

Evolution from Data Silos to FAIR Digital Objects - (Two) Decade(s) of Data Management at IPK Gatersleben

Danuta Schüler, Daniel Arend, Anne Fiebig, Jochen C. Reif, Elena Rey-Mazón, Uwe Scholz, Matthias Lange

Research Group **Bioinformatics and Information Technology (BIT)**

Leibniz Institute of Plant Genetics and Crop Plant Research (**IPK**), Gatersleben, Germany

19th International Digital Curation Conference - The Hague, the Netherlands

Leibniz Institute of Plant Genetics and Crop Plant Research



Staff 500 – 200 scientists incl. PhD students – 30 research groups



The Federal ex-situ Genebank at IPK – Genotypes and Phenotypes



2,913 species
152,359 accessions
(as of July 2023)

Distribution of
material since 1946:
>1,000,000 samples

Genomes:

*for comparison:
human genome size
3 giga-base-pairs
(Gbp)*

Wheat

2014



16 Gbp

Barley

2017



5 Gbp

And further crop genomes:

- **Rye** 2021 (7 Gbp)
- **Oat** 2022 (11 Gbp)
- **Faba bean** 2023 (13 Gbp)
- ...



- Genotyping
- Generation of pan-genomes
- High throughput phenotyping



**Conservation, Activation and
Digitalisation of
Plant Genetic Resources**

IPK - Partner in National and International Infrastructures and Initiatives



More than Twenty Years Back – History of Data Management at IPK*

* perspective of research group Bioinformatics and Information Technology (BIT)

2003: Research Group Bioinformatics (BI)

Tasks: data analysis, data management (file-based and/or in databases) and data publication on webserver

Characteristics: mostly completely new implementations



2008: Research Group Bioinformatics and Information Technology (BI + IT = BIT)

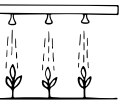
Tasks: data management, data publication and IT administration for the entire IPK

Characteristics: IT systems are closely linked to data management

Issue: no general and customizable system for data management

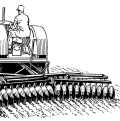


2008: Increase the awareness at the IPK: internal workshops and talks about data management



2011: Requirement analysis Laboratory Information Management Systems and strategy development

Features: customizable system for data management closely connected to IT infrastructure



2012: Roll out of the LIMS systems at IPK

2015: Start the implementation of the sequencing management process using LIMS



2016: Continuous adaption sequence data management processes in the LIMS with constant involvement of users to curate the meta-data



2021: Implementing FAIR data management using LIMS: <https://dx.doi.org/10.1093/bib/bbab010>

