# **Digital Resources for Plant Genomics & Phenomics at IPK Gatersleben**

Daniel Arend, Uwe Scholz, Matthias Lange

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, 06466 Seeland, Germany

As the host of the largest European gene bank for cultivated plants, the IPK Gatersleben offers a rich basis for research of plant biodiversity at the structural and functional level. Using high-throughput technologies such as sequencing and phenotyping, millions of data points are captured and persisted in the IPK research data infrastructure. The concept for the FAIRification of digital plant genetic resources and the underlying research data management (RDM) is based on four pillars: (I) a core bioinformatics organisational structure that operates a long-term committed and uniform RDM technology stack, (II) the generalization and coordination of data structures across data acquisition projects, (III) the harmonisation of data acquisition, documentation and analysis processes across IPK, and (IV) a data publication concept and systems that is highly integrated into national and international frameworks and standards.

dit <u>V</u>iew <u>F</u>avourites <u>L</u>ayouts <u>T</u>ools <u>M</u>odules <u>W</u>indow

projects Samples LJ: Heterosis >

1. 👽 💫 🚡 🐛 🗢 🐂 🐼 🤚 🗮 🖓 🐂 🛤 😒 🐄 💷

# (I) - Central Bioinformatics Working Group & Administration

- IT-Infrastructure
- Laboratory Information Management System: LIMSOPHY
- Scalable Storage Management: HSM
- Scalable Hardware: Oracle Data Appliance
- Frontend Technologies (OS-native, Web, Mobile Devs)

## Processes & Communication

- Minimum Information Standards: MIAPPE, MCPD • Standards: material naming, IDs, file formats, storage
- paths
- Lab-Processes & policies: sequencing, chemicals, ...
- User Training & Support
- FAIR Data Management
- Data Publication: DataCite
- Researcher Identifier: ORCID-DE
- Registries & Repositories: re3Data
- RESTful API's: Breeding API
- Research Data Management Plans: RDMO tool



# (II) - Laboratory Information & Management System

A generic relational database schema for the management of structured, unstructured and binary research data was implemented in the form of a laboratory information management system (LIMS) [1]. All experimental data can be depicted in a bundled database. For a high degree of interoperability, data records follow strict system of cross-linked, unique identifier and detailed metadata. The data can be accessed according to individual project requirements. Furthermore all components are documented and may be accessed interactively using the LIMS GUI or using the embedded machine-readable interfaces.





(III) - Harmonization of Data Acquisition

The harmonization of data acquisition is an ongoing process. Based on minimal information standards, MIAPPE, the MCPD or DataCite an increasing number of data acquisition and curation processes are defined, documented in Research Data Management Plans and stored in RDMP-Repositories like the RDMO tool.



									1												=
	Н	leer	0	0	0	R077	2474	7,19	7,19	R104	3466	10,07	5,04	R077	3581	10,41	3,47	Q180	2941	8,55	4
	G	Q179	4809	13,98	4,66	R104	1321	3,84	3,84	R103	1181	3,43	3,43	R007	2038	5,92	5,92	C24	1210	3,52	1
	F	Q024	3624	10,53	3,51	Q060	3222	9,36	4,68	R007	1948	5,66	2,83	Q230	2103	6,11	6,11	R173	2340	10,26	3
	E	R114	4478	13,01	4,34	R096	2831	8,23	4,11	R114	5465	15,88	5,29	Q068	2005	5,83	2,91	R094	2706	7,86	3
	D	Q056	5287	15,36	5,12	R104	5259	15,28	5,09	R077	1499	4,36	2,18	Q024	3655	10,62	5,31	R096	4988	14,5	7
	С	R103	1755	5,1	5,1	Q009	3028	8,8	4,4	Q180	3097	9	3	Q024	4520	13,14	4,38	R173	5269	15,31	5
	В	Q068	7601	22,09	7,36	R033	2192	6,37	3,19	R007	3941	11,45	5,73	Q188	4748	13,8	6,9	Q050	4587	11,87	5
										the second se											

