



Sveriges lantbruksuniversitet  
Swedish University of Agricultural Sciences

SLU-biblioteket

SLU ID: SLU.ua.2018.2.3.2.IÄ-1

# Hantering av datamängder, programvaror och rådata vid SLU

## – Enkät forsknings- och miljöanalysdata

Karl Pettersson, Mikaela Asplund, Olof Frank

SLU-biblioteket

SLU-bibliotekets rapportserie, 8

Uppsala 2018

## Hantering av datamängder, programvaror och rådata vid SLU – Enkät forsknings- och miljöanalysdata

Karl Pettersson            Sveriges lantbruksuniversitet, Data Curation Unit,  
karl.pettersson@slu.se

Mikaela Asplund            Sveriges lantbruksuniversitet, Data Curation Unit,  
mikaela.asplund@slu.se

Olof Frank                    Sveriges lantbruksuniversitet, Data Curation Unit,  
olof.frank@slu.se

**Utgivningsort:**            Uppsala

**Utgivningsår:**            2018

**Serietitel:**                SLU-bibliotekets rapportserie

**Delnummer i serien:**      8

**Elektronisk publicering:** <https://pub.epsilon.slu.se>

**Bibliografisk referens:** Pettersson, K., Asplund, M. & Frank, O. (2018). *Hantering av datamängder, programvaror och rådata vid SLU - Enkät forsknings- och miljöanalysdata*. Uppsala: Sveriges lantbruksuniversitet. (SLU-bibliotekets rapportserie, 8).

**Nyckelord:**                forskningsdata, miljöanalysdata, datamängder, datahantering, arkivering, öppen tillgång, öppna myndighetsdata, enkät

# Innehåll

|  |           |
|--|-----------|
| <b>Sammanfattning</b>                                    | <b>4</b>  |
| <b>Abstract</b>  | <b>4</b>  |
| <b>Om rapporten</b>                                      | <b>4</b>  |
| <b>1 Bakgrund</b>  | <b>4</b>  |
| <b>2 Fråga 1</b>   | <b>5</b>  |
| 2.1 Fråga 1a: Datatyper                                  | 5         |
| 2.2 Fråga 1b: Programvaror                               | 7         |
| 2.3 Fråga 1c: Format                                     | 13        |
| 2.4 Fråga 1d: Datamängder                                | 16        |
| 2.5 Fråga 1e: Känsliga data                              | 17        |
| 2.6 Fråga 1f: Arkivkrav                                  | 17        |
| 2.7 Fråga 1g: Externa samarbeten                         | 18        |
| 2.8 Fråga 1h: Versionshantering                          | 18        |
| <b>3 Fråga 2</b>   | <b>19</b> |
| 3.1 Fråga 2a: Erfarenheter av tillgängliggörande         | 19        |
| 3.2 Fråga 2b: Inställning till att öppet dela data       | 19        |
| 3.3 Fråga 2c: Tillgängliggörande                         | 21        |
| 3.4 Fråga 2d: Tillgängliggörande                         | 21        |
| 3.5 Fråga 2e: Tillgängliggörande                         | 23        |
| 3.6 Fråga 2f: Datahanteringsplan                         | 25        |
| <b>4 Fråga 3</b>   | <b>25</b> |
| 4.1 Fråga 3a: Institution                                | 25        |
| 4.2 Fråga 3b: Tillhörighet                               | 26        |
| 4.3 Fråga 3c: Tillhörighet                               | 27        |
| 4.4 Fråga 3d: Titel                                      | 27        |
| 4.5 Fråga 3e: Extern finansiering                        | 27        |
| 4.6 Fråga 3f: Övriga synpunkter                          | 27        |
| <b>5 Sammanfattande diskussion och behov av åtgärder</b> | <b>28</b> |
| 5.1 Fråga 1: datatyper, känsliga data och arkivering     | 28        |
| 5.2 Fråga 2: öppna data och delning av data              | 28        |
| <b>A Uppföljningsenkät</b>                               | <b>30</b> |
| A.1 Datamängder  | 30        |
| A.2 Fördelning rådata/bearbetade data                    | 30        |
| A.3 Kommentarer  | 30        |
| <b>B Enkäten</b>   | <b>31</b> |

## Sammanfattning

En enkät skickades till forskare vid Sveriges lantbruksuniversitet (SLU) med frågor rörande deras hantering av forskningsdata, exempelvis vilka datatyper och mjukvaror de använder och hur stora datavolymer de hanterar, deras kunskaper om juridiska aspekter på datahantering och deras attityder till att göra data öppet tillgängliga. I rapporten diskuteras svaren på enkätfrågorna och deras implikationer för framtida arbete med datakurering vid universitetet.

*Nyckelord:* forskningsdata, miljöanalysdata, datamängder, datahantering, arkivering, öppen tillgång, öppna myndighetsdata, enkät

## Abstract

A survey was sent to researchers at the Swedish University of Agricultural Sciences (SLU) concerning their research data management, e.g. data types and software products used and data volumes, their knowledge about legal aspects of data management, and their attitudes towards making data openly available. The report discusses the survey answers and their implications for future work with data curation at the university.

*Keywords:* research data, environmental data, data volumes, data management, archiving, open access, open government data, survey

## Om rapporten

Mikaela Asplund har gjort sammanställning av svaren på fråga 2b och 2c och skrivit bakgrundsavsnittet. Olof Frank har gjort sammanställning av svaren på fråga 2a och 2d. Tomas Lundén har gjort sammanställning av svaren på fråga 2e. Karl Pettersson har gjort sammanställning av svaren på fråga 1 och 3 och sammanställt rapporten som helhet.

## 1 Bakgrund

Den här rapporten är framtagen av Sekretariatet för arkivering och publicering av forskningsdata (på engelska Data Curation Unit, DCU) vid Sveriges lantbruksuniversitet som en del i arbetet med att utveckla Tilda, SLU:s nya system för publicering och arkivering av forsknings- och miljöanalysdata. Tilda ska fungera som den centrala platsen för att synliggöra och långtidsbevara data skapad inom forskning och miljöanalys på SLU.

Under utvecklingsprojektet och inför lanseringen av Tilda återstod flera frågor som projektgruppen inte kunde besvara på egen hand. Några av dessa frågor var vilka slags data som SLU:s forskare och miljöanalytiker framställer, bearbetar och analyserar avseende både storlek och dataformat. Likaså var det viktigt att inför lansering ta reda på de anställdas kunskap om och inställning till arkivering och publicering av data i syfte att kunna utforma lämpliga informations- och utbildningsinsatser.

Sekretariatet konstruerade en enkät (se bilaga B) som huvudsakligen bestod av två delar, en del som berörde datatyper, känsliga data och arkivering samt en del som var inriktad mot publicering av öppna data. Enkäten skickades ut i november 2017 via e-post till institutionerna vid SLU, och 1441 personer fick den. De flesta

frågorna har besvarats av ca 20 procent av respondenterna. Exakt antal svar beskrivs i respektive avsnitt. Den närmare sammansättningen hos de som besvarat enkäten, med avseende på befattning och institutionstillhörighet, beskrivs under avs. 4. I diskussionen kommer de omväxlande att benämnas "respondenterna" eller "forskarna", eftersom de som regel har någon forskar- eller doktorandtjänst.

## 2 Fråga 1

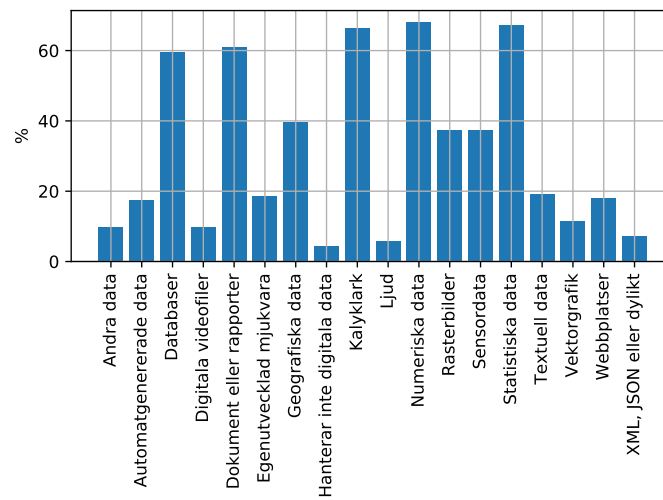
Svaren på fråga 1b och 1c, om mjukvaror och format, har normaliserats för att kunna bearbetas och redovisas med diagram. I anslutning till dessa ges en kort diskussion om de vanligaste svaren, med fokus på vad dessa innebär när det gäller förutsättningar för öppen vetenskap.

