Delivering the 3b generation of LNMO cells for the xEV market of 2025 and beyond

Data Management Plan

Horizon 2020 | LC-BAT-5-2019
Research and innovation for advanced Li-ion cells (generation 3b)
GA # 875033

Deliverable No.	D9.2	
Deliverable Title	Data Management Plan	
Due Date	30-06-2020	
Deliverable Type	Report	
Dissemination level	Public	
Written By	Boschidar GANEV (AIT)	16-06-2020
Checked by	3beLiEVe Steering Committee	18-06-2020
Approved by	Michele DE GENNARO (AIT)	24-06-2020
Status	Final (1st version)	24-06-2020

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no. 875033.

Revision History

Version	Date	Who	Changes
1	16.6.2020	B. Ganev	First version
1	24.06.2020	B. Ganev	Finalized first version (further versions to follow after periodic updates of this DMP)

Project Abstract

3beLiEVe aims to strengthen the position of the European battery and automotive industry in the future xEV market by delivering the next generation of battery cells, designed and made in Europe, for the electrified vehicles market of 2025 and beyond. The project activities are focused on three domains:

- Development of automotive battery cells that are highly performant (high energy density, fast charge capability, long cycle life) and free of critical raw materials such as cobalt and natural graphite;
- Development and integration of sensors into and onto the cells to enable smart, adaptive operating strategies and advanced diagnostics in order to extend the useful life of the battery in first and second life applications and improve safety;
- A comprehensive manufacturing approach that is designed from the outset for a circular economy and industrial volumes. This encompasses green manufacturing processes for cell, module and pack, as well as recyclability assessment of the components, and a target lifecycle cost of 90 €/kWh at scale.

The project will deliver two 12kWh-demonstrator battery packs at TRL6 and MRL8. These aim at demonstrating the 3beLiEVe technology performance for applications in light duty (i.e. passenger cars, freight vehicles) and commercial vehicles (i.e. city buses and trucks) in fully electric/plug-in hybrid (BEV/PHEV) configurations.

The strong and complementary consortium of 21 partners from 10 different European countries representing industrial companies, SMEs, RTOs and academia is coordinated by AIT Austrian Institute of Technology. 3beLiEVe is scheduled to run from January 1st, 2020 to June 30th, 2023, for a total duration of 42 months and has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no. 875033. A full list of partners and funding can be found at: https://cordis.europa.eu/project/id/875033.

