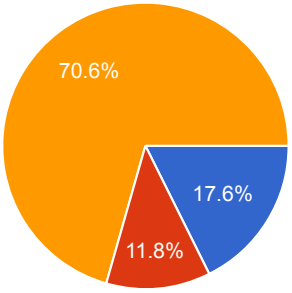
34 responses

Summary

Part 1. Organization profile

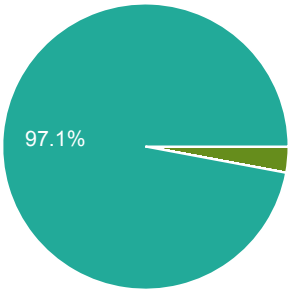
How big is your organisation?



Small (1-50 employees)	**6**	17.6%
Medium (50-250 employees)	**4**	11.8%
Large (250+ employees)	**24**	70.6%

Part 1. Organization profile

Where is your organisation located?



Afghanistan	**0**	0%
Akrotiri	**0**	0%
Albania	**0**	0%
Algeria	**0**	0%
American Samoa	**0**	0%

Andorra	**0**	0%
Angola	**0**	0%
Anguilla	**0**	0%
Antarctica	**0**	0%
Antigua and Barbuda	**0**	0%
Argentina	**0**	0%
Armenia	**0**	0%
Aruba	**0**	0%
Ashmore and Cartier Islands	**0**	0%
Australia	**0**	0%
Austria	**0**	0%
Azerbaijan	**0**	0%
Bahamas, The	**0**	0%
Bahrain	**0**	0%
Bangladesh	**0**	0%
Barbados	**0**	0%
Bassas da India	**0**	0%
Belarus	**0**	0%
Belgium	**0**	0%
Belize	**0**	0%
Benin	**0**	0%
Bermuda	**0**	0%
Bhutan	**0**	0%
Bolivia	**0**	0%
Bosnia and Herzegovina	**0**	0%
Botswana	**0**	0%
Bouvet Island	**0**	0%
Brazil	**0**	0%
British Indian Ocean Territory	**0**	0%
British Virgin Islands	**0**	0%
Brunei	**0**	0%
Bulgaria	**0**	0%
Burkina Faso	**0**	0%
Burma	**0**	0%
Burundi	**0**	0%
Cambodia	**0**	0%
Cameroon	**0**	0%
Canada	**0**	0%
Cape Verde	**0**	0%
Cayman Islands	**0**	0%
Central African Republic	**0**	0%
Chad	**0**	0%
Chile	**0**	0%
China	**0**	0%

Christmas Island	**0**	0%
Clipperton Island	**0**	0%
Cocos (Keeling) Islands	**0**	0%
Colombia	**0**	0%
Comoros	**0**	0%
Congo, Democratic Republic of the	**0**	0%
Congo, Republic of the	**0**	0%
Cook Islands	**0**	0%
Coral Sea Islands	**0**	0%
Costa Rica	**0**	0%
Cote d'Ivoire	**0**	0%
Croatia	**0**	0%
Cuba	**0**	0%
Cyprus	**0**	0%
Czech Republic	**0**	0%
Denmark	**0**	0%
Dhekelia	**0**	0%
Djibouti	**0**	0%
Dominica	**0**	0%
Dominican Republic	**0**	0%
Ecuador	**0**	0%
Egypt	**0**	0%
El Salvador	**0**	0%
Equatorial Guinea	**0**	0%
Eritrea	**0**	0%
Estonia	**0**	0%
Ethiopia	**0**	0%
Europa Island	**0**	0%
Falkland Islands (Islas Malvinas)	**0**	0%
Faroe Islands	**0**	0%
Fiji	**0**	0%
Finland	**0**	0%
France	**0**	0%
French Guiana	**0**	0%
French Polynesia	**0**	0%
French Southern and Antarctic Lands	**0**	0%
Gabon	**0**	0%
Gambia, The	**0**	0%
Gaza Strip	**0**	0%
Georgia	**0**	0%
Germany	**1**	2.9%
Ghana	**0**	0%
Gibraltar	**0**	0%
Glorioso Islands	**0**	0%

Greece	**0**	0%
Greenland	**0**	0%
Grenada	**0**	0%
Guadeloupe	**0**	0%
Guam	**0**	0%
Guatemala	**0**	0%
Guernsey	**0**	0%
Guinea	**0**	0%
Guinea-Bissau	**0**	0%
Guyana	**0**	0%
Haiti	**0**	0%
Heard Island and McDonald Islands	**0**	0%
Holy See (Vatican City)	**0**	0%
Honduras	**0**	0%
Hong Kong	**0**	0%
Hungary	**0**	0%
Iceland	**0**	0%
India	**0**	0%
Indonesia	**0**	0%
Iran	**0**	0%
Iraq	**0**	0%
Ireland	**0**	0%
Isle of Man	**0**	0%
Israel	**0**	0%
Italy	**0**	0%
Jamaica	**0**	0%
Jan Mayen	**0**	0%
Japan	**0**	0%
Jersey	**0**	0%
Jordan	**0**	0%
Juan de Nova Island	**0**	0%
Kazakhstan	**0**	0%
Kenya	**0**	0%
Kiribati	**0**	0%
Korea, North	**0**	0%
Korea, South	**0**	0%
Kuwait	**0**	0%
Kyrgyzstan	**0**	0%
Laos	**0**	0%
Latvia	**0**	0%
Lebanon	**0**	0%
Lesotho	**0**	0%
Liberia	**0**	0%
Libya	**0**	0%

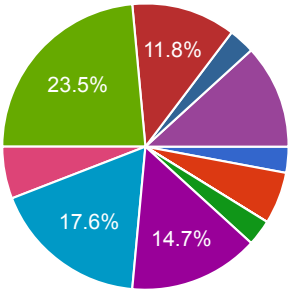
Liechtenstein	**0**	0%
Lithuania	**0**	0%
Luxembourg	**0**	0%
Macau	**0**	0%
Macedonia	**0**	0%
Madagascar	**0**	0%
Malawi	**0**	0%
Malaysia	**0**	0%
Maldives	**0**	0%
Mali	**0**	0%
Malta	**0**	0%
Marshall Islands	**0**	0%
Martinique	**0**	0%
Mauritania	**0**	0%
Mauritius	**0**	0%
Mayotte	**0**	0%
Mexico	**0**	0%
Micronesia, Federated States of	**0**	0%
Moldova	**0**	0%
Monaco	**0**	0%
Mongolia	**0**	0%
Montenegro	**0**	0%
Montserrat	**0**	0%
Morocco	**0**	0%
Mozambique	**0**	0%
Namibia	**0**	0%
Nauru	**0**	0%
Navassa Island	**0**	0%
Nepal	**0**	0%
Netherlands	**33**	97.1%
Netherlands Antilles	**0**	0%
New Caledonia	**0**	0%
New Zealand	**0**	0%
Nicaragua	**0**	0%
Niger	**0**	0%
Nigeria	**0**	0%
Niue	**0**	0%
Norfolk Island	**0**	0%
Northern Mariana Islands	**0**	0%
Norway	**0**	0%
Oman	**0**	0%
Pakistan	**0**	0%
Palau	**0**	0%
Panama	**0**	0%

Papua New Guinea	**0**	0%
Paracel Islands	**0**	0%
Paraguay	**0**	0%
Peru	**0**	0%
Philippines	**0**	0%
Pitcairn Islands	**0**	0%
Poland	**0**	0%
Portugal	**0**	0%
Puerto Rico	**0**	0%
Qatar	**0**	0%
Reunion	**0**	0%
Romania	**0**	0%
Russia	**0**	0%
Rwanda	**0**	0%
Saint Helena	**0**	0%
Saint Kitts and Nevis	**0**	0%
Saint Lucia	**0**	0%
Saint Pierre and Miquelon	**0**	0%
Saint Vincent and the Grenadines	**0**	0%
Samoa	**0**	0%
San Marino	**0**	0%
Sao Tome and Principe	**0**	0%
Saudi Arabia	**0**	0%
Senegal	**0**	0%
Serbia	**0**	0%
Seychelles	**0**	0%
Sierra Leone	**0**	0%
Singapore	**0**	0%
Slovakia	**0**	0%
Slovenia	**0**	0%
Solomon Islands	**0**	0%
Somalia	**0**	0%
South Africa	**0**	0%
South Georgia and the South Sandwich Islands	**0**	0%
Spain	**0**	0%
Spratly Islands	**0**	0%
Sri Lanka	**0**	0%
Sudan	**0**	0%
Suriname	**0**	0%
Svalbard	**0**	0%
Swaziland	**0**	0%
Sweden	**0**	0%
Switzerland	**0**	0%
Syria	**0**	0%

