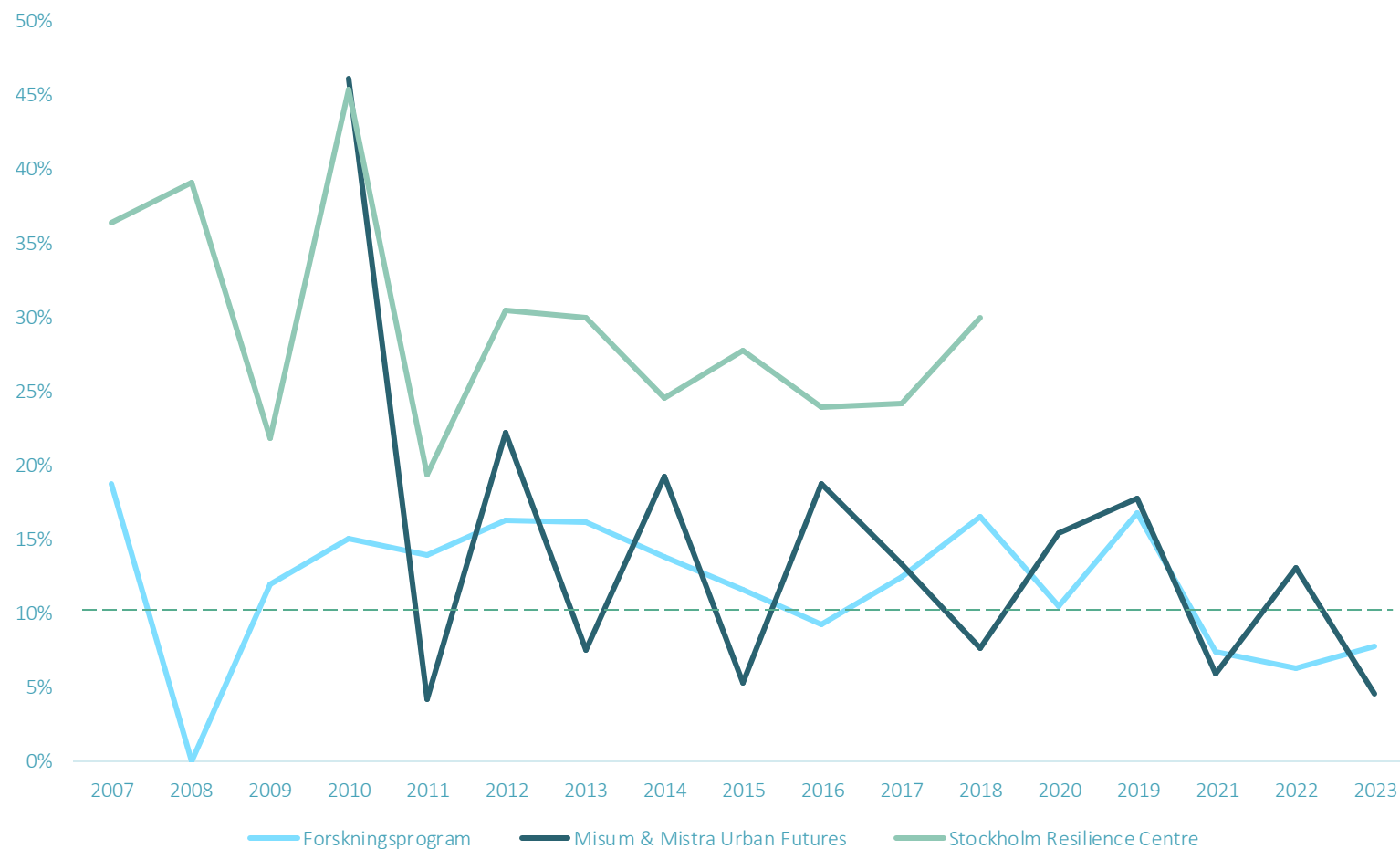
* En mer detaljerad beskrivning av andel toppciteringar finns i Appendix.

Publikationer från Misum och Urban Futures följer liknande utveckling som de från forskningsprogrammen

Figuren visar en tydlig skillnad mellan centrumbildningarna Misum och Urban Futures samt forskningsprogrammen, och SRC. Andelen publikationer som tillhör de topp 10 procent mest citerade är med undantag för år 2010 konstant högre än för centrumbildningarna och forskningsprogrammen.

Andelen publikationer som tillhör de topp 10 procent mest citerade är relativt stabil för forskningsprogrammen – under majoriteten av åren är andelen över 10 procent. En något nedgående trend noteras under perioden.

Under den andra halvan av SRCs period stabiliseras andelen något mellan 24-30 procent av publikationerna tillhörande de topp 10 procent mest citerade inom respektive forskningsfält.



* En mer detaljerad beskrivning av andel toppciteringar finns i Appendix.