### 2.1 Fråga 1a: Datatyper

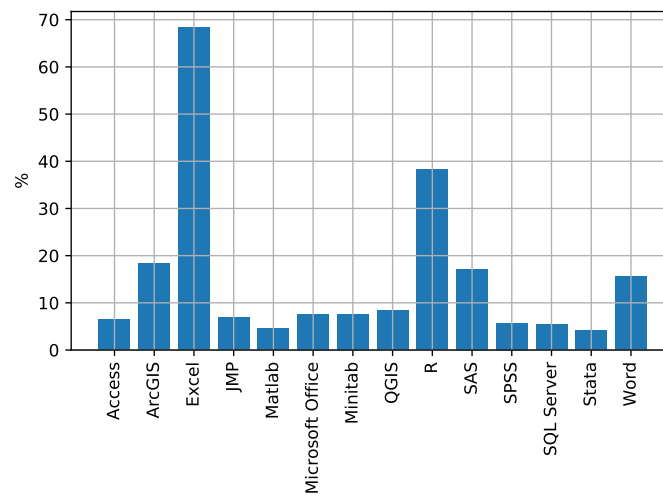
**Fråga** Vilka datatyper arbetar du med? Respondenterna gavs 16 alternativ för olika datatyper, plus alternativet "arbetar inte med digitala data" och "andra datatyper, ange gärna vilka", där de kunde specificera datatyp i fritextfält. Mer än en datatyp kunde anges.

Det är 278 som svarat. Fördelningen av svaren visas av fig. 1.

Majoriteten av de 27 fritextsvaren handlar om sekvensdata, DNA-sekvensdata eller liknande. Det var ett alternativ som inte fanns med, och det passar inte på något uppenbart sätt in bland de definierade kategorierna.



Figur 1: Andel forskare som uppgett olika datatyper av alternativen på fråga 1a.



Figur 2: Andel av alla som svarat på 1b som uppgett programvarorna på x-axeln, för program med minst 10 svar.

## 2.2 Fråga 1b: Programvaror

**Fråga** Vilka programvaror använder du i din datahantering?

Det har getts 262 svar på denna fritextfråga, där de flesta angett mer än en programvara. Efter att namnen på programvarorna normaliserats är det totalt 194 olika programvaror som rapporterats. En del respondenter har gett ospecifika svar, som "[o]lika analysprogram", "statistikprogram", "[b]ioinformatics tools" och liknande. Många programvaror har endast uppgetts av en respondent. Det är 14 programvaror som uppgetts av minst 10 respondenter var. Fördelningen av dessa visas av fig. 2. Samtliga programvaror som angetts visas i tab. 1, med korta kommentarer fokuserade på förutsättningarna att hantera öppna format, för de vanligare programvaror som visas i fig. 2.

Tabell 1: Uppgivna mjukvaror.

| Mjukvara                 | Antal | Frekvens (%) | Kommentar   |
|--------------------------|-------|--------------|---|
| Access                   | 17    | 6,5          | Proprietär relationsdatabas från Microsoft. Kan exportera data till SQL eller textformat. |
| Acrobat                  | 4     | 1,5          |   |
| Agisoft<br>photoscan     | 1     | 0,4          |   |
| Agisoft<br>photoscan pro | 1     | 0,4          |   |
| Aimms                    | 1     | 0,4          |   |
| Alba                     | 1     | 0,4          |   |
| Applications-<br>master  | 1     | 0,4          |   |
| ArcGIS                   | 48    | 18,3         | Proprietär mjukvara för geografiska informationssystem. Kan hantera data i öppna format.  |
| Arview                   | 2     | 0,8          |   |
| Ascii                    | 1     | 0,4          |   |
| Asreml                   | 2     | 0,8          |   |
| Athena                   | 1     | 0,4          |   |
| Balancelink              | 1     | 0,4          |   |
| Bash                     | 1     | 0,4          |   |
| Bin                      | 1     | 0,4          |   |
| Biorad                   | 1     | 0,4          |   |
| Boris                    | 1     | 0,4          |   |
| C                        | 2     | 0,8          |   |
| C++                      | 2     | 0,8          |   |
| Can-eye                  | 1     | 0,4          |   |
| Canoco                   | 5     | 1,9          |   |
| Ccpn                     | 1     | 0,4          |   |
| Chemstation              | 1     | 0,4          |   |
| Clc genomic              | 1     | 0,4          |   |
| Clc genomic<br>workbench | 1     | 0,4          |   |
| Clc main                 | 1     | 0,4          |   |
| Cloudcompa-<br>re        | 1     | 0,4          |   |

| Mjukvara           | Antal | Frekvens (%) | Kommentar   |
|--------------------|-------|--------------|---|
| Codoncode          | 1     | 0,4          |   |
| Creative suite     | 1     | 0,4          |   |
| Ctd-sond           | 1     | 0,4          |   |
| Dat                | 1     | 0,4          |   |
| Datagraph          | 1     | 0,4          |   |
| Datem              | 2     | 0,8          |   |
| summit             |       |              |   |
| Delphi             | 1     | 0,4          |   |
| Diamond 2.1e       | 1     | 0,4          |   |
| Distance           | 1     | 0,4          |   |
| Djview             | 1     | 0,4          |   |
| Dmu                | 1     | 0,4          |   |
| Dnastar            | 1     | 0,4          |   |
| Dropbox            | 1     | 0,4          |   |
| Ebi services       | 1     | 0,4          |   |
| Endnote            | 1     | 0,4          |   |
| Er60               | 1     | 0,4          |   |
| Erdas              | 1     | 0,4          |   |
| Erdas imagine      | 1     | 0,4          |   |
| Esrisk             | 1     | 0,4          |   |
| Eviews             | 1     | 0,4          |   |
| Exafspak           | 1     | 0,4          |   |
| Excel              | 179   | 68,3         | Proprietärt kalkylprogram från Microsoft. Numera används ett XML-baserat format med öppen specifikation (XLSX) som standardformat. Kan även exportera CSV och andra avgränsade textfiler. |
| Fastq              | 1     | 0,4          |   |
| Filemaker          | 3     | 1,1          |   |
| Fme                | 6     | 2,3          |   |
| Fortran            | 1     | 0,4          |   |
| Fugroviewer        | 1     | 0,4          |   |
| Fuji               | 1     | 0,4          |   |
| Fusion             | 3     | 1,1          |   |
| Gams               | 1     | 0,4          |   |
| Gap light analyzer | 1     | 0,4          |   |
| Gel blot imaging   | 1     | 0,4          |   |
| Gnumeric           | 1     | 0,4          |   |
| Gnxas              | 1     | 0,4          |   |
| Google earth       | 2     | 0,8          |   |
| Google maps        | 1     | 0,4          |   |
| Gps plus           | 1     | 0,4          |   |
| Grapher            | 1     | 0,4          |   |
| Graphpad           | 1     | 0,4          |   |
| Grass              | 1     | 0,4          |   |
| Grassgis           | 1     | 0,4          |   |



| Mjukvara         | Antal | Frekvens (%) | Kommentar   |
|------------------|-------|--------------|---|
| Gretl            | 1     | 0,4          |   |
| Gromacs          | 1     | 0,4          |   |
| Gsea             | 1     | 0,4          |   |
| Hadoop           | 1     | 0,4          |   |
| Heureka          | 2     | 0,8          |   |
| Heyex 2          | 1     | 0,4          |   |
| Hobo             | 1     | 0,4          |   |
| Horos            | 2     | 0,8          |   |
| Illumina         | 1     | 0,4          |   |
| Illustrator      | 7     | 2,7          |   |
| Imagej           | 5     | 1,9          |   |
| Iphoto           | 1     | 0,4          |   |
| Irfanview        | 1     | 0,4          |   |
| Isomap           | 1     | 0,4          |   |
| JMP              | 18    | 6,9          | Proprietär mjukvara för statistik. Det finns ett skriptspråk (JSL) som kan användas för att reproducera analys, där skript sparas som text och kan läsas och redigeras (men inte nödvändigtvis köras) utan proprietär mjukvara. Data i programmets nativa format kan importeras i R. Programmet kan också exportera data till textformat.       |
| Jags             | 1     | 0,4          |   |
| Java             | 2     | 0,8          |   |
| Kurvlr           | 1     | 0,4          |   |
| Labchart pro     | 1     | 0,4          |   |
| Lastools         | 1     | 0,4          |   |
| Latex            | 1     | 0,4          |   |
| Libreoffice      | 2     | 0,8          |   |
| Lims             | 1     | 0,4          |   |
| Maestro          | 1     | 0,4          |   |
| Mathematica      | 1     | 0,4          |   |
| Matlab           | 12    | 4,6          | Proprietär mjukvara för beräkningar. Det finns ett skriptspråk som kan användas för att reproducera analys, där skript sparas som text och kan läsas och redigeras (men inte nödvändigtvis köras) utan proprietär mjukvara. Programmet kan också exportera data till textformat.  |
| Mega             | 1     | 0,4          |   |
| Mendeley         | 1     | 0,4          |   |
| Mestrenova       | 1     | 0,4          |   |
| Microsoft Office | 20    | 7,6          | Vissa respondenter har angett Microsoft Office utan att närmare specificera vilka Officeprogram som används. Det är troligt att de flesta som angett detta svar använder Excel eller Access för datahantering, eftersom det bland de program som ingår i Microsoft Office i första hand är dessa som är avsedda att användas för datahantering. |
| Microstation     | 2     | 0,8          |   |