Taiwan	**0**	0%
Tajikistan	**0**	0%
Tanzania	**0**	0%
Thailand	**0**	0%
Timor-Leste	**0**	0%
Togo	**0**	0%
Tokelau	**0**	0%
Tonga	**0**	0%
Trinidad and Tobago	**0**	0%
Tromelin Island	**0**	0%
Tunisia	**0**	0%
Turkey	**0**	0%
Turkmenistan	**0**	0%
Turks and Caicos Islands	**0**	0%
Tuvalu	**0**	0%
Uganda	**0**	0%
Ukraine	**0**	0%
United Arab Emirates	**0**	0%
United Kingdom	**0**	0%
United States	**0**	0%
Uruguay	**0**	0%
Uzbekistan	**0**	0%
Vanuatu	**0**	0%
Venezuela	**0**	0%
Vietnam	**0**	0%
Virgin Islands	**0**	0%
Wake Island	**0**	0%
Wallis and Futuna	**0**	0%
West Bank	**0**	0%
Western Sahara	**0**	0%
Yemen	**0**	0%
Zambia	**0**	0%
Zimbabwe	**0**	0%

Part 1. Organization profile

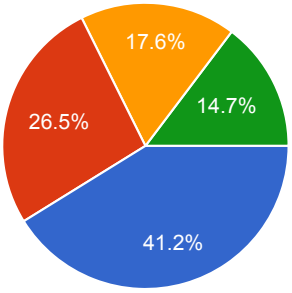
What is your function in the organization?



		9%
		9%
		0%
		9%
Head of bioinformatics / data science department	**5**	14.7%
Head of research group	**6**	17.6%
Research scientist (non-computational)	**2**	5.9%
Research scientist (computational)	**8**	23.5%
Project manager	**4**	11.8%
Marketing Director / Manager	**1**	2.9%
Other	**4**	11.8%

Part 1. Organization profile

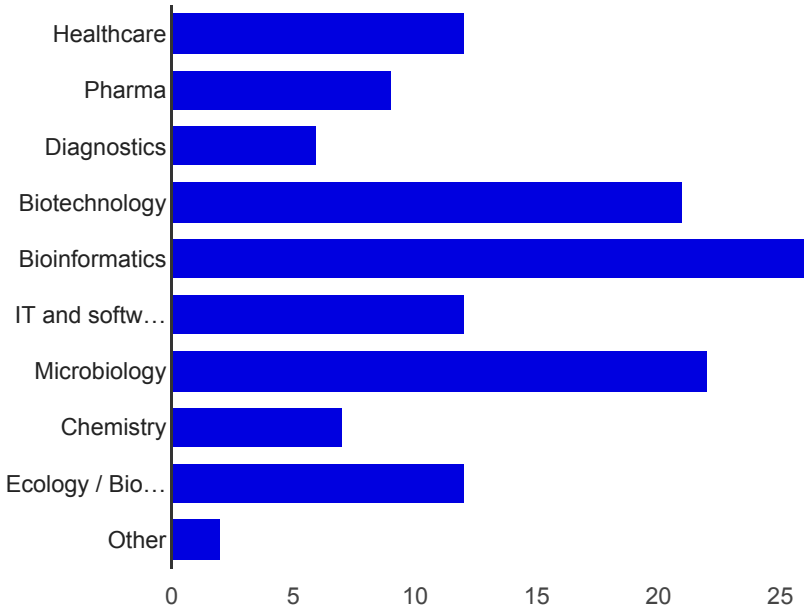
Which of the following describes your organisation best?



Academic institution / University	**14**	41.2%
Industry	**9**	26.5%
Research institute	**6**	17.6%
Service provider / Contract research organization (CRO)	**5**	14.7%
Collaborative venture / umbrella organisation	**0**	0%
Other	**0**	0%

Part 1. Organization profile

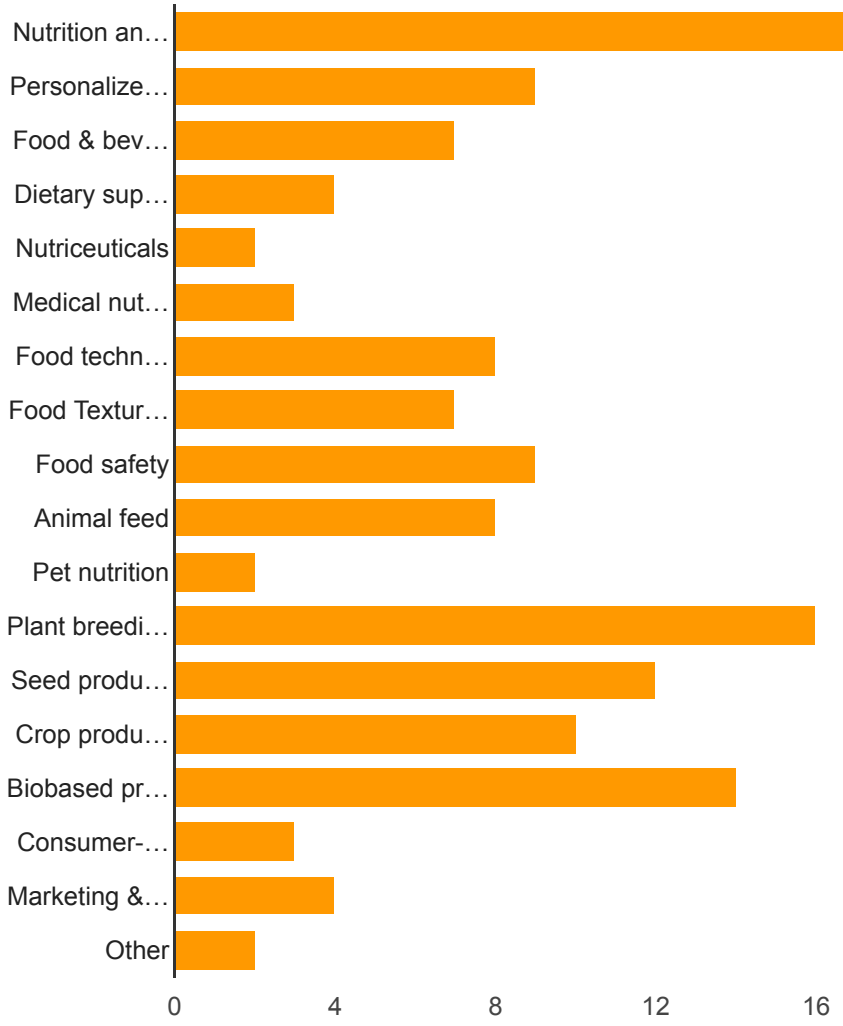
In addition to agri & food, in which other research areas does your organisation also operate?



Pharma	**9**	29%
Diagnostics	**6**	19.4%
Biotechnology	**21**	67.7%
Bioinformatics	**28**	90.3%
IT and software development	**12**	38.7%
Microbiology	**22**	71%
Chemistry	**7**	22.6%
Ecology / Biodiversity	**12**	38.7%
Other	**2**	6.5%

Part 1. Organization profile

Within the agri & food research area, what are the main research topics in your organisation?



Seed production optimisation	**12**	36.4%
Crop production optimisation	**10**	30.3%
Biobased products	**14**	42.4%
Consumer-driven supply chain	**3**	9.1%
Marketing & business intelligence	**4**	12.1%
Other	**2**	6.1%

Part 1. Organization profile

How many employees in your organisation are working in research?

5000

1000

2000

100

150

400

10

0

30

6000

700

800

380

4500

20000

19

300

2500

3

15

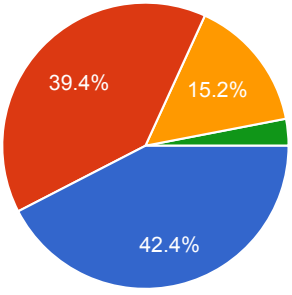
4

90

50

Part 1. Organization profile

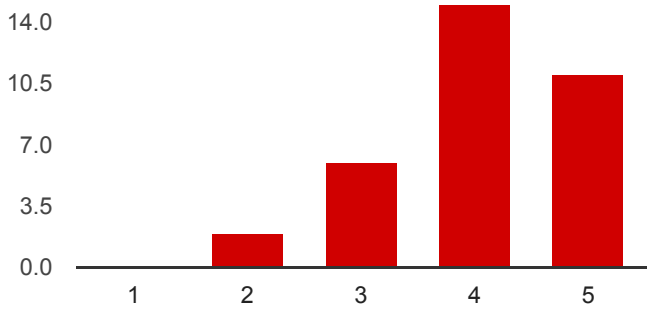
What is the main funding resource for research in your organisation?