Källa: ADC, Scopus

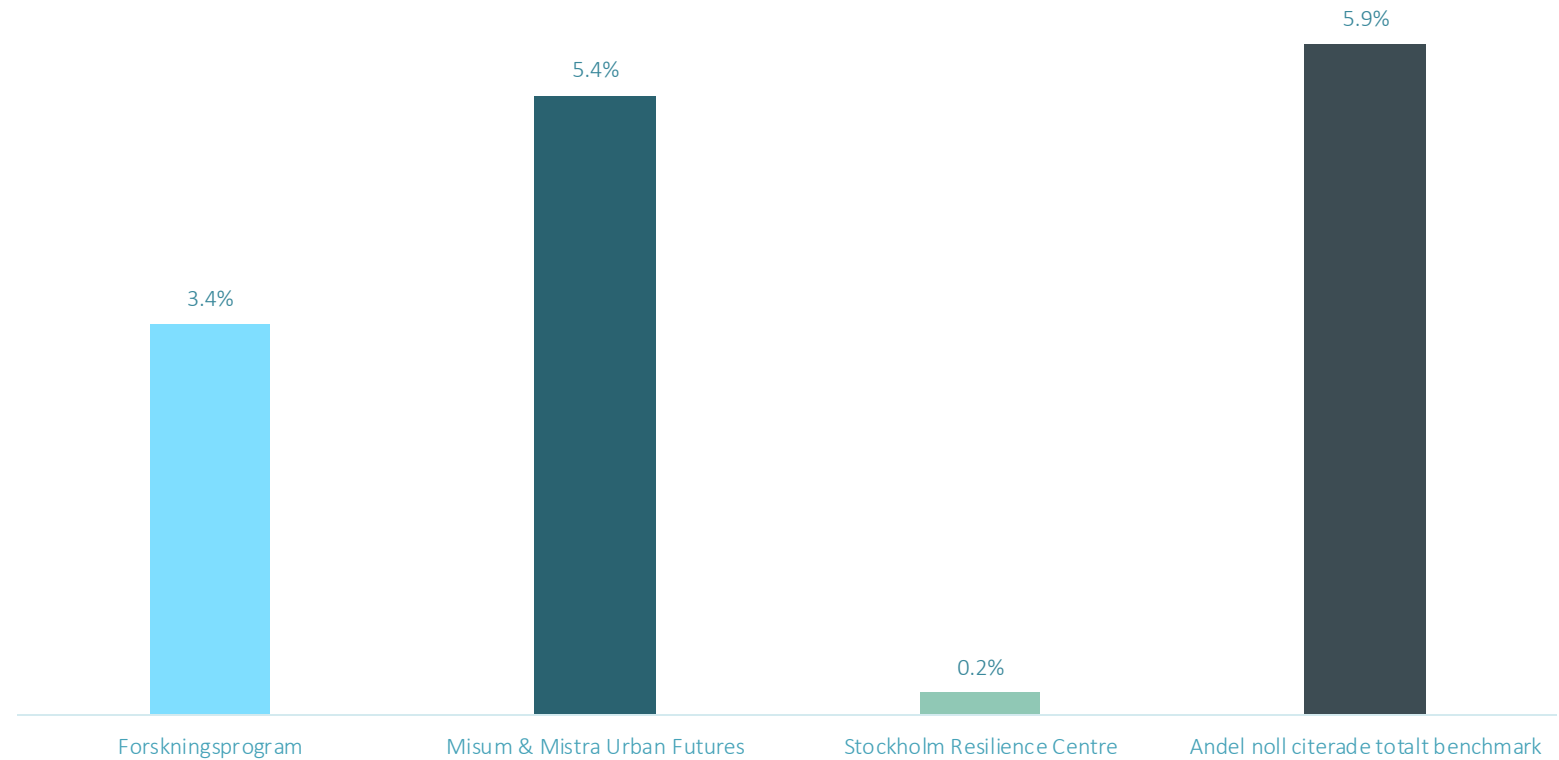
Mistrafinansierad forskning är i låg utsträckning nollciterad

Figuren till höger visar andel nollciterade Mistrafinansierade publikationer jämfört med ett benchmark. Benchmarket utgörs av andel nollciterade artiklar från samtliga tidskrifter Mistrafinansierade publikationer har publicerats i.

Samtliga typer av Mistrasatsningar visar på en lägre andel nollciterade publikationer jämfört med benchmark.

Bland publikationer producerade av Stockholm Resilience Centre är mycket få nollciterade- endast 0,2 procent.

Andelen nollciterade publikationer producerade av forskningsprogrammen ligger på 3,4 procent.