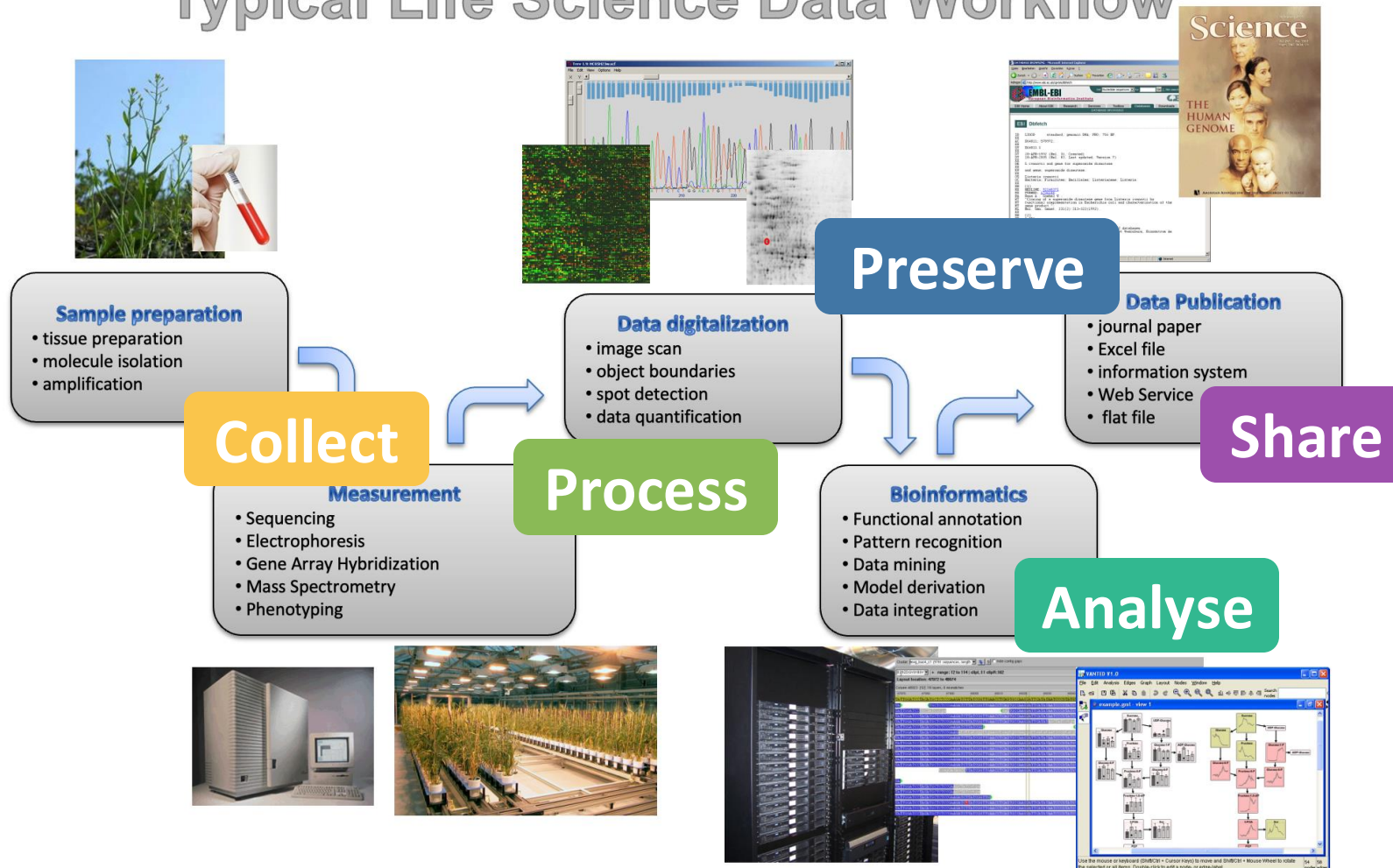
2024: LIMS-based generation of sequence data sets to submit as FAIR digital objects to repositories



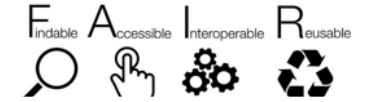
Developments at IPK are Connected to FAIR and the Data Life Cycle

2008: IPK internal presentation by Matthias Lange

Typical Life Science Data Workflow



2016:



Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

Scientific Data 3, Article number: 160018 (2016) | [Cite this article](#)

340k Accesses | 2931 Citations | 1912 Altmetric | [Metrics](#)

2021:



ELIXIR (2021) Research Data Management Kit. A deliverable from the EU-funded ELIXIR-CONVERGE project (grant agreement 871075). <https://rdmkit.elixir-europe.org>