Internal resources	**14**	42.4%
Public resources, national	**13**	39.4%
Public resources, EU	**5**	15.2%
Other	**1**	3%

Part 2. Current data practices

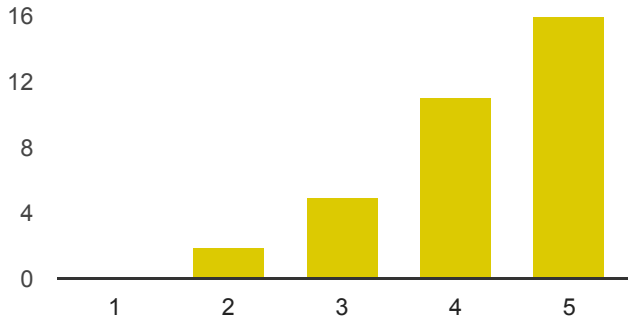
How familiar are you with your organization's use of or plans for data storage, handling and analysis?



Not familiar: 1	**0**	0%
2	**2**	5.9%
3	**6**	17.6%
4	**15**	44.1%
Fully familiar: 5	**11**	32.4%

Part 2. Current data practices

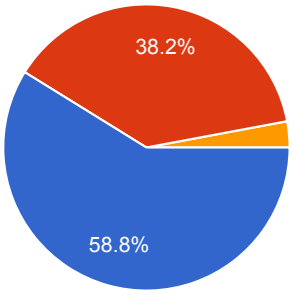
Would you describe research within your organisation as data-intensive?



Not at all data-intensive: 1	**0**	0%
2	**2**	5.9%
3	**5**	14.7%
4	**11**	32.4%
Very data-intensive: 5	**16**	47.1%

Part 2. Current data practices

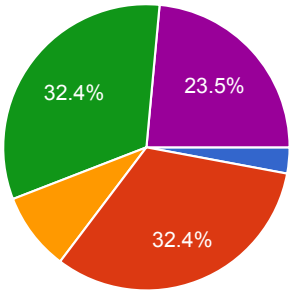
Is there a dedicated team or department within your organisation responsible for data storage, handling and/or analysis tasks?



Yes	**20**	58.8%
No	**13**	38.2%
Other	**1**	2.9%

Part 2. Current data practices

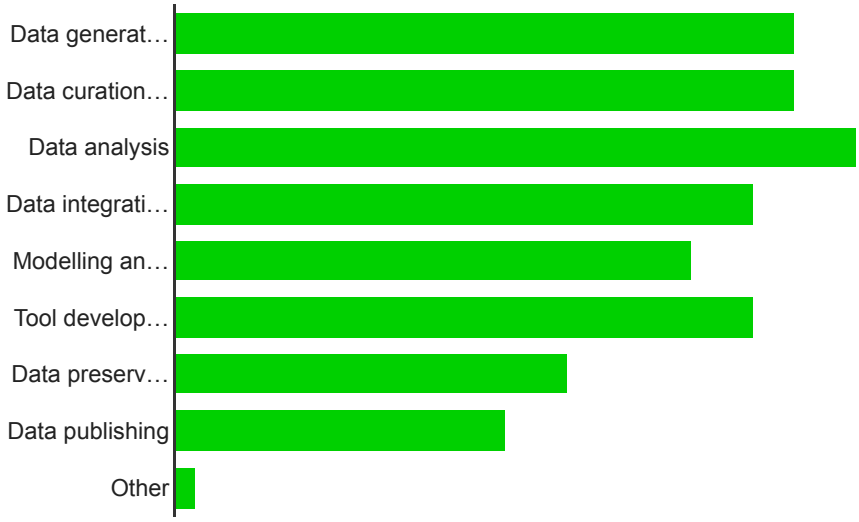
How many employees in your organisation are currently working exclusively on data storage, handling and/or analysis tasks?



No employees have exclusively data-related tasks	**1**	2.9%
< 5	**11**	32.4%
5 - 10	**3**	8.8%
> 10	**11**	32.4%
Not sure	**8**	23.5%

Part 2. Current data practices

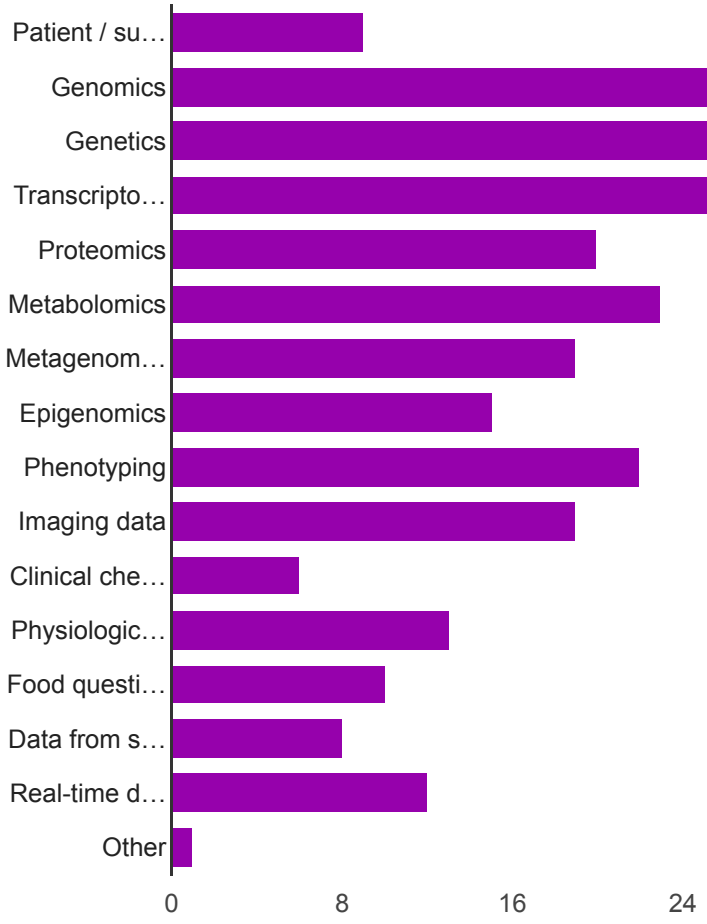
Which data-related aspects are covered by your organisation's research?



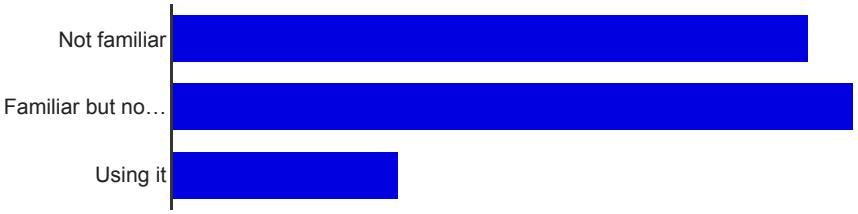
Other	**1**	2.9%

Part 2. Current data practices

What type of data is generated and/or used within your organisation's research activities?



ArrayExpress [Are you familiar with following public repositories?]



Not familiar	**14**	41.2%
Familiar but not using it	**15**	44.1%
Using it	**5**	14.7%

Gene Expression Omnibus (GEO) [Are you familiar with following public repositories?]

Not familiar	**16**	47.1%
Familiar but not using it	**8**	23.5%
Using it	**10**	29.4%

PRoteomics IDEntifications (PRIDE) database [Are you familiar with following public repositories?]

Not familiar	**19**	55.9%
Familiar but not using it	**12**	35.3%
Using it	**3**	8.8%

PeptideAtlas [Are you familiar with following public repositories?]

Not familiar	**18**	52.9%
Familiar but not using it	**13**	38.2%
Using it	**3**	8.8%

Human Protein Atlas [Are you familiar with following public repositories?]

Not familiar	**13**	38.2%
Familiar but not using it	**15**	44.1%
Using it	**6**	17.6%

European Genome-phenome Archive (EGA) [Are you familiar with following public repositories?]

Not familiar	**26**	76.5%
Familiar but not using it	**5**	14.7%
Using it	**3**	8.8%

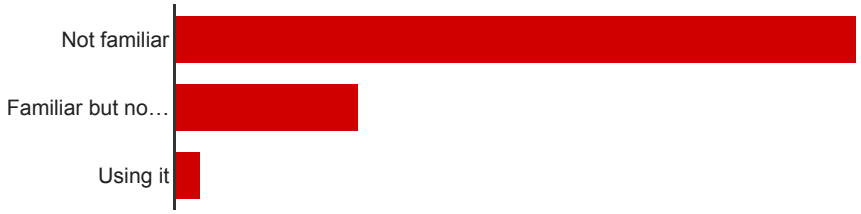
ClinVar [Are you familiar with following public repositories?]

Not familiar	**28**	82.4%
Familiar but not using it	**6**	17.6%
Using it	**0**	0%

The database of Genotypes and Phenotypes (dbGaP) [Are you familiar with following public repositories?]