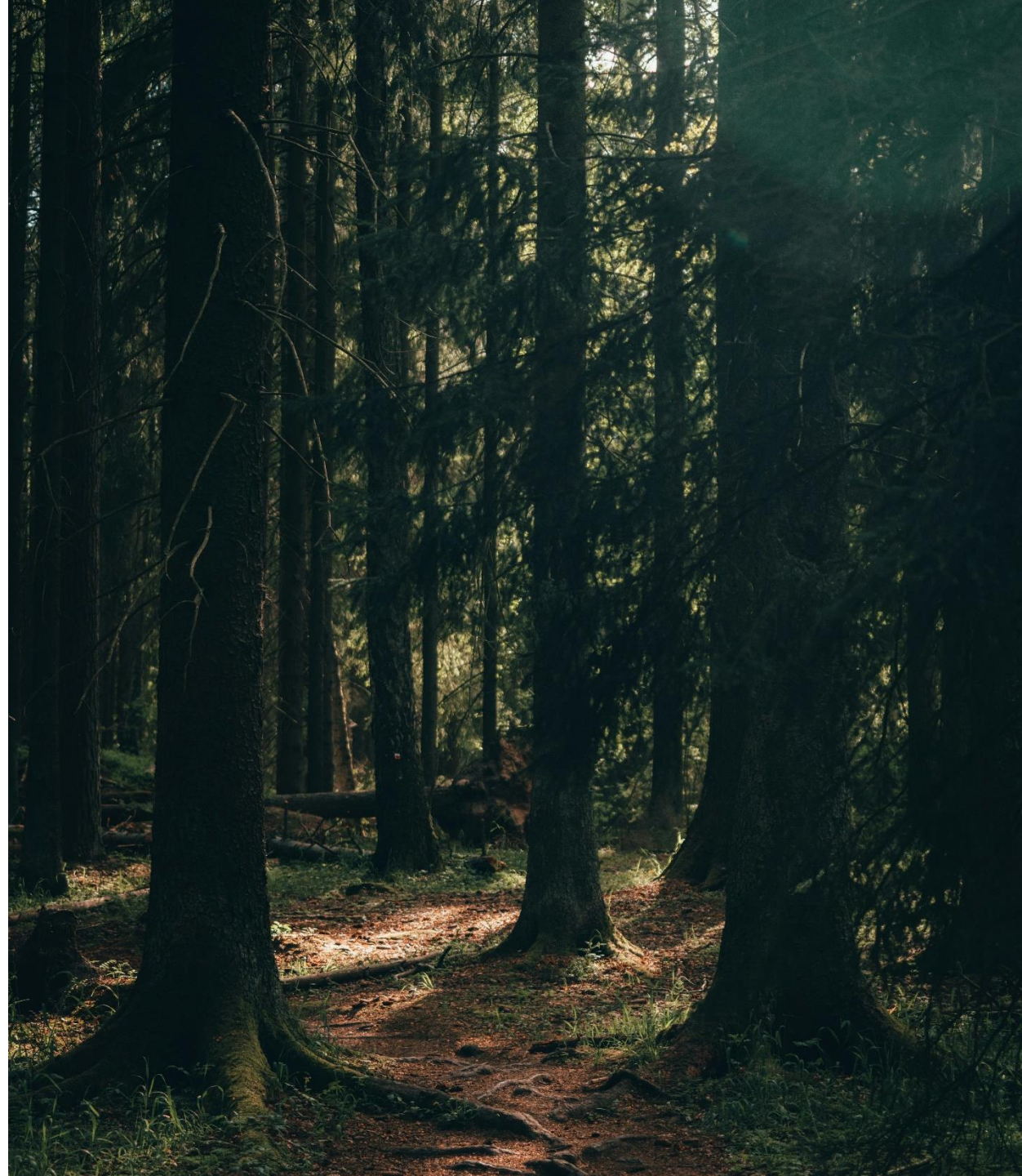
Appendix



Appendixinnehåll

A Indikatorer.....[31](#)

B Klusteranalys.....[33](#)



A: Indiktorer



Bibliometriska indikatorer

Andel toppciterade publikationer

Indikerar världsledande forskning med högt genomslag (kvalitet). En toppciterad publikation definieras som en publikation som är bland de 10% mest citerade publikationerna inom ett forskningsområde och år. Detta är samma definition som Vetenskapsrådet använder i sin vetenskapsbarometer. I enlighet med Vetenskapsrådets metod omfattar analysen enbart publikationer av typen Articles och Reviews. Andelen toppciterade publikationer bestäms genom att beräkna hur många av publikationerna som har tillräckligt med citeringar för att vara bland de 10% mest citerade publikationerna inom sitt fält.

Publicering i topprankade tidskrifter

Visar hur många publikationer inom det aktuella området som publiceras i de bästa och mest inflytelserika tidskrifterna. Vi använder Scimago Journal Rank (SJR) för att ranka tidskrifterna.

$$\frac{\text{Antal mistrafinansierade publikationer som är publicerade i topp 1 \% rankade tidskrifter}}{\text{Totalt antal mistrafinansierade publikationer}}$$

Internationell samverkan

Internationell samverkan beskriver i vilken grad lärosätet publicerar tillsammans med utländska organisationer. Publikationer i internationell samverkan har bland annat identifierats som viktigt för högt vetenskapligt genomslag. Data baseras på programmets publikationer vilka identifieras genom databasen Scopus. Genom databasen identifieras de organisationer som kan kopplas till varje enskild publikation. Organisationerna klassificeras sedan som antingen svenska eller internationella. Utifrån detta kan andelen publikationer i internationell samverkan beräknas.