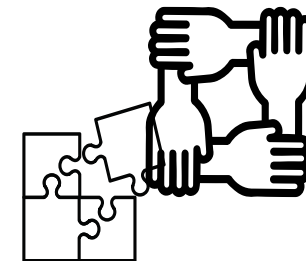
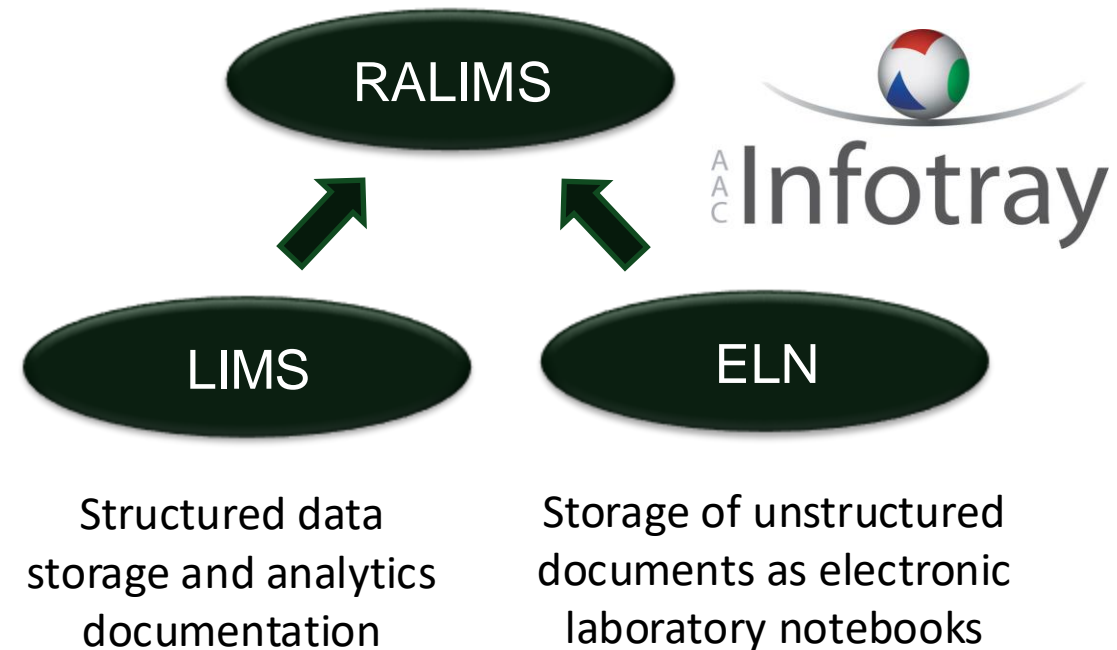
LIMS - Laboratory Information Management System at IPK

The system:

- Combination of LIMS and ELN
- Combination of relational database and external storage
- Customizable, tabular user interface
- Freely programmable import and export interfaces
- Central installation as a remote app (MS-Windows Server)
- Extensible data schema
- Auditing and data protection

LIMS system administrator:

- Set up in parallel with the system
- Engineer of Biotechnology
- Wide practical knowledge in the laboratory
- Experience in interacting with colleagues of all areas



LIMS - Laboratory Information Management System at IPK

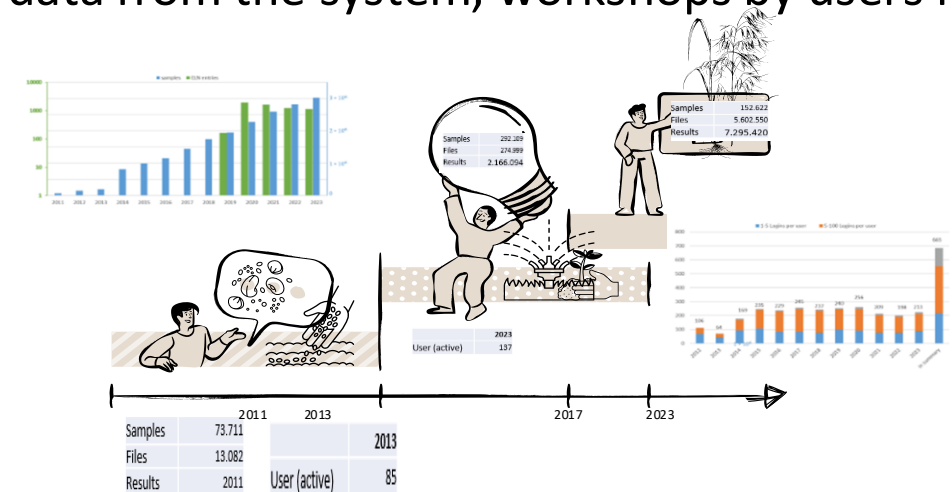
LIMS System:



- Commercial System with con-current user license
- Maintenance and service costs are calculated based on the number of modules required
- Consistent project partner on the part of the company AAC Infotray

