Storage of genomic data

Task Force on documentation software for genebanks

Tina Flisar

UNIVERSITY OF LJUBLJANA | BF Biotechnical Faculty

ERFP European Regional Focal Point for Animal Genetic Resources

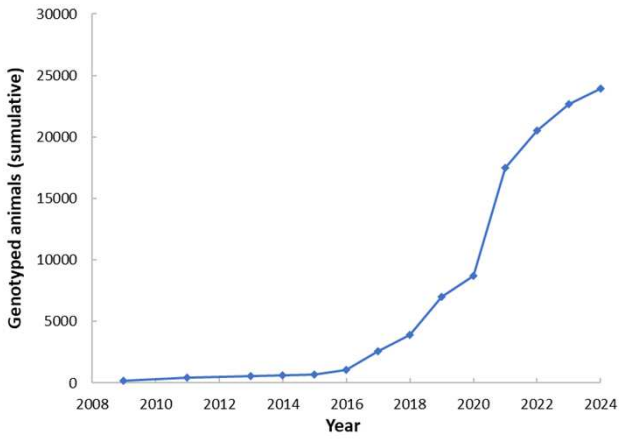
THE RECOVERY AND RESILIENCE PLAN

Funded by the European Union NextGenerationEU

1

Genotyping – within breeding programmes for cattle breeds

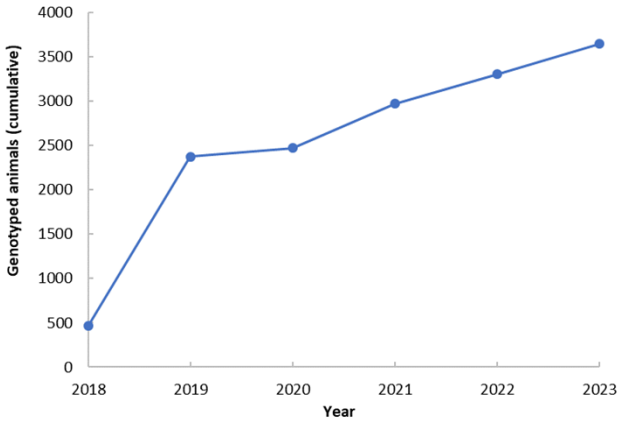
Breed	#
Aberdeen Angus	42
Brown cattle	6793
Charolais	263
Cika cattle	384
Holstein (black and white)	12742
Limousin	380
Red Holstein	21
Simmental	3150
Crossbred	188
Total	**23953**



