

Rutiner för överföring av data till Datavårdskapet för Kust och Hav

Dokumentrevideringar

Utgåva	Datum	Kommentar

Innehåll

1	Basfakta	3
1.1	Syfte med dokumentet	3
1.2	Definitioner	3
1.3	Bilagor	4
2	Leverans av data till den nationella datavärden	4
2.1	Varför leverera data till datavärden?	4
2.2	Vad kan rapporteras?	5
2.3	Vem levererar data?	5
2.4	När kan en dataleverans göras?	5
2.5	Hur rapporteras data?	5
3	Detaljerad information om dataleveranser	6
3.1	Rapporteringsmallar	6
3.2	Hur data formateras enligt rapporteringsmallen	6
3.3	Kvalitetssäkring av data	7
3.4	Kvalitetsmärkning av data	7
3.5	Mätgränser	8
3.6	Avrundningar	8
3.7	Kommentarer	8
3.8	Profiler från CTD	8
3.9	Metodbeskrivningar	8
4	Standardisering	9
4.1	Laboratorier (LABO: RLABO, ALABO, SLABO, ORDERER)	9
4.2	Projekt / mätprogram / undersökning (PROJ)	9
4.3	Arter	10
4.4	Storleksklasser	10
4.5	Fartyg eller provtagningsplattform (SHIPC)	10
4.6	Stationer	11

1 Basfakta

1.1 Syfte med dokumentet

SMHI driver på Naturvårdsverkets uppdrag Datavärdskapet för Kust och Hav, programområde Oceanografi och marinbiologi. I uppdraget ingår att samla in, kvalitetsgranska, lagra och tillhandahålla oceanografiska och marinbiologiska data insamlade inom nationella och regionala miljöövervakningsprogram.

Syftet med detta dokument är att i korthet beskriva hur främst regionala data överförs från en utförare till datavärden. Denna information är tänkt att underlätta en dataleverantörs formatering och leverans av data till datavärden.

Kraven på datavärden ändras ofta. För att leva upp till nya kravbilder förändras även rapporteringsmallar, kodlistor och manualer. Använd därför alltid den senaste version som du hittar på nätet: www.smhi.se och klicka vidare via Datavärdskap Oceanografi och marinbiologi.

1.2 Definitioner

Beställare Den organisation som skriver utförarkontrakt med provtagande/analyserande laboratorium.

Utförare Den organisation som samlar in prover (kod: *SLABO*) och som analyserar prover (kod: *ALABO*) enligt kontrakt med beställare.

Dataleverantör Den som slutligt ansvarar för leverans av data till datavärd. För regionala data är detta beställaren och för nationella data är detta utföraren.

Datavärd Nationell datavärd för oceanografi och marinbiologi, SMHI, som tar emot data från en dataleverantör och sedan granskar, lagrar och presenterar data på webben.

SHARK Förkortning för *Svenskt HavsARKiv*. Detta är det datalager som SMHI håller för marina data, och där ingår datavärdskapet som en del.

Nationella data Data där Naturvårdsverket är beställare.

Regionala data Data där Länsstyrelse eller Kommun är beställare. Inkluderar även data från Vattenvårdsförbund.

ICES Internationella Havsforskningsrådet. Hit levererar datavärden data som lagras och används på uppdrag av Helsingforskommissionen (HELCOM) och Oslo-Pariskonventionen (OSPAR).

EEA European Environmental Agency, Europeiska miljöbyrån. Sverige har skyldighet att rapportera metadata om miljöövervakningen till EEA. Datavärden bidrar till rapporteringen.

Datatyp En gruppering av data som bygger vidare på den indelning som ICES använder. För närvarande hanteras följande datatyper: makrofauna, växtplankton, djurplankton, vegetationsklädda bottenar, primärproduktion, bakterioplankton, sedimentation, temperatur-salthalt, närsaltskoncentrationer, Secchi-djup, alkalinitet, pH, humus, partikulärt material, organiskt kol, syre och svavelvätekoncentrationer.

