Taylor & Francis
Taylor & Francis Group

Science and Technology of Advanced Materials: Methods

ISSN: 2766-0400 (Online) Journal homepage: www.tandfonline.com/journals/tstm20

Designing a unified data structure for solid-electrolyte research projects

Yasuhiro Fujiwara, Kota Suzuki, Satoshi Hori, Yoshiyuki Inaguma, Akitoshi Hayashi, Masaki Azuma, Yasutoshi Iriyama, Hirotoshi Yamada, Tetsuhiro Katsumata, Kazunori Takada, Ryoji Kanno & Masahiko Demura

To cite this article: Yasuhiro Fujiwara, Kota Suzuki, Satoshi Hori, Yoshiyuki Inaguma, Akitoshi Hayashi, Masaki Azuma, Yasutoshi Iriyama, Hirotoshi Yamada, Tetsuhiro Katsumata, Kazunori Takada, Ryoji Kanno & Masahiko Demura (2025) Designing a unified data structure for solid-electrolyte research projects, Science and Technology of Advanced Materials: Methods, 5:1, 2568365, DOI: 10.1080/27660400.2025.2568365

To link to this article: https://doi.org/10.1080/27660400.2025.2568365

Published online: 29 Oct 2025.

SCIENCE AND TECHNOLOGY OF ADVANCED MATERIALS: METHODS
2025, VOL. 5, NO. 1, 2568365
https://doi.org/10.1080/27660400.2025.2568365

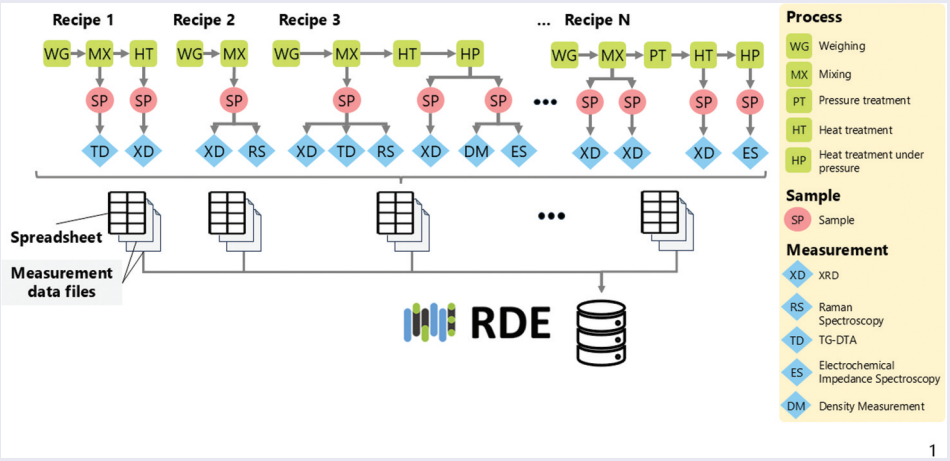
Designing a unified data structure for solid-electrolyte research projects

Yasuhiro Fujiwara[a], Kota Suzuki[b], Satoshi Hori[b], Yoshiyuki Inaguma[c], Akitoshi Hayashi[d], Masaki Azuma[e], Yasutoshi Iriyama[f], Hirotoshi Yamada[g], Tetsuhiro Katsumata[h], Kazunori Takada[i], Ryoji Kanno[b] and Masahiko Demura[a]

[a]Research Network and Facility Services Division, National Institute for Materials Science, Ibaraki, Japan; [b]Research Center for All-Solid-State Battery, Institute of Integrated Research, Institute of Science Tokyo, Kanagawa, Japan; [c]Department of Chemistry, Faculty of Science, Gakushuin University, Tokyo, Japan; [d]Department of Applied Chemistry, Osaka Metropolitan University, Osaka, Japan; [e]Institute of Integrated Research, Institute of Science Tokyo, Kanagawa, Japan; [f]Department of Materials Design Innovation Engineering, Graduate School of Engineering, Nagoya University, Aichi, Japan; [g]Graduate School of Engineering, Nagasaki University, Nagasaki, Japan; [h]Department of Chemistry, School of Science, Tokai University, Kanagawa, Japan; [i]Research Center for Energy and Environmental Materials, National Institute for Materials Science, Ibaraki, Japan

ABSTRACT

We designed a data structure to promote the reuse of experimental data collected from multiple research institutions in a research project on solid electrolytes for next-generation batteries. Leveraging data from multiple research institutions is extremely challenging because the data formats and items differ, preventing their integration and joint use. Therefore, a data structure is required that meets the following criteria: (1) it is readily usable by researchers, (2) it introduces commonality into the data, and (3) it is easily understood by third parties. By analysing experimental data from several institutions, we decomposed each process into 'elementary process units' and proposed a 'process-chain' data structure created by concatenating these units. Furthermore, by adopting RDE – the cloud-based data acquisition system developed by NIMS – and linking the tabular format to the RDE format, we streamlined data collection and management. The proposed data structure is not limited to solid electrolytes but can also be applied to the management of experimental data for other materials, making it useful as a cross-disciplinary method for accumulating material data.



ARTICLE HISTORY
Received 26 June 2025
Revised 11 September 2025
Accepted 26 September 2025

KEYWORDS
solid electrolytes; process chain; data structure; data model; data management; Research Data Express (RDE)

IMPACT STATEMENT

Elementary-process chains unify heterogeneous battery-electrolyte datasets from multiple institutions, converting spreadsheets into a shared RDE repository and providing a scalable template for AI-ready materials curation.