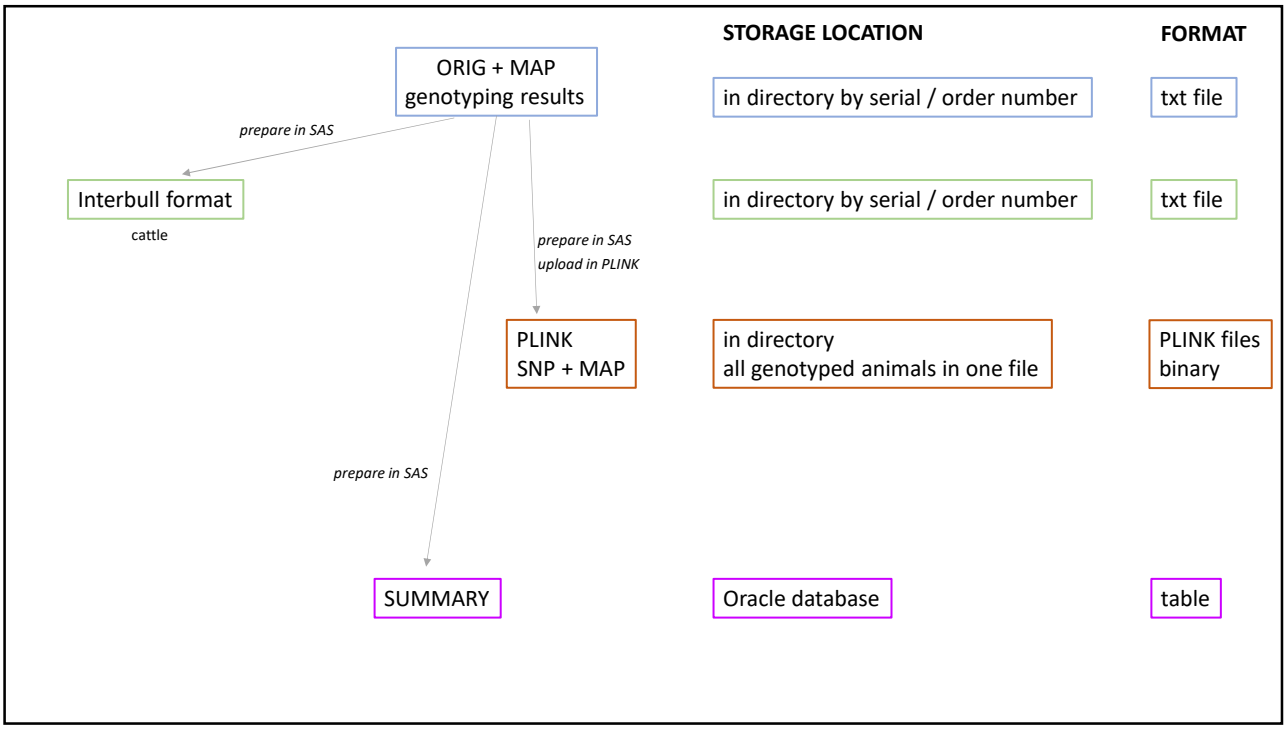
2

Genotyping – within national programme for conservation AnGR

Species	Breed	#
Horse	Lipizzan horse	102
Horse	Posavje horse	229
Horse	Slovenian cold bloded horse	11
Sheep	Bela Krajina pramenka	90
Sheep	Bovec sheep	94
Sheep	Istrian pramenka	70
Sheep	Jezersko-Solčava sheep	120
Sheep	Improved Jezersko-Solčava sheep	40
Pig	Krškopolje pig	1736
Goat	Drežnica goat	876
Poultry	Styrian hen	57
	Total	**3425**