Format Alla datatyper samlar inte in samma sorts kringdata och mätdata. Därför struktureras informationen på olika sätt, dvs. vissa fält och koder är unika för en del datatyper. Hur en viss typ av data struktureras och rapporteras till datavärden beskrivs av formatet.

Rapporteringsmall För att underlätta formateringen av data till ett visst format, så har datavärden skapat specifika rapporteringsmallar för de olika formaten.

Kodlista För att kunna jämföra provtagningar från olika dataleverantörer krävs i vissa fall klassning av data eller kringinformation till ett begränsat antal bestämda värden. Dessa förkortas ofta till koder som presenteras i en kodlista. Den senaste kodlistan finns på webben:

http://www.smhi.se/oceanografi/oce_info_data/shark_web/codelist.html

men koderna presenteras även på en flik i rapporteringsmallarna. Om olika data formateras enligt samma koder ökar chanserna att kunna jämföra data från olika platser.

Station En station (provtagningsstation, lokal) definieras av ett unikt namn och en unik position. Beställaren beslutar om stationens position och nominella namn. Datavärden rapporterar metadata om nationella stationer till meta-databasen VISS (VattenInformationsSystem Sverige).

1.3 Bilagor

Bilaga	Dokumentnamn	Utgåva, datum
1	Checklista för lyckad dataleverans	Version 2009-02-16

2 Leverans av data till den nationella datavärden

2.1 Varför leverera data till datavärden?

Genom att data levereras till datavärden kommer data att lagras (säkras för framtiden) och göras tillgängliga (utsökningsbara) på webben tillsammans med en stor mängd liknande data från hela Sverige. Genom att lagra och presentera data från ett stort antal utförare hos en gemensam datavärd ökar chanserna för en användare att finna de data som behövs för att analysera miljötillståndet i havet.

2.2 Vad kan rapporteras?

Datavärden för oceanografiska och marinbiologiska data hanterar följande typer av data: djurplankton, växtplankton, bakterioplankton, mjukbottenfauna (bottenhugg), epibenthos (vegetationsklädda bottenar, makrovegetation), temperatur och salthalt (inklusive CTD-data), siktdjup, pH och alkalinitet, svavelväteförekomst, primärproduktion, sedimentationshastighet samt koncentrationer av närsalter, klorofyll, syre, partikulärt kol + kväve och löst organiskt kol. Även sälldata kommer framöver att hanteras.

Se rapporteringsmallen för respektive dataformat för att se vilka fält och parametrar som datavärden tar emot:

http://www.smhi.se/oceanografi/oce_info_data/shark_web/how_to_report.html

Observera att all information som levereras till datavärden kommer att göras publik på webben. Således skall endast sådan information som dataleverantören äger och vill göra offentlig levereras, samt sådan information som dataleverantören på uppdrag av en beställare skall rapporteras till datavärden levereras. Det ankommer på dataleverantören att kontrollera att data som levereras till datavärden får publiceras på webben.

2.3 Vem levererar data?

Inom datavärdskapet hanteras både nationella data där utföraren finansieras direkt av Naturvårdsverket och regionala data där utföraren finansieras av en Länsstyrelse (även om Naturvårdsverket indirekt kan finansiera vissa mätningar), Kommun eller Vattenvårdsförbund.

Datavärden önskar att regionala data levereras av den som beställt data av en utförare. I vissa fall kan data levereras direkt från utföraren, men i egenskap av beställare av undersökningen kommer datavärden att be beställaren om ett sista godkännande av dataleveransen.

Data från nationell miljöövervakning levereras enligt avtal uppsatta av Naturvårdsverket.

2.4 När kan en dataleverans göras?

Data bör överföras regelbundet (månadsvis, kvartalsvis eller årsvis). Leverantörer av data bör kontakta datavärden för att komma överens om lämplig frekvens och tid på året.