RALIMS

Carrier         Complete         id_tag         Measurement Label         Pot no         Node         Node         Tmestamp         Water         Weight Before           1391         TRUE         1948ST001         1948ST         1         N0005         28.112019         0         9987           1170         TRUE         1948ST002         1948ST         2         N0176         N0031         28.112019         0         10093           1025         TRUE         1948ST002         1948ST         3         N0026         N0331         28.112019         0         10022           1025         TRUE         1948ST004         1948ST         4         N0324         N0338         28.112019         0         10022           1358         TRUE         1948ST004         1948ST         4         N0324         N0388         28.112019         0         10113           1494         TRUE         1948ST005         1948ST         5         N0396         R0312         28.112019         0         10103           1202         TRUE         1948ST005         1948ST         6         N0106         R0315         28.112019         0         10054           1202         TRUE		r the measurements of the initial humidity a period of max. 1h is sufficient.										
1391         TRUE         1948ST001         1948ST         1         N0002         N0005         28.11.2019 10.53.47         0         9987           1170         TRUE         1948ST002         1948ST         2         N0170         N0391         28.11.2019         0         10093           1025         TRUE         1948ST003         1948ST         3         N0082         N0233         28.11.2019         0         10022           1358         TRUE         1948ST003         1948ST         4         N0324         N0338         28.11.2019         0         10122           1358         TRUE         1948ST005         1948ST         4         N0324         N0388         28.11.2019         0         10113           1149         TRUE         1948ST005         1948ST         5         N0395         N0396         28.11.2019         0         10103           1202         TRUE         1948ST005         1948ST         6         N0100         N0315         28.11.2019         0         10054           1202         TRUE         1948ST001         1948ST         1         N0002         N0035         28.11.2019         0         10054           1391         T	C	Carrier D	Complete	id_tag	Measurement Label	Pot no	Node A	Node B	Timestamp	Water Amount	Weight Before	Weigh
1170         TRUE         1948ST002         1948ST         2         N0176         N0391         28.11.2019         0         10093           1025         TRUE         1948ST002         1948ST         3         N0082         N0233         28.11.2019         0         10022           1358         TRUE         1948ST004         1948ST         4         N0324         N038         28.11.2019         0         10022           1358         TRUE         1948ST004         1948ST         4         N0324         N038         28.11.2019         0         10113           1149         TRUE         1948ST005         1948ST         5         N0396         0.81.1.2019         0         10108           1202         TRUE         1948ST006         1948ST         6         N0100         N0375         28.11.2019         0         10054           1202         TRUE         1948ST006         1948ST         6         N0100         N0375         28.11.2019         0         10054           1391         TRUE         1948ST002         1948ST         1         N0002         R0005         28.11.2019         0         14490           1170         TRUE         1948ST002 </td <td>13</td> <td>391</td> <td>TRUE</td> <td>1948ST001</td> <td>1948ST</td> <td>1</td> <td>N0002</td> <td>N0005</td> <td>28.11.2019 10:53:47</td> <td>0</td> <td>9987</td> <td>9987</td>	13	391	TRUE	1948ST001	1948ST	1	N0002	N0005	28.11.2019 10:53:47	0	9987	9987
1025         TRUE         1948ST003         1948ST         3         N0082         N0233         28.11.2019         0         10022           1358         TRUE         1948ST004         1948ST         4         N034         N0358         28.11.2019         0         10131           1358         TRUE         1948ST005         1948ST         4         N0358         28.11.2019         0         10113           1149         TRUE         1948ST005         1948ST         5         N0395         N0396         28.11.2019         0         10108           1202         TRUE         1948ST005         1948ST         6         N0100         N0315         28.11.2019         0         10054           1202         TRUE         1948ST001         1948ST         1         N0002         N0035         28.11.2019         0         10054           1391         TRUE         1948ST001         1948ST         1         N0002         N0005         28.11.2019         0         14490           1170         TRUE         1948ST02         1948ST         2         N0176         N0391         28.11.2019         0         14600	1	170	TRUE	1948ST002	1948ST	2	N0176	N0391	28.11.2019 10:54:10	0	10093	1005
Instruction         Instruction <thinstruction< th=""> <thinstruction< th=""></thinstruction<></thinstruction<>	10	025	TRUE	1948ST003	1948ST	3	N0082	N0233	28.11.2019	0	10022	T
Image:	1	358	TRUE	1948ST004	1948ST	4	N0324	N0358	28.11.2019	0	10113	
149         INUE         19485100         194851         3         NO39         NO39         20.12.019         0         NO36           1202         TRUE         19485100         194851         6         N0100         N0315         28.11.2019         0         10054           1391         TRUE         194851001         194851         1         N0002         N0050         28.11.2019         0         14490           1170         TRUE         19485102         194851         2         N0176         N0391         28.11.2019         0         14600	-	140	TRUE	10400000	1049CT	6	N0305	N0208	10:54:57	0	10109	
1202         TRUE         1948ST006         1948ST         6         N0100         N0315         28.11.2019         0         10054           1391         TRUE         1948ST001         1948ST         1         N0002         N0005         28.11.2019         0         14490           1170         TRUE         1948ST002         1948ST         2         N0176         N0391         28.11.2019         0         14600	Ľ	149	INCE	194031003	194031	5	140393	140390	10:55:20	0	10108	1
1391         TRUE         1948ST001         1948ST         1         N0002         N0005         28.11.2019 11.27.38         0         14490           1170         TRUE         1948ST02         1948ST         2         N0176         N0391         28.11.2019         0         14600	1:	202	TRUE	1948ST006	1948ST	6	N0100	N0315	28.11.2019 10:55:44	0	10054	
1170 TRUE 1948ST002 1948ST 2 N0176 N0391 28.11.2019 0 14600	13	391	TRUE	1948ST001	1948ST	1	N0002	N0005	28.11.2019 11:27:38	0	14490	1.
	1	170	TRUE	1948ST002	1948ST	2	N0176	N0391	28.11.2019	0	14600	14