1. Introduction

Solid electrolytes are inorganic materials that are expected to be used in next-generation storage battery technology [1,2]. Currently, active research is being conducted worldwide on various materials, focusing on synthesis processes and property evaluation [3–5]. In Japan as well, multiple research institutions participate in solid-electrolyte projects, synthesizing

CONTACT Masahiko Demura DEMURA.Masahiko@nims.go.jp Research Network and Facility Services Division, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan

diverse solid-electrolyte materials through various processes and evaluating their properties [6,7]. For example, ALCA-SPRING (ALCA-Specially Promoted Research for Innovative Next Generation Batteries) was launched in September 2013. This is a research project that was launched as a special priority technology area of the Advanced Low Carbon Technology Research and Development Program (ALCA) promoted by the Japan Science and Technology Agency (JST). This project is comprising more than 50 institutions and 120 principal investigators [8]. More recently, the GteX program (Innovative GX Technology Creation Project), which promotes research on hydrogen, batteries, and bio-manufacturing in support of green transformation, was launched by JST in October 2023 [9].

In such large-scale projects, it is extremely difficult to leverage data from multiple institutions, because the data formats and items vary, and the data cannot be integrated for collective use. Thus, effective reuse of the data requires that they be recorded in a unified format. This challenge resides at a layer distinct from the data-ingestion methods advocated by various data platforms (e.g [10–15]) and concerns how to document the diverse research data generated in multiple laboratories. Several conceptual frameworks have been proposed for describing research data. One approach uses directed graphs to describe the relationships among samples [16–18]. Another effort records material state changes by combining directed graphs with property descriptions for each graph node [19]. However, to implement these concepts, concrete examples tailored to each specific research field are required. Examples have been examined in the fields of polymers [20] and metals [21], and for charge – discharge data in the battery field [22]; yet no concrete data structure and practical accumulation example have been reported for uniformly describing data from multiple laboratories on solid-electrolyte materials [3].

In this study, we therefore developed a data description method focused on the solid-electrolyte domain with the aim of enabling data reuse; three considerations were emphasized: (1) Provide a simple mechanism that researchers can readily use. (2) Impart commonality to the data. (3) Ensure that the data remain comprehensible to other researchers. To address these issues, we first analysed the characteristics of experimental data obtained from multiple research institutions. Based on insights derived from this analysis, we examined ways to represent experimental processes and view the experimental data as a whole and devised a shared data concept (data structure) [23] that could be commonly used by multiple laboratories. Furthermore, by adopting RDE, the data-collection system developed by NIMS [12,24], we implemented this data structure and created concrete data formats for accumulating and recording data.

2. Analysis of data structures

To clarify practical issues in recording and reusing solid-electrolyte data, we first selected 'Laboratory A', which has a proven record in data accumulation and utilization [25,26], as a case study. This laboratory records experimental results in a unified spreadsheet format. As illustrated in Figure 1, the data sheet

Preprocessing Weighing raw powders to achieve the target composition | Mixing of raw powders | XRD measurement ← #1

Experiment No.	Date	Target Composition	1st Material		Hand milling			Ball milling			Vibration milling	XRD after pretreatment	
			Chemical formula	Weight [mg]	Mixing time [min]	...	...	1st type of ball Diameter[mm]	Qty. [pieces]	...	...	Temperature	File path
MS1	2020/ 7/29	Li3.84(Ge0.6V0.4)0.9 Ca0.1O4	Li2O	**.*	15			**	**				
MS2	2020/ 8/ 9	Li3.84(Ge0.6V0.4)0.9 Ba0.1O4	Li2O	**.*	15			**	**				

← #2 ← #3

Synthesis Heat treatment of mixed raw powders | XRD measurement ← #1

Experiment No.	Date	solid-phase syntesis		High-pressure synthesis			firing in vacuum sealing		XRD	
		Time [h]	Temperature [℃]	Pressure [GPa]	Time [h]	Temperature [℃]	Time [h]	Temperature [℃]	Temperature	File path
MS1	2020/ 8/ 6	12	1000							
MS25	2021/ 1/ 8	12	900							

← #2 ← #3

Conductivity Shape of heat-treated pellets and its molding method | Conductivity measurement ← #1

Experiment No.	Date	Pellet Information						Results		
		General data		Sintered	Spark plasma sintering			Conductivity [mS/cm] @25℃	Activation energy [eV]	File path
		Pellet size [mm]	Thickness [mm]	Temperature [℃]	Pressure [MPa]	Temperature [℃]	Time [min]			
MS6	2021/ 2/ 5	5.95	1	900						
MS7	2021/ 2/ 8	5.95	1.14	950						

← #2 ← #3

Where Area #1 shows **Process Units**, Area #2 shows **Individual Experimental Processes**, and Area #3 shows **Process Parameters**, respectively.

Figure 1. Example of data-recording table.

consists of three individual sheets: Pre-processing, Synthesis, and Conductivity. The Pre-processing sheet corresponds to powder preparation, the Synthesis sheet to reactive synthesis of the target compound, and the Conductivity sheet to ionic-conductivity measurement of the product. Each sheet is generally organized so that the experimental steps proceed from left to right. For example, in the Pre-processing sheet, the leftmost columns record the experiment number and date, followed by the target composition and weighing values of raw powders, after which the mixing method is documented. Thus, one can read the experimental flow of defining the target composition, weighing raw powders, and mixing them. Closer inspection shows that three kinds of mixing – hand mixing, ball-mill mixing, and vibration mixing – are listed, but not all three are necessarily carried out in every experiment. In the 'Synthesis' sheet, solid-state reaction, high-pressure synthesis, and vacuum sealing are listed from left to right, but it appears that only one of them is selected for execution. Although the Conductivity sheet is intended for property measurement, it is noteworthy that it also contains entries for pellet fabrication procedures required for conductivity testing. Hereafter, we refer to the flow of these experiments as 'processes'. Next, we examine the table structure. Data items are described across the first three rows, and actual data are recorded starting from the fourth row. Some columns have a simple structure in which rows 1–3 are merged into a single header (e.g. experiment number, date, target composition, synthesis method, pellet type). Other columns form blocks in which row 1 spans multiple columns and row 2 is further merged to create sub-blocks. These observations indicate that the data are treated hierarchically.