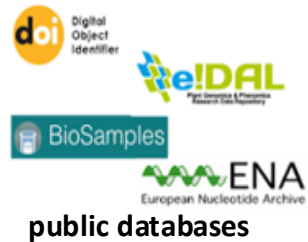
LIMS User:

- Not introduced as mandatory
- System was made known to all employees through projects such as hazardous substance management
- Advantages of the system through e.g. website design with data from the system, workshops by users for users constantly spread throughout the institute
- Own module designs for individual use
- Regular training and visibility of the administrator in the individual laboratories

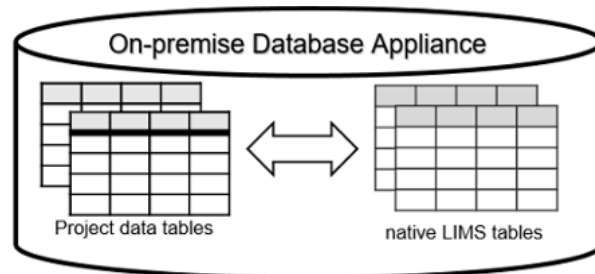
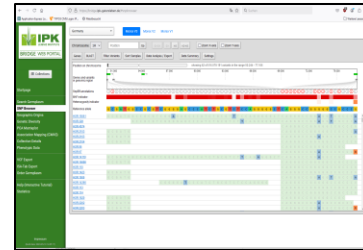


Development of Technology, IT System Integration, User and Data Storage

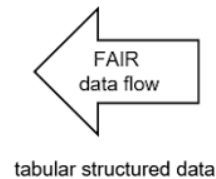
- Increase of current users
- Extension of LIMS system by ELN component
- Renewal of Windows Terminal Server
- Increase of storage capacity
- Extension of machine-readable access and SQL interfaces
- Development of Apex application for web-based access



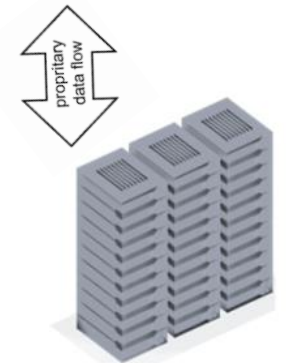
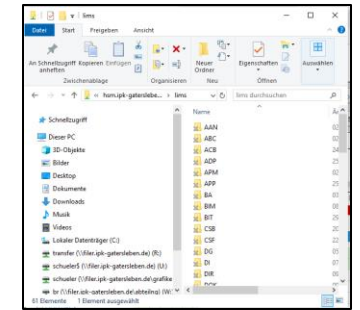
Json,ISA-Tab-Export
machine actionable,
remote access
API, SQL,CSV