| Mjukvara             | Antal | Frekvens (%) | Kommentar   |
|----------------------|-------|--------------|---|
| Minitab              | 20    | 7,6          | Proprietär mjukvara för statistik. Det finns ett skriptspråk (MAC), som kan användas för att reproducera analys, där skript sparas som text och kan läsas och redigeras (men inte nödvändigtvis köras) utan proprietär mjukvara. Programmet kan också exportera data till textformat. |
| Modde                | 2     | 0,8          |   |
| Mysql                | 1     | 0,4          |   |
| N-logit              | 1     | 0,4          |   |
| Ncss                 | 1     | 0,4          |   |
| Nfts                 | 2     | 0,8          |   |
| Notepad              | 1     | 0,4          |   |
| Nudist               | 1     | 0,4          |   |
| Nvivo                | 8     | 3,1          |   |
| Opals                | 1     | 0,4          |   |
| Openbugs             | 1     | 0,4          |   |
| Openoffice           | 1     | 0,4          |   |
| Opus wire            | 1     | 0,4          |   |
| Origin               | 2     | 0,8          |   |
| Origin pro           | 1     | 0,4          |   |
| Osirix               | 1     | 0,4          |   |
| Ospecificerad        | 3     | 1,1          |   |
| Ospecificerad        | 1     | 0,4          |   |
| adobe                |       |              |   |
| Ospecificerad        | 1     | 0,4          |   |
| analysis             |       |              |   |
| Ospecificerad        | 6     | 2,3          |   |
| bioinfo              |       |              |   |
| Ospecificerad        | 2     | 0,8          |   |
| browser              |       |              |   |
| Ospecificerad        | 1     | 0,4          |   |
| dictaphone           |       |              |   |
| Ospecificerad        | 2     | 0,8          |   |
| dnaseq               |       |              |   |
| Ospecificerad        | 1     | 0,4          |   |
| gene                 |       |              |   |
| Ospecificerad        | 1     | 0,4          |   |
| genomicalig-<br>ners |       |              |   |
| Ospecificerad        | 1     | 0,4          |   |
| genomicana-<br>lysis |       |              |   |
| Ospecificerad        | 3     | 1,1          |   |
| gis                  |       |              |   |
| Ospecificerad        | 1     | 0,4          |   |
| graphics             |       |              |   |
| Ospecificerad        | 1     | 0,4          |   |
| instr                |       |              |   |

| Mjukvara                 | Antal | Frekvens (%) | Kommentar  |
|--------------------------|-------|--------------|--|
| Ospecificerad labinstr   | 2     | 0,8          |  |
| Ospecificerad microscopy | 1     | 0,4          |  |
| Ospecificerad microsoft  | 2     | 0,8          |  |
| Ospecificerad office     | 1     | 0,4          |  |
| Ospecificerad public     | 1     | 0,4          |  |
| Ospecificerad self       | 3     | 1,1          |  |
| Ospecificerad sensor     | 1     | 0,4          |  |
| Ospecificerad spread     | 1     | 0,4          |  |
| Ospecificerad sql        | 5     | 1,9          |  |
| Ospecificerad stat       | 5     | 1,9          |  |
| Ospecificerad unix       | 1     | 0,4          |  |
| Ospecificerad video      | 1     | 0,4          |  |
| Ospecificerad webapp     | 1     | 0,4          |  |
| Pdf                      | 1     | 0,4          |  |
| Perl                     | 1     | 0,4          |  |
| Photoshop                | 6     | 2,3          |  |
| Plink                    | 1     | 0,4          |  |
| Pls toolbox              | 1     | 0,4          |  |
| Postgres                 | 1     | 0,4          |  |
| Powerpivot               | 2     | 0,8          |  |
| Powerpoint               | 3     | 1,1          |  |
| Premiere                 | 1     | 0,4          |  |
| Prism                    | 1     | 0,4          |  |
| Python                   | 7     | 2,7          |  |
| QGIS                     | 22    | 8,4          | Fri mjukvara för geografiska informationssystem. Kan hantera data i öppna format.  |
| Qtmodeller               | 2     | 0,8          |  |
| Quick terrain modeller   | 1     | 0,4          |  |
| R                        | 100   | 38,2         | R betecknar dels ett programmeringsspråk med inriktning på statistik, dels den dominerande implementeringen av detta språk, som är fri mjukvara. Programkod sparas i textfiler. Programmet kan också exportera data till textformat. |

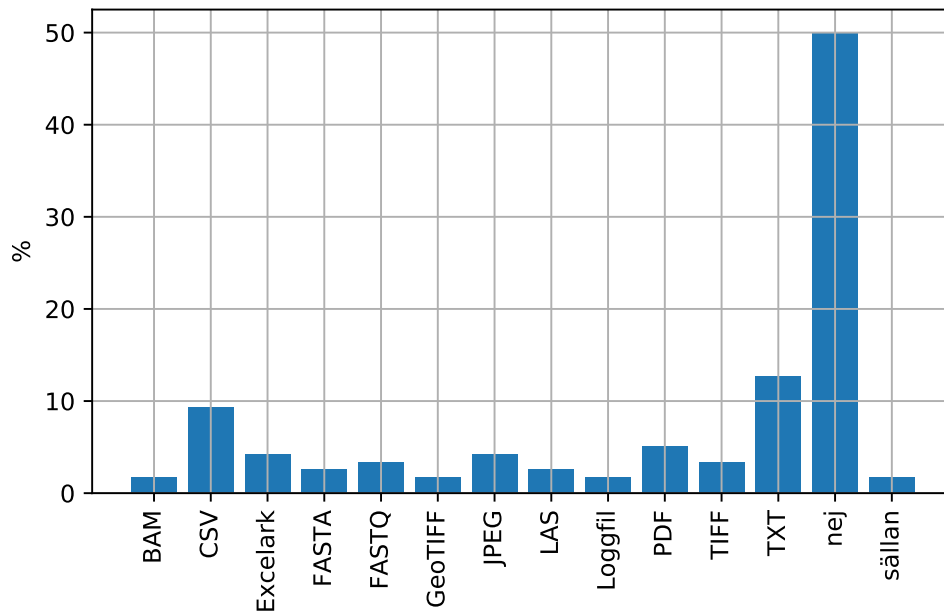
| Mjukvara            | Antal | Frekvens (%) | Kommentar   |
|---------------------|-------|--------------|---|
| Raspberry pi camera | 1     | 0,4          |   |
| Rengis              | 1     | 0,4          |   |
| Rsem                | 1     | 0,4          |   |
| Rust                | 1     | 0,4          |   |
| SAS                 | 45    | 17,2         | Proprietär mjukvara för statistik. Det finns ett skriptspråk som kan användas för att reproducera analys, där skript sparas som text och kan läsas och redigeras (men inte nödvändigtvis köras) utan proprietär mjukvara. Data i programmets nativa format kan importeras i R. Programmet kan också exportera data till textformat. |
| SPSS                | 15    | 5,7          | Proprietär mjukvara för statistik. Data i programmets nativa format kan importeras i R. Programmet kan också exportera data till textformat.  |
| SQL Server          | 14    | 5,3          | Proprietär relationsdatabas från Microsoft. Kan exportera data till SQL eller textformat.   |
| Sharepoint          | 2     | 0,8          |   |
| Sigmaplot           | 3     | 1,1          |   |
| Simca               | 2     | 0,8          |   |
| Sonar5 pro          | 1     | 0,4          |   |
| Sqlite              | 2     | 0,8          |   |
| Stata               | 11    | 4,2          | Proprietär mjukvara för statistik. Det finns ett skriptspråk som kan användas för att reproducera analys, där skript sparas som text och kan läsas och redigeras (men inte nödvändigtvis köras) utan proprietär mjukvara. Data i programmets nativa format kan importeras i R. Programmet kan också exportera data till textformat. |
| Statgraphics        | 1     | 0,4          |   |
| Statistica          | 2     | 0,8          |   |
| Statview            | 1     | 0,4          |   |
| Steplr              | 1     | 0,4          |   |
| Summit evolution    | 1     | 0,4          |   |
| Superbase           | 1     | 0,4          |   |
| Syntech             | 1     | 0,4          |   |
| Tapestation         | 1     | 0,4          |   |
| Terramodeler        | 1     | 0,4          |   |
| Terrascan           | 2     | 0,8          |   |
| Topspin             | 1     | 0,4          |   |
| Trimble realworks   | 1     | 0,4          |   |
| Trinity             | 1     | 0,4          |   |
| Ugene               | 1     | 0,4          |   |
| Unscrambler         | 2     | 0,8          |   |
| Visualstudio        | 1     | 0,4          |   |
| Windows             | 1     | 0,4          |   |

| Mjukvara                  | Antal | Frekvens (%) | Kommentar  |
|---------------------------|-------|--------------|--|
| Windows dokumenthanterare | 1     | 0,4          |  |
| Windows media player      | 1     | 0,4          |  |
| Winstat                   | 1     | 0,4          |  |
| Wizard                    | 1     | 0,4          |  |
| Word                      | 41    | 15,6         | Proprietärt ordbehandlingsprogram från Microsoft. Numera används ett XML-baserat format med öppen specifikation (DOCX) som standardformat. Kan även exportera textfiler. Används sannolikt oftare för att framställa rapporter och liknande än för lagring eller analys av data. |
| Zen black                 | 1     | 0,4          |  |
| Ztree                     | 1     | 0,4          |  |
| @risk                     | 1     | 0,4          |  |

### 2.3 Fråga 1c: Format

**Fråga** Sparar du data i andra format än de som är standard i de programvaror du angett i förra frågan?

Bland de respondenter som svarat på 1b har 118 svarat på denna fritextfråga, varav hälften angett svaret ”nej”. Efter normalisering är det 52 olika svar som getts. Fördelningen av de svar som getts av mer än en respondent visas av fig. 3. Samtliga format som angetts visas i tab. 2, med kommentarer fokuserade på formatens öppenhet för de vanligare format som visas i fig. 3.