We then performed trial data entry in other laboratories based on Laboratory A's data sheet to identify potential issues. The trials revealed that while many items are common, the order and types of processes exhibit greater variations. Figure 2 charts the experimental flows collected from the various laboratories. First, Hot Press in pattern (1) involves applying uniaxial pressure during heating; this step is absent from Laboratory A's data sheet. Patterns (2)–(4), comprising weighing, hand mixing, ball-mill mixing, pressing, firing, and spark-plasma sintering (SPS), share the same components as Laboratory A's sheet, but their order and number of repetitions differ. Furthermore, to evaluate samples, similar processes were sometimes repeated. For example, in case K0011 firing is performed twice: the sample is first fired and crushed for XRD measurement and then refired to form a pellet for conductivity testing. In the exceptional case K0022, a material once prepared by hot pressing is re-sintered. In other instances, such as K0031, conductivity is measured by directly pelletizing the mixture without firing, intentionally omitting specific processes. Such intentional omission frequently occurs in control experiments aimed at assessing the influence of a particular process. When samples are measured with various analytical instruments, some undergo multiple identical or different measurements, whereas others are not subjected to any measurement at all. For example, in solid-electrolyte evaluation, conductivity tests are performed only on selected samples after material quality is assessed using some of XRD, Raman spectroscopy, and/or TG-DTA. Hence, among data from multiple institutions, there are numerous variations within the same process category (mixing, synthesis, etc.), resulting in diverse process parameters to be recorded. Additional variations

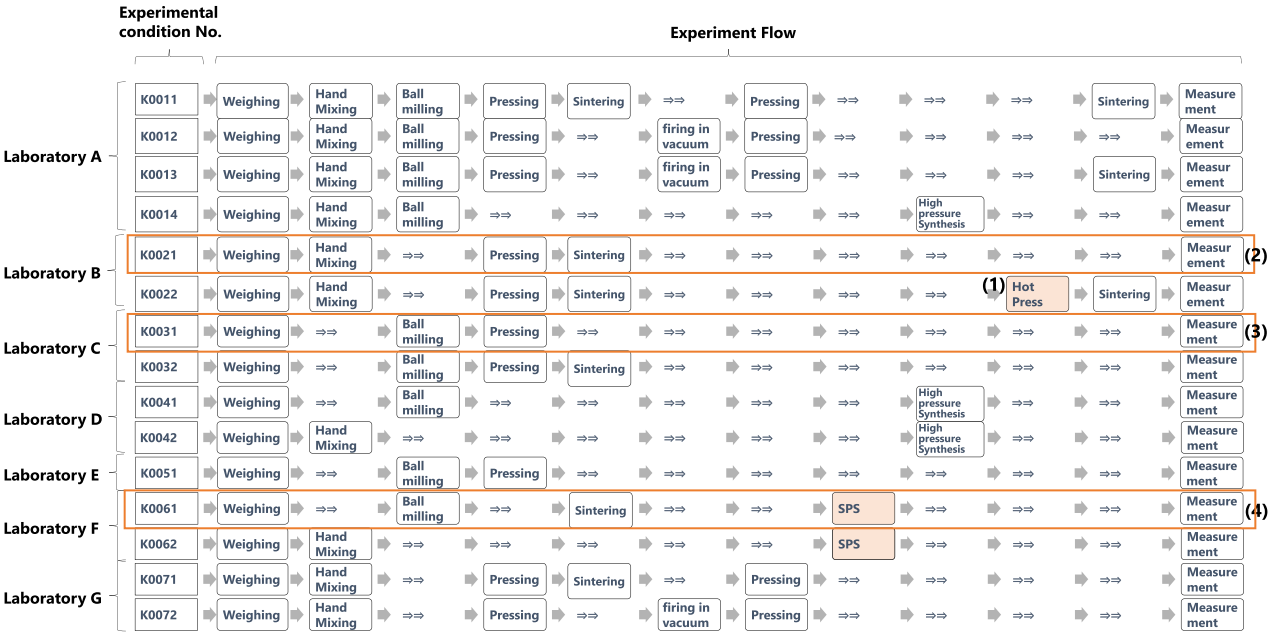


Figure 2. List of solid-electrolyte experiments.

include repetition of the same process, omission of processes, repeated or multiple measurements on the same sample, and omission of measurements. Attempting to record all these variations in a single data sheet would require a huge table that encompasses every pattern. Such a table would contain so many columns that users would find it difficult to locate those relevant to their own processes. Moreover, each time a new process configuration is added, the table would need to be rebuilt. This reconstruction is time-consuming and demands considerable effort to transform existing data sheets for consistency with prior records [27].