# (IV) - FAIR Data Access & Publication

The data sharing and the on-site data access is featured by community accepted APIs, e.g. the Breeding API [2], and data exchange formats, such as ISA-TAB, VCF. Furthermore a publication services of DOI citable high volume research data is implemented using the e!DAL software infrastructure [3]. In addition sequence data will be submitted along with inks to the LIMS stored metadata to the European Nucleotide Sequence Archive (ENA). Furthermore the IPK is involved in national and international infrastructure programs, like de.NBI, ECPGR, ELIXIR, Emphasis, DivSeek, ORCID, DataCite etc. to supply global data re-use and value creation workflows.

Use a Knowledge Representation Language

R3: (Meta)data meet domain-relevant standards

## Example: FAIR NGS @ IPK

### Findability

- F1a: Identifier uniqueness
- F1b: Identifier persistence
- F2: Machine Readability of Metadata
- F3: Resource Identifier in Metadata F4: Must be registered in searchable resource

### Accessible

A1: Access Authorization - Check for HTTP return codes A2: Metadata Longevity (Metadata should outlive the data) -Check the presence of a policy document

## F1a <u>F1b F2 F3 F4 A1 A2 I1 I2 I3 R1 R2</u>

nteroperable

Reusable

Use FAIR vocabularies

Use aualified references

Accessible usage license

R2: Detailed Provenance





[1] Ghaffar et al. (2020). Programmatic Access to FAIRified Digital Plant Genetic Resources. Journal of Integrative Bioinformatics, 16(4). 2016. [2] Abbeloos et al. BrAPI - an application programming interface for plant breeding applications. Bioinformatics, 2019. [3] Arend et al. PGP repository: a plant phenomics and genomics data publication infrastructure. Database





www.ipk-gatersleben.de

LEIBNIZ INSTITUTE OF PLANT GENETICS AND CROP PLANT RESEARCH