# How Logilab ELN helps organizations to follow FAIR Principles

## Importance of Data Sharing

In the modern industrial era, data sharing provides a great stimulus to scientific advancement by enhancing transparency, improving collaboration, accelerating research and driving better decision-making. In areas such as public health, data sharing is very vital and critical during emergency situations such as outbreaks of infectious diseases. It is very important to have a set of guidelines that need to be following while preparing for data sharing.

## About FAIR Principles

The FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable), published on Scientific Data in 2016, are a set of guiding principles proposed by a consortium of scientists and organizations to support the reusability of digital assets. It has since been adopted by research institutions worldwide. The guidelines are timely as we see unprecedented volume, complexity, and creation speed of data.

To make data findable
- Data and supplementary materials should have sufficiently rich metadata as well as a unique and persistent identifier such as DOI.

To be accessible
- Metadata and data should be understandable to humans and machines; and data must be stored in a trusted repository.

To be interoperable
- metadata should use a formal, accessible, shared, and broadly applicable language for knowledge representation.

To be reusable
- data and collections should have a clear usage license and provide accurate information on provenance.

# How the process of following FAIR principles benefit data handlers and data users?

The users of data are now relying more on computational support from systems and machines, FAIR data can enable computational systems to find, access, interoperate, and reuse data with no or minimal human intervention.

On another level, the FAIR Data Principles provide a data management framework to help researchers manage their data assets. Additionally, by sharing data from the system that are FAIR principles compliant, researchers facilitate knowledge discovery and increase the chance of possible collaboration, which are beneficial most for new and experienced analysts and research personnel.

# The Three Point FAIRification Framework

**Findable**

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

F1. (Meta) data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta) data are registered or indexed in a searchable resource

**Accessible**

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

A1. (Meta) data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

**Interoperable**

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta) data use vocabularies that follow FAIR principles

I3. (Meta) data include qualified references to other (meta) data

**Reusable**

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta) data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta) data are released with a clear and accessible data usage license

R1.2. (Meta) data are associated with detailed provenance

R1.3. (Meta) data meet domain-relevant community standards

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).

# ROLE OF LOGILAB ELN IN FAIR

## ABOUT LOGILAB ELN

Logilab ELN is Agaram Technologies' generic Electronic Lab Notebook to enable the lab users to document protocols & procedures, enter lab results, scientific and research observations, notes and other data & perform calculations in paperless electronic format.

Logilab ELN is a proven and dependable system that provides a fully configurable sheet template with test- based workflow design to meet the needs of QA/QC and R&D operations for industries including pharmaceutical, life sciences, biopharmaceutical, chemical, petrochemical, environmental, food, feed, milling, and dairy.

Logilab ELN is designed to capture data in a spreadsheet like template called as Labsheet or lab notebook. Labsheet templates can be designed by scientists by dragging and dropping generic fields into the Labsheet and creating a form like input template depending on the type of test, experiment or research task. This makes the ELN software usable for any digital data capture application for lab personnel, research scientists, process and process research personnel.

Labsheets can be designed with data fields like text, numeric, dop-down list, date, time, formula, image, hyperlinks etc.,

Logilab ELN has been designed and developed with a unique feature of protocol management by which set of laboratory procedures and instructions can be configured and corresponding results can also be captured in the same. It has very wide variety of rich features namely data input, research information, tables, images and charts preparation, drawing chemical diagrams, tagging of fields, etc.

Logilab ELN is designed to capture data from any analytical instrument that has RS232 and TCP/IP and also PC-based instruments that can either export results in ASCII, Excel, CSV formats or has the capability to can print a result report.

The system is a fully scalable, flexible, enterprise system designed to streamline test or experiment along with instrument data capture (via SDMS) in a controlled environment, calculate and integrate with external system. It helps laboratories to adopt paperless processes.

Our ELN software ensures easy adaptability, time-savings due to faster configuration and operations and reduced complexity to use resulting in better customer experience.

# LOGILAB ELN vs FAIR PRINCIPLES

Let us now discuss how Logilab ELN helps organizations to fulfil FAIR principles

## F - Findable

F1 - (Meta) data are assigned globally unique and persistent identifiers

Globally unique and persistent identifiers remove ambiguity in the meaning of your published data by assigning a unique identifier to every element of metadata and every concept/measurement in your dataset.