Figur 3: Andel av alla som svarat på 1c som uppgett givet format, för format med mer än ett svar.

Tabell 2: Uppgivna format.

| Format         | Antal | Frekvens (%) | Kommentar   |
|----------------|-------|--------------|---|
| BAM            | 2     | 1,7          | En binär version av SAM-formatet, som används av SAMtools, en uppsättning fria verktyg för DNA-sekvensering.  |
| BIN            | 1     | 0,8          |   |
| BMP            | 1     | 0,8          |   |
| CONFOCA-LIMAGE | 1     | 0,8          |   |
| CSV            | 11    | 9,3          | Kommaseparerade textfiler. Ett öppet format som används av många program, men detaljer i implementeringen (t.ex. decimalavgränsare och hantering av textsträngar) kan variera.  |
| CVS            | 1     | 0,8          |   |
| EDR            | 1     | 0,8          |   |
| Excelark       | 5     | 4,2          | Nativt format i Microsoft Excel. Det finns både ett äldre binärt format (filnamnstillägget .XLS) och ett nyare XML-baserat format (tillägget .XLSX). Utifrån svaren har det inte varit möjligt att skilja mellan dessa. |
| FASTA          | 3     | 2,5          | Ett textbaserat format för sekvensering av DNA och proteiner. Används av det fria mjukvarupaketet FASTA.  |
| FASTQ          | 4     | 3,4          | Ett textbaserat format som används för att lagra information om biologiska sekvenser tillsammans med gradering av deras kvalitet. Kan hanteras med hjälp av fria program (t.ex. FASTX-Toolkit).                         |
| FILECONVERTER  | 1     | 0,8          |   |
| GENANALYSIS    | 1     | 0,8          |   |
| GENSTAT        | 1     | 0,8          |   |
| GRO            | 1     | 0,8          |   |
| GeoTIFF        | 2     | 1,7          | Ett format för att bädda in geografiska koordinater i TIFF-filer. Kan hanteras med hjälp av fria program.   |
| JOURNALS       | 1     | 0,8          |   |
| JPEG           | 5     | 4,2          | Ett öppet format för komprimerad rastergrafik.  |
| LABBOOKS       | 1     | 0,8          |   |
| LARGE          | 1     | 0,8          |   |
| LAS            | 3     | 2,5          | Ett öppet format för molndata.  |
| LaTeX          | 1     | 0,8          |   |
| Loggfil        | 2     | 1,7          | Några respondenter har refererat till loggfiler. De specifika formaten för dessa kan dock variera och har inte angetts närmare. Ofta, men inte alltid, är loggfiler textbaserade.                                       |
| MAPINFO        | 1     | 0,8          |   |
| MDB            | 1     | 0,8          |   |
| MINITAB        | 1     | 0,8          |   |
| MP4            | 1     | 0,8          |   |
| MTS            | 1     | 0,8          |   |

| Format            | Antal | Frekvens (%) | Kommentar   |
|-------------------|-------|--------------|---|
| NETCDF            | 1     | 0,8          |   |
| ODF               | 1     | 0,8          |   |
| OLAP CUBE         | 1     | 0,8          |   |
| PDF               | 6     | 5,1          | PDF-formatet används sannolikt mer för rapporter och liknande än för rena forskningsdata.   |
| PX                | 1     | 0,8          |   |
| R                 | 1     | 0,8          |   |
| RAR               | 1     | 0,8          |   |
| RAWIN-STRUMENT    | 1     | 0,8          |   |
| RDS               | 1     | 0,8          |   |
| SAM               | 1     | 0,8          |   |
| SEQDATA           | 1     | 0,8          |   |
| SFF               | 1     | 0,8          |   |
| SHP               | 1     | 0,8          |   |
| SQL               | 1     | 0,8          |   |
| SQL               | 1     | 0,8          |   |
| SERVER            |       |              |   |
| TIFF              | 4     | 3,4          | Ett öppet format för rastergrafik.  |
| TXT               | 15    | 12,7         | Textfiler med data kan följande en rad olika standarder när det gäller fältavgränsning och liknande. Ur bevarandesynpunkt är det viktigt att se till att dessa finns tydligt dokumenterade. |
| XML               | 1     | 0,8          |   |
| XTC               | 1     | 0,8          |   |
| ZIP               | 1     | 0,8          |   |
| nej               | 59    | 50,0         |   |
| nya format behövs | 1     | 0,8          |   |
| ospecificerat     | 1     | 0,8          |   |
| spreadsheet       | 1     | 0,8          |   |
| sällan            | 2     | 1,7          |   |

## 2.4 Fråga 1d: Datamängder

**Fråga** Hur stora datamängder behöver du kunna publicera och arkivera? Det gavs fem graderade alternativ, ”1–500 MB”, ”500 MB–1 GB”, ”1–500 GB”, ”500 GB–1 TB”, plus ett alternativ ”ange i fritext om inget av ovanstående alternativ är lämpligt”. Flera alternativ kunde markeras.

Det har getts 231 svar. Fördelningen av dessa visas av tab. 3.

Tabell 3: Uppgivna datamängder.

| Intervall | Antal | Frekvens (%) |
|-----------|-------|--------------|
| 1–500 GB  | 74    | 32,0         |
| 1–500 MB  | 69    | 29,9         |



| Intervall   | Antal | Frekvens (%) |
|-------------|-------|--------------|
| 500 GB–1 TB | 23    | 10,0         |
| 500 MB–1 GB | 45    | 19,5         |
| Över 1 TB   | 43    | 18,6         |

Det har getts 39 fritextsvar. En del specificerar datamängder, ”ca 4 TB”, ”>100 TB” upp till ”PB”. Några har anmärkt att frågan är otydlig, att det inte är specificerat om det är datamängd per projekt eller totalt som avses, eller att de är osäkra på om vi avser arbetsmaterial eller rådata. Dessa svar pekar på ett behov av utbildning när det gäller kriterierna för vad som skall arkiveras.

## 2.5 Fråga 1e: Känsliga data

**Fråga** Hanterar du känsliga data? Fem alternativ fanns: ”nej, jag hanterar inte känsliga data”, ”osäker på om jag hanterar känsliga data”, ”ja, sekretessbelagda data” (där en länk till offentlighets- och sekretesslagen bifogades), ”ja, personuppgifter” och ”ja, annan känslig data, ange gärna vad för slag”, där respondenterna kunde specificera i fritext. Flera alternativ kunde markeras.

Det har getts 270 svar. Fördelningen av dessa visas av tab. 4.

Tabell 4: Hantering av känsliga data.

| Typ av känsliga data    | Antal | Frekvens (%) |
|-------------------------|-------|--------------|
| Andra känsliga data     | 36    | 13,3         |
| Inga                    | 159   | 58,9         |
| Osäker                  | 47    | 17,4         |
| Personuppgifter         | 34    | 12,6         |
| Sekretessreglerade data | 26    | 9,6          |

Det har getts 32 fritextsvar. Respondenterna har i de flesta av dessa fall gjort närmare specifikation av vad för slags känsliga data de hanterar, t.ex. ”djurförsök”, ”[p]roduktionsuppgifter som kan ha bäring för företags konkurrenskraft”, ”geografiska data, som kan kopplas till enskilda markägare”, ”positioner för skyddade arter”. Någon uttrycker osäkerhet kring vad som räknas som personuppgifter.

## 2.6 Fråga 1f: Arkivkrav

**Fråga** I vilken grad känner du till de arkivkrav en myndighet ska uppfylla och hur dessa påverkar din hantering av data? Fyra graderade alternativ fanns: ”inte alls”, ”i liten grad”, ”i viss grad” och ”stor grad”.

Det har getts 274 svar. Fördelningen av dessa visas av tab. 5.

Tabell 5: Kännedom om arkivkrav.

| Grad av kännedom | Antal | Frekvens (%) |
|------------------|-------|--------------|
| I liten grad     | 125   | 45,6         |

| Grad av kännedom | Antal | Frekvens (%) |
|------------------|-------|--------------|
| I stor grad      | 49    | 17,9         |
| I viss grad      | 74    | 27,0         |
| Inte alls        | 26    | 9,5          |

Av svarsfördelningen framgår att det krävs utbildning av forskare när det gäller arkivkraven.

## 2.7 Fråga 1g: Externa samarbeten

**Fråga** Deltar du i externa samarbeten eller samarbetsavtal? Alternativen var ”ja” och ”nej”.

Det har getts 274 svar. Fördelningen av dessa visas av tab. 6.

Tabell 6: Deltagande i externa samarbeten.

| Deltar i externa samarbeten | Antal | Frekvens (%) |
|-----------------------------|-------|--------------|
| Ja                          | 217   | 79,2         |
| Nej                         | 57    | 20,8         |

Inga fritextfält finns i anslutning till denna fråga.

## 2.8 Fråga 1h: Versionshantering

**Fråga** Använder du något system för versionskontroll av data? Alternativen var ”nej, använder inget sådant system”, ”ja, manuell hantering genom namngivning av filer eller kataloger” och ”ja, system för automatiserad versionskontroll (Git, SVN eller liknande), ange gärna vilket/vilka”, där respondenterna kunde specificera system i fritext. Flera alternativ kunde markeras.