Data som uppfyller kraven i check-listan (bilaga 1) kan levereras till datavärden när som helst. Kontakta datavärden i god tid om det är viktigt att data snabbt databasläggs så att den nås från webben. Detta sker normalt inom en vecka efter det att en dataleverans passerat datavärdens granskning.

2.5 Hur rapporteras data?

Gör så här för att sammanställa och leverera data till datavärden. Gå till webbsidan http://www.smhi.se/oceanografi/oce_info_data/shark_web/how_to_report.html

1. Sök upp det dataformat som passar dina data och ladda ner rapporteringsmallen.

-
2. Formatera data enligt instruktionerna på webbsidan och i rapporteringsmallen.
 3. Kontrollera data enligt checklistan (bilaga 1).
 4. Sänd formaterade data tillsammans med information om dataleveransen (en "fraktsedel" som bl.a. innehåller information om vad dataleveransen innehåller och vem som är kontaktperson) till den person som anges för det aktuella dataformatet på webbsidan.
 5. Datavärden bekräftar att data tagits emot.
 6. Datavärden granskar data och sammanställer eventuella frågor till dataleverantören.
 7. När datavärden granskat data och fått svar på ev. frågor sammanställs data till en fil som importerar till datavärdens databas. Detta gör data synliga och nerladdningsbara från webben.

Om någon finner fel i data, så rapporteras dessa till dataleverantören (ofta genom datavärden). Dataleverantören kan då korrigera felen och leverera data som en ny fil med ett nytt versionsnamn och ny fraktsedels-information. Alla datafiler arkiveras hos datavärden. När den nya filen levererats till datavärden kan datavärden radera den felaktiga filen från databasen och istället importera den upprättade datafilen.

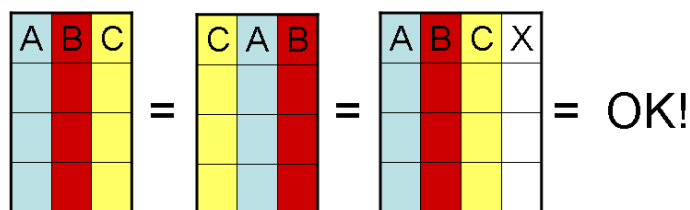
3 Detaljerad information om dataleveranser

3.1 Rapporteringsmallar

Data från nationell miljöövervakning ska rapporteras vidare till internationella organ (HELCOM och OSPAR genom ICES, metadata till EEA genom VISS). Därför måste dessa uppfylla de krav som ICES ställer och datavärdens databas är till stor del uppbyggd efter kraven att kunna rapportera till ICES. Eftersom samma krav inte gäller regionala data, så presenteras i många fall en förenklad version av rapporteringsmallen, men det går bra att använda rapporteringsmallen för nationella data även för regionala data.

3.2 Hur data formateras enligt rapporteringsmallen

Ställ upp data kolumnvis enligt rapporteringsmallen. Observera: ordningen på kolumnerna är utan betydelse, men varje kolumns rubriknamn måste anges enligt rapporteringsmallen (se figur nedan). Det går också att leverera kolumner som inte datavärden frågar efter och inte kommer att lagras i databasen eller visa upp på webben, men som kommer att arkiveras och kunna laddas ner som primärdata. För att sådan information ska vara användbar är det viktigt att kolumnhuvudet tydligt talar om vad kolumnen innehåller.



Figuren visar schematiskt att ordningen på olika kolumner inte spelar någon roll så länge fältnamnet anges enligt rapporteringsmallarna (fältkolumner A, B, C i figuren). Även information som datavärden idag inte hanterar kan rapporteras (fältkolumn X i figuren).

Viss information kan saknas hos dataleverantören (i synnerhet för historiska data), men data kan ändå vara av värde. Fyll i de fält där det finns information och lämna övriga tomma.

Om du önskar leverera information som inte hanteras i rapporteringsmallen men som bör presenteras på webben bör du kontrollera om du kan hitta något datafält inom något annan datatyp. Kontakta annars datavärden.