It must be globally unique (i.e., someone else could not reuse/reassign the same identifier without referring to your data). You can obtain globally unique identifiers from a registry service that uses algorithms guaranteeing the uniqueness of newly minted identifiers.

It must be persistent. It takes time and money to keep web links active, so links tend to become invalid over time. Registry services guarantee resolvability of that link into the future, at least to some degree.

F2 - Data are described with rich metadata

In creating FAIR digital resources, metadata can (and should) be generous and extensive, including descriptive information about the context, quality and condition, or characteristics of the data. Rich metadata allow a computer to automatically accomplish routine and tedious sorting and prioritising tasks that currently demand a lot of attention from researchers.

F3 - Metadata clearly and explicitly include the identifier of the data they describe

The metadata and the dataset they describe are usually separate files. The association between a metadata file and the dataset should be made explicit by mentioning a dataset's globally unique and persistent identifier in the metadata.

F4 - (Meta) data are registered or indexed in a searchable resource

Identifiers and rich metadata descriptions alone will not ensure 'findability' on the internet. Perfectly good data resources may go unused simply because no one knows they exist. If the availability of a digital resource such as a dataset, service or repository is not known,

then nobody (and no machine) can discover it. There are many ways in which digital resources can be made discoverable, including indexing.

## How Logilab ELN fulfils this principle

- ✓ The sample ID is unique in Logilab ELN
- ✓ All the required meta data information about (of) sample are be available in Logilab ELN. For example, customer id, date of collection, type of sample, etc. This will enrich the sample (Data)
- ✓ There is a  logical link between meta data and identifier in Logilab ELN
- ✓ Search functionality in Logilab ELN is available in such a way that meta data and identifier are easily searchable using certain search criteria like date, user, etc.
- ✓ Final sheet or protocol after data collection is stored as GUID based JSON file

## A - Accessible

A1 - (Meta) data are retrievable by their identifier using a standardised communication protocol

FAIR data retrieval should be mediated without specialised or proprietary tools or communication methods. This principle focuses on how data and metadata can be retrieved from their identifiers*.

A 1.1 - The protocol is open, free and universally implementable

To maximise data reuse, the protocol should be free (no-cost) and open (-sourced) and thus globally implementable to facilitate data retrieval. Anyone with a computer and an internet connection can access at least the metadata. Hence, this criterion will impact your choice of the repository where you will share your data.

A 1.2 - The protocol allows for an authentication and authorisation where necessary

Accessibility is specified in such a way that a machine can automatically understand the requirements, and then either automatically execute the requirements or alert the user to the requirements. It often makes sense to request users to create a user account for a repository. This allows to authenticate the owner (or contributor) of each dataset, and to potentially set user-specific rights. Hence, this criterion will also affect your choice of the repository where you will share your data.

A2 - Metadata should be accessible even when the data is no longer available

- Agaram Technologies | 2 - Dec - 2021

Datasets tend to degrade or disappear over time because there is a cost to maintaining an online presence for data resources. When this happens, links become invalid and users waste time hunting for data that might no longer be there. Storing the metadata generally is much easier and cheaper. Hence, principle A2 states that metadata should persist even when the data are no longer sustained.

## How Logilab ELN fulfils this principle

- ✓ Using web-browser standard http protocol, in Logilab ELN, meta data and data must be retrieved without having to install any additional software or tool
- ✓ Logilab ELN uses a standard protocol like http or https and must not use proprietary communication software like skype, outlook, etc.
- ✓ Logilab ELN uses https protocol which allows authorisation (userid and password) to use the application based on user level access controls and password policies setup in User Management module.
- ✓ In Logilab ELN, even when samples are deactivated or rejected, the meta data (for example, department, customer, etc.) associated with it is still available and searchable.
- ✓ Final sheet or protocol after data collection is stored as GUID based JSON file.
- ✓ Logilab ELN has a very flexible and powerful search engine

## Interoperable

I1 - (Meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation

Humans should be able to exchange and interpret each other's data (so preferably do not use dead languages). But this also applies to computers, meaning that data that should be readable for machines without the need for specialised or ad hoc algorithms, translators, or mappings. Interoperability typically means that each computer system at least has knowledge of the other system's data exchange formats. For this to happen and to ensure automatic findability and interoperability of datasets, it is critical to use (1) commonly used controlled vocabularies, ontologies, thesauri (having resolvable globally unique and persistent identifiers, see F1) and (2) a good data model (a well-defined framework to describe and structure (meta)data).