Det har getts 275 svar. Fördelningen av dessa visas av tab. 7.

Tabell 7: System för versionshantering.

| Versionshantering | Antal | Frekvens (%) |
|-------------------|-------|--------------|
| Automatiserat     | 28    | 10,2         |
| Inget             | 142   | 51,6         |
| Manuellt          | 119   | 43,3         |

Det har getts 21 fritextsvar. Respondenter har ombetts ange vilket/vilka system för versionskontroll de använder, om de använder ett sådant system. Majoriteten har svarat att de använder Git (vissa specificerar förrådet, GitHub eller GitLab). En respondent har angett git-annex, som används för att hålla reda på filer med Git med hjälp av länkar till filerna snarare än själva filerna (används ofta för stora filer). Några har angett Subversion. Mercurial och Fossil har angetts av en respondent var.

## 3 Fråga 2

I fråga 2 har huvudsakligen attityder till och erfarenheter av öppna data undersökts. Syftet med frågorna har varit ta reda på hur utbredd publicering av data är vid SLU samt hur forskarna ställer sig till ett utökat tillgängliggörande av forskningsdata. Vi har också varit intresserade av att veta vad som motiverar forskarna att tillgängliggöra och vad de ser som problem i sammanhanget.

### 3.1 Fråga 2a: Erfarenheter av tillgängliggörande

**Fråga** Har du delat eller tillgängliggjort forskningsdata/miljöanalysdata?

Fem alternativ fanns (varav mer än ett kunde anges): ”ja, jag har delat med nära forskarkollegor (informellt)”, ”ja, jag har delat data direkt vid förfrågan från andra än nära kollegor”, ”ja, jag har tillgängliggjort via dataportal (datacenter) eller dataarkiv. Om ja, ange gärna i vilken tjänst.”, ”ja, jag har tillgängliggjort på webbplats (forskningsprojekt eller personlig)”, ”nej”.

Bland de 265 forskare som svarade har 162 uppgett att de delat informellt med nära forskningskollegor. 99 respondenter har delat vid förfrågan med andra än nära kollegor och 77 uppger sig ha tillgängliggjort data via olika dataportaler. Bland de svarande uppgav 57 att de tillgängliggjort data via en webbplats och 46 uppgav att de inte har delat data. Fördelningen av svaren visas av fig. 4.

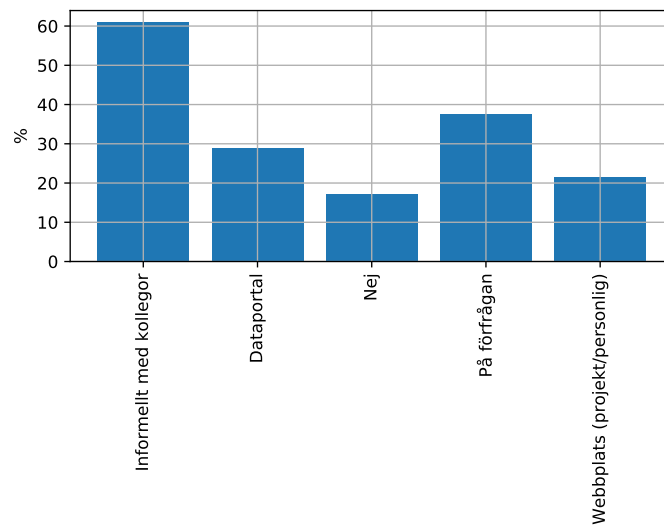
Det gavs 77 svar angående var forskarna har tillgängliggjort sina data. Den tjänst som flest nämnde som en plats där de delat var Dryad (12 svar), sex personer uppgav NCBI, 5 Artdataportalen, 4 European Nucleotide Archive, 3 Github, 3 Zenodo, 2 Figshare och 2 Genbank. Utöver dessa nämns en del av de plattformar fortlöpande miljöanalys vid SLU tillhandahåller för data och en uppsjö olika mindre tjänster.

### 3.2 Fråga 2b: Inställning till att öppet dela data

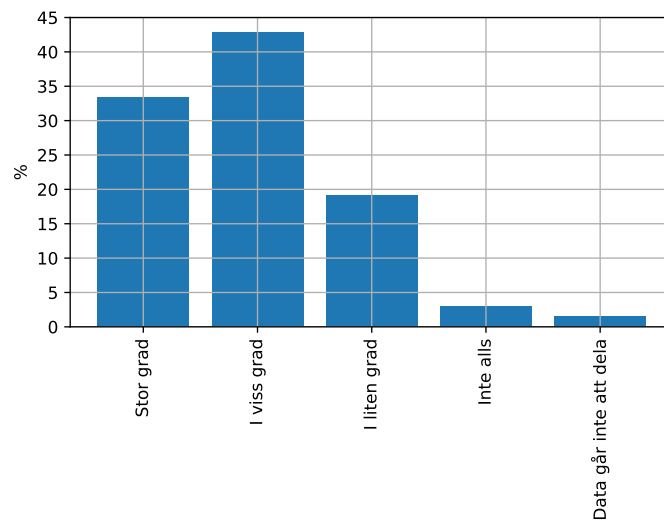
**Fråga** I vilken grad kan du tänka dig att öppet dela den data du arbetar med?

Fem alternativ fanns: ”Stor grad (allt eller så mycket som möjligt)”, ”I viss grad (större utvalda delar)”, ”I liten grad (mindre utvalda data)”, ”Inte alls”, ”Data går inte att dela p.g.a. sekretess eller liknande”.

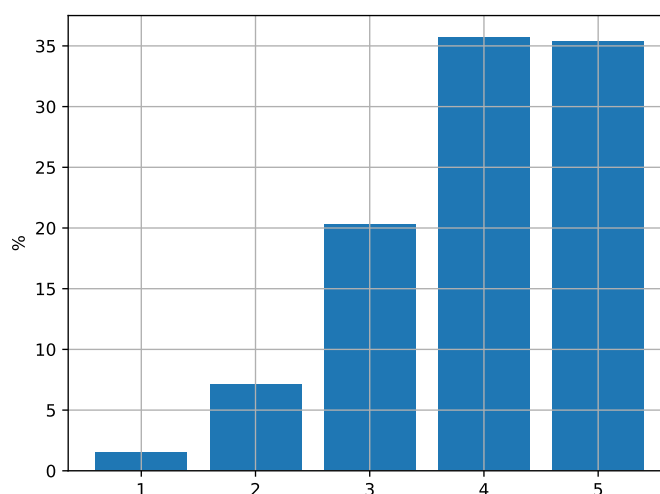
Bland de 267 respondenter som svarade har 114 uppgett att de i viss grad kan öppet dela data, 89 kan dela data i stor grad, 52 respondenter i liten grad, 8 att de inte alls kan dela data och 4 som anser att data inte går att dela. Fördelningen av svaren visas av fig. 5.



Figur 4: Fördelning av svar på fråga 2a, ”har du delat eller tillgängliggjort forskningsdata/miljöanalysdata”.



Figur 5: Fördelning av svar på fråga 2b, ”i vilken grad kan du tänka dig att öppet dela den data du arbetar med”.



Figur 6: Fördelning av svar på fråga 2c, ”anser du att det är viktigt att tillgängliggöra data till forskare och till allmänheten”.

### 3.3 Fråga 2c: Tillgängliggörande

**Fråga** Anser du att det är viktigt att tillgängliggöra data till forskare och till allmänheten?

Respondenterna ombads välja svarsalternativ enligt en femgradig svarsskala där 1 motsvarade ”Absolut inte” och 5 motsvarade ”Mycket viktigt”.

Bland de 267 respondenterna som svarade har 96 uppgett ”4”, 94 anser att det är ”mycket viktigt” d.v.s. ”5”, 54 har svarat ”3”, 19 har svarat ”2” och 4 har svarat ”Absolut inte” dvs alternativ ”1”. Fördelningen av svaren visas av fig. 6.

Medelvärdet på frågan om inställningen till att tillgängliggöra data för forskare och till allmänheten är 3,7.

### 3.4 Fråga 2d: Tillgängliggörande

**Fråga** Motivera gärna ditt val i föregående fråga

I fråga 2d ombads respondenterna ge en motivering av sitt svar på fråga 2c, om de anser att det är viktigt att tillgängliggöra data till forskare och allmänhet. Det gavs 145 fritextsvar på frågan. Svaren har kategoriserats och indelats i sådana som är positiva till delning, vilket visas i tab. 8 och sådana som anger olika problem, vilket visas i tab. 9. En respondent kan ha gett svar i flera kategorier.

De som svarat att de är positiva till delning har som främsta anledning att det främjar forskningen på något sätt (46 svar). Deras kommentarer berör saker som att data behöver utnyttjas av fler än enskilda forskargrupper och att forskningsdata har stor potentiell användning som inte utnyttjas till fullo idag. Data kan användas i olika syften och för olika typer av undersökningar och därför bör den tillgängliggöras, menar en stor del av respondenterna. Gemensamt för svaren är att de uttrycker att mer kunskap kan skapas om data är tillgänglig

och att detta är forskningens uppgift. Dessa svar relaterar till en annan kategori som nämner återanvändning av data eller utnyttjande av så kallad "big data" som en positiv effekt av att dela med sig av sina forskningsdata (5 svar).