Not familiar	**23**	67.6%
Familiar but not using it	**9**	26.5%
Using it	**2**	5.9%

Nutritional Phenotype database (dbNP) [Are you familiar with following public repositories?]



Not familiar	**26**	76.5%
Familiar but not using it	**7**	20.6%
Using it	**1**	2.9%

MetaboLights [Are you familiar with following public repositories?]

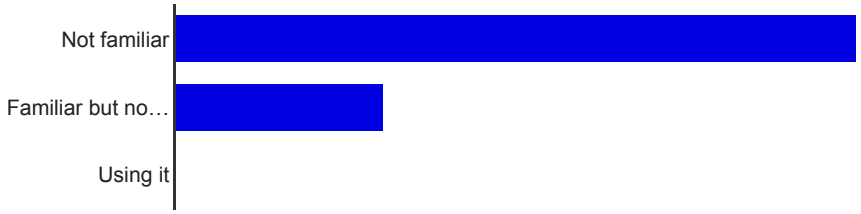
Not familiar	**25**	75.8%
Familiar but not using it	**6**	18.2%
Using it	**2**	6.1%

MassBank [Are you familiar with following public repositories?]

Not familiar	**25**	73.5%
Familiar but not using it	**8**	23.5%
Using it	**1**	2.9%

FigShare [Are you familiar with following public repositories?]

Dryad [Are you familiar with following public repositories?]



Not familiar	**26**	76.5%
Familiar but not using it	**8**	23.5%
Using it	**0**	0%

GigaDb [Are you familiar with following public repositories?]

Not familiar	**25**	73.5%
Familiar but not using it	**7**	20.6%
Using it	**2**	5.9%

Nature Scientific Data [Are you familiar with following public repositories?]

Not familiar	**17**	50%
Familiar but not using it	**12**	35.3%
Using it	**5**	14.7%

Dataverse [Are you familiar with following public repositories?]

Not familiar	**29**	87.9%
Familiar but not using it	**4**	12.1%
Using it	**0**	0%

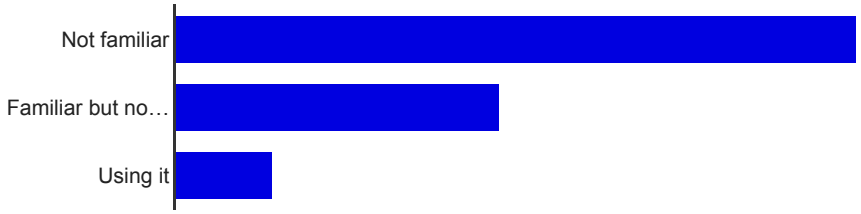
Mendeley Data [Are you familiar with following public repositories?]

Not familiar	**21**	61.8%
Familiar but not using it	**9**	26.5%
Using it	**4**	11.8%

Zenodo [Are you familiar with following public repositories?]

Not familiar	**28**	84.8%
Familiar but not using it	**4**	12.1%
Using it	**1**	3%

DataHub [Are you familiar with following public repositories?]



Not familiar	**21**	61.8%
Familiar but not using it	**10**	29.4%
Using it	**3**	8.8%

Data Archiving and Networked Services (DANS) [Are you familiar with following public repositories?]

Not familiar	**22**	64.7%
Familiar but not using it	**9**	26.5%
Using it	**3**	8.8%

EUDat [Are you familiar with following public repositories?]

Not familiar	**25**	73.5%
Familiar but not using it	**9**	26.5%
Using it	**0**	0%

NEUROVAULT [Are you familiar with following public repositories?]

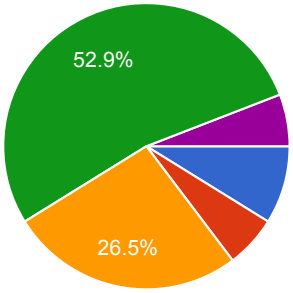
Not familiar	**31**	91.2%
Familiar but not using it	**3**	8.8%
Using it	**0**	0%

The Image & Data Archive (IDA) [Are you familiar with following public repositories?]

Not familiar	**26**	78.8%
Familiar but not using it	**5**	15.2%
Using it	**2**	6.1%

Part 2. Current data practices

Do you use external data from public repositories in your research?



No, there is no need for it	**3**	8.8%
Not yet, but I am interested in knowing more about it	**2**	5.9%
Yes, sporadically	**9**	26.5%
Yes, regularly	**18**	52.9%
Not sure	**2**	5.9%

Part 2. Current data practices

Which public repositories do you use data from?

NCBI ENA UNIPROT

GEO, PRIDE/ProteomeXchange, PeptideAtlas, HPA, UniProt

NCBI, Ensembl, SRA

Metacyc - molar volumes of proteins BRENDA - turnover data of proteins Biomodels.net - models plus stoichiometric matrices

PDB, BMRB, Uniprot

Uniprot Genbank EMBL Etc

GEO, NCBI, Genevestigator

many. For example. Km PLotter, Omero

Academic publications, academic genome browser websites

PubMed, Genbank

from challenges

metabolights genbank amdis

sequences, ncbi genetic maps assembled genome expression data

GeneSys: passport data of PGR

tcga, hapmap, geo

uniprot, metacyc, kegg, DSZM, wikidata, reactome

NCBI, EBI, different genome sequencing consortia.

ENA DSMZ GOLD

ENA EBI NCBI SRA

Genome sequence, annotation and variation data from ensemble, UCSC, NCBI Data from SRA, ENA

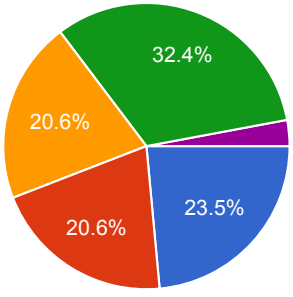
Statline

NCBI, EBI, various specialized databases (MIPS, DIP, Reactome, IntAct, KEGG, SGD, TAIR...)

MEDLINE NCBI taxonomy ChEBI Silva GooglePatents GEO ArrayExpress

Part 2. Current data practices

Do you submit / share any data generated within your organisation through public repositories?



No, there is no need for it	**8**	23.5%
Not yet, but I am interested in knowing more about it	**7**	20.6%
Yes, sporadically	**7**	20.6%
Yes, regularly	**11**	32.4%
Not sure	**1**	2.9%

Part 2. Current data practices

To which public repositories do you submit / share data?

ENA

ProteomeXchange/MassIVE

DANS - imaging data with experimental collaborator and simulation (meta)data

Google code - source code

PDB, BMRB

Genbank

GEO, SOL genomics

genbank

metabolights

EURISCO: PGR passport data and phenotypic data

geo

EBI genomes

EBI ENA

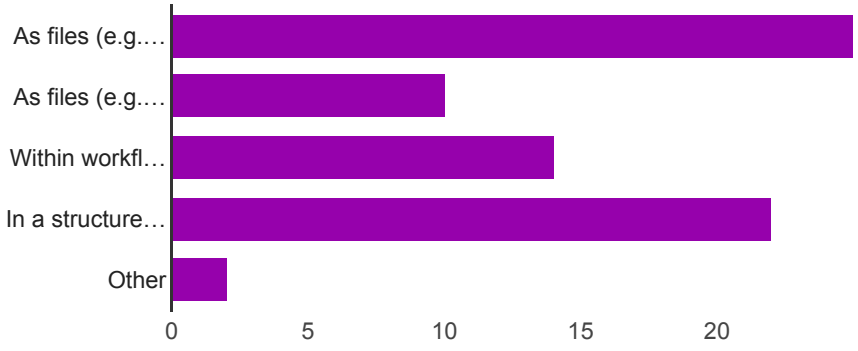
ENA Genbank NCBI dbSNP

NCBI, EBI

NCBI genome database GEO

Part 2. Current data practices

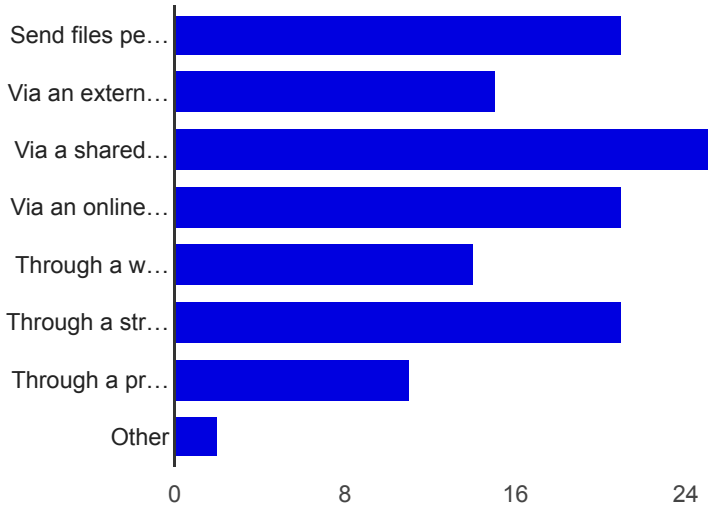
How do you internally store data?