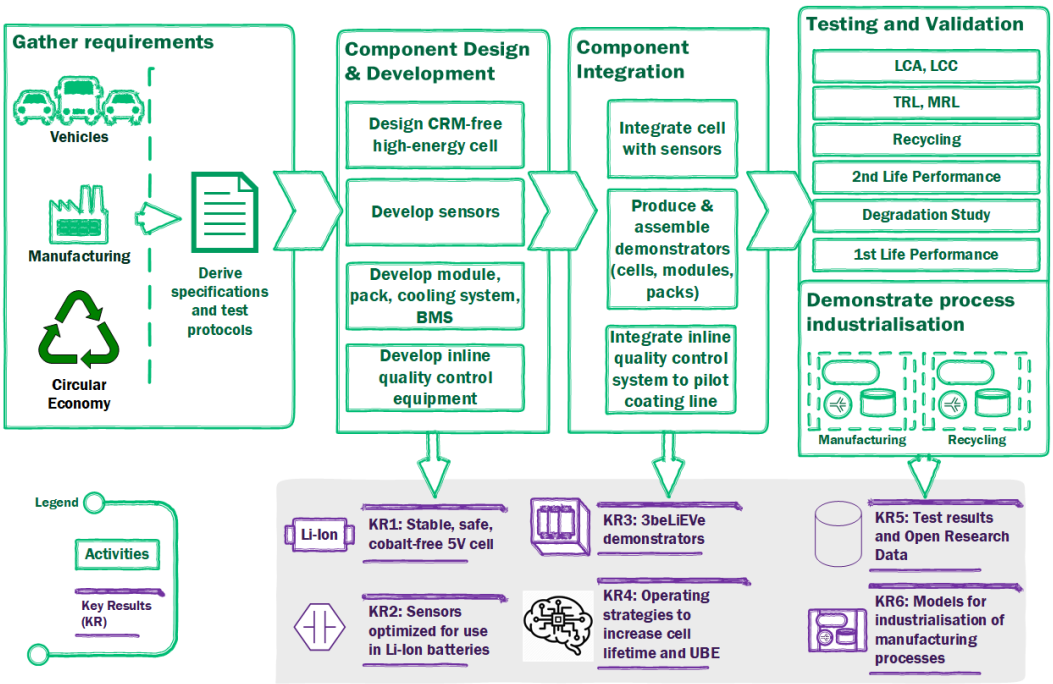

Figure 1: Overview of major 3beLiEVe project steps

Executive Summary

3beLiEVe has chosen to opt-in to the Open Data Research Pilot (ORDP). For many 3beLiEVe project members, however, participation in the ORDP is new territory. In this first version of the Data Management Plan for 3beLiEVe, the focus is therefore on providing background on the ORDP in the context of H2020. The basis and requirements of the ORDP are briefly summarized in Section 1. The FAIR data principles are described, and other resources that may be useful at later stages of the data management process are also listed as references. This is intended to provide an overall orientation and introduction to the subject matter.

Section 2 outlines the stepwise approach that 3beLiEVe will take towards further developing the DMP. The process entails capturing all the anticipated possibilities in a general, low-detail first step, and then progressively specifying and detailing information as visibility emerges around data that will be produced during the project. A standard template is adopted for this purpose. Data is a result, and thus IPR management is also applied to datasets as part of the process.

Finally, section 2 also gives a first indicative view of the data may be generated in each work package of the project. At this stage it is high-level, to be further specified in subsequent steps.

The DMP itself is a living document that is updated periodically, typically as new information becomes available, as well as before every project review.

Table of Contents

List of figure and tables

List of abbreviations

Acronym / Short Name	Meaning
DMP	Data Management Plan
EC	European Commission
H2020	Horizon 2020
LCA	Lifecycle analysis
LCC	Lifecycle costing
ORDP	Open Research Data Pilot
WP	Work Package

1. Introduction

In Horizon 2020 the European Commission has launched a flexible pilot for open access to research data (ORD pilot, or ORDP). The pilot aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects, taking into account the need to balance openness and protection of scientific information; commercialisation and IPR; privacy concerns; security; and data management and preservation questions [1]. One of the guiding principles for the ORDP is 'as open as possible, as closed as necessary'. It is therefore possible to opt out of research data sharing at any stage – before or after the signature of the grant agreement, but reasons must be provided.

1.1. H2020 contractual basis for ORDP participation

The legal requirements for participating projects (i.e., those that did not opt out of the ODRP) are set out in article 29.3 of the Model Grant Agreement, included by default in the Grant Agreement unless the project is opted out. The article states that for actions participating in the Open Research Data Pilot, regarding the digital research data generated in the action ('data'), the beneficiaries must deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

- the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
- other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan'.

Link between Open Access and ORD

Open access can be defined as the practice of providing on-line access to scientific information that is free of charge to the reader. In the context of R&D, open access typically focuses on access to 'scientific information' or 'research results', which refers to two main categories:

- Peer-reviewed scientific research articles (primarily published in academic journals)
- Research data.

Under Horizon 2020, each beneficiary *must* ensure open access to all peer-reviewed scientific publications relating to its results. Beneficiaries can choose between the most appropriate route towards open access for them, either "green" open access or "gold" open access.

In light of this, the data underpinning scientific research articles should be considered in the context of the ORDP.

1.2. Scope and types of data covered by the ORDP

Scope: The Open Research Data Pilot applies primarily to the data needed to validate the results presented in scientific publications. Other data *can* also be provided by the beneficiaries on a voluntary basis.

Types of data covered by the ORDP

1. 'underlying data' (the data needed to validate the results presented in scientific publications), including the associated metadata (i.e. metadata describing the research data deposited), as soon as possible;

2. any other data (for instance curated data not directly attributable to a publication, or raw data), including the associated metadata, as specified and within the deadlines laid down in the DMP – that is, according to the individual judgement by each project/grantee.

Beneficiaries must also provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

1.3. Requirements of the ORDP

Projects must meet the following requirements [1]:

1. They must deposit the research data described above, preferably in a research data repository. These are online research data archives, which may be subject-based/thematic, institutional or centralised. The Registry of Research Data Repositories offers useful listings of repositories (cf. section 1.6). The Open Access Infrastructure for Research in Europe (OpenAIRE) provides additional information and support on linking publications to underlying research data. Some repositories like Zenodo (an OpenAIRE and CERN collaboration), allows researchers to deposit both publications and data, while providing tools to link them. Zenodo and some other repositories as well as many academic publishers also facilitate linking publications and underlying data through persistent identifiers and data citations.