Den näst största kategorin av svar uttrycker en positiv inställning till att dela data eftersom det ger upphov till en allmän nytta (34 svar). I dessa svar ges inte närmare beskrivningar av vad för nytta utan uttrycker mest en generell positiv hållning till att dela data i. Enskilda svar uttrycker dock att denna nytta består i att forskning sprids till större delar av samhället. Andra respondenter anger liknande anledningar till en positiv inställning men uttrycker sig mer principiellt då de menar att data bör vara tillgängliga för alla eftersom allmänheten finansierar verksamheten via skattemedel (21 svar).

Möjligheten att kontrollera resultat och granska publicerad forskning ges också som ett skäl till att dela data (25). Respondenterna menar att granskningen av forskningsresultat är svår eller omöjlig att genomföra om underbyggande data inte är tillgänglig. I dessa svar uttrycks en uppfattning att tillgängliggörande av data höjer forskningens kvalitet och bidrar till en ökad öppenhet. Dessa svar hör i ihop med de två svar som menar att tillgängliggörande av data ökar transparensen inom forskningen och de svar som menar att en säkrad arkivering ökar forskningens långsiktighet (2 svar).

Resterande kategori av svar kretsar del kring en ökad samarbetspotential mellan forskare genom delade data (7 svar).

Tabell 8: Kategorier med positiva synpunkter på tillgängliggörande av data.

| Kategori                 | Beskrivning  | Antal |
|--------------------------|--|-------|
| främja forskning         | Tillgängliggörande av data ses som något som främjar forskningen. Kan avse tempo, möjligheten att svara på komplicerade frågor m.m.              | 46    |
| allmän nytta             | Tillgängliggörande av data genererar en allmän nytta. Kategorin beskriver en generellt positiv hållning till delande utan närmare konkretisering | 34    |
| kontroll av resultat     | Tillgängliggörandet ses som en garant för att se till att forskning genomförts på lämpligt sätt eller förhindrar forskningsfusk.                 | 25    |
| skattefinansierad        | Positiv inställning till tillgängliggörande baserat på principen att data är allmänt finansierade och därför bör vara allmänt tillgängliga.      | 21    |
| samarbete                | Tillgängliggörande beskrivs främja samarbeten inom forskningen.  | 7     |
| återanvändning, big data | Möjligheten att återanvända data samt aggregera stora mängder data för att svara på forskningsfrågor.  | 5     |
| mandat                   | Benägenhet att tillgängliggöra därför att det krävs av olika aktörer.  | 3     |
| långsiktighet            | Framtida användbarhet av data.   | 2     |
| transparens              | Bidrar till ökad transperans inom forskningen.   | 2     |

Tabell 9: Kategorier med negativa synpunkter på tillgängliggörande av data.

| Kategori                        | Beskrivning  | Antal |
|---------------------------------|--|-------|
| timing, post-pub                | Kommentarer som berör när tillgängliggörande ska ske. Oftast betonas att tillgängliggörande inte bör ske innan publicering av resultat.        | 18    |
| risk för fel                    | Vill ogärna tillgängliggöra eftersom det finns en risk att datat misstolkas.   | 14    |
| typ av data                     | Kommentarer som berör vilken typ av data som bör tillgängliggöras. Oftast en ovilja att dela rådata – aggregerade data beskrivs som ok.        | 13    |
| sekretess, känslig              | Avvaktande pga att man jobbar med sekretessreglerade eller känsliga data: persondata, provpunkter etc.   | 12    |
| resurser                        | Att tillgängliggöra är resurskrävande och tar tid och pengar på bekostnad av andra uppgifter.  | 7     |
| metadata                        | Uppfattningen att data måste beskrivas väldigt väl och att detta tar för mycket tid.   | 6     |
| scooped                         | Farhåga att tillgängliggörande kan leda till att andra forskare använder ens data och får erkännande innan man själv hunnit slå mynt av datat. | 6     |
| konkurrensfördel, finansiering  | Ovilja att dela med sig av data då den ses som ens främsta konkurrensfördel och grunden till att man får finansiering.                         | 6     |
| allmänheten kan inte tolka data | Avvaktande inställning till tillgängliggörande eftersom man anser att allmänheten kan misstolka data.  | 4     |
| citering, meritering            | Vill få något tillbaka för att tillgängliggöra. Citering av data måste blir norm eller så bör det vara meriterande inom forskarsamhället.      | 4     |
| definition av data              | Anmärkningar på bristfällig definition av data i fråga.  | 3     |
| målgrupp                        | Tveksam inställning till tillgängliggörande, där man vill kunna kontrollera vilka grupper som får tillgång till datat.                         | 3     |
| ovanligt inom disciplin         | Delning är ovanligt inom disciplin.  | 1     |
| stora dataset                   | Delning är svårt när det gäller stora datamängder.   | 1     |

### 3.5 Fråga 2e: Tillgängliggörande

**Fråga** Vad skulle få dig att börja, eller i större utsträckning än idag, tillgängliggöra data?

Det har getts 266 svar. Fördelningen av dessa visas av tab. 10.

Tabell 10: Svar på tillgängliggörande.

| Svar                                | Antal | Frekvens (%) |
|-------------------------------------|-------|--------------|
| Efterfrågan forskare                | 124   | 46,6         |
| Mer kunskap                         | 103   | 38,7         |
| Mer stöd                            | 87    | 32,7         |
| Mätning av genomslag för öppna data | 71    | 26,7         |
| Policy                              | 57    | 21,4         |
| Tillgängliggörande ses som merit    | 103   | 38,7         |
| Tillgång till verktyg/plattform     | 122   | 45,9         |
| Vet ej                              | 28    | 10,5         |
| Övrigt                              | 32    | 12,0         |

Det gavs 31 fritextsvar på frågan. De har kategoriserats enligt vad som visas i tab. 11. En respondent kan ha gett svar i flera kategorier.

Tabell 11: Kategorier med synpunkter på tillgängliggörande.

| Kategori                | Beskrivning  | Antal |
|-------------------------|--|-------|
| resurser                | Vill ha mer tid eller ekonomiska resurser.   | 14    |
| meritering              | Vill att det blir meriterande att dela data (citering etc.).   | 5     |
| datakureringsstöd       | Vill ha stöd från enhet för datakurering.  | 3     |
| infrastruktur           | Vill ha bättre infrastruktur (lagringsutrymme, bandbredd etc.).  | 3     |
| bottom-up               | Motsätter sig styrning uppifrån.   | 1     |
| dela om andra           | Det finns risk för konkurrensnackdel om bara SLU-forskare beordras dela.   | 1     |
| fortsatt eget forskande | Vill delta i framtida projekt som använder datat.  | 1     |
| inte sekretess          | Det krävs att källorna släpper sekretesskrav.  | 1     |
| kontroll, risk för fel  | Vill ha kontroll över användning och tolkning för att motverka feltolkning.  | 1     |
| krav                    | Har det redan som krav och söker skapa plattformar.  | 1     |
| allmän åtkomst          | Låta alla som vill använda data som samlas in.   | 1     |
| öppna standarder        | Data måste vara tillgängliga och uppladdningsbara för användare av fri programvara; det måste finnas en öppen standard för datalagring | 1     |



### 3.6 Fråga 2f: Datahanteringsplan

**Fråga** Har du en datahanteringsplan för din verksamhet?

Det har getts 264 svar. Fördelningen av dessa visas av tab. 12

Tabell 12: Svar på datahanteringsplan.

| Svar                                  | Antal | Frekvens (%) |
|---------------------------------------|-------|--------------|
| Ja                                    | 73    | 27,7         |
| Nej                                   | 114   | 43,2         |
| Vet inte vad en datahanteringsplan är | 77    | 29,2         |

## 4 Fråga 3

### 4.1 Fråga 3a: Institution

**Fråga** Vilken är din primära hemmahörande institution/organisation?

Det har getts 262 svar på frågan. Fördelningen av dessa visas i tab. 13.