As files (e.g. Excel) on a shared network / intranet drive	**25**	75.8%
As files (e.g. Excel) on a local harddrive, USB stick or external drive	**10**	30.3%
Within workflow or other data analysis software tool	**14**	42.4%
In a structured database or data repository	**22**	66.7%
Other	**2**	6.1%

Part 2. Current data practices

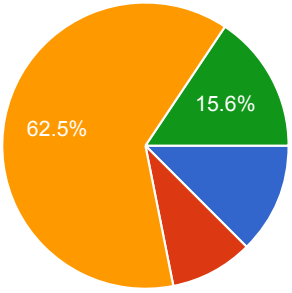
How do you internally share data?



Send files per e-mail	**21**	61.8%
Via an external medium (e.g. USB stick, external harddrive, DVD)	**15**	44.1%
Via a shared network / intranet drive	**32**	94.1%
Via an online service (e.g. DropBox, WeTransfer)	**21**	61.8%
Through a workflow / software tool	**14**	41.2%
Through a structured database or data repository	**21**	61.8%
Through a programmatic interface (e.g. API)	**11**	32.4%
Other	**2**	5.9%

Part 2. Current data practices

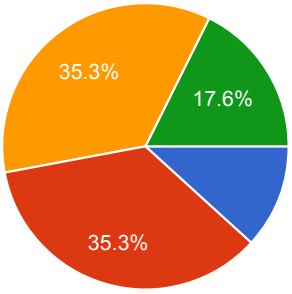
If a structured database or data repository is used for storing and/or sharing data within your organisation, which solution do you use?



	4	12.5%
	3	9.4%
None of above	**20**	62.5%
Other	**5**	15.6%

Part 2. Current data practices

Is data within your organisation stored and/or shared in standardized formats?



No, there is no need for it	**4**	11.8%
No, but I am interested in knowing more about it	**12**	35.3%
Yes	**12**	35.3%
Not sure	**6**	17.6%

Part 2. Current data practices

Which standardized formats (e.g. CDISC, ISA-Tab, RDF) do you use?

RDF FASTA FASTQ EMBL

other, typically XML based, community-developed

RDF, XML

fastq, bam, vcf, gff3, ...

GAR (Genalice Aligned Reads) GVM (Genalice Variant Map) VCF

MCPD (Multi-Crop Passport Descriptor List)

RDF

RDF, own rules.

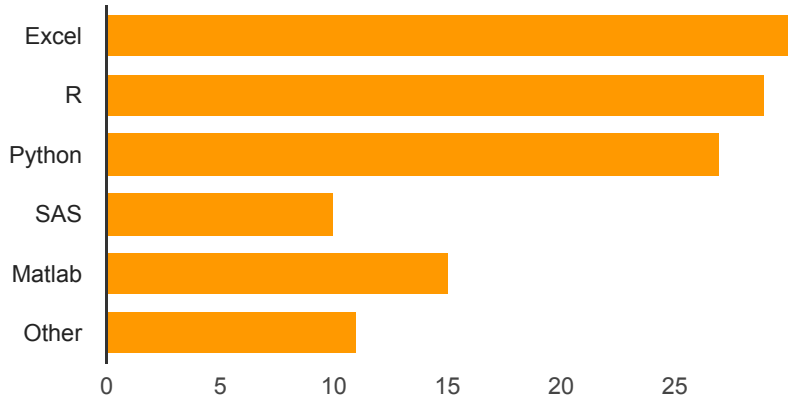
depends on type of data

BIOM format

RDF CSV XML TSV

Part 2. Current data practices

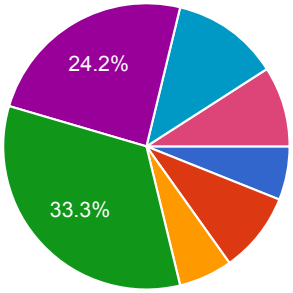
Which data analysis tools / platforms are used within your organisation?



Excel	**30**	88.2%
R	**29**	85.3%
Python	**27**	79.4%
SAS	**10**	29.4%
Matlab	**15**	44.1%
Other	**11**	32.4%

Part 2. Current data practices

Do you use workflow platforms within your organisation?

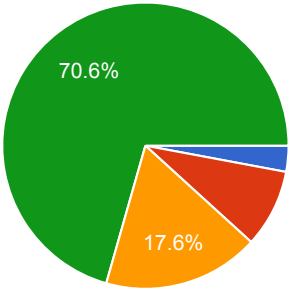


No, there is no need for it	**2**	6.1%
No, but I am interested in knowing more about it	**3**	9.1%
Yes, commercial workflow tools (e.g. RapidMiner)	**2**	6.1%
Yes, open source workflow tools (e.g. Taverna, Galaxy, Knime)	**11**	33.3%

Yes, a customly designed solution	**8**	24.2%
Not sure	**4**	12.1%
Other	**3**	9.1%

Part 2. Current data practices

Do you perform data integration in your research?



No, there is no need for it	**1**	2.9%
Not yet, but I am interested in knowing more about it	**3**	8.8%
Yes, sporadically	**6**	17.6%
Yes, regularly	**24**	70.6%

Part 2. Current data practices

What types of data do you integrate?

RNAseq Genome \ EMBL

genomics and proteomics, also transcriptomics and proteomics, *omics and phenotypic (clinical) data...

metabolomic/stoichiometric matrices

experimental raw data, processed data, analyzed data, publications, etc

Px Tx Mx Modeling Analytical Fermentation

transcriptomics, genome sequences, proteomics, metabolomics, phenotypes

patient, genetical, biochemical, drugs

Genotype, phenotype, envirotype

genomic, genetic, phenotypic, climate data

All data generated within projects

genomic, metabolomic, transcriptomic, proteomic, phenotypic

NGS sequences genome assemblies public sequence and protein data genome annotations expression data metabolomics genetic maps physical maps

DNA/RNA/phenotypical data

ERP data combined with R&D sequence data

GIS data (climatological and soil data)

genomic and transcriptomic data, with clinical variables

everything related to bacterial genome annotation data

Genomic/transcriptomic/phenomic data and environmental data

Genome annotation Expression Knockout Growth

divers

sequences, variation, annotation, expression data

Data from different kind of sensors and administrations such as farm accountancy data

Genomics, transcriptomics, interactions, networks

Genotype phenotype

Genomics Proteomics Microbiomics Metabolomics SubjectMetadata

Part 2. Current data practices

What tools / platforms / infrastructure do you use for data integration?

Blazegraph Java

Taverna, R, make, ...

custom

Linux PC, windows laptop

Matlab Spotfire Lims

R based

many. Bust mostly: our brains

R, custom built

R and databases with API's

Excel, matlab, R, spotfire, python

in house

oracle, mongodb, redhat, postgresql, python, R, java, API, data warehousing

GENALICE LINK

R, Python

R

my own implemented in R

RDF triples stores and galaxy

proprietary tools

In-house developed software RDF SPARQL BlazeGraph

custom python and perl scripts or open source tools if applicable (e.g. from UCSC)

traditional statistical analysis and simple database structures.

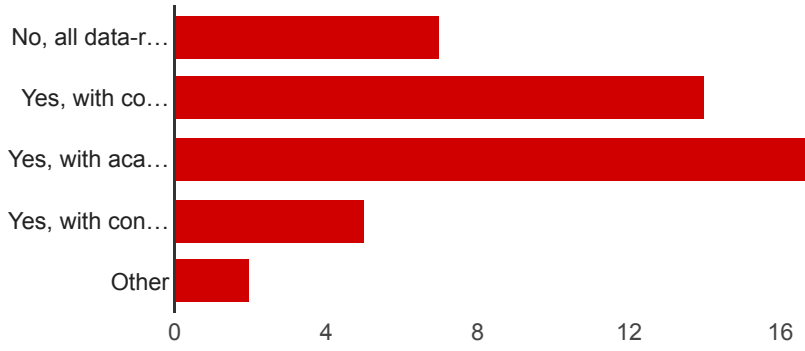
Python, Perl, Matlab, R

In house bioinformatics platform

R Python MySQL

Part 2. Current data practices

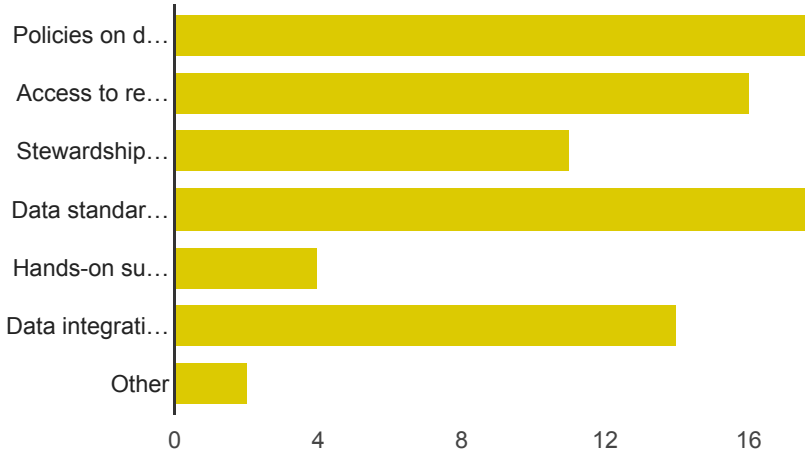
Does your organisation collaborate with external parties in performing data storage, handling and/or analysis tasks?