Kontakta datavärden om data lagras hos dataleverantören i ett format som inte lätt kan anpassas till datavärdens rapporteringsmall. Datavärden formaterar inte data, men om det rör sig om mycket stora datamängder formaterade på annat sätt (t.ex. som flera nycklade filer eller XML-format) så kan datavärden möjligtvis skapa ett unikt inläsnings-skript för dessa specifika data.

3.3 Kvalitetssäkring av data

De data som levereras från olika myndigheter, konsulter och/eller vattenvårdsförbund förutsätts vara kvalitetssäkrade och granskade, först och främst av analyserande laboratorium, men även av berörda tillsynsmyndigheter. Vid överföring till datavärden kommer ytterligare en kvalitetsgranskning att genomföras innan materialet görs tillgängligt på webben. Denna granskning omfattar för fysikalisk-kemiska data bl.a. en rimlighetsanalys och jämförelse med tidigare undersökningsresultat. Inga ändringar av materialet kommer dock att ske utan samråd med dataleverantören. För biologiska data görs än så länge inga automatiska rimlighetskontroller utan endast manuell kontroll.

3.4 Kvalitetsmärkning av data

Data skall efter kvalitetskontroll märkas enligt tabellen nedan. Kvalitetsflaggan noteras i rapporteringsmallen för respektive parameter. Vissa märkningar kräver en kommentar i rapporteringsmallens fält COMNT_PARAM. Övrig märkning som utförs av provtagande, analyserande eller rapporterande laboratorium skall om möjligt anges i en separat kolumn eller införas i kommentarsfälten.

Beskrivning	Q-Flagga	Kommentar
	QFLAG	COMNT_PARAM
Ingen kvalitetskontroll utförd	Blank	Valfritt
Accepterat värde. Data kontrollerad och OK	A	Valfritt
Dåligt/ felaktigt värde	B	Obligatorisk

Värdet under detektionsområdet	<	Valfritt
Värdet överstiger detektionsområdet	>	Valfritt
Värde saknas	M	Valfritt
Misstänkt värde - anses misstänkt baserat på kvalitetskontroll	S	Obligatorisk
Misstänkt extremvärde som kontrollerats och funnits korrekt.	E	Obligatorisk
Uppskattat värde (genom interpolation eller extrapolation)	R	Obligatorisk

3.5 Mätgränser

Observationer som är lägre (eller högre) än eventuella detektionsgränser skall inte bytas ut mot något subjektivt valt värde. I stället skall "<"- eller ">"-tecken anges vid leveransen. Det är användaren som måste besluta hur sådana observationer skall hanteras i analysen.

Datavärden håller på att utveckla hanteringen av mätgränser och sådan information kommer att efterfrågas framöver.

3.6 Avrundningar

Data bör bestå av ett rimligt antal värdesiffror. Omotiverade avrundningar bör däremot inte ske. Avrundningar bör ske vid dataanalys och inte på grunddata. För fysikalisk-kemiska data anges i rapporteringsmallen antalet värdesiffror för respektive parameter som datavärden tar emot.

3.7 Kommentarer

Det finns specifika fält i rapporteringsmallen för kommentarer som rör mätvärden, provtagning etc. Vissa kvalitetsflaggor kräver en kommentar. För fysikalisk-kemiska data måste parameter och djup anges för varje kommentar.

3.8 Profiler från CTD

Data insamlade med s.k. CTD-sond i en profil mellan ytan och botten kan levereras till datavärden antingen som enskilda värden för respektive standarddjup eller så kan data levereras som en separat fil för varje CTD-profil. Samråd med datavärden om lämpligt filformat, grad av upplösning och eventuell förbehandling av CTD-data.

3.9 Metodbeskrivningar

Data utan beskrivning av hur data samlats in och analyserats är i det närmaste värdelösa. Rapporteringsmallarna har fält där referenser till metodbeskrivningar anges. Datavärden måste lagra alla metodbeskrivningarna och kommer att begära en kopia av de metodbeskrivningar som inte redan lagras hos datavärden.