3. Construction of the data model

The foregoing analysis highlighted two principal challenges. First is how to represent the wide variety of sample-fabrication processes. A process is composed of a combination of multiple elements. There are multiple possible combinations of these elements, and as research progresses, elements may be added or omitted, changing the patterns [19]. Thus, a representation method that flexibly accommodates the increasing variations in sample-fabrication processes is required. Second is how to associate samples with measurement data. The relationships among processes, the output data of processes, and related processes or samples can become complex [28]. For example, with non-destructive analysis and measurements, a sample fabricated by a certain process may be characterized by XRD and subsequently subjected to conductivity measurement using the same sample. Conversely, the conductivity of a different sample fabricated by the same method as the XRD-characterized sample may be measured. Although both cases appear to provide the same information about a specific process in terms of structure and properties, differences in sampling introduce variations that must be considered. Similarly, multiple measurements of the same sample and measurements of multiple samples fabricated in the same manner represent different sampling strategies. Therefore, a representation method capable of distinguishing such sampling differences is required. In view of these issues, this study proposes a more flexible and unified data model for organizing, integrating, and utilizing research data.

3.1. Representing processes

To flexibly capture diverse workflows, we decompose each process into standardised building blocks that can be concatenated. These building blocks are termed ‘elementary process units’. Analysis of solid-electrolyte synthesis routes revealed five such units, denoted by two-letter labels:

- Weighing (WG): measuring masses of precursor powders.
- Mixing (MX): homogenizing raw materials using a mortar or ball mill, etc.
- Pressure treatment (PT): compacting powders using a hydraulic press.
- Heat treatment (HT): thermally processing materials in a furnace.
- Heat treatment under pressure (HP): simultaneously applying heat and pressure, as in hot pressing or SPS.

Figure 3 shows that all workflows analysed in Figure 2 map onto one or more of these five units. Each unit possesses its own parameter set; for heat treatment (HT), for instance, relevant parameters include ramp rate, dwell time, cooling schedule, and furnace atmosphere. For pressure treatment (PT), the applied pressure, die material, and geometry are principal parameters. A unit may encompass multiple techniques whose parameter lists differ—e.g. mixing (MX) covers hand milling in an agate mortar and ball-milling in a planetary mill. Such techniques are treated as sub-categories under the parent unit. A complete workflow is expressed by chaining the unit labels, yielding a process chain. For example, Figure 4(a) represents weighing followed by mixing; Figure 4(b) denotes weighing, mixing, pressing, and heating; Figure 4(c) comprises weighing, mixing, and pressing only. When the same unit appears multiple times, numeric suffixes are appended in the order of execution (e.g. MX1 for hand mixing, MX2 for ball-milling). Crucially, the suffix is an ordinal, not an indicator of specific conditions.[1]

Because all laboratories apply the same labelling rule, heterogeneous workflows can be systematically represented, enabling researchers to discern differences and commonalities at a glance from the process-chain string. Downstream analyses can filter datasets by identical process chains or by the presence of particular units. The scheme thus standardises data description without constraining experimental creativity.

3.2. Data model for describing the entire dataset

Beyond process information, a complete research dataset also encompasses the analytical and evaluative results obtained for the resulting samples. A coherent framework must therefore link all related entities. We classify the dataset into four item types as follows:

(1) Processes: procedures that synthesise samples.
(2) Samples: physical outputs generated by those processes.

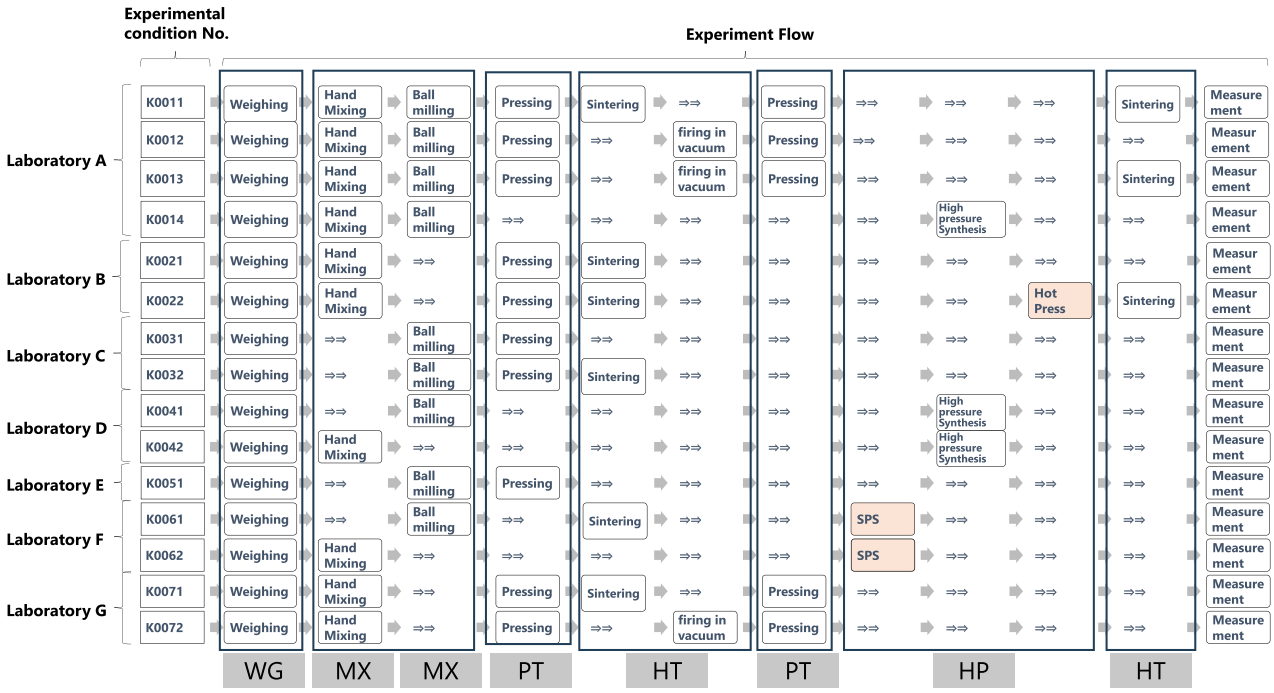


Figure 3. Example of organized experimental flow using process-element units.