No, all data-related tasks are currently performed in-house	**7**	20.6%
Yes, with companies providing data-related services	**14**	41.2%
Yes, with academic or research institutions	**18**	52.9%
Yes, with contract research organisations (CROs)	**5**	14.7%
Other	**2**	5.9%

Part 2. Current data practices

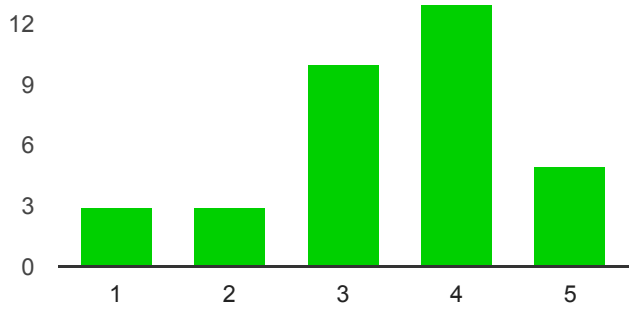
In which of the following tasks would support be most beneficial for your organisation?



Policies on data storage, handling and analysis	**18**	52.9%
Access to resources (tools, databases, standards)	**16**	47.1%
Stewardship and curation of data	**11**	32.4%
Data standardisation and interoperability	**19**	55.9%
Hands-on support in data analysis tasks	**4**	11.8%
Data integration	**14**	41.2%
Other	**2**	5.9%

Part 2. Current data practices

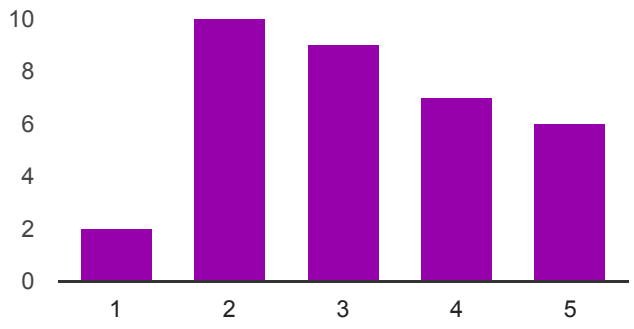
How easy is for you to find or get access to data experts in or outside of your organisation?



Extremely difficult : 1	**3**	8.8%
2	**3**	8.8%
3	**10**	29.4%
4	**13**	38.2%
Very easy: 5	**5**	14.7%

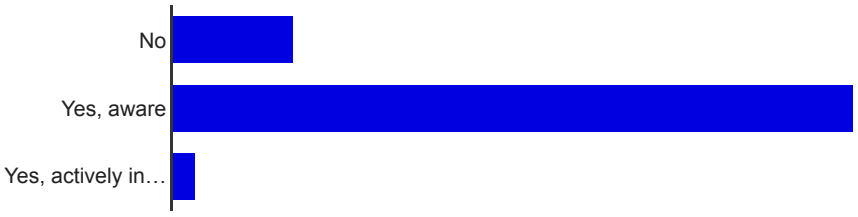
Part 2. Current data practices

How easy is for you to find or get access to training in data expertise?



Part 2. Current data practices

ELIXIR [Are you aware of the following projects and/or umbrella initiatives?]



No	**5**	14.7%
Yes, aware	**28**	82.4%
Yes, actively involved	**1**	2.9%

BBMRI [Are you aware of the following projects and/or umbrella initiatives?]

No	**15**	44.1%
Yes, aware	**16**	47.1%
Yes, actively involved	**3**	8.8%

ISBE [Are you aware of the following projects and/or umbrella initiatives?]

Yes, aware	**6**	17.6%
Yes, actively involved	**1**	2.9%

PRACE [Are you aware of the following projects and/or umbrella initiatives?]

No	**33**	97.1%
Yes, aware	**1**	2.9%
Yes, actively involved	**0**	0%

GÉANT [Are you aware of the following projects and/or umbrella initiatives?]

No	**33**	97.1%
Yes, aware	**1**	2.9%
Yes, actively involved	**0**	0%

EU-DAT [Are you aware of the following projects and/or umbrella initiatives?]

No	**24**	70.6%
Yes, aware	**10**	29.4%
Yes, actively involved	**0**	0%

BD2K [Are you aware of the following projects and/or umbrella initiatives?]

No	**32**	94.1%
Yes, aware	**2**	5.9%
Yes, actively involved	**0**	0%

BioCADDIE [Are you aware of the following projects and/or umbrella initiatives?]

No	**32**	97%
Yes, aware	**1**	3%
Yes, actively involved	**0**	0%

COSMOS [Are you aware of the following projects and/or umbrella initiatives?]

No	**28**	84.8%
Yes, aware	**5**	15.2%

Yes, actively involved	**0**	0%

EOSC [Are you aware of the following projects and/or umbrella initiatives?]

No	**33**	97.1%
Yes, aware	**1**	2.9%
Yes, actively involved	**0**	0%

Euro-BioImaging [Are you aware of the following projects and/or umbrella initiatives?]

No	**28**	82.4%
Yes, aware	**6**	17.6%
Yes, actively involved	**0**	0%

Part 2. Current data practices

What would you want TKI Agri&Food, TIFN and DTL to help realise in terms of strengthening data handling and stewardship in the agri & food sector?

Show case not only political blabla. Show real hands on applied methods to achieve this in the field of biology.

Provide, or lobby for, funding for the development of common data analysis and sharing resources and tools. Courses and training given by experts. Less talk, more action.

Data format standardization Data sharing

Enter talks with the performers and tune to store structured the captured data.

awareness with researchers, financial support for data storage, data management, stewardship, workshops, ...

Provide secure data infrastructure, guidelines an best practices.

standardized formats, easy to use cost effective data storage facilities

Nothing. I think a super structure will slow things down and make them expensive.

Provide guidelines from best-practice

Promote interoperability of data streams and standardization, i.e. end the 'Excel era'.

Best practices in data management.

Provide tools, support, training and services

technology - driven projects

x

suport, help and training and funding.

broad support, including the development of bioinformatics tools and interfaces

help people share data and expertise across areas of study - I work with human genomic and transcriptomic profiles. Methods I use might also be useful in other applications, and vice-versa.

Have discussion/work group, so I know which other people are working with data integration and what they are doing.

help with infrastructure development/maintenance

Give proper big data hands-on practicals. How do I store 10.000 bacterial genomes and in what manner?

standaardization

easy access to HPC and cost effective mass storage. Standardised analyses pipelines

Awareness that FAIR principle is also needed in R&D environments and the academic world.

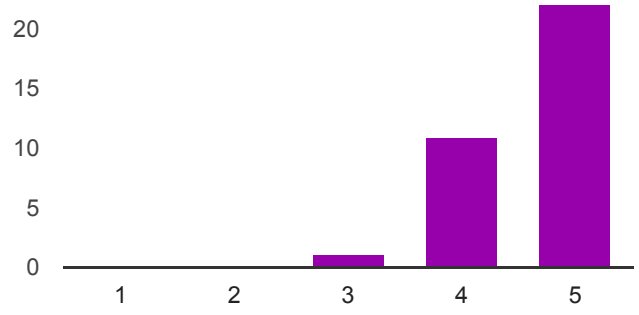
Raise awareness that sharing data benefits everyone in the long run Work on phenotype data standards

Start a public initiative for one crop/ plant species as an example how data can be stored, presented, analyzed, made useful to the breeding community. Invest more money in bioinformatics in general. Now very often no money is available for the bioinformatics part of projects.