4 Standardisering

En viktig del av datavärdens arbete är att försöka samordna olika begrepp som används för att beskriva samma sak. Ett sätt att göra detta på är att ge ett begränsat antal alternativ för att beskriva något. Beskrivningen kan vara längre namn (t.ex. för arter och stationer) eller kortare koder. Datavärden anpassar sig om möjligt till koder som används internationellt. ICES använder sig ofta av kodbeteckningar för olika begrepp, och datavärden använder sig till stor del av samma koder men skapar även nya vid behov.

4.1 Laboratorier (LABO: RLABO, ALABO, SLABO, ORDERER)

Det är viktigt för en dataanvändare att känna till vem som levererat data (*RLABO* = reporting laboratory), vem som samlat in provet (*SLABO* = sampling laboratory) och vem som analyserat provet (*ALABO* = analyzing laboratory). Därför måste dessa uppgifter rapporteras in till datavärden som i sin tur kan publicera dem för användare på webben. Även koder för beställare (*ORDERER*) hanteras här. Den gemensamma förkortningen på laboratoriekoderna *RLABO*, *SLABO* och *ALABO* och *ORDERER* är *LABO*.

ICES kodlista innehåller många internationella dataleverantörers laboratorie-koder. Detta är en upp till sex tecken lång förkortning av laboratoriets namn. Datavärden följer denna standard och adderar vid behov *LABO*-koder till sin lista.

Hur rapporteras data med laboratoriekod?

- a) Om laboratoriet/ organisationen finns med i kodlistan för *LABO*: använd koden när data registreras i rapporteringsmallen och rapportera enligt rapporteringsmallen.
- b) Om laboratoriet inte finns i listan: kontakta datavärden som meddelar ny unik laboratoriekod och för in den i kodlistan. Datavärden kan i vissa fall behöva ytterligare information om laboratoriet. Rapportera sedan som vanligt.

4.2 Projekt / mätprogram / undersökning (PROJ)

Begreppen projekt, mätprogram och undersökning betecknar i en del fall ungefär samma sak. *PROJ* kan vara ett regionalt kustkontrollprogram hos ett vattenvårdsförbund, eller mätprojekt/kampanj som är begränsat i tiden eller omfattar ett begränsat område. För att kunna hämta ut samlade data från ett specifikt projekt är det nödvändigt att märka varje provtagning med projektets namn. Eftersom vissa projekt har mycket långa namn, så gör datavärden om projektets namn till en mer lätthanterlig kod. I kodlistan finns de projekt som någon gång rapporterat data till datavärden. Flera projekt kan rapporteras för samma prov.

Hur rapporteras data med projektkod (PROJ)?

Kontrollera om projektet redan finns med i kodlistan för *PROJ*.

- a) projektet finns med: använd koden i rapporteringsmallens fält *PROJ* och rapportera som vanligt.
- b) om projektet inte finns: skriv projektets namn i fältet *PROJ_NAME*, så kommer datavärden att skapa en ny unik *PROJ*-kod. Datavärden kan i vissa fall behöva ytterligare information om projektet för att kunna skapa koden.

4.3 Arter

För att undvika problem med bl.a. ogiltiga namn på arter eller att arter som rapporterats med synonyma namn inte går att söka ut i samma tidsserie, så går datavärden över till att använda en standardlista på gällande artnamn (i den här texten används begreppet "artnamn" men även andra taxonomiska nivåer än arter hanteras på samma sätt). Listan skapas årligen från information lagrad i ArtDatabankens databas DynamiskTAXA.

Datavärden lagrar det rapporterade artnamnet och organismens taxon-id enligt DynamiskTAXA. Taxon-id gör att en användare kan hämta alla data om en viss art, oberoende om den bytt namn flera gånger under den utsökta tidsperioden. Funktionen planeras att tas i drift våren 2009. Mer information om arthanteringen, samt senaste artlistan finns att hämta från sidan http://www.smhi.se/oceanografi/oce_info_data/oce_data.html

Hur rapporteras art-information?