2. As far as possible, projects must then take measures to enable third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) this research data.

One straightforward and effective way of doing this is to attach Creative Commons Licences (CC BY or CC0) to the data deposited. The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of adequate license for research data (→1.6 Miscellaneous resources).

At the same time, projects should provide information via the chosen repository about the tools available to the beneficiaries that are needed to validate the results, e.g. specialised software or software code, algorithms and analysis protocols. Where possible, they should provide these instruments themselves.

1.4. Data Management Plans

DMPs are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR → see section 1.5), a DMP should include information on:

- the handling of research data during & after the end of the project
- what data will be collected, processed and/or generated
- which methodology & standards will be applied
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project).

Participating in the Open Research Data Pilot does not necessarily mean opening up all research data. Rather, the focus of the Pilot is on encouraging good data management as an essential element of research best practice.

DMPs during the project lifecycle

Once a project has started, a first version of the DMP (as a deliverable) must be submitted within the first 6 months of the project. The Commission provides a DMP template, the use of which is recommended but voluntary. The DMP needs to be updated over the course of the project whenever significant changes arise, such as (but not limited to):

- new data
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

The DMP should be updated as a minimum in time with the periodic evaluation/assessment of the project.

- If there are no other periodic reviews foreseen within the grant agreement, then such an update needs to be made in time for the final review at the latest.
- Furthermore, the consortium can define a timetable for review in the DMP itself.

1.5. FAIR data principles

The aim of the FAIR principles is to improve the findability, accessibility, interoperability, and reuse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data [2].

The FAIR acronym stands for Findable, Accessible, Interoperable, Reusable, and each of these elements should meet certain criteria:

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.

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 needs to know how can they 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).

1.6. Miscellaneous resources

The following is a listing of resources that may be useful in the data management process.

Repositories

- **Registry of Research Data Repositories**: https://www.re3data.org/
- **Zenodo**: https://zenodo.org/ allows researchers to deposit both publications and data, while providing tools to link them.
- **Metadata Standards Directory**: http://rd-alliance.github.io/metadata-directory/ can be searched for discipline-specific standards and associated tools
- **Example datasets/repositories/useful resources:**
    - https://arxiv.org/ - open access archive
    - https://catalog.data.gov/dataset/li-ion-battery-aging-datasets
    - https://data.nasa.gov/dataset/Li-ion-Battery-Aging-Datasets/uj5r-zjdb
    - https://www.researchgate.net/post/Lithium_Battery_SoC_and_SoH_dataset
    - https://web.calce.umd.edu/batteries/data.htm
    - https://ieee-dataport.org/documents/automotive-li-ion-cell-usage-data-set
    - Data repositories: https://www.cesaer.org/content/5-operations/2020/20200213-white-paper-rdm.pdf, page 12; RDM tools on page 13

3beLiEVE may also deposit its research data in a proprietary database, since the development of this has been contracted with the company providing the media package (logo, website, social media management etc) for the project. We will also consider submitting data to the BIG-MAP project (LC-BAT-12), which at time of writing has however not yet kicked off.

Other tools

- **Template for DMP**: https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx is a template for DMPs provided by the EC
- **DMP Online**: https://dmponline.dcc.ac.uk/ helps you to create, review, and share data management plans that meet institutional and funder requirements.
- **European Data Portal**: https://www.europeandataportal.eu/
- **License selector tool**: https://eudat.eu/services/userdoc/license-selector#UserDocumentation-LicenseSelector-Locatingthetool is provided to ease the selection of the correct license to attach to your data-set or software package without requiring expert knowledge of every available license.
- https://choosealicense.com/ - another license selector
- https://osf.io/ - "a free, open platform to support your research and enable collaboration"
- https://www.openaire.eu/guides - OpenAire guide to Open Science.

1.7. Publications and IPR considerations

The data expected to be generated in the 3beLiEVe project will have important links to publications and questions of IPR. For scientific publications, it will be necessary to consider which datasets will be used to support them. For such research datasets, a decision should be made about whether these are deposited in a publicly accessible repository, and what the terms and conditions for access should be (e.g. free or restricted) – cf. Figure 2. However, the same consideration can hold true for "other" types of datasets as well.