B: Kluster



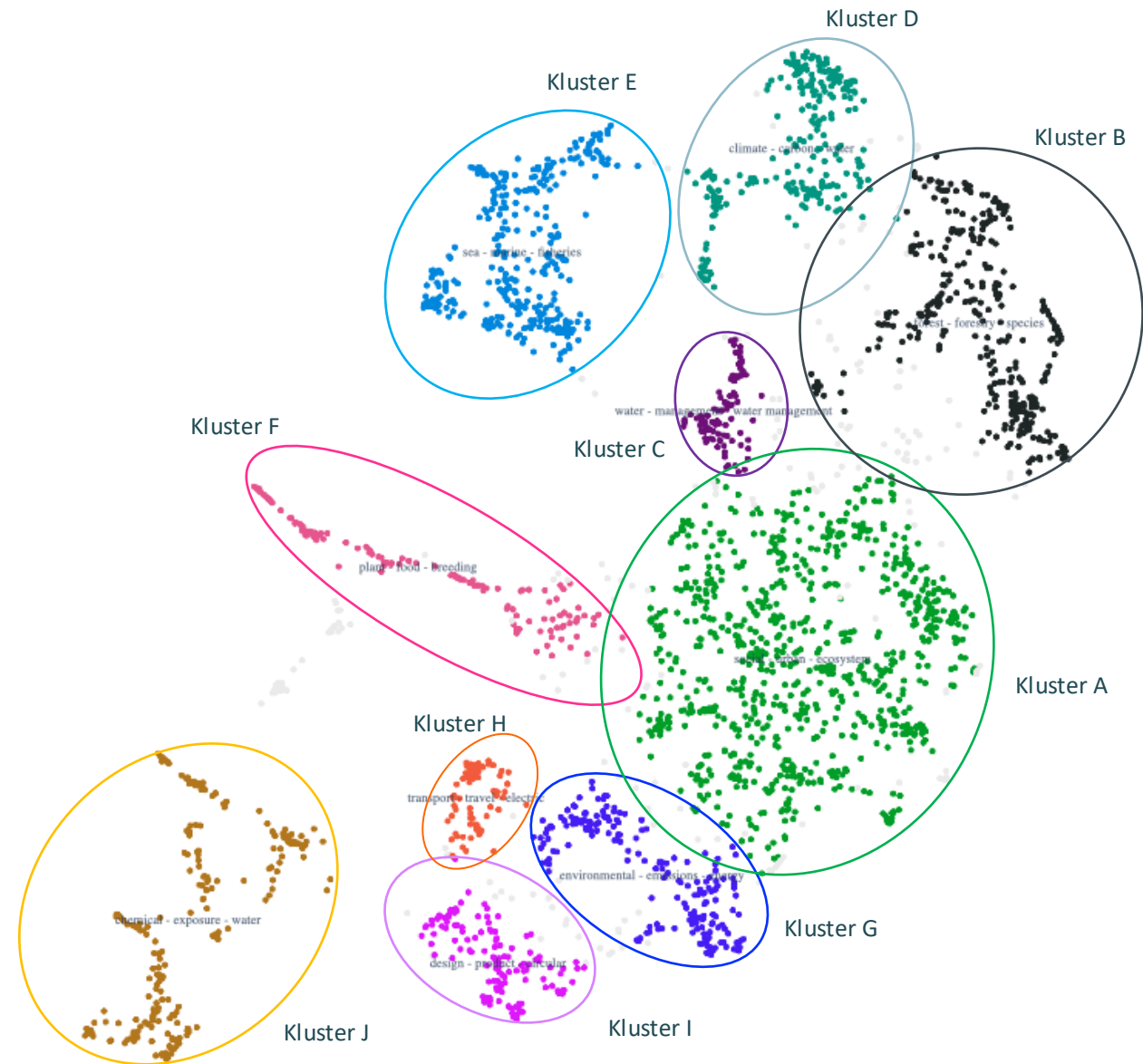
Klusterområden

Klusteranalys är en statistisk metod som används för att hitta likheter mellan punkter i data, där mer liknande punkter grupperas ihop. Gruppering/kluster kan användas för att kategorisera och förklara data.

Här används metoden för att ge ett nytt perspektiv på den av Mistra finansierade forskningen. Publikationerna under den studerade perioden har grupperats i kluster baserat på likheter i artikteltitlar och innehåll i artiklarnas abstracts. Publikationer med tematiskt liknande titlar och abstract kommer således att grupperas nära varandra.

Publikationerna delas upp i tio huvudsakliga kluster varav kluster A är överlägset störst. De flesta klustren är separerade och kompakta vilket tyder på hög grad av likhet inom klustren sett till innehåll i abstract och titelnamn.

Utöver de tio klustren finns en grupp som är icke-kategoriserad. Den icke-kategoriserade gruppen utgörs av publikationer som inte kan kategoriseras till någon av de tio etablerade klustren. Dessa publikationer är således för tematiskt annorlunda för att kunna klassificeras in i något av de tio klustren.