I2 - (Meta) data use vocabularies that follow the FAIR principles

The controlled vocabulary used to describe datasets needs to be documented and

resolvable using globally unique and persistent identifiers. This documentation needs to be easily findable and accessible by anyone who uses the dataset.

I3 - (Meta) data include qualified references to other (meta)data

A qualified reference is a cross-reference that explains its intent. you should specify if one dataset builds on another data set, if additional datasets are needed to complete the data, or if complementary information is stored in a different dataset. In particular, the scientific links between the datasets need to be described. Furthermore, all datasets need to be properly cited (i.e., including their globally unique and persistent identifiers)

I4 - (Meta) data are richly described with a plurality of accurate and relevant attributes

The data publisher should provide not just metadata that allows discovery, but also metadata that richly describes the context under which the data was generated. This may include the experimental protocols, the manufacturer and brand of the machine or sensor that created the data, the species used, the drug regime, etc. Moreover, R1 states that the data publisher should not attempt to predict the data consumer's identity and needs. We chose the term 'plurality' to indicate that the metadata author should be as generous as possible in providing metadata, even including information that may seem irrelevant.

## How Logilab ELN fulfils this principle

- ✓ Logilab ELN allows meta data to use a formal, accessible, shared and broadly applicable knowledge base
- ✓ Logilab ELN allows meta data to use vocabularies that follow FAIR Principles
- ✓ Logilab ELN can link qualified references to other meta data

## R – Reusable

R1.1 - (Meta) data are released with a clear and accessible data usage license

What usage rights do you attach to your data? This should be described clearly. Ambiguity could severely limit the reuse of your data by organisations that struggle to comply with licensing restrictions. Clarity of licensing status will become more important with automated searches involving more licensing considerations. The conditions under which the data can be used should be clear to machines and humans.

R1.2 - (Meta) data are associated with detailed provenance

For others to reuse your data, they should know where the data came from (i.e., clear story of origin/history, see R1), who to cite and/or how you wish to be acknowledged. Include a description of the workflow that led to your data: Who generated or collected it? How has it been processed? Has it been published before? Does it contain data from someone else that you may have transformed or completed? Ideally, this workflow is described in a machine-readable format.

R1.3 - (Meta) data meet domain-relevant community standards

It is easier to reuse data sets if they are similar: same type of data, data organised in a standardised way, well-established and sustainable file formats, documentation (metadata) following a common template and using common vocabulary.

## How Logilab ELN fulfils this principle

- ✓ Logilab ELN is licensed based on number of users

- ✓ Logilab ELN has workflow configuration using which the users can review and approve. The reviewing and approving users can view the details of data generation i.e., who created it, at what time and other details.

- ✓ Logilab ELN is a generic electronic Lab Notebook with out-of-box functionalities which can be used across multiple industries

# CONCLUSION

For an organization that has laboratory recording, maintaining and sharing of data will be finding fulfilling of FAIR principles with the help of Logilab ELN very easy in terms of the four principles namely Findability, Accessibility, Interoperability and Reuse.

The following are the benefits that organization can reap by implementing Logilab ELN.

- ✓ Better user experience and confidence
- ✓ Time savings in terms of implementation and roll-out
- ✓ Operational time savings and cost due to automated and paperless processes

✓ Low cost of ownership and reduced overhead costs due to standard technology architecture.

✓ Low compliance cost due to adherence to regulatory and industry standards of data integrity

✓ Minimum errors due to automated processes and consistent result generation.

For more information about Agaram Technologies' Logilab ELN, please refer to the website page: https://www.agaramtech.com/product/logilab-eln-software/

# REFERENCES

1) FAIR Principles: https://www.go-fair.org/fair-principles/
2) https://en.wikipedia.org/wiki/FAIR_data
3) GOFAIR: https://www.go-fair.org/