IPK central tabular data store



MS Windows Terminalserver Cluster

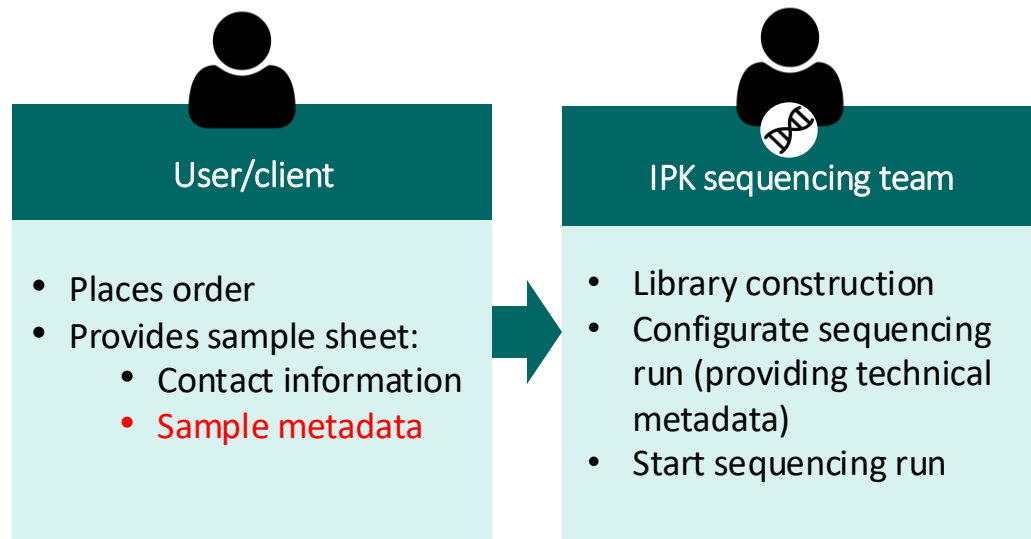


Hierarchical Mass File Archive
(SSD/HDD/Tape)

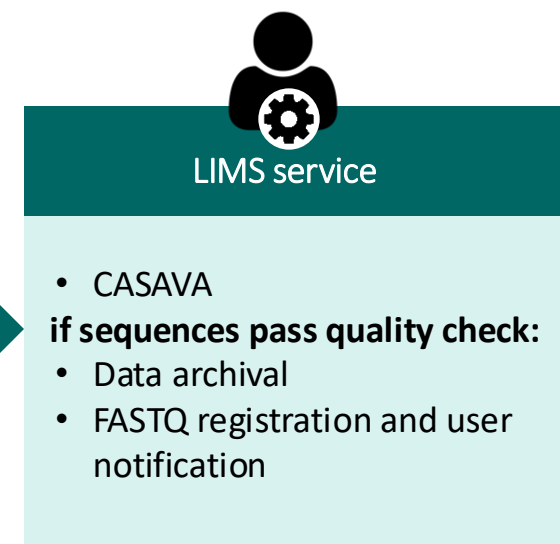


IPK sequencing workflow integrates FAIR principles

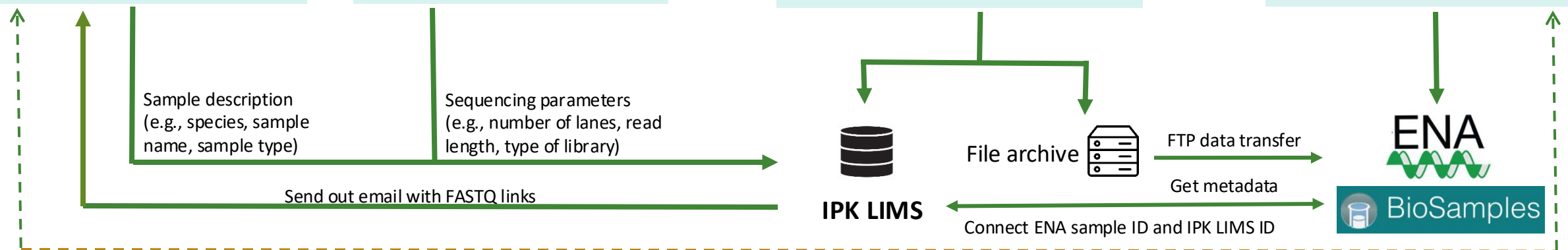
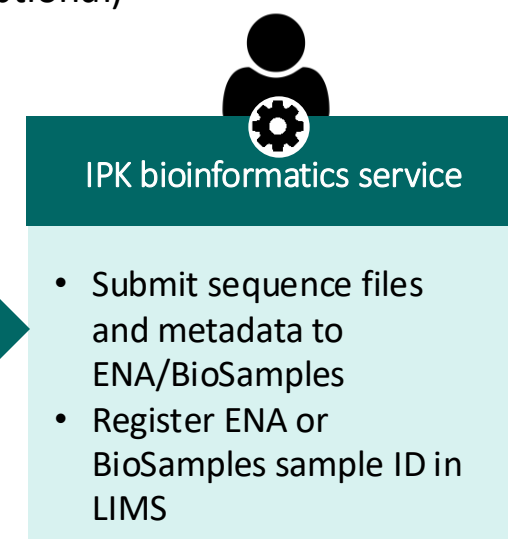
Step 1: Order and sequencing process (wet lab)



Step 2: Data transfer to LIMS



Step 3: Submission to EMBL-ENA (optional)