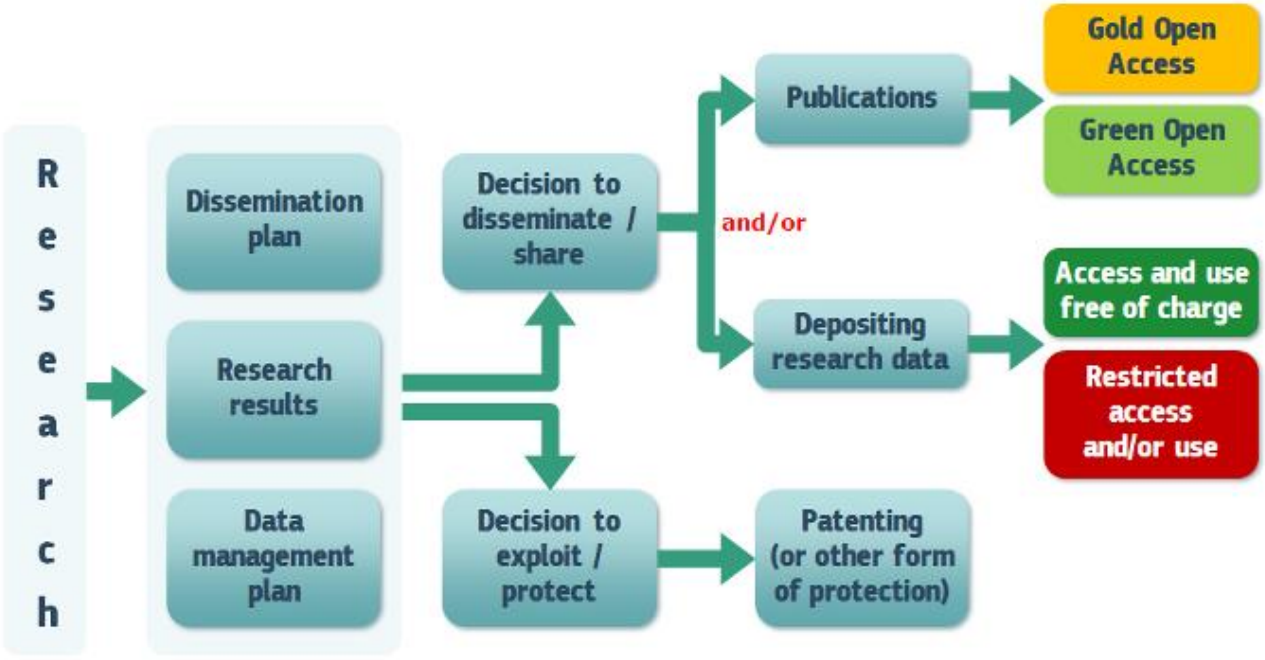

Figure 2: Dissemination and Exploitation pathways for research results. Source: [3]

2. The 3beLiEVe Data Management Plan (DMP)

2.1. Approach to establishing the DMP

3beLiEVe will proceed in consecutive steps to establish the project DMP, as illustrated in Figure 3. The stepwise approach is due to the fact that not all information needed for the full DMP is available at this early stage of the project (M6 at time of writing). At present, there is no full visibility as to the exact nature of the datasets that will be generated during the project. Rather, only indicative information is available at this time. Therefore, the basic idea is to start with the information that is available in the early stages of the project and build on it by specifying and adding further detail as the needed information becomes available.

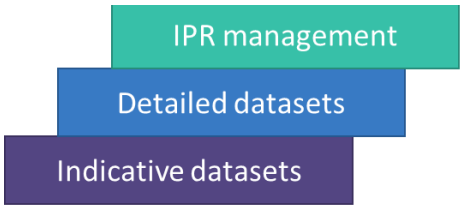

Figure 3: Stepwise approach to establishing the DMP

Indicative datasets – Some indications and ideas concerning datasets that will likely be generated are available at the start of the project. In most cases these are rough outlines derived from forward thinking about possible data outcomes from the topics and activities in the project. The listing at this stage is very broad, to ensure that all ideas and conceivable data are captured. Such an indicative compilation has been gathered and is given in section 2.3. It is possible that in subsequent reviews, some datasets may be removed from the indicative list if they do not materialize as expected.

Detailed datasets – in this stage, information available from the previous step can be further specified, for instance in terms of the type of data that will be generated, the format, an estimate of data volume, and so forth. The growing body of results in the project will provide increasing clarity also on the data. Further datasets that were not captured in the previous stage may also be added in this stage, providing as much detail as available.