Topp 10 mest frekvent förekommande ord inom respektive kluster

A	B	C	D	E	F	G	H	I	J
Social	Forest	Water	Climate	Sea	Plant	Environmental	Transport	Design	Chemical
Urban	Forestry	Management	Carbon	Marine	Food	Emissions	Travel	Product	Exposure
Ecosystem	Species	Water management	Water	Fisheries	Breeding	Energy	Electric	Circular	Water
Ecological	Sweden	Use	Model	Management	Aphid	Carbon	Vehicles	Life	Surface
Change	Management	Water supply	Soil	Baltic	Plants	Corporate	Mobility	Life cycle	Acid
Development	Forests	Governance	Stream	Ecosystem	Potato	Policy	Car	Cycle	Drug
Environmental	Tree	Crop	Boreal	Baltic sea	Genetic	Firms	Charging	Circular economy	Assessment
Management	Climate	Irrigation	Catchment	Fish	Gene	Performance	Vehicle	Products	Activity
Governance	Change	Supply	Models	Aquaculture	Resistance	Financial	Accessibility	Economy	High
Sustainability	Forest management	Assessment	Temperature	Ocean	Genome	Market	Public	Business	Properties



Klusteruppdelning och vilka forskningsprogram som utgör störst andel av respektive kluster

Kluster A

Stockholm Resilience Centre – 53%
 Mistra Urban Futures – 15%
 Misum – 7%
 Övrigt – 25%
Publikationer totalt – 829

Kluster B

Mistra Future Forests – 69%
 Stockholm Resilience Centre – 17%
 Mistra-SWECIA – 6%
 Övrigt – 8%
Publikationer totalt – 334

Kluster C

Stockholm Resilience Centre – 78%
 Mistra InfraMaint – 12%
 Misum – 2%
 Övrigt – 8%
Publikationer totalt – 89

Kluster D

Mistra Future Forests – 43%
 Mistra –Swecia – 29 %
 Stockholm Resilience Centre – 26%
 Övrigt – 2%
Publikationer totalt – 249

Kluster E

Stockholm Resilience Centre – 88%
 AquaAgri – 5%
 Mistra-SWECIA – 1%
 Övrigt – 6%
Publikationer totalt – 313

Kluster F

Mistra Biotech – 60 %
 PlantComMistra – 17%
 Stockholm Resilience Centre – 7%
 Övrigt – 16%
Publikationer totalt – 139

Kluster G

Misum – 33%
 Sirp – 20%
 Mistra Carbon Exit – 12%
 Övrigt – 35%
Publikationer totalt – 197

Kluster H

Mistra Sams – 30%
 Mistra Carbon Exit – 26%
 Mistra InfraMaint – 12%
 Övrigt – 32%
Publikationer totalt – 77

Kluster I

Mistra REES – 44%
 Mistra Future Fashion – 17%
 Mistra STEPS – 9%
 Övrigt – 30%
Publikationer totalt – 139

Kluster J

MistraPharma – 23%
 Mistra SafeChem – 21%
 Mistra STEPS – 15%
 Övrigt – 41%
Publikationer totalt – 258