Improved access to external data sources Community building around data science

Part 3. Data strategy

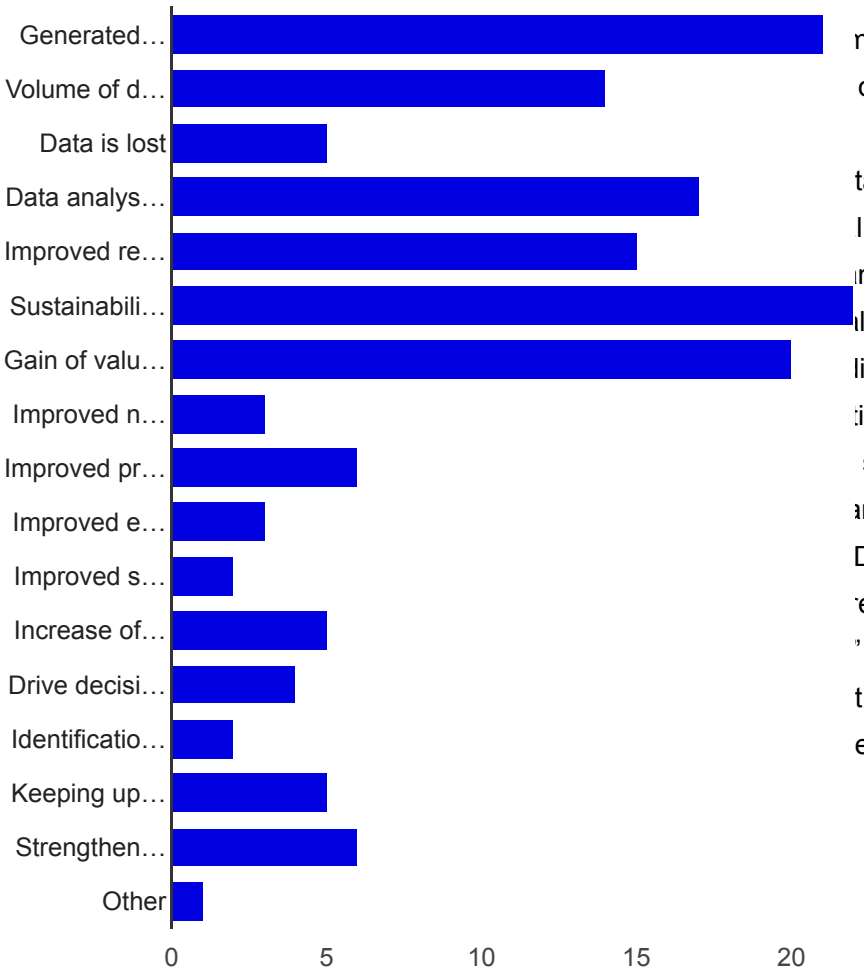
How relevant are optimal data storage, handling and/or analysis solutions for your organisation?



Not relevant: 1	**0**	0%
2	**0**	0%
3	**1**	2.9%
4	**11**	32.4%
Essential: 5	**22**	64.7%

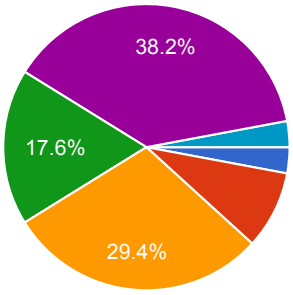
Part 3. Data strategy

What are the main motivation factors for implementing solutions for data storage, handling and analysis within your organisation?



nally exploited	21	61.8%
overwhelming	14	41.2%
Data is lost	5	14.7%
tandardisation	17	50%
lity of analysis	15	44.1%
nd knowledge	22	64.7%
lytic platforms	20	58.8%
lient discovery	3	8.8%
ion processes	6	17.6%
safety of food	3	8.8%
rketing efforts	2	5.9%
D investments	5	14.7%
e investments	4	11.8%
' in the market	2	5.9%
th competitors	5	14.7%
e organisation	6	17.6%
Other	1	2.9%

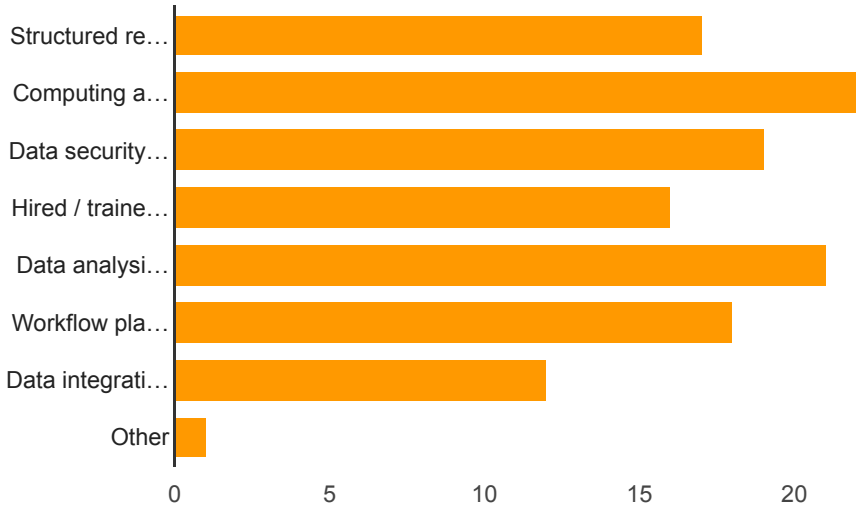
Overall, what is the current status of your organization regarding solutions for data storage, handling and analysis?



Not yet considering implementation	1	2.9%
Getting familiar with existing solutions	3	8.8%
Planning strategy for implementation	10	29.4%
Implementing	6	17.6%
Using	13	38.2%
Not sure	1	2.9%

Part 3. Data strategy

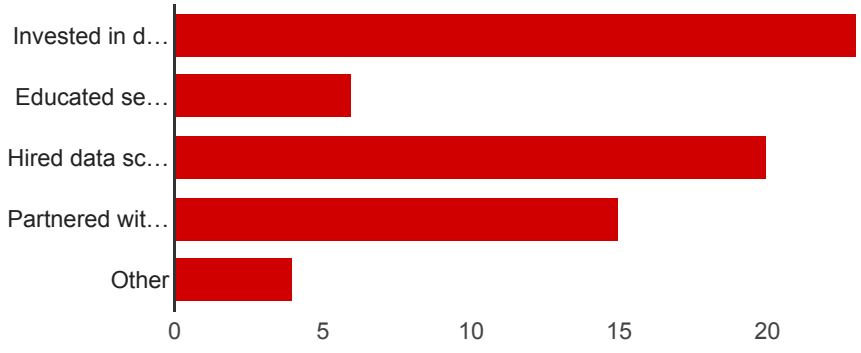
Which aspect of data storage, handling and analysis has already been implemented or is planned to be implemented?



Structured repository for data storage and/or sharing	**17**	51.5%
Computing and storage infrastructure	**22**	66.7%
Data security aspects	**19**	57.6%
Hired / trained dedicated personnel	**16**	48.5%
Data analysis solutions	**21**	63.6%
Workflow platforms	**18**	54.5%
Data integration solutions	**12**	36.4%
Other	**1**	3%

Part 3. Data strategy

How did your organization prepare for opportunities arising from data-driven research?

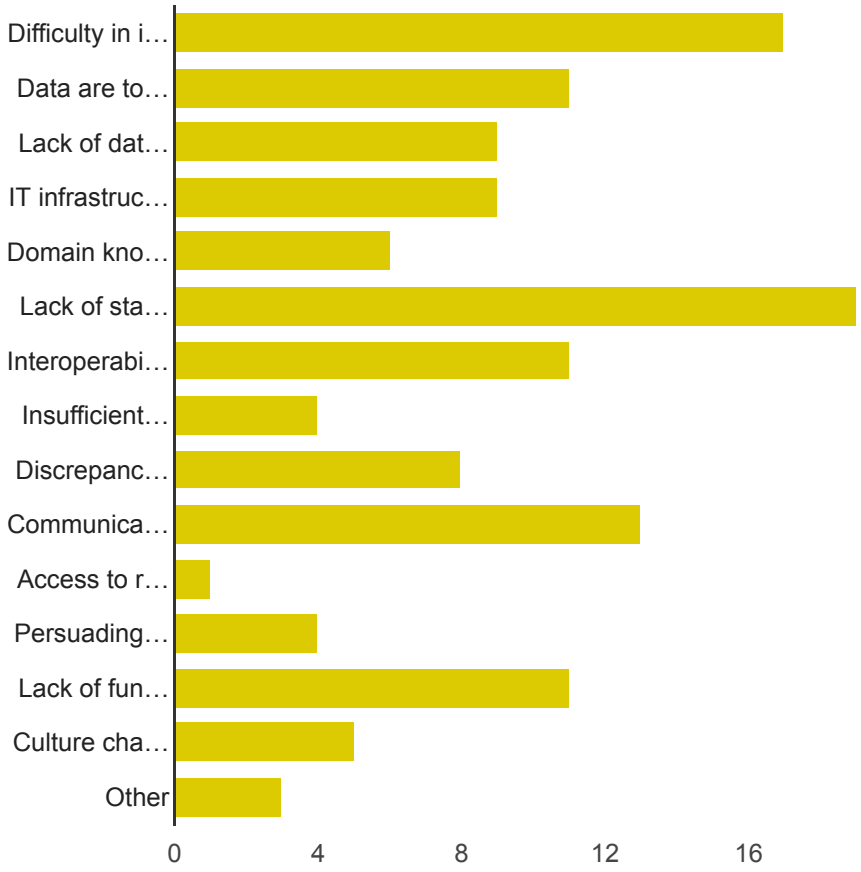


Invested in data storage, handling and/or analysis solutions	**23**	69.7%
Educated senior managers	**6**	18.2%

Hired data scientists	**20**	60.6%
Partnered with external parties with relevant expertise	**15**	45.5%
Other	**4**	12.1%

Part 3. Data strategy

What are the biggest challenges your organization is facing to seize the opportunities arising from data-driven research?