Tabell 13: Uppgivna institutioner.

| Institution   | Antal | Frekvens (%) |
|---|-------|--------------|
| Annan   | 9     | 3,4          |
| Institutionen för akvatiska resurser (SLU Aqua)                     | 116   | 44,3         |
| Institutionen för arbetsvetenskap, ekonomi och miljöpsykologi (AEM) | 34    | 13,0         |
| Institutionen för biomedicin och veterinär folkhälsovetenskap       | 31    | 11,8         |
| Institutionen för biosystem och teknologi                           | 31    | 11,8         |
| Institutionen för ekologi   | 35    | 13,4         |
| Institutionen för ekonomi   | 20    | 7,6          |
| Institutionen för husdjurens utfodring och vård                     | 3     | 1,1          |
| Institutionen för husdjursgenetik                                   | 9     | 3,4          |
| Institutionen för kliniska vetenskaper                              | 17    | 6,5          |
| Institutionen för landskapsarkitektur, planering och förvaltning    | 2     | 0,8          |

| Institution   | Antal | Frekvens (%) |
|---|-------|--------------|
| Institutionen för mark och miljö                          | 16    | 6,1          |
| Institutionen för molekylära vetenskaper                  | 8     | 3,1          |
| Institutionen för norrländsk jordbruksvetenskap           | 4     | 1,5          |
| Institutionen för skogens biomaterial och teknologi (SBT) | 3     | 1,1          |
| Institutionen för skogens ekologi och skötsel             | 9     | 3,4          |
| Institutionen för skoglig genetik och växtfysiologi       | 5     | 1,9          |
| Institutionen för skoglig mykologi och växtpatologi       | 6     | 2,3          |
| Institutionen för skoglig resurshushållning               | 29    | 11,1         |
| Institutionen för stad och land                           | 11    | 4,2          |
| Institutionen för sydsvensk skogsvetenskap                | 11    | 4,2          |
| Institutionen för vatten och miljö                        | 6     | 2,3          |
| Institutionen för vilt, fisk och miljö                    | 10    | 3,8          |
| Institutionen för växtbiologi                             | 6     | 2,3          |
| Institutionen för växtförädling                           | 7     | 2,7          |
| Institutionen för växtproduktionsekologi                  | 1     | 0,4          |
| Institutionen för växtskyddsbiologi                       | 19    | 7,3          |
| Skogsmästarskolan   | 2     | 0,8          |

## 4.2 Fråga 3b: Tillhörighet

**Fråga** Tillhör du [alternativ]?

Respondenterna har tillfrågats om de tillhör ett forskningscentrum eller en miljöanalysverksamhet. Det har getts 160 svar på frågan. Fördelningen av dessa visas i tab. 14.

Tabell 14: Uppgivna verksamheter.

| Institution           | Antal | Frekvens (%) |
|-----------------------|-------|--------------|
| Annan verksamhet      | 56    | 35,0         |
| Forskningscentrum     | 62    | 38,8         |
| Miljöanalysverksamhet | 42    | 26,2         |

### 4.3 Fråga 3c: Tillhörighet

**Fråga** Ange vilket forskningscentrum alternativt miljöanalysverksamhet

Det har getts 32 svar på denna fritextfråga. Artdatabanken, Riksskogstaxeringen och Sötvattenslaboratoriet är exempel på svar som förekommit.

### 4.4 Fråga 3d: Titel

**Fråga** Vilken titel beskriver din tjänst bäst?

Det har getts 263 svar på frågan. Fördelningen av dessa visas i tab. 15.

Tabell 15: Uppgivna titlar.

| Titel                                     | Antal | Frekvens (%) |
|---|-------|--------------|
| Annan tjänst                              | 34    | 12,9         |
| Doktorand                                 | 22    | 8,4          |
| Lektor, postdoc eller annan forskartjänst | 158   | 60,1         |
| Professor                                 | 49    | 18,6         |

### 4.5 Fråga 3e: Extern finansiering

**Fråga** Är din nuvarande verksamhet helt eller delvis externfinansierad?

Det har getts 259 svar på frågan. Fördelningen av dessa visas i tab. 16.

Tabell 16: Andelar som angett extern finansiering.

| Institution | Antal | Frekvens (%) |
|-------------|-------|--------------|
| Ja          | 215   | 83,0         |
| Nej         | 44    | 17,0         |

### 4.6 Fråga 3f: Övriga synpunkter

**Fråga** Övriga synpunkter

Här kunde respondenterna ange egna synpunkter. Det har getts 22 sådana synpunkter. Vissa har uttryckt oro för att verksamheten med publicering och arkivering kommer att bli en ytterligare börda för forskarna. Andra har uttryckt önskan om mer utbildning.

## 5 Sammanfattande diskussion och behov av åtgärder

Enkäten skickades, som anges i avs. 1, till 1441 mottagare. På de flesta av frågorna har knappt 20 procent av dessa svarat. Denna relativt låga svarsfrekvens innebär att resultaten måste tolkas med viss försiktighet. De svarande kan inte betraktas som representativa för forskare vid SLU i allmänhet: det kan tänkas att de exempelvis är mer engagerade när det gäller tillgängliggörande av data och mer aktiva när det gäller hantering av forskningsdata. Å andra sidan är kanske just dessa forskare mer representativa för de som kommer att bli de mest frekventa användarna av Tilda.

### 5.1 Fråga 1: datatyper, känsliga data och arkivering

Forskare vid SLU arbetar med flera olika datatyper, bland annat olika typer av numeriska och statistiska data och gensekvensdata. För detta använder de en rad olika mjukvaror och filformat, både fria och proprietära. Generellt verkar dock dessa mjukvaror erbjuda möjlighet till export till öppna format. Det är nödvändigt att utforma de s.k. leveransöverenskommelserna, de avtal som görs mellan SLU och dataproducenterna om villkor för leverans och hantering av data, så att informationen kvalitetssäkras när det gäller exempelvis format.

De datamängder forskarna angett varierar. Det finns ett behov av ytterligare kommunikation med institutioner där forskare rapporterat stora datamängder. Initiativ till detta har tagits med en mindre uppföljningsenkät, som beskrivs i bilaga A. Enkäten har inte tagit hänsyn till i vilken mån forskare arbetar med överlappande datamängder, vilket gör det problematiskt att tolka summeringar av datamängderna.

En del forskare har angett sig ha olika typer av känsliga data, men uttrycker samtidigt en osäkerhet kring vad som räknas som känsliga data och personuppgifter. Andra är osäkra på hur rådata skiljer sig från bearbetade data och de självrapporterade kunskaperna om arkivkrav varierar. Författarna uppfattar att det finns ett behov av utbildning kring gällande regelverk för hantering av data, såsom kommande forskningsdatalag, arkivlagen och offentlighets- och sekretesslagen.

Majoriteten har angett att de deltar i externa samarbeten. Utbildningar riktade mot SLU:s forskare behöver lägga vikt vid att hanteringen av data i sådana samarbetsprojekt måste vara förenliga med de krav som gäller enligt t.ex. offentlighetsprincipen.

Relativt få av de forskare som svarat på enkäten har angett system för automatiserad versionshantering, men flera av de vanligaste systemen, framför allt Git, finns representerade.

### 5.2 Fråga 2: öppna data och delning av data

De flesta forskare har rapporterat någon grad av erfarenhet när det gäller delning och tillgängliggörande av data. Ofta rör det sig om informell delning med nära kollegor, men ca 30 procent av respondenterna har angett delning via dataportaler.

Många forskare har rapporterat en positiv syn på tillgängliggörande. ”Tillgängliggörande” som används i frågeformuleringen måste här tolkas som att det syftar på avsiktlig publicering, snarare än den generella möjlighet till utlämnande, som styrs av lagkrav. Vanliga problem som lyfts fram handlar om att det måste vara möjligt att styra tidpunkt för tillgängliggörande, att data kan vara känsliga, att det finns risk för feltolkningar av data och att tillgängliggörande är resurskrävande. Här ser rapportförfattarna ett behov av information till användarna om de funktioner som utvecklas i Tilda för hantering av embargo och sekretess. Det är också viktigt att underlätta och automatisera arbetsflöden för forskarna så mycket som möjligt.

När det gäller frågan vad som skulle öka tillgänglighöret av data är ökade resurser och att det skulle vara mer meriterande för forskare de vanligaste svaren. Vi kan behöva framhålla den funktionalitet i Tilda som är avsedd att underlätta dessa saker, som DOI:er (beständiga identifierare som gör det möjligt att referera till digitala resurser på ett sätt som överlever förändringar av bakomliggande URL).

DCU bör även driva på användningen av datahanteringsplaner inom SLU så att dataset dokumenteras löpande under deras tillkomst. Detta skulle underlätta publicering och arkivering av data då tillförlitliga beskrivningar av data vad gäller insamling, tillkomst, metodik och bearbetning redan finns på plats när forskarna ska publicera och/eller arkivera sina data. Skulle dessa uppgifter dokumenteras löpande sänks tröskeln och resursbehovet för att lägga in data i Tilda.

DCU arbetar för närvarande med att utforma utbildningar för Tilda. Här läggs stor vikt vid de punkter som diskuterats ovan där enkätsvaren visat på behov av utbildning. DCU utformar också standardiserade leveransöverenskommelser som tar hänsyn till rapporterade programvaror och filformat.

DCU planerar att publicera underliggande data för enkäten i Tilda när systemet blir tillgängligt. Tills vidare kan data lämnas ut på begäran.

## Bilaga A Uppföljningsenkät

En uppföljningsenkät skickades till institutioner där forskare angett datamängder över 1 TB på fråga 1d. Syftet med uppföljningsenkäten var att få en mer precis uppfattning av datamängderna hos dessa för att bättre kunna uppskatta lagringsbehoven för Tilda. Enkäten innehöll tre frågor med fritextalternativ. Det inkom 70 svar på denna enkät.

### A.1 Datamängder

**Fråga** Hur mycket data genererar du uppskattningsvis under ett år? (svara med så god precision som möjligt i hela Mb, Gb eller Tb)

Svaren visar på variation mellan 10 MB och 80 TB. Några har angett att de är osäkra. Det är möjligt att flera av de som svarat arbetar med överlappande datamängder.

### A.2 Fördelning rådata/bearbetade data

**Fråga** Hur är dessa data fördelade mellan rådata och bearbetade data? Med rådata menar vi data som inte genomgått någon bearbetning eller analys och är det direkta resultatet av din datainsamlingsmetod. (uppskatta i procent, ex. 60% rå 40% aggregerade)

Svaren varierar mellan 100 procent rådata och 100 procent bearbetade data. Någon generell tendens är svår att urskilja. Några har angett osäkerhet när det gäller fördelningen.