1a) Om arten finns med i artlistan: ange arten när data registreras i rapporteringsmallen och rapportera som vanligt.

1b) Om arten inte finns med: spara om möjligt organismen eftersom det hänt att experter önskat granska nyfunna arter. Rapportera därefter som vanligt där du använder den nya artens namn. Datavärden låter vid behov expertis kommentera fyndet innan det registreras i ArtDatabankens databas DynamiskTAXA. I den processen kan datavärden behöva ställa frågor till dataleverantören. Därefter för datavärden upp arten på den nationella miljöövervakningsartlistan.

Observera att datavärden screenar inkommande data mot artlistan, så dataleverantören behöver inte nödvändigtvis jämföra att varje rapporterad art finns med i artlistan innan data levereras. Om datavärden finner att synonymer eller ogiltiga artnamn rapporterats, så frågar datavärden dataleverantören om dessa namn för att klargöra vilken art som åsyftas.

4.4 Storleksklasser

Storlek på växtplankton rapporteras enligt den indelning som gjorts av HELCOM och som kan laddas ner från <http://www.ices.dk/env/repfor/index.asp> med namnet PEG_BVOL[årtal].

4.5 Fartyg eller provtagningsplattform (SHIPC)

För vissa datatyper är det viktigt att registrera vilket fartyg som varit ute och tagit provet. Detta gäller främst data för den fria vattenmassan, där kombinationen av fartygskod, provtagningsår, serienummer ibland används för att identifiera vilka prover som tagits vid samma tillfälle. SMHI har sedan länge använt sig av koder för olika fartyg och mätplattformar för att koppla ihop olika prover.

Hur rapporteras data från ett fartyg eller annan mätplattform?

a) Om fartyget finns med i kodlistan för SHIPC: använd koden när data registreras i rapporteringsmallen och rapportera enligt rapporteringsmallen.

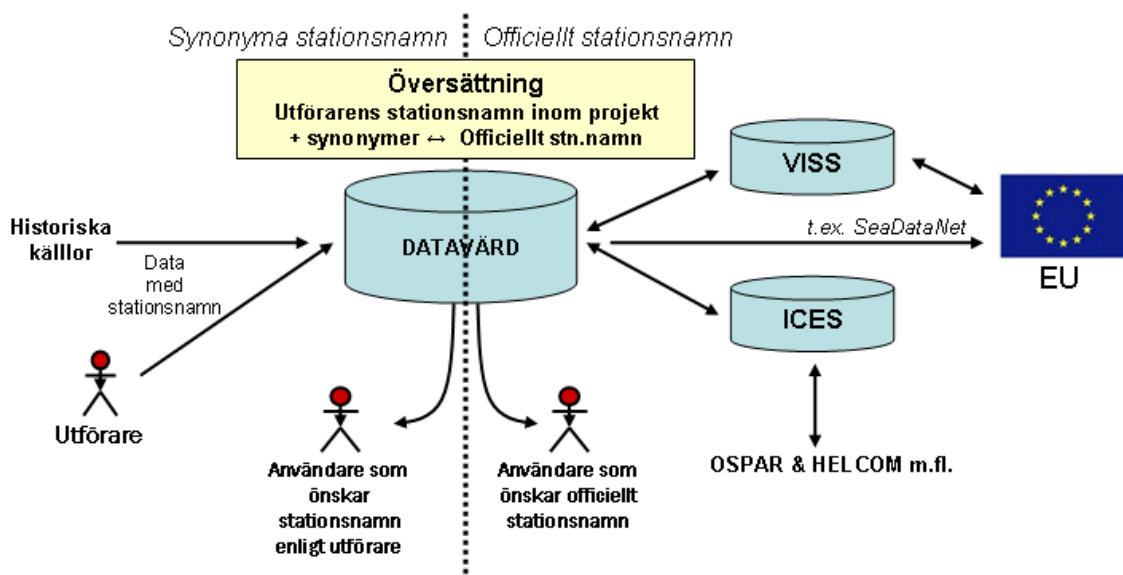
b) Om fartyget inte finns med: kontakta datavärden som meddelar ny unik fartygskod. Datavärden kan då behöva ytterligare information om fartyget. Leverera därefter data med den nya koden.