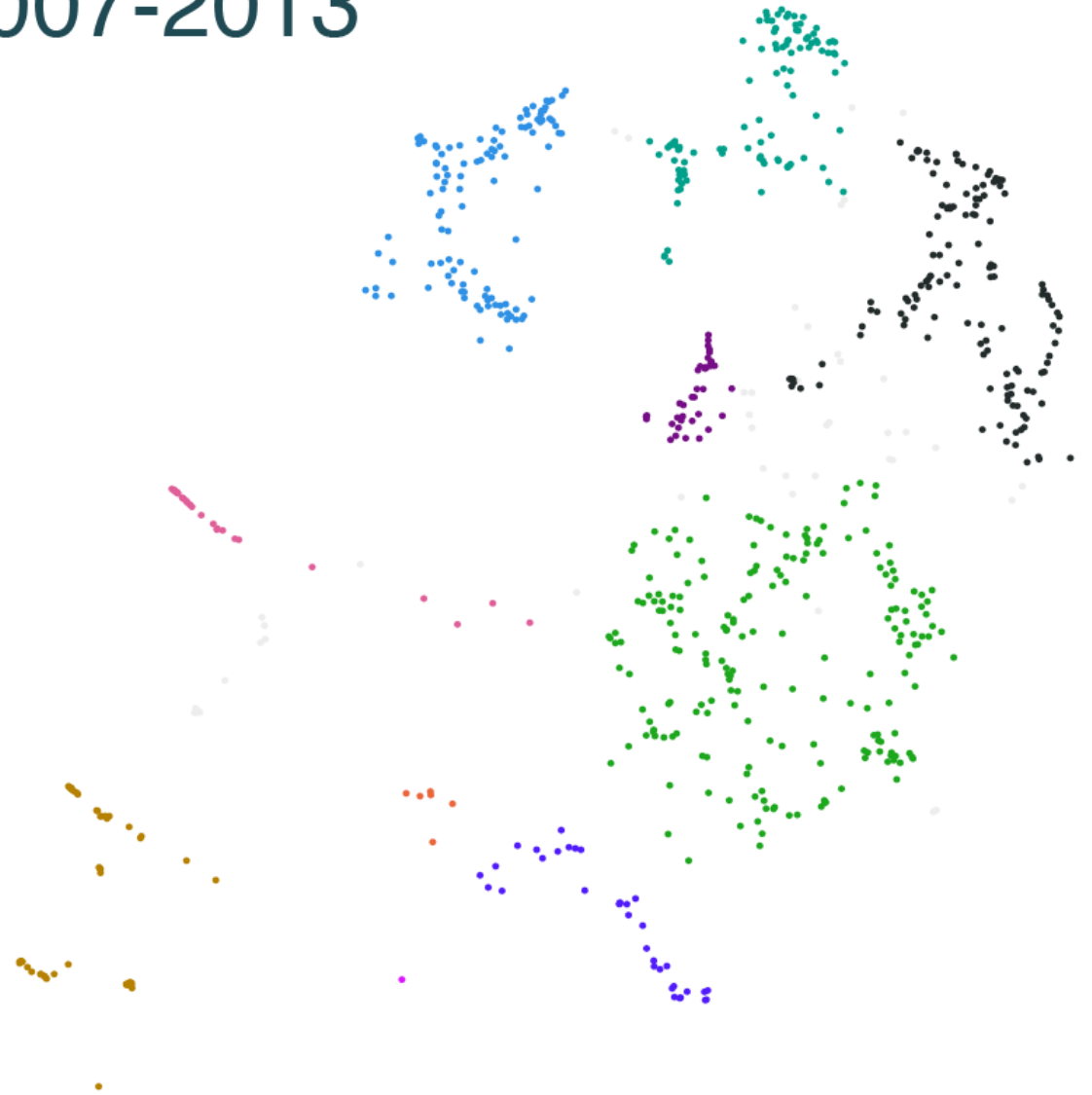


Utveckling över tid

Till höger visas en animation av klustren över tid, uppdelad i perioderna 2007-2013, 2014-2017 och 2018-2024. Perioderna är valda för att säkerställa att det finns ungefär lika många publikationer i varje period.

- De tematiska områden förändras över tid. Under perioden 2007-2013 var de största klustren A, B, C, D och E.
- Mellan 2014-2017 växer de tidigare nämnda klustren, samtidigt som nya tematiska områden framträder inom klustren F, G, H, I och J.
- I perioden 2018-2023 har de flesta klustren som dominerade perioden 2007-2013 försvunnit och de nya områden som uppstod 2014-2017 intar nu en dominerande position. Enbart kluster A behåller en konsekvent storlek genom hela perioden.
- Utvecklingen över tid speglar i hög grad de tematiska inriktningarna i de olika satsningarna.

2007-2013

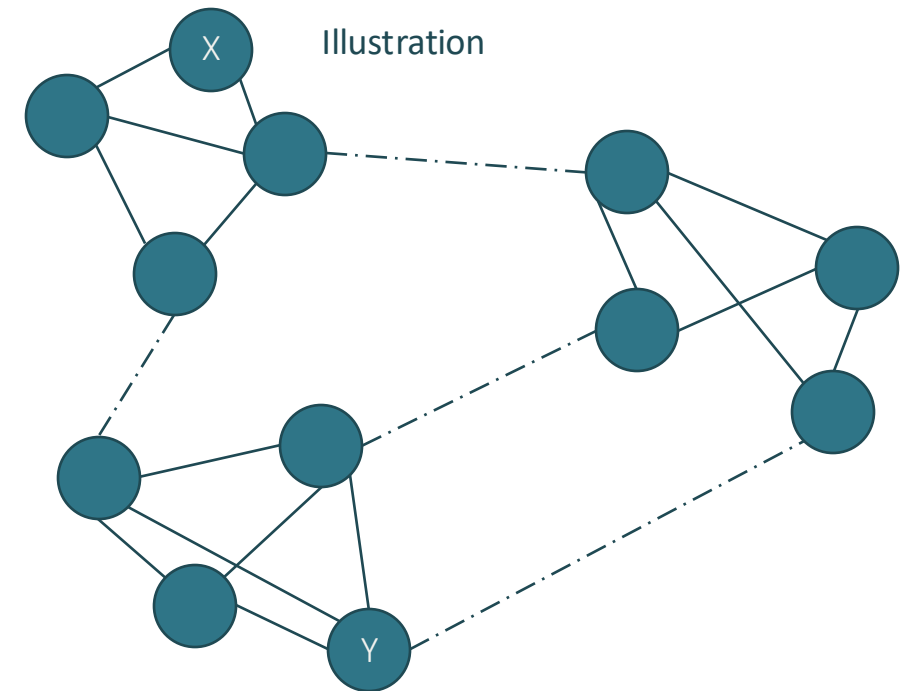


Metodbeskrivning för klusteranalys

Klusteranalysen ämnar att gruppera publikationer utifrån semantiska liknelser i publikationernas titlar och abstrakt. Detta görs genom att använda en språkmodell för att hitta tematiskt liknande publikationer. Publikationer anses till olika hög grad vara närbesläktade beroende på liknelser i deras titlar och abstrakt. Klustertillhörigheten bestäms genom att estimeras sannolikheten att en publikation tillhör ett specifikt kluster. En grupp publikationer som är semantiskt lika kommer således att grupperas tätt.