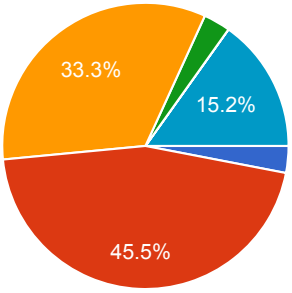


Difficulty in integrating different types of data from diverse sources	**17**	50%
Data are too large / complex	**11**	32.4%
Lack of data science expertise	**9**	26.5%
IT infrastructure is insufficient for the large data volume	**9**	26.5%
Domain knowledge to select the required data to be collected	**6**	17.6%
Lack of standards	**19**	55.9%
Interoperability between technologies	**11**	32.4%
Insufficient maturity of methodology and tools to effectively perform the analysis	**4**	11.8%
Discrepancy between requirements of IT support and research	**8**	23.5%
Communication gap between statisticians / data scientists and life scientists	**13**	38.2%
Access to relevant education and training	**1**	2.9%

Persuading higher management about the needs / benefits	**4**	11.8%
Lack of funding	**11**	32.4%
Culture change / scepticism toward innovation	**5**	14.7%
Other	**3**	8.8%

Part 3. Data strategy

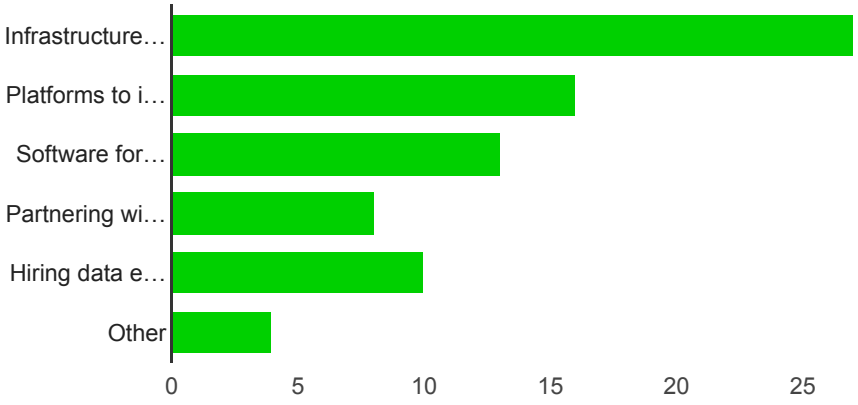
Compared to this year, do you expect your organisation's budget for data storage, handling and/or analysis in 2017 to:



Increase by more than 50%	**1**	3%
Increase by more than 10%	**15**	45.5%
Stay the same	**11**	33.3%
Decrease by more than 10%	**1**	3%
Decrease by more than 50%	**0**	0%
Not sure	**5**	15.2%

Part 3. Data strategy

In 2016/2017, budget for data storage, handling and/or analysis will be used for:



Infrastructure to improve data storage	**27**	84.4%

Platforms to improve data analytics	**16**	50%
Software for data analysis, visualization and/or integration	**13**	40.6%
Partnering with data science service providers	**8**	25%
Hiring data experts	**10**	31.3%
Other	**4**	12.5%

Part 3. Data strategy

Can you describe in 1-2 sentences your personal vision on data opportunities in your expertise field?

A community effort on updating collaborating and sharing data from a single entity

More funding must be provided to experts in the scientific domains to be able to hire data scientists/programmers to work directly with the scientific expertise generating, sharing, reusing and understanding the data and the technology. Too much work thus far has done by data scientists in vacuum, resulting in tools and services that are not needed and/or cannot be used.

In my own field the key opportunity is in inference of dynamic models from big data

Data storage is essential for future retrieval. Finding back data produced by former early-stage researchers is hardly possible, but often required.

Integral use if models and data will speed up product and process development and lead to higher quality of results for food, feed, pharma innovation, including personal health.

FAIR ;)

A lot can be gained from data integration. I see many opportunities in this aspect. Especially with the upcoming technologies which make it easier to obtain data

everybody uses databases daily. Databases are great to produce leads but you still have to do all the experiments yourself. It is simply impossible to incorporate external data in your research without proper testing and reproduction.

Data will be driving more and more research and management decisions than it is currently the case.

In the field of plant breeding I believe the biggest challenge is to overcome old traditions. Technology is revolutionizing the shift from selection to genetics and data driven decision making is exactly what genetics is all about. The challenge is the find the data scientists to make it happen.

data is key!

Footprint reduction and faster analysis tools are required to tackle the NGS big data deluge. Smart storage and easy-to-query solutions for multi-sample analysis are key.

Opportunities are within combining iternal sources with sources from outside the organisation, perform more prediction analysis, what if scenario's.

it is essential to store and share data in a structured way.

Genomics will change the way we work, but we do not know how to get ready for this.

Data should be made publicly available, in raw format as well as pre-processed, after publication. Scripts/workflows used to produce results should also be made public, so that procedures become transparent and reproducibility is guaranteed.

Use semantic web technologies as a standard for data integration and use existing opensource solutions for data stewardship problems.

Most important is data integration that will allow to draw conclusions from different data sources. The switch is made from univariate analysis to multivariate where factors from different fields are taken into account for better prediction.

Crucial to implement.

Data driven farming systems will contribute to feeding the world and preserving the earth. However a goof cooperation is needed between all kind of experts, including domain, IT and analytical experts.

The opportunities are to be found in better communication between end-user and the data scientist and focusing on getting the right questions. In my opinion 80% of data related problems can be solved with the current infrastructure and technology.

Part 3. Data strategy

Can you describe in 1-2 sentences your personal vision on data challenges in your expertise field?

Data silo's

The key challenge is "what to do with all the data" and "data standardization". A lot of the data in my opinion remains underused; it could be put to much better use by integrating it with mechanistic models.

It will be hardly possible to streamline everybody in the organization into a single data workflow. Automatic backing up local systems seems quite challenging

Take along R&D community in process. Standardization and storage of all data / sample and data tracking.

convincing collegues for the need to handle data FAIR

Re-use of data should be more facilitated by using proper standards. Infrastructure to work with huge datasets should become better / more accessible

The biggest challenge is that large sums of money are diverted to 'structure' and 'standardize' data streams, leading to standards that will be outdated in 5 years and that severely limit our possibilities. Better use the money to hire some more hands on the workfloor. I am a firm believer in open-source, non-profit tools. The web is full of them.

Combining the data and convincing the sceptics will be the major challenge.

Interoperability is key. Data should be easily accesible from any tool. Also, tools that are shipped with API's to the common scripting languages (Java, Python, R, Julia) will

hopefully become the new standard soon.

storage space, tools, standards and funding

Combining/integrating data with different file formats and make sure the business is willing to change and rely more on data.

Small investments can create very large gains. However, our field of research is not attractive enough to find the funds.

Data storage: large, safe but confidentiality assured if need be, enabling efficient sharing of progress between physical locations. More structure needed: teamwerk stores all files at the same level, whilst folders would be extremely desirable. Different pages in the same project could be visually clear. Sharing of scripts: not just as packages, but as the entire script just as in vignettes from BioConductor. Tools exist to enable this, but a cultural change needs to take place.

Data handling is more and more difficult. Data conversion to compatible formats is time consuming.

Many to face, depending on the market needs

Agriculture is not an 'open' sector, so the biggest challenge is to open up and become aware that data sharing is beneficial to all.

The tools and expertise is there; the main bottleneck is getting it to users in accessible and affordable ways.

Generating the right data and integrating them to get results on which basis you can take decisions.

The challenge is in communication and building high performance teams consisting of data-experts, statisticians and domain experts. Technological/IT challenges are less relevant, most of it works.

Number of daily responses