Add sample metadata if not available in LIMS



Briefings in Bioinformatics, 22(5), 2021, 1–14

<https://doi.org/10.1093/bib/bbab010>
Case Study

Implementing FAIR data management within the German Network for Bioinformatics Infrastructure (de.NBI) exemplified by selected use cases

Gerhard Mayer[†], Wolfgang Müller[†], Karin Schork, Julian Uszkoreit, Andreas Weidemann, Ulrike Wittig, Maja Rey, Christian Quast, Janine Felden, Frank Oliver Glöckner, Matthias Lange, Daniel Arend, Sebastian Beier, Astrid Junker, Uwe Scholz, Danuta Schüler, Hans A. Kestler, Daniel Wibberg, Alfred Pühler, Sven Twardziok, Jürgen Eils, Roland Eils, Steve Hoffmann, Martin Eisenacher and Michael Turewicz

D1500–D1507 Nucleic Acids Research, 2022, Vol. 50, Database issue
<https://doi.org/10.1093/nar/gkab1046>

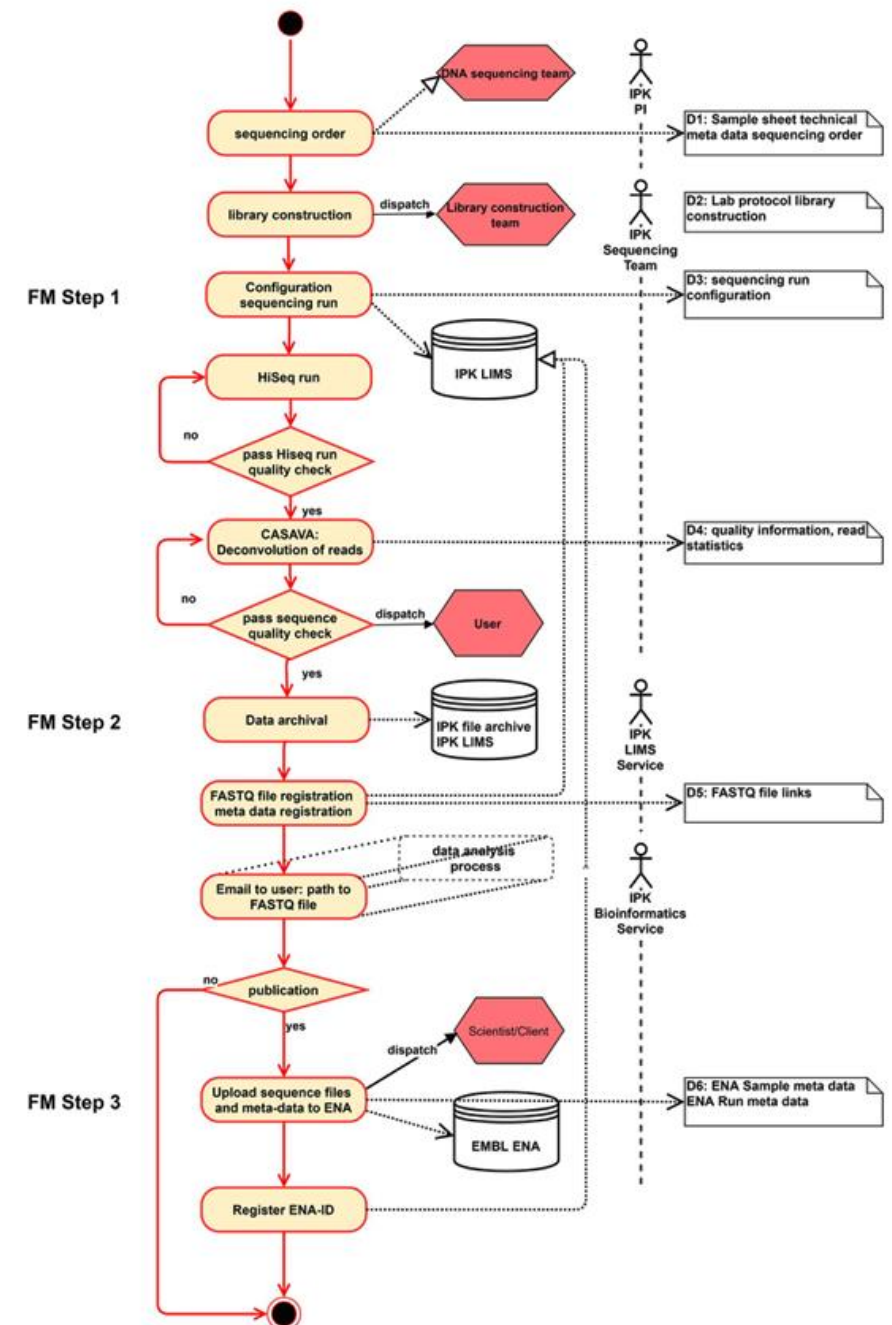
Published online 8 November 2021

BioSamples database: FAIRer samples metadata to accelerate research data management

Mélanie Courtot[✉], Dipayan Gupta[✉], Isuru Liyanage[✉], Fuqi Xu[✉] and Tony Burdett[✉]

European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK

Received September 15, 2021; Revised October 13, 2021; Editorial Decision October 14, 2021; Accepted October 14, 2021



Impact of Collaboration in International Networks

ELIXIR Implementation Study

FONDUE = FAIR-ification of Plant Genotyping Data and its linking to Phenotyping using ELIXIR Platforms

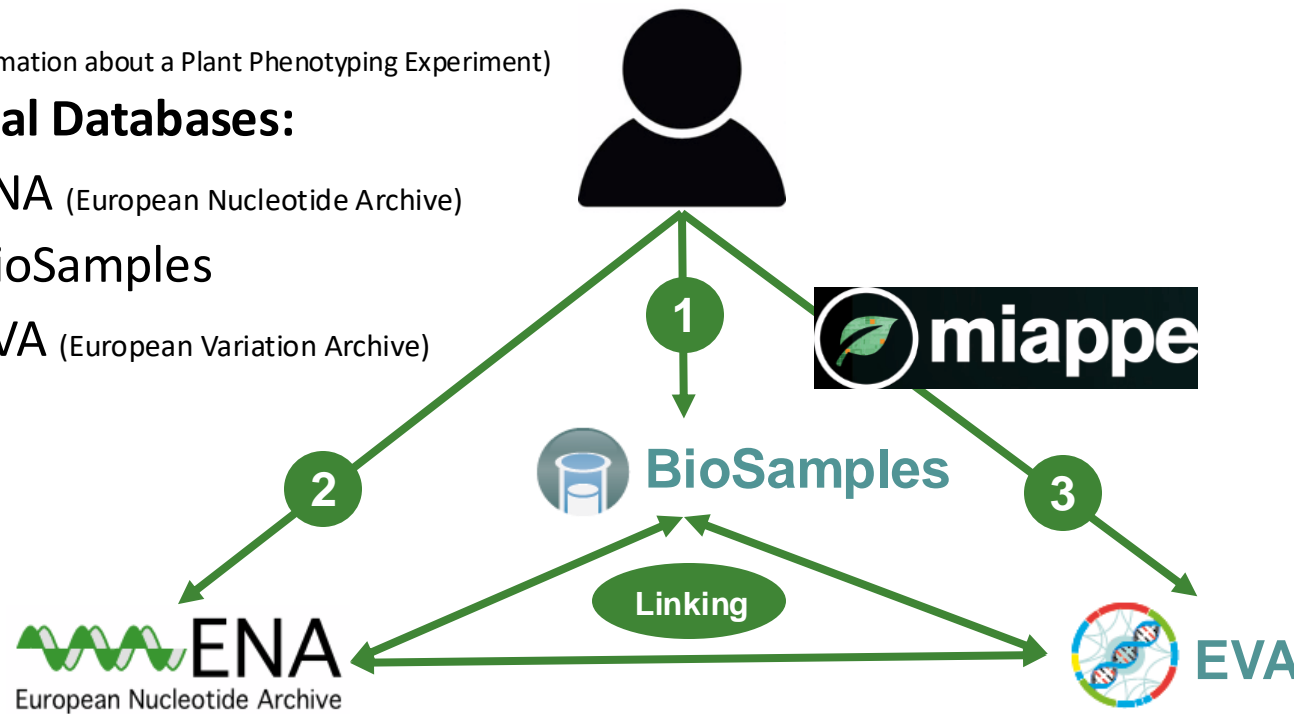


Standards:

- MIAPPE
(Minimal Information about a Plant Phenotyping Experiment)

International Databases:

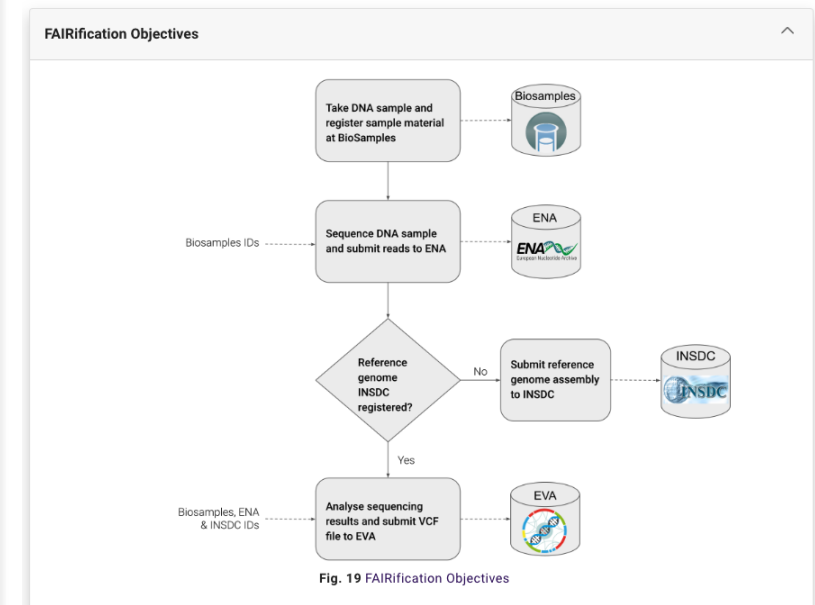
- EMBL-ENA (European Nucleotide Archive)
- EMBL-BioSamples
- EMBL-EVA (European Variation Archive)



Plant genomic and genetic variation data submission to EMBL-EBI databases


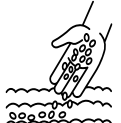
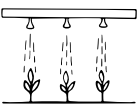
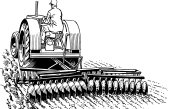





Graphical overview of the FAIRification Objectives



<https://faircookbook.elixir-europe.org/content/recipes/reusability/miappe.html>

Conclusions and Outlook

- 
- 
- 
- 
- 
- 
- 
- 20 years ago: single databases/systems → data silos
 - 15 years ago: establishment of LIMS (laboratory information management system)
 - Embedded in the overall IT infrastructure
 - Additional staff: LIMS manager and administrator
 - No mandatory use → voluntary and recommendation by the management
 - Labs who are willing to use it and understand the added value
 - Regular field reports from users for users
 - External developments provide a boost, e.g. NFDI for IPK internal expansions
 - Continuous support from the institute management (investment and personnel)
 - Anchoring data management in the institute's strategy and manifested in official documents:
 - Guideline for handling IPK research data
 - IPK research data policy
 - Performing a RISE self-evaluation RISE → Research data management strategy 2024 - 2026

Continuous adaptation of the system + processes regarding new requirements!

Training of the users and data curation are permanent challenges!

Acknowledgements

- All IPK colleagues
- Felipe Wettstein (AAC Infotray)



External collaborators:



Funding:



Thank you for your attention!



<https://twitter.com/IPKGatersleben/status/1394697082200539148>

See you next year in Gatersleben:



Gatersleben Research Conference

19th International Symposium on

Integrative Bioinformatics

10–12 September 2025 / Gatersleben



SCAN ME

<https://meetings.ipk-gatersleben.de/grc-ib2025/>