Beroende på de parametrar som används för respektive steg i modellen tillåts lösare och lösare kopplingar för att skapa ett kluster. Det kan därför också vara ett större antal steg mellan två publikationer trots att de tillhör samma kluster. I illustrationen till höger exemplifieras detta genom att tre grupper av publikationer (det kan exempelvis vara tre olika författare som har skrivit fyra publikationer var) grupperas till ett enda kluster genom de gruppöverskridande kopplingarna (streckade linjer). Det är nu fyra steg mellan exempelvis publikation X och publikation Y men de tillhör fortfarande samma kluster.

En mer teknisk beskrivning av klustringsmetoden följer på följande två sidor.



Metodbeskrivning för klusteranalys

Vi skapar kluster genom att analysera semantiska liknelser mellan publikationernas titlar och abstract. Processen sker i följande fem steg:

1. Databearbetning

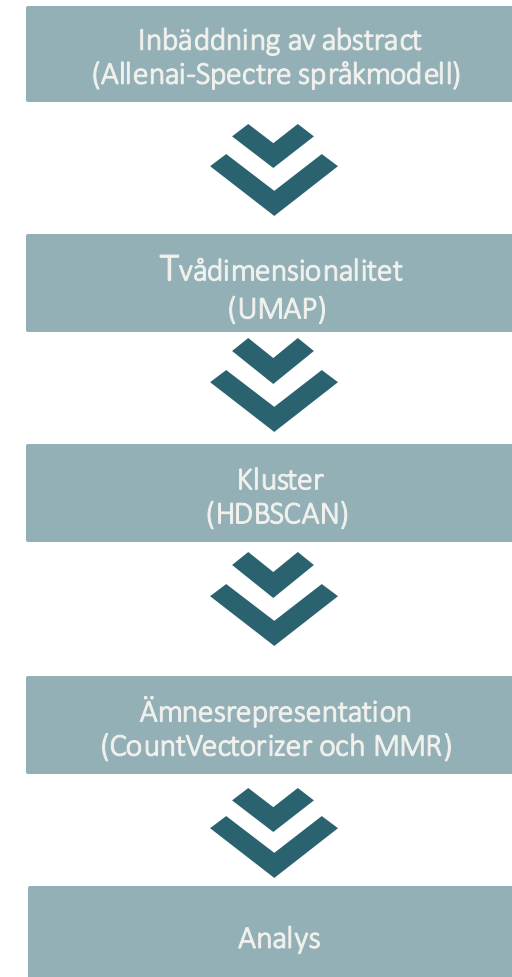
Datamaterialet består av publikationerna (artiklar och reviews) från samtliga forskningsprogram och centrubildningar. Vi sammanställer dessa publikationer till ett dataset som utgör grunden för klusteranalysen. Data är berikat med variabler såsom abstract och titel från publikationsdatabasen Scopus.

2. Språkmodellen och gruppering av publikationer på basen av titel och abstract

För att gruppera publikationerna använder vi oss av språkmodellen "Allenai-Spectre." Allenai-Spectre är en variant av en språkmodell baserad på Bidirectional Encoder Representations from Transformers (BERT), som är specifikt tränad på vetenskapliga publikationer för att identifiera likheter mellan publikationer. BERT grundar sig på transformerarkitekturen för maskininlärning och utgör grunden för modern Natural Language Processing (NLP). Praktiskt innebär detta att varje publikation representeras av en vektor i en n-dimensionell rymd. Värdet av vektorn för varje publikation skapas utifrån dess titel och abstract. Publikationerna kommer att placeras med olika avstånd från varandra beroende på likheten mellan dem. Denna inbäddning möjliggör således jämförelser av hur lika publikationerna är baserat på vad som framkommer av abstract och titel.

3. Uniform Manifold Approximation and Projection (UMAP) för dimensionsreducering

För att visualisera inbäddningarna och presentera klustren reduceras publikationsnätverket till ett tvådimensionellt plan. För detta ändamål tillämpar vi UMAP. UMAP är en icke-linjär dimensionsreduceringsteknik som används för att skapa en tvådimensionell representation av inbäddningarna samtidigt som dess ursprungliga struktur och egenskaper bevaras.