3



4

Original and map file

in directory by serial / order number txt file

```
[Header]
GSGT Version,2.0.4
Processing Date,2/13/2024 5:16 PM
Content,,WSBv1_20006795x352140_A1.bpm
Num SNPs,49745
Total SNPs,52523
Num Samples,7
Total Samples,2976
[Data]
SNP Name,Sample ID,Allele1 - Forward,Allele2 - Forward,Allele1 - AB,Allele2 - AB,Allele1 - Top,Allel
ARS-BFGL-BAC-10172,SI35838522,G,G,B,B,G,G,0.9305,0.019,0.846,1.0000,0.1284
ARS-BFGL-BAC-1020,SI35838522,G,G,B,B,G,G,0.9492,0.020,0.780,1.0000,0.2765
ARS-BFGL-BAC-10245,SI35838522,C,C,B,B,G,G,0.7857,0.124,1.427,1.0000,-0.2320
ARS-BFGL-BAC-10345,SI35838522,C,C,B,B,C,C,0.8315,0.009,0.853,1.0000,-0.4383
ARS-BFGL-BAC-10375,SI35838522,A,G,A,B,A,G,0.9341,0.429,0.529,0.5822,0.1476
ARS-BFGL-BAC-10591,SI35838522,A,G,A,B,A,G,0.9169,0.376,0.472,0.5510,0.5688
ARS-BFGL-BAC-10867,SI35838522,G,C,A,B,C,G,0.9481,0.317,0.387,0.4867,1.0137
ARS-BFGL-BAC-10919,SI35838522,A,G,A,B,A,G,0.8192,0.867,0.876,0.5165,-0.0095
ARS-BFGL-BAC-10952,SI35838522,A,G,A,B,A,G,0.8037,0.924,0.802,0.5376,0.1580
ARS-BFGL-BAC-10960,SI35838522,G,G,B,B,G,G,0.7388,0.037,0.241,0.9906,-0.0116
ARS-BFGL-BAC-10972,SI35838522,G,C,A,B,C,G,0.9504,0.643,0.813,0.4837,0.2044
ARS-BFGL-BAC-10975,SI35838522,A,A,A,A,A,A,0.8119,1.247,0.054,0.0080,-0.0225
ARS-BFGL-BAC-10986,SI35838522,G,G,B,B,C,C,0.9367,0.002,0.726,1.0000,0.2165
ARS-BFGL-BAC-10993,SI35838522,C,C,B,B,G,G,0.7857,0.044,1.371,0.9909,-0.2404
ARS-BFGL-BAC-11000,SI35838522,T,G,A,B,A,C,0.9282,0.553,0.558,0.4394,1.2035
```

```
Index,Name,Chromosome,Position,GenTrain Score,SNP,ILMN Strand,Customer Strand,NormID
3,ARS-BFGL-BAC-10172,14,6371334,0.8937,[A/G],TOP,TOP,1
4,ARS-BFGL-BAC-1020,14,7928189,0.9158,[T/C],BOT,TOP,1
5,ARS-BFGL-BAC-10245,14,31819743,0.7812,[T/C],BOT,BOT,1
6,ARS-BFGL-BAC-10345,14,6133529,0.8106,[A/C],TOP,TOP,1
7,ARS-BFGL-BAC-10375,14,6616434,0.8977,[A/G],TOP,TOP,1
8,ARS-BFGL-BAC-10591,14,17544926,0.8795,[A/G],TOP,TOP,1
9,ARS-BFGL-BAC-10867,14,34639444,0.9144,[G/C],BOT,BOT,101
10,ARS-BFGL-BAC-10919,14,31267746,0.8023,[A/G],TOP,TOP,1
11,ARS-BFGL-BAC-10952,10,18882288,0.7924,[A/G],TOP,TOP,1
12,ARS-BFGL-BAC-10960,10,20609250,0.7934,[A/G],TOP,TOP,1
13,ARS-BFGL-BAC-10972,10,20792754,0.9173,[G/C],BOT,BOT,101
14,ARS-BFGL-BAC-10975,10,21225382,0.7976,[A/G],TOP,TOP,1
15,ARS-BFGL-BAC-10986,10,26527257,0.9006,[A/C],TOP,BOT,1
16,ARS-BFGL-BAC-10993,10,78512500,0.7812,[A/G],TOP,BOT,1
17,ARS-BFGL-BAC-11000,10,79252023,0.8912,[T/G],BOT,BOT,1
18,ARS-BFGL-BAC-11003,10,80410977,0.9073,[T/C],BOT,BOT,1
19,ARS-BFGL-BAC-11007,10,80783719,0.9209,[T/C],BOT,BOT,0
20,ARS-BFGL-BAC-11025,10,84516867,0.8846,[T/G],BOT,BOT,1
21,ARS-BFGL-BAC-11028,10,85649939,0.7960,[A/G],TOP,TOP,1
22,ARS-BFGL-BAC-11034,1,12839893,0.8964,[A/G],TOP,BOT,1
23,ARS-BFGL-BAC-11039,11,21276136,0.9470,[A/G],TOP,BOT,1
24,ARS-BFGL-BAC-11042,1,132983082,0.8152,[A/G],TOP,TOP,1
25,ARS-BFGL-BAC-11044,1,12805406,0.9334,[T/C],BOT,BOT,1
```

5

Interbull format 706

in directory by serial / order number txt file

```
706;BSW;SVN;M;000035838522;BFRO;a4f73df1-7c28-47cc-ab22-b17c14f8f054;Weatherbys Ireland;SI35838522;;49745_a;1001202120120101200102
706;BSW;SVN;F;000035838065;BFRO;a4b6ff6e-7aa7-4f0a-851b-785e6ebeb9c2;Weatherbys Ireland;SI 85838065;;49745_a;2001012200220021210202
706;BSW;SVN;F;000055838063;BFRO;7b6816bd-3176-427d-9499-85b422c3778f;Weatherbys Ireland;SI 55838063;;49745_a;2001112201211111201111
706;BSW;SVN;M;000035796967;BFRO;6273b46c-298a-4ac4-9dfb-5963bd48ec5c;Weatherbys Ireland;SI 35796967;;49745_a;2001211201211020202202
706;BSW;SVN;M;000045796966;BFRO;8cbcc447-d8c5-4601-86b7-d9ee1cc67afd;Weatherbys Ireland;SI 45796966;;49745_a;2000022201200121211102
706;BSW;SVN;F;000075838061;BFRO;06796e9e-350a-4bd7-bc90-051b0e0ba240;Weatherbys Ireland;SI 75838061;;49745_a;2001121221111021202202
706;BSW;SVN;M;000065094901;BFRO;7c248524-5bc1-441b-a7ed-2b5f8d2843aa;Weatherbys Ireland;SI 65094901;;49745_a;2002022120221120200202
```