(3) Measurements: analytical actions to characterise the samples.
(4) Measurement results: numerical or qualitative outputs from those measurements.

Table 1 summarizes the input-output relationships among these items. A process is defined as an action-type item that accepts a sample as input and outputs another sample. A measurement is likewise an action-type item that accepts a sample as input and outputs a measurement result. Whereas a process transforms the input sample irreversibly – thereby being destructive – a measurement leaves the sample intact and is therefore non-destructive. A sample is an object-type item obtained as the output of a process and used as input for subsequent processes or measurements. A measurement result is an object-type item obtained by supplying a sample to a measurement. Figure 5(a) integrates these relationships. Among the four items, processes are expressed as process chains composed of elementary process units, as proposed in Section 3.1. When a sample is inserted into this representation, as shown in Figure 6(a), a sample follows each elementary unit and serves as the input to the next unit. This representation, however, can be verbose. In particular, when the entire sample is passed to the next process and fully transformed, explicitly depicting that sample adds no information. In such cases, the intermediate sample items may be omitted, and the chain can be described simply as 'a sample produced through process chain A→B→C→ . . . ,' as in Figure 6(b).

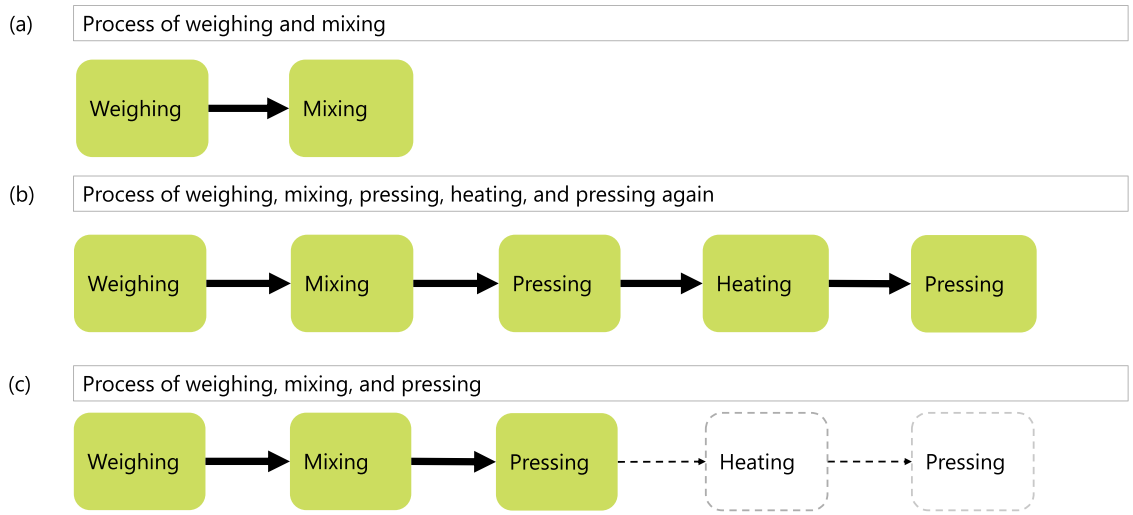


Figure 4. Schematic image of element process – chain.

Table 1. Relationships among the four items that make up the research data.

Operation	Input	Output	Description
Process	Sample	Sample	Destroy sample Irreversible
Measurement	Sample	Measurement Result	Does not destroy sample Can be performed again

Conversely, if a portion of the sample is retained for a different process, the sample item must be explicitly represented. We next consider measurements. Because a measurement and its result form an inseparable pair, the result item need not be shown explicitly; Thus, as in Figure 5(b), it suffices to depict the measurement as acting on a sample produced by a process. Accordingly, the challenges associated with describing solid-electrolyte experiments can be resolved by combining the notions of process chains – assembled from elementary units – and arrows that link the three items: process, sample, and measurement. These ideas enable diverse combinations of pressing, heat treatment, and concurrent heat-and-pressure treatment to be represented together with the various measurements performed on the resultant samples. Based on these ideas, we propose the data model presented here for representing solid-electrolyte experimental data.

To uniquely identify the processes, samples, and measurements stored according to this model, the following identifiers (IDs) are introduced.

- Composition ID (prefix CM + number): identifies the target composition.
- Process ID (prefix PR + number): identifies the synthesis process.
- Sample ID (prefix SP + number): identifies a sample produced by a process.
- Measurement ID (e.g. XD, TD, RS, ES + number): identifies measurement data for each sample.

When managing data across multiple laboratories, a laboratory ID should be added. Examples of these IDs are provided in Table 2. As shown in Figure 7, Laboratory ID, Composition ID, Process ID, Sample ID, and Measurement ID are arranged hierarchically. Therefore, to uniquely identify a process, sample, or measurement, the IDs of higher levels are prepended. For example, a process can be designated by concatenating the Laboratory ID, Composition ID, and Process ID. Adopting this hierarchical ID scheme eliminates the need to check for duplication when assigning IDs at each level. Furthermore, the number of digits should be large enough to accommodate the amount of experimental data in the research project but should also be adjusted appropriately to ensure visibility for users. However, in order to stabilize mechanical processing, the number of digits should be clearly indicated using zero padding or other methods. This approach should also be applied to subsequent ID numbers.