### A.3 Kommentarer

**Fråga** Ytterligare kommentarer (icke-obligatorisk)

Vissa har påtalat att de arbetar med andras datamängder. Någon har ansett att det vore mer relevant att fråga om den totala mängden data som lagras och hur mycket lagringsbehoven kan förväntas öka de kommande åren. Någon har uppgett att data lagras i egna SQL-databaser och inte kommer att belasta Tilda.

## Bilaga B Enkäten



## Enkät om hantering av forsknings- och miljöanalysdata vid SLU

MIKAELA ASPLUND

PDF

SÖK

HJÄLP

Den här enkäten är en del i arbetet med att utveckla Tilda, SLUs nya system för publicering och arkivering av forsknings- och miljöanalysdata. Tilda kommer att fungera som den centrala platsen för att synliggöra och långtidsbevara data skapad på SLU. Alla data deponerade i Tilda förses med beständig identifierare (DOI) och tydliga licenser vilket underlättar spridning och citering av dina data. Alla data arkiveras också enligt gällande svensk arkivpraxis vilket borgar för långsiktig datakvalitet och säkerhet.

Tilda är inte tänkt som en arbetsyta där data kan mellanlagras.

För att Tilda och den nya stödfunktionen DCUs\* arbete ska blir så välunderbyggt som möjligt vill vi att du svarar på några korta frågor om vad du arbetar med för data och hur du hanterar dem.

Svaren behandlas anonymt. Vill du veta mer om Tilda eller har frågor om datahantering, tveka inte att kontakta DCU\* ([dcu@slu.se](mailto:dcu@slu.se) (<mailto:dcu@slu.se>)). Information om Tilda finns också på webbsidan (<https://internt.slu.se/stod-service/admin-stod/it/tjanster/systemutveckling/verktyg-och-system-utveckling/tilda/>).

*\*) Sekretariatet för arkivering och publicering av forskningsdata, Data Curation Unit*

*(DCU)*





## 1. Frågor om forsknings- och miljöanalysdata

### 1 a) Vilka datatyper arbetar du med?

Det går att markera flera svarsalternativ.

- Arbetar inte med digitala data
- Automatgenererade data från datorapplikation
- Databaser
- Digitala fotografier och andra rasterbilder
- Digitala ljudfiler
- Digitala videofiler
- Dokument eller rapporter
- Egenutvecklad mjukvara inom projekt
- Geografiska data
- Kalkylark
- Numeriska data
- Sensordata eller data insamlade mha instrument
- Statistisk data
- Textuell data
- Vektorgrafik/ritningar
- Webplatser
- XML, JSON eller dylikt
- Andra datatyper, ange gärna vilka \_\_\_\_\_

### 1 b) Vilka programvaror använder du i din datahantering?

---

---

---

---

---

### 1 c) Sparar du data i andra format än de som är standard i de programvaror du angett i förra frågan? Ange i sådana fall vilka.

---

---

---

---

---



### 1 d) Hur stora datamängder behöver du kunna publicera och arkivera?

Det går att markera flera svarsalternativ.

- 1-500 MB
- 500 MB - 1 GB
- 1 - 500 GB
- 500 GB - 1 TB
- > 1TB
- Ange i fritext om inget av ovanstående alternativ är lämpligt:

\_\_\_\_\_

### 1 e) Hanterar du känsliga data?

Det går att markera flera svarsalternativ.

- Osäker på om jag hanterar känsliga data.
- Nej, jag hanterar inte känsliga data.
- Ja, sekretessbelagda data. *Se definition nedan.*
- Ja, personuppgifter.
- Ja, annan känslig data. *Ange gärna vad för slag.* \_\_\_\_\_

### 1 f) I vilken grad känner du till de arkivkrav en myndighet ska uppfylla och hur dessa påverkar din hantering av data?

- Stor grad
- I viss grad
- I liten grad
- Inte alls

### 1 g) Deltar du i externa samarbeten eller samarbetsavtal?

- Ja
- Nej

### 1h) Använder du något system för versionskontroll av data?

Det går att markera flera svarsalternativ.

- Nej, använder inget sådant system
- Ja, manuell hantering genom namngivning av filer eller kataloger
- Ja, system för automatiserad versionskontroll (Git, SVN eller liknande). Ange gärna vilket/vilka

\_\_\_\_\_

Definition av **sekretessbelagd** uppgift: [http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/offentlighets--och-sekretesslag-2009400\\_sfs-2009-400](http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/offentlighets--och-sekretesslag-2009400_sfs-2009-400)  
([http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/offentlighets--och-sekretesslag-2009400\\_sfs-2009-400](http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/offentlighets--och-sekretesslag-2009400_sfs-2009-400))

#### Language

- English
- Svenska

## 2. Öppna data

MIKAELA ASPLUND



SÖK

HJÄLP



### 2 a) Har du delat eller tillgängliggjort forskningsdata/miljöanalysdata?

Det går att markera flera svarsalternativ.

- Nej.
- Ja, jag har delat med nära forskarkollegor (informellt).
- Ja, jag har delat data direkt vid förfrågan.
- Ja, jag har tillgängliggjort på webbplats (forskningsprojekt eller personlig).
- Ja, jag har tillgängliggjort via dataportal (datacenter) eller dataarkiv. Om ja, ange gärna i vilken tjänst. \_\_\_\_\_

### 2 b) I vilken grad kan du tänka dig att öppet dela den data du arbetar med?

- Stor grad (allt eller så mycket som möjligt)
- I viss grad (större utvalda delar)
- I liten grad (mindre utvalda data)
- Inte alls
- Data går inte att dela pga sekretess eller liknande

### 2 c) Anser du att det är viktigt att tillgängliggöra data till forskare och till allmänheten?

- Absolut inte
- 2
- 3
- 4
- Mycket viktigt

### 2 d) Motivera gärna ditt val i föregående fråga (2c)

---

---

---

---

---

**2 e) Vad skulle få dig att börja, eller i större utsträckning, använda tillgänglig data?**

*Det går att markera flera svarsalternativ.*

- Efterfrågan från andra forskare i mitt ämne
- Mer kunskap om hur data kan delas
- Mer stöd
- Tillgång till verktyg/plattform för att dela data
- Tillgängliggörande av data skulle ses som en merit
- Mätning av genomslag (impact) för den öppna datan
- Policy eller rektorsbeslut vid SLU för öppna data
- Vet ej
- Annat \_\_\_\_\_

**2 f) Har du en datahanteringsplan för din verksamhet?**

- ja
- nej
- jag vet inte vad en datahanteringsplan är



SÖK

HJÄLP



### 3. Kontextuella frågor



#### 3 a) Vilken är din primära hemmahörande institution/organisation:

- Institutionen för akvatiska resurser (SLU Aqua)
- Institutionen för anatomi, fysiologi och biokemi
- Institutionen för arbetsvetenskap, ekonomi och miljöpsykologi (AEM)
- Institutionen för biomedicin och veterinär folkhälsovetenskap
- Institutionen för biosystem och teknologi
- Institutionen för ekologi
- Institutionen för ekonomi
- Institutionen för energi och teknik
- Institutionen för husdjurens miljö och hälsa
- Institutionen för husdjurens utfodring och vård
- Institutionen för husdjursgenetik
- Institutionen för kliniska vetenskaper
- Institutionen för landskapsarkitektur, planering och förvaltning
- Institutionen för mark och miljö
- Institutionen för molekylära vetenskaper
- Institutionen för norrländsk jordbruksvetenskap
- Institutionen för skogens biomaterial och teknologi (SBT)
- Institutionen för skogens ekologi och skötsel
- Institutionen för skogens produkter
- Institutionen för skoglig genetik och växtfysiologi
- Institutionen för skoglig mykologi och växtpatologi
- Institutionen för skoglig resurshushållning
- Institutionen för skogsekonomi
- Institutionen för stad och land
- Institutionen för sydsvensk skogsvetenskap
- Institutionen för vatten och miljö
- Institutionen för vilt, fisk och miljö
- Institutionen för växtbiologi
- Institutionen för växtförädling
- Institutionen för växtproduktionsekologi
- Institutionen för växtskyddsbiologi
- Skogsmästarskolan
- Annan \_\_\_\_\_

#### 3 b) Tillhör du:

- ett forskningscentrum
- en miljöanalysverksamhet
- annan verksamhet, ange gärna vilken \_\_\_\_\_

**3 c) Ange vilket forskningscentrum alternativt miljöansvarig verksamhet**

SÖK

HJÄLP

---

---

---

---

---

**3 d) Vilken titel beskriver din tjänst bäst?**

- Professor
- Doktorand
- Lektor, postdoc eller annan forskartjänst
- Annan tjänst \_\_\_\_\_

**3 e) Är din nuvarande verksamhet helt eller delvis externfinansierad?**

- nej
- ja (specificera gärna vilken/vilka finansiär) \_\_\_\_\_

**3 f) Övriga synpunkter?**

---

---

---

---

---

**Tack för din medverkan!**

Vid frågor om datapublicering och -arkivering mejla [dcu@slu.se](mailto:dcu@slu.se)

Om du har frågor eller funderingar finns vårt Support Center som hjälp.