One value for each SNP (both alleles)

BB→0
AB→1
AA→2
'unknown'→5

6

PLINK format

in directory
all genotyped animals in one file

PLINK files

```
1 1 BSWSVNF000035838065 0 0 2 -9 1 2 1 2 1 1 2 2 1 1 1 2 2 2 2 2 1 2 1 1 2 1 2 1 1 1 2 2 2 2 1 2 1 2 2 2 1 1 1 2 1
2 1 BSWSVNF000055838063 0 0 2 -9 1 2 1 2 1 1 1 2 1 1 1 1 2 2 2 2 2 2 1 1 2 2 2 2 1 1 1 2 2 2 2 2 1 2 1 1 1 2 1 1 1 1
3 1 BSWSVNF000075838061 0 0 2 -9 1 1 1 2 1 1 2 2 1 2 2 2 2 2 2 2 2 1 2 1 1 1 1 2 2 2 1 2 1 2 2 2 1 2 1 1 1 1 2 2 1 2 1
4 1 BSWSVNM000035796967 0 0 1 -9 2 2 1 1 1 1 2 2 1 1 1 1 2 2 2 2 2 1 2 1 1 1 1 1 2 1 2 1 2 2 2 2 2 1 2 1 1 2 2 1 2 1 1
5 1 BSWSVNM000045796966 0 0 1 -9 1 2 1 1 1 1 2 2 1 1 2 2 2 2 2 2 2 1 1 1 1 1 2 2 2 1 1 1 1 1 2 2 2 1 1 2 2 2 1 1 1 1
6 1 BSWSVNM000065094901 0 0 1 -9 1 2 1 2 1 1 2 2 1 2 2 2 2 2 2 2 2 1 2 1 1 1 2 1 2 1 2 1 1 2 2 1 1 1 1 1 1 2 2 1 1 1 1
```

```
1 14 ARS-BFGL-BAC-10172 0 6371334
2 14 ARS-BFGL-BAC-1020 0 7928189
3 14 ARS-BFGL-BAC-10245 0 31819743
4 14 ARS-BFGL-BAC-10345 0 6133529
5 14 ARS-BFGL-BAC-10375 0 6616434
6 14 ARS-BFGL-BAC-10591 0 17544926
7 14 ARS-BFGL-BAC-10867 0 34639444
8 14 ARS-BFGL-BAC-10919 0 31267746
9 10 ARS-BFGL-BAC-10952 0 18882288
10 10 ARS-BFGL-BAC-10960 0 20609250
11 10 ARS-BFGL-BAC-10972 0 20792754
12 10 ARS-BFGL-BAC-10975 0 21225382
13 10 ARS-BFGL-BAC-10986 0 26527257
14 10 ARS-BFGL-BAC-10993 0 78512500
15 10 ARS-BFGL-BAC-11000 0 79252023
16 10 ARS-BFGL-BAC-11003 0 80410977
17 10 ARS-BFGL-BAC-11007 0 80783719
18 10 ARS-BFGL-BAC-11025 0 84516867
19 10 ARS-BFGL-BAC-11028 0 85649939
20 1 ARS-BFGL-BAC-11034 0 12839893
21 11 ARS-BFGL-BAC-11039 0 21276136
22 1 ARS-BFGL-BAC-11042 0 132983082
23 1 ARS-BFGL-BAC-11044 0 12805406
24 11 ARS-BFGL-BAC-11047 0 21681893
25 11 ARS-BFGL-BAC-11050 0 22201316
26 1 ARS-BFGL-BAC-11057 0 134030804
27 11 ARS-BFGL-BAC-11090 0 4730528
```