IPR management – Once enough detail on the data and metadata aspects is available, the next step is to perform IPR management for these data sets. Data can be considered a project result, which means that it will be necessary to determine ownership (e.g. is there a sole owner, or joint ownership?). Decisions will need to be made as to whether the data should remain confidential to enable exploitation by the consortium, or whether they can be disseminated (cf. also Figure 2). Such IPR considerations will be performed in close coordination with the project's *T8.3 IPR Management* and will be recorded in the DMP. All other actions in relation to compiling, detailing, and publishing data run under *T9.3 Quality, data and risk management*.

Two or more of these steps may be executed simultaneously (e.g. adding a new dataset, mapping the IPR, and deciding on whether to disseminate or exploit). The main thing is that at least the top three steps are performed systematically – whether separately or all at once is incidental.

2.2. DMP Structure

The 3beLiEVe DMP is based structurally on the Template for the Data Management Plan provided by the EC (→ section 1.6). It will attempt to provide an answer to as many of the questions contained in Table 1 as possible.

Table 1: Summary of issues to be addressed in the DMP. Source: extract from DMP template provided by EC (section 1.6)

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable

3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

2.3. Indicative datasets

Figure 4 provides an indicative overview of the types of data anticipated to emerge from each 3beLiEVe work package. At this first, indicative level, the project takes a broad view of the data that may be generated.

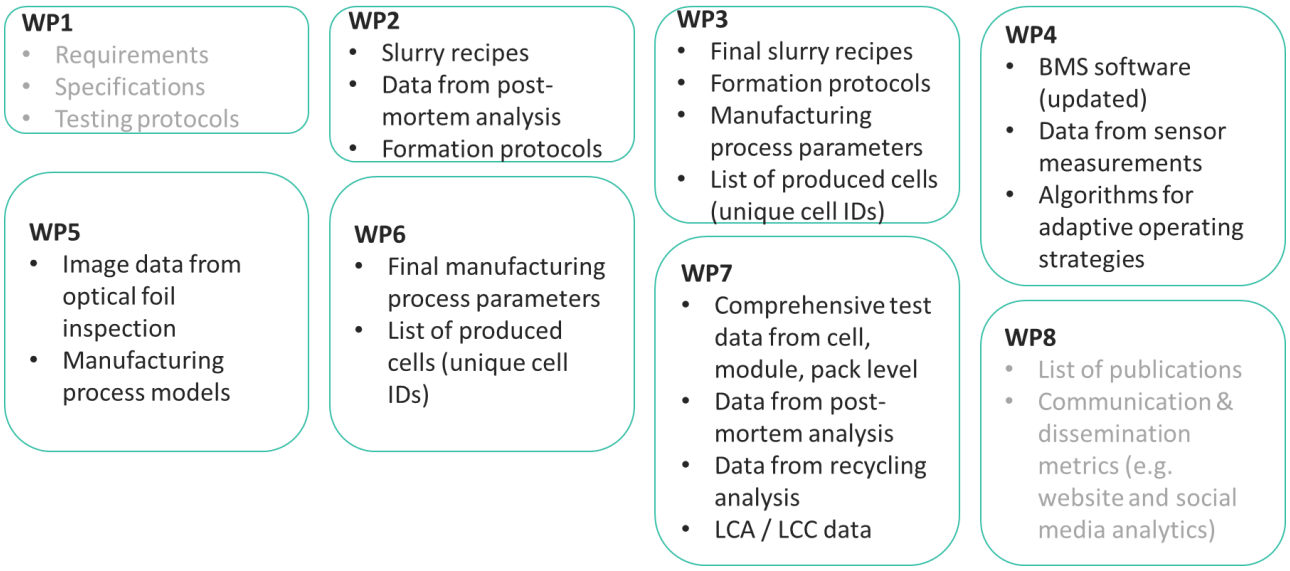

Figure 4: Indicative overview of expected data from each 3beLiEVe work package

Where possible, these indicative data are described in more detail in the following subsections. At this stage, detail is not given for every WP.

2.3.1. From WP1 - Technical requirements and specifications

- Requirements for the cell, module and pack including OEM requirements, circular economy requirements, and those from the LC-BAT-5-2019 call
- Specifications for the cell, module and pack in order to meet the requirements
- Testing protocols for cells, modules and packs.