4.6 Stationer

Inom många marina provtagningsuppdrag används begreppet "station". Station används synonymt med begreppet provtagningslokal. En station definieras av en unik position och ett unikt nominellt stationsnamn. Den verkliga provtagningen kanske inte görs exakt på den nominella positionen, utan stationen kan ha en viss utbredning.

På nationell nivå råder i dag till viss del en begreppsförvirring kring miljöövervakningens stationsnamn. En station kan t.ex. av historiska skäl ha många synonyma namn (utföraren använder ett namn, beställaren ett annat, osv.). För andra stationer har man kanske aldrig definierat någon position, utan det finns en mängd provtagningar som gjorts på något olika platser, men som rapporterats med samma stationsnamn. Åter andra stationer har samma namn, fast de ligger på helt olika platser och används i olika undersökningar.

Ett nominellt stationsbegrepp behövs när en beställare önskar att insamling av prover görs vid en viss station, eller om man på en karta vill se från vilka stationer som det finns vissa typer av data. Internationellt används begreppet "station" på samma sätt, och vid svensk rapportering till EEA, ICES, HELCOM, OSPAR anges unika namn och positioner på stationer. Det är viktigt att samma stationsnamn används till samtliga organisationer vid internationell rapportering av svenska data. Om olika namn används vid rapporteringen till datavärden, så kan datavärden behöva översätta namnet till det internationellt gällande namnet. Det är beställaren som bestämmer vad en station heter och var den ligger.



Figuren ovan visar hur olika synonyma stationsnamn utåt måste översättas till samma namn för att externa organisationer ska kunna kommunicera.

När data lagras hos den nationella datavärden **måste namn och position vara nationellt unika för varje station**. Därför måste inrapporterade stationer i en del fall

justeras avseende namn. För att kunna hantera olika stationer håller datavärden på att skapa en central stationslista i databasen. Denna kommer att publiceras på webben. Till dess att den publicerats går det bra att rapportera de stationsnamn som används av beställaren, så kontrollerar datavärden dem mot en intern stationslista och återkommer vid behov med frågor.

Utifrån informationen som finns i stationslistan kan datavärden presentera stationsdata på webben på ett sätt som gör det lättare att överblicka och hämta hem data. Stationer som inte registrerats i stationslistan kan inte korrekt visas på detta sätt.

Det är värt att komma ihåg att alla typer av data inte använder sig av ett stationsbegrepp. Exempelvis används det inte i fallet där döda sälar rapporteras in. Där är endast positionen för fyndet av intresse.

CHECKLISTA FÖR LYCKAD DATALEVERANS:

Data har kvalitetsgranskats av dataleverantören.

Data har formaterats enligt aktuell rapporteringsmall som hämtats från webben. Gamla rapporteringsmallar skall inte användas eftersom mallarna kan ha ändrats av datavärden.

Kontakta datavärd om det saknas LABO-kod (för rapportering och beställande organisation, provtagande och analyserande laboratorium) eller andra koder såsom SHIPC-kod (för fartyg som inte finns med i kodlistan), så skapar datavärden en sådan kod.

Fraktsedel-informationen finns med (t.ex. i e-postmeddelande). Vad som ingår anges i varje rapporteringsmall.

Sänd data till datavärden enligt tabellen på webbsidan med rapporteringsmallar (grovt uppdelat: fysikalisk/kemiska data till jan.szaron@smhi.se och biologiska data till anders.olsson@smhi.se med kopia till lars.hansson@smhi.se).

Datavärden granskar data och återkommer med eventuella frågor. När data granskats och eventuellt kompletterats, så lagras data i databasen så att den visas på webben.