For each SNP – 2 values, for 2 alleles

7

Oracle database

table

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	ID_LOCAL	VARCHAR2(30 BYTE)	No	(null)	1	(null)
2	ID_ORIGINAL	VARCHAR2(30 BYTE)	No	(null)	2	(null)
3	IDANIM	VARCHAR2(19 BYTE)	Yes	(null)	3	(null)
4	BREED	VARCHAR2(3 BYTE)	Yes	(null)	4	(null)
5	SEX	VARCHAR2(1 BYTE)	Yes	(null)	5	(null)
6	ORIGIN	VARCHAR2(3 BYTE)	Yes	(null)	6	(null)
7	ANIMAL	VARCHAR2(12 BYTE)	Yes	(null)	7	(null)
8	STATUS	NUMBER(4,0)	Yes	(null)	8	(null)
9	DTSENT	DATE	Yes	(null)	9	(null)
10	DTACC	DATE	Yes	(null)	10	(null)
11	NSNP	NUMBER(9,0)	Yes	(null)	11	(null)
12	CHIP_NAME	VARCHAR2(30 BYTE)	Yes	(null)	12	(null)
13	CHIP_DENSITY	VARCHAR2(15 BYTE)	Yes	(null)	13	(null)
14	HOMOZYG	NUMBER(7,5)	Yes	(null)	14	(null)
15	CALL_RATE	NUMBER(7,4)	Yes	(null)	15	(null)
16	IDRUN	NUMBER(5,0)	No	(null)	16	(null)
17	BREEDER_NAME	VARCHAR2(50 BYTE)	Yes	(null)	17	(null)
18	OWNER_NAME	VARCHAR2(50 BYTE)	Yes	(null)	18	(null)
19	REMARK	VARCHAR2(72 BYTE)	Yes	(null)	19	(null)
20	CALL_RATE_C	NUMBER(7,4)	Yes	(null)	20	(null)
21	KEY	VARCHAR2(30 BYTE)	Yes	(null)	21	(null)
22	DTTAKE	DATE	Yes	(null)	22	(null)
23	IDSEZNAM	NUMBER	Yes	(null)	23	(null)
24	LAB_NAME	VARCHAR2(50 BYTE)	Yes	(null)	24	(null)

8

App on the web

Summary of genotyped animals by serial number,
data are storaged in Oracle database

serial number of genotypization | number of animals | sent | received results | chip | density | number of snp | lab | breed

Pregled genotipizacij

Serija	Št. Živali	Poslano	Prejeto	Čip	Gostota	N_snp	Laboratorij	Pasme
9999	100	-	-	-	-	-	-	HOL
1348	7	-	15.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	BSW
1347	1	-	19.02.2024	IDB_versa50K	50K	49750	Weatherbys Ireland	LIM
1346	4	-	15.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	BSW
1345	9	-	15.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	BSW
1344	2	-	15.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	BSW
1343	14	-	15.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	CHA-LIM
1342	9	-	15.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	CHA-LIM
1341	14	-	15.02.2024	IDB_versa50K	50K	49749	Weatherbys Ireland	CHA-LIM
1340	2	-	15.02.2024	IDB_versa50K	50K	49750	Weatherbys Ireland	BSW
1339	1	-	15.02.2024	IDB_versa50K	50K	49750	Weatherbys Ireland	SIM
1338	162	-	12.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	AAN-BSW-CHA-CIK-HOL-LIM-SIM-UUU
1337	21	-	12.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	HOL-SIM-UUU
1336	15	-	12.02.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	SIM
1335	39	-	05.02.2024	IDB_versa50K	50K	49749	Weatherbys Ireland	CHA-HOL-LIM
1334	137	-	29.01.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	BSW-CIK-HOL-UUU
1333	426	-	29.01.2024	IDB_versa50K	50K	49745	Weatherbys Ireland	AAN-BSW-CHA-HOL-LIM-SIM-UUU

9

App on the web

Search by breed, id, gender, serial number, status
data are storaged in Oracle database

animal id | id | sent | received results | number of snp | chip | density | homozigosity | reliability | serial number of genotypization | number of repeats | Lab_id