2.3.2. From WP2 - Materials selection for 3beLiEVe battery cells

From T2.1, T2.2:

- Materials characterization (in particular LNMO, Si-graphite composite, electrolyte): XRD, NMR, EM images, chemical composition, TGA, XPS, etc.
- Slurry recipes
- Cycling data from various cells with different materials combinations
- Cycling and formation protocols

From T2.3 Definition of the final 3beLiEVe cell chemistry:

- Cycling data from various cells with different materials combinations
- Cycling and formation protocols

From T2.4 Materials degradation assessment:

- Images of electrodes, post-mortem characterizations (see non exhaustive list of techniques above)
- Analysis of gas composition

2.3.3. From WP5 - Advanced manufacturing methods and equipment

At time of writing, T5.1 has the highest level of detail regarding the expected data output.

From T5.1 Optical and algorithmic approaches and equipment for inline quality inspection:

AIT-VAC is providing image data of inspected battery foils and generally saves them in a lossless compressed widely compatible image format, e.g. PNG (Portable network graphics) or TIF (Tagged Image Format).

Our current setup uses a camera with width 2336 pixels and 10 bit acquisition intensity precision and 16 bit image intensity precision, which results in approx. 4,7 KB/Line for a 50 um/px resolution setup. For one millimetre of battery foil of 100 mm width, approximately 96 KB of storage space are needed per image. Correspondingly, a length of 1000mm foil would result in an image size of approximately 93MB.

We will illuminate the battery foil from 4 different directions, which increases the storage demands by a factor of 4 for acquired raw image data. In our 1000 mm foil example, would generate roughly 372 MB of raw image data.

The following data will be delivered:

1. Acquired images, from 4 light directions in PNG/TIF format.
2. Surface normal map generated from acquired images in PNG/TIF format.
3. An image mask, whose pixels encode detected defects as different colours, in PNG/TIF format.
4. A text file, possibly CSV format:

Lines encoding detected defects and computed quality metrics at positions denoted as millimetres of foil since the start of the acquisition.

Example content:

- Pos [mm], Measurement type, Value,…
- 10, Pinhole,
- 10, Roughness, 0.1
- 20, Roughness, 0.3
- …
- 200, Crack,
- 200, Roughness, 0.5

5. [Optional] A pointcloud formatted in the PLY format for qualitative analysis containing the same information as (1), (2) and (3).

2.3.4. From WP6 - Production of the 3beLiEVe cells, modules and demonstrator packs

- Information about the cells like electrode composition, balancing and capacity
- A list of produced electrodes and connect these electrodes with the IDs of the cells they are used in
- Provide cells with unique IDs and all information about the formation + check-up.

2.4. Revisions of the DMP

The first version of the DMP will be submitted by M6 of the project, that is end of June 2020.

Revisions to the DMP will be made in the course of the project as visibility increases around the nature, volume, formats, etc. of the data.

At minimum, the DMP should be reviewed and updated, if necessary, before review meetings, and once close to or at the end of the project. That means after M18, after M30, and at or shortly before M42.

3. Conclusions

This first version of the DMP for 3beLiEVe focused on laying the foundation for understanding the ORDP. The aim was to provide project members with a relatively compact overview of the scope and requirements relating to the Open Data Research Pilot under Horizon 2020. In addition to ORDP context, it also undertook to list some major resources that ought to be helpful in the further elaboration of subsequent versions of this DMP. Examples include the FAIR principles, which appear to be an accepted approach to making data findable, accessible, interoperable and reusable. The document also gives a list of resources that the project team may later refer to when specifying the data and applying the FAIR principles.

This document also described the stepwise approach to establishing the DMP for 3beLiEVe, going from general information to more and more specific information as this becomes available. IPR management is an important part of this process and is considered in conjunction with the IPR management (T8.3 in this project) for other (non-data) results.

Finally, a first pass at identifying likely data outputs from each work package was also undertaken, giving the reader an indicative view of what may be expected in later stages.

As a living document, this DMP will be updated in future iterations with progressively more detail and IPR considerations for the various datasets, as well as documentation on how and where data will be or was published.

4. References

[1] European Commission, “Horizon 2020 Online Manual - Open Access & Data Management,” [Online]. Available: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm. [Accessed 15 04 2020].

[2] GO FAIR Initiative, “FAIR Principles,” [Online]. Available: https://www.go-fair.org/fair-principles/. [Accessed 16 04 2020].

[3] European Commission, “Horizon 2020 Online Manual - Open Access,” [Online]. Available: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm. [Accessed 24 04 2020].

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no. 875033.

This publication reflects only the author's view and the Innovation and Networks Executive Agency (INEA) is not responsible for any use that may be made of the information it contains.