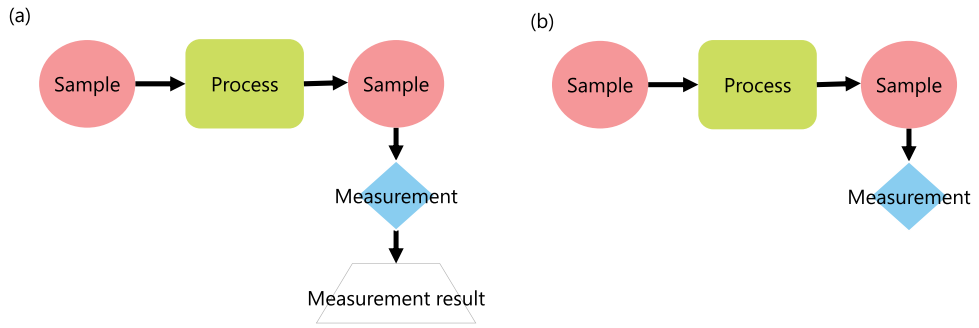


Figure 5. Input – output relationships for process, sample, measurement and measurement-result items.

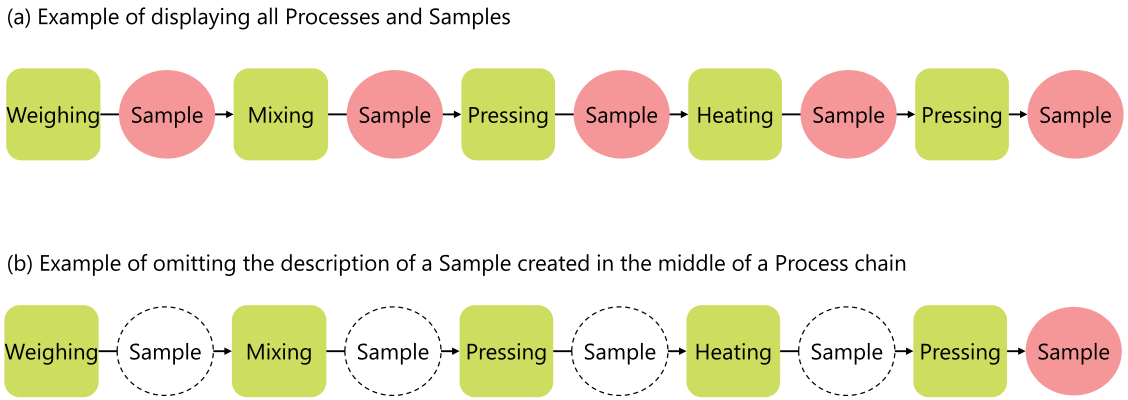


Figure 6. Detailed description of a combined Process -Sample chain.

Table 2. List of IDs.

Items	ID Category	ID Examples
Laboratory	Laboratory	K001, K002, ...
Composition	Target composition	CM0001, CM0002, ...
Process	Process	PR001, PR002, ...
Sample	Specimen	SP01, SP02, ...
Measurement	X-ray diffractometry	XD01, XD02, ...
Measurement	Raman spectroscopy	RS01, RS02, ...
Measurement	Electrochemical Impedance Spectroscopy, EIS	ES01, ES02, ...
Measurement	TG-DTA	TD01, TD02, ...
Measurement	Density	DM01, DM02, ...

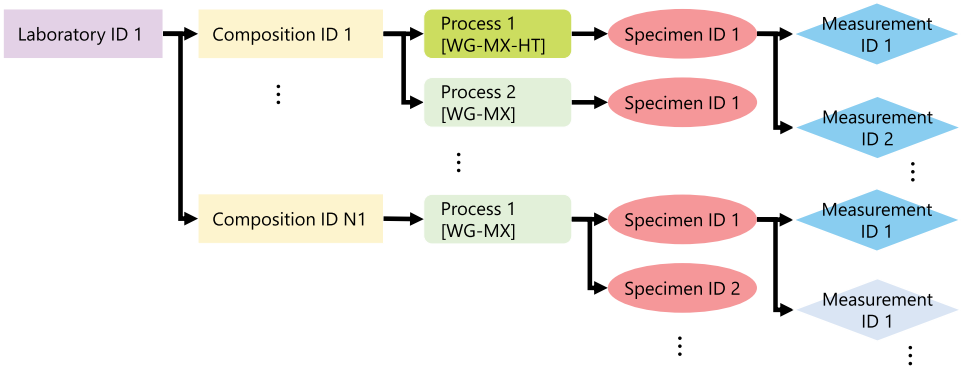


Figure 7. Relationship between the IDs assigned to each data record.

(1) Composition ID
A Composition ID is constructed by combining the prefix 'CM' with numerals (e.g. a four-digit number). Within each Laboratory ID, numbers are assigned sequentially from 1; numbering restarts for different laboratories. No verification is required against compositions from other laboratories. • Even within a laboratory, composition identity is not checked; IDs such as CM0001, CM0002, and so forth are assigned sequentially. As illustrated in Figure 8, when an experiment is repeated for reproducibility, a new ID is issued without referring to past IDs. A user-friendly alias (sample name) may be appended after the Composition ID, for example, using square brackets. For instance, it may be written as CM0001[LS_Si05P05].

(2) Process ID
A Process ID is constructed by combining the prefix 'PR' with numerals (e.g. a three-digit number). As shown in Figure 9, numbers are assigned sequentially within each Composition ID; numbering restarts for different compositions. For example, as shown in Figure 9(a), the identity of the process is not referenced. To help users grasp the structure of the process chain, it is desirable to append the process chain as a label to clarify the meaning of the Process ID. This label may be also enclosed in square brackets such as [process-chain].