Metodbeskrivning för klusteranalys

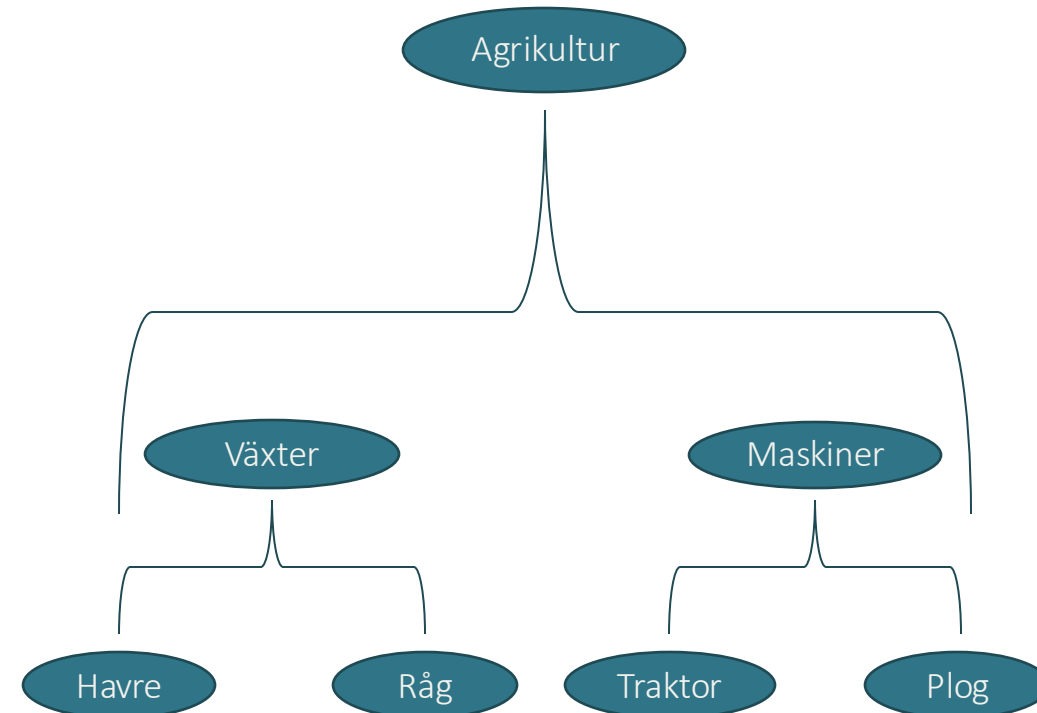
4. HDBSCAN för klustring av publikationerna

HDBSCAN används för att framställa klustren på ett tvådimensionellt plan. HDBSCAN är en hierarkisk och densitetsbaserad klustermetod. HDBSCAN fungerar genom att kategorisera publikationer i kluster baserat på avståndet mellan dem, där avståndet även används för att skapa hierarki mellan publikationerna. Alla publikationer har en lägstanivåshierarki med den publikation som den är mest lik. Den resulterande hierarkin används sedan för att bilda kluster på den näst högsta nivån. T.ex. om två publikationer bildar en hierarki så kommer den att paras ihop med en annan hierarki av två publikationer och på det sättet bilda nästa steg av hierarkin (illustrerat i figuren till vänster). Klustertillhörigheten bestäms utifrån ett minimumantal publikationer som behövs för att bilda ett kluster samt utifrån ett minimumantal "grannar" en publikation har och hur långt ifrån en "granne" får vara. Publikationer som möter detta minimum benämns som kärnpunkter. Kärnpunkter som är kopplade till varandra bildar ett kluster. Publikationer som inte är kärnpunkter men är placerade inom en kärnpunkts radie (dvs en "granne" till en kärnpunkt) kommer att tillhöra samma kluster som kärnpunkten, men publikationer som ligger i deras radie och inte är kopplade till någon kärnpunkt kommer inte att tillhöra klustret. Publikationer som inte tillhör något kluster kommer att tillhöra ett brus-kluster, där alla publikationer som inte kan kategoriserat till något av de andra tematiska områden placeras. Förutom kärnpunkter kommer också publikationerna bilda en hierarki av liknelser till varandra.

5. Ämnesrepresentation

CountVectorizer används för att skapa en "Document-Term Matrix" (DTM), vilket är en matris som omfattar alla ord som förekommer inom ett kluster. Raderna i DTM representerar publikationerna, medan kolumnerna representerar samtliga ord som används i klustren. För att effektivisera beräkningstiden och utesluta irrelevanta ord tillämpas en omfattande lista över "stop-words" från Natural Language Toolkit (NLTK). Denna lista inkluderar exempelvis ord såsom "at," "some," "the," osv. MMR används för att balansera förhållandet mellan relevans och mångfald av ord inom ett kluster. Målet är att optimera representationen med hänsyn till övriga abstrakt i klustret.

Hierarkisk klusteranalys





where it all adds up

Copenhagen

Overgaden Oven Vandet 58A
1415 Copenhagen
Denmark

Stockholm

Götgatan 22A
118 46 Stockholm
Sweden

Amsterdam

De Ruijterkade 7
1013 AA Amsterdam
The Netherlands

Contact

Daniel Ekström
Country lead
+46 7 39853575
daniel@adc-consulting.com