Pregled

Išči | Država Vse | Pasma živali Vse | Spol Vse | Serija Vse | Status Vse

Št. prikazanih 100 | Go | Reset

ID	Žival	Status	Poslano	Sprejeto	N_snp	Čip	Gostota	Homozig.	Točnost	Serija	Ponovitev	LabID
SI04927378	BSWSVNF000004927378	SNP	25.03.2021	23.04.2021	49745	IDB_versa50K	50K	0,7049	0,9945	1175	1	473779012
SI04927426	BSWSVNF000004927426	SNP	12.07.2021	19.08.2021	49745	IDB_versa50K	50K	0,6773	0,9907	1208	1	470216318
SI04927828	BSWSVNF000004927828	SNP	14.05.2021	01.07.2021	49745	IDB_versa50K	50K	0,6846	0,9752	1196	1	466446316
SI04929868	BSWSVNF000004929868	SNP	30.09.2019	30.09.2019	49706	IDB_versa50K	50K	0,7016	0,9907	1125	1	SI04929868
SI04929875	BSWSVNF000004929875	SNP	14.05.2021	01.07.2021	49745	IDB_versa50K	50K	0,6645	0,9487	1195	1	494616916
SI04930095	BSWSVNF000004930095	SNP	01.06.2022	07.07.2022	49749	IDB_versa50K	50K	0,7006	0,9942	1268	1	473954815
SI04930507	BSWSVNF000004930507	SNP	13.08.2018	13.08.2018	52445	IDB_IDBv3	53K	0,7165	0,9941	1103	1	SI04930507
SI04931247	BSWSVNF000004931247	SNP	12.05.2021	23.06.2021	49745	IDB_versa50K	50K	0,6853	0,9971	1191	1	467008214
SI04931618	BSWSVNF000004931618	SNP	03.05.2022	06.06.2022	49745	IDB_versa50K	50K	0,6909	0,9922	1258	1	473394811
SI04931900	BSWSVNF000004931900	SNP	21.08.2017	15.09.2017	52445	IDB_IDBv3	53K	0,7189	0,9922	1079	1	SI04931900
SI04931931	BSWSVNF000004931931	SNP	12.01.2018	15.02.2018	52445	IDB191_IDBv3	53K	0,7236	0,9934	1092	1	SI04931931
SI04933270	BSWSVNF000004933270	SNP	14.05.2021	01.07.2021	49745	IDB_versa50K	50K	0,6827	0,9644	1195	1	508960512
SI04939867	BSWSVNF000004939867	SNP	03.06.2021	05.08.2021	49745	IDB_versa50K	50K	0,7166	0,9916	1203	1	471888318
SI04939874	BSWSVNF000004939874	SNP	03.06.2021	05.08.2021	49745	IDB_versa50K	50K	0,7202	0,9867	1203	1	475754514
SI04941693	BSWSVNF000004941693	SNP	12.01.2018	23.02.2018	52445	IDB191_IDBv3	53K	0,7039	0,9461	1093	1	SI04941693
SI04942508	BSWSVNF000004942508	SNP	07.06.2022	01.07.2022	49749	IDB_versa50K	50K	0,6997	0,9814	1264	1	498698316

10

Livestock databases

40 different breeding programmes: 8, 17, 5, 5, 4,

Kmetijski inštitut Slovenije — CPZ Cattle, Bee

BF UNIVERSITY OF LJUBLJANA Biotechnical Faculty — Small Ruminants, Pigs, Genetic evaluation for cattle, Genetic evaluation for horse, Conservation of AnGR

VF UNIVERSITY OF LJUBLJANA Veterinary Faculty — Horses

LIPICA — Lipizzan horse

REPUBLIKA SLOVENIJA MINISTRSTVO ZA KMETIJSTVO, GOZDARSTVO IN PREHRANO — central register, ARSKTRP

NOO | THE RECOVERY AND RESILIENCE PLAN

Funded by the European Union
NextGenerationEU

11

The Recovery and Resilience Plan

The goal is to develop a unified modern information system that will provide:

- Interoperability of databases
- Efficient support for carrying out public service tasks related to livestock farming and breeding programs.
- Assistance to farmers in managing herds and collaboration with relevant professional services.
- Supporting public administration in managing and monitoring agricultural policies at both national and local levels, particularly for the Ministry of Agriculture, Forestry, and Food (MKGP) and its affiliated bodies, as well as other users of this data

DigŽiv

Kmetijski inštitut Slovenije | BF UNIVERZA V LJUBLJANI Biotehniška fakulteta | VF UNIVERZA V LJUBLJANI Veterinarska fakulteta

Digitalizacija podatkovnih zbirk v živinoreji

12