(3) Sample ID
A Sample ID is constructed by combining the prefix 'SP' with numerals (e.g. a two-digit number). If the Process ID differs, numbering restarts from 1 (Figure 10(a)). When multiple samples are produced simultaneously from one process, numbers are assigned sequentially within that Process ID (Figure 10(b)).

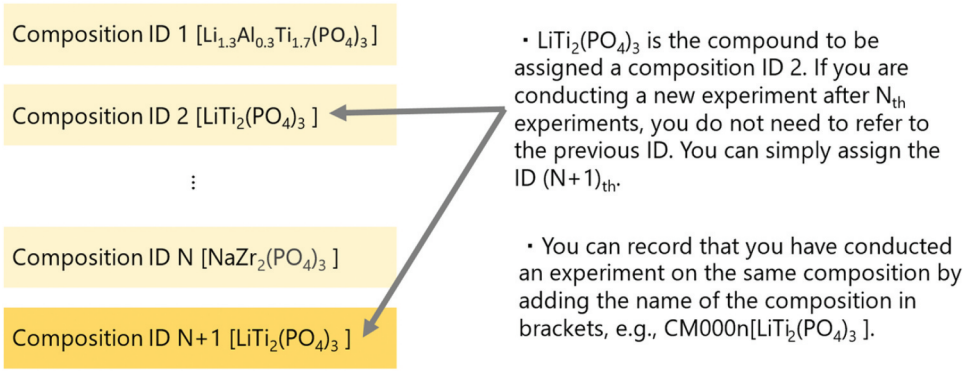


Figure 8. Assignment of composition IDs.

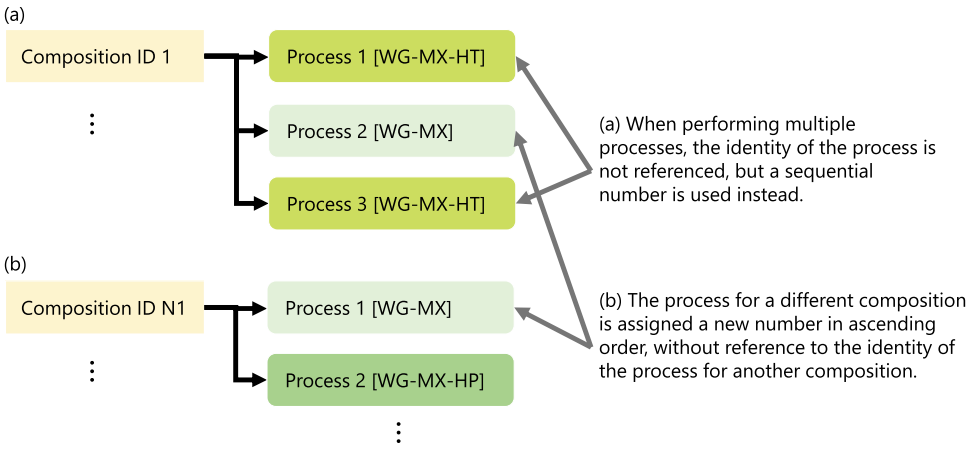


Figure 9. Relationship between Composition ID and Process ID.

(4) Measurement ID
A Measurement ID consists of a two-letter instrument code followed by a number (e.g. two digits). In this study XD denotes X-ray diffraction, TD thermogravimetric analysis, RS Raman spectroscopy, ES electrochemical impedance, and DM density measurement. Instrument codes with similar names are chosen carefully to avoid collisions. Numbers are assigned sequentially from 1 for each sample; numbering restarts for different Sample IDs. When the same measurement type is performed multiple times on a sample, IDs such as TD1, TD2, and so forth are used (Figure 11(a)). When different measurement types are performed on the same sample, distinct Measurement IDs are assigned (Figure 11(b)).

Table 3 shows examples of IDs created under these rules. For instance, experiment S1 denotes the first powder sample obtained by weighing composition LS_Si05P05 under condition WG1; This is expressed as CM001[LS_Si05P05] -PR01[WG1] -SP1. Measurement M1 indicates that the apparent bulk density of that weighed powder was recorded: Accordingly, it is written as CM0001[LS_Si05P05] -PR001[WG] -SP01 -DM01. Measurement M4 represents electrochemical impedance data obtained for a pellet produced by weighing, mixing, pressing, firing, and heat-and-pressure treatment of the same composition. Therefore, it is written as CM0001[LS_Si05P05] -PR001 [WG-MX-PT-HT-HP] -SP01 -ES01. Other samples and measurement data can be documented following the same rules. Combining the process-chain structure with hierarchical item IDs yields the data model illustrated in Figure 12. This model enables the consistent description of relationships such as composition→process→sample→measurement – for example, elemental-ratio data of a weighed powder or conductivity data of a sintered pellet.

Although the process-chain and directed-graph approaches adopted here resemble those of previous studies [16], our model differs by presenting explicit elementary process units and concrete ID schemes for uniquely recording experimental data. Moreover, approaches that treat all measurement conditions and experimental processes uniformly as process nodes [19] do not distinguish destructive processes from non-destructive tests, whereas our model separates instrument-generated measurement data from manually recorded process

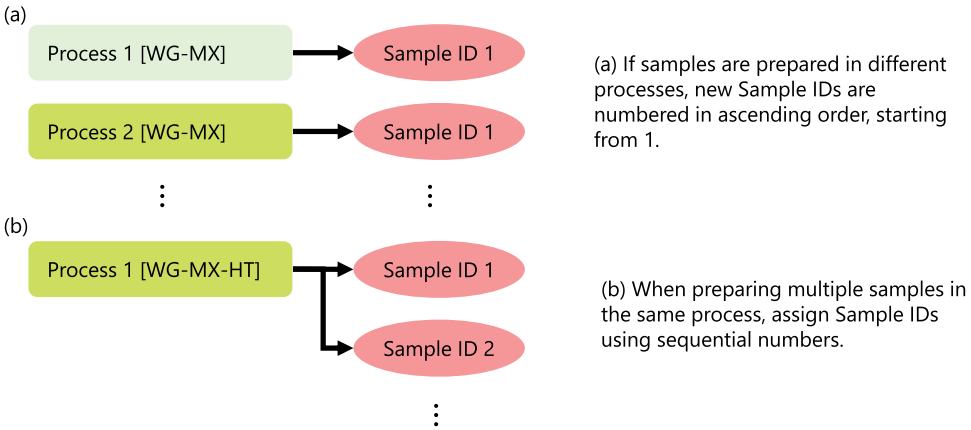


Figure 10. Relationship between Process ID and Sample ID.

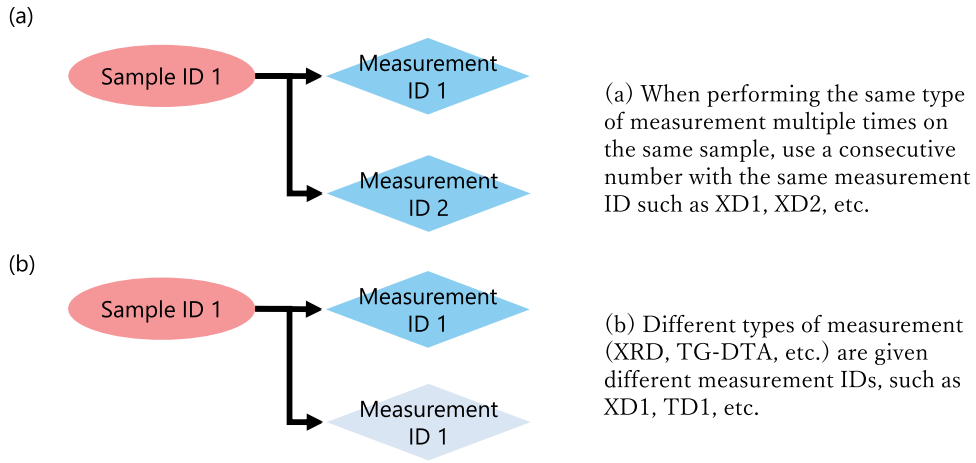


Figure 11. Relationship between Sample ID and Measurement ID.

data, thereby enhancing data integrity and facilitating flexible reuse across diverse research scenarios. Thus, our data model offers a clear separation while maintaining comprehensive traceability. This distinction reinforces data reliability and promotes broad reusability across research domains.

4. Implementing data formats for the data model

This chapter examines data formats [23] for implementing the data model designed in the previous chapter. Methods to implement a data-format data model include XML-based approaches [29], relational-database-based approaches [30], and, as adopted in this study, a tabular data format [22]. Spreadsheet formats are familiar to researchers and well suited to personal data logging; however, when integrating data from multiple researchers, it is necessary to organize the data items and resolve issues such as inconsistencies in naming conventions, as noted earlier [27].

4.1. Data format (1): spreadsheet

In this study, we employ six sheets arranged along the experimental workflow, as exemplified in Figure 13: (1) Target Composition, (2) Process – Weighing, (3) Process, (4) Sample, (5) Property, and (6) Measurement. When users enter data according to the header items of each sheet, related IDs such as Composition ID, Weighing ID, and Process ID are generated automatically. The information to be entered on each sheet is as follows:

(1) Target Composition: manages compositions by entering the local material name, target formula, material category (oxide, sulfide, etc.), and constituent elements.
(2) Weighing: records weighing date, raw-material name, weight, part number, etc., and automatically assigns a Weighing ID linked to the target composition.
(3) Process: records, in execution order, conditions such as mixing, calcination, and sintering, and automatically assigns a Process ID by linking these to the local material name and Weighing ID.

Table 3. Examples of experiment ID containing Process-chain strings.

No	A sequence of IDs	Description
S1	CM0001[LS_Si05P05]-PR001[WG]-SP01	The first raw powder obtained by weighing condition WG to obtain the material with the target composition of LS_Si05P05.
S2	CM0001[LS_Si05P05]-PR001[WG-MX]-SP01	The first mixture obtained by mixing condition MX for the above raw powder.
S3	CM0001[LS_Si05P05]-PR001[WG-MX-PT]-SP01	The first pressurized compact obtained by following the pressing condition PT for the above mixture.
S4	CM0001[LS_Si05P05]-PR001[WG-MX-PT-HT]-SP01	The first heat-treated sample obtained by following the heating condition HT for the above pressurized compact.
S5	CM0001[LS_Si05P05]-PR001[WG-MX-PT-HT-HP]-SP01	The first sintered product obtained by following the sintering conditions HP for the above heat-treated sample.
M1	CM0001[LS_Si05P05]-PR001[WG]-SP01-DM01	The first apparent density data obtained by performing measurement DM01 on the first raw powder.
M2	CM0001[LS_Si05P05]-PR001[WG-MX]-SP01-TG01	The first Thermogravimetric analysis data obtained by performing measurement TD01 on the first mixture.
M3	CM0001[LS_Si05P05]-PR001[WG-MX-PT-HT]-SP01-XD01	The first X-ray diffraction data obtained by performing measurement XD01 on the first heat-treated sample.
M4	CM001[LS_Si05P05]-PR001[WG-MX-PT-HT-HP]-SP01-ES01	The first Electrochemical Impedance Spectroscopy data obtained by performing measurement ES01 on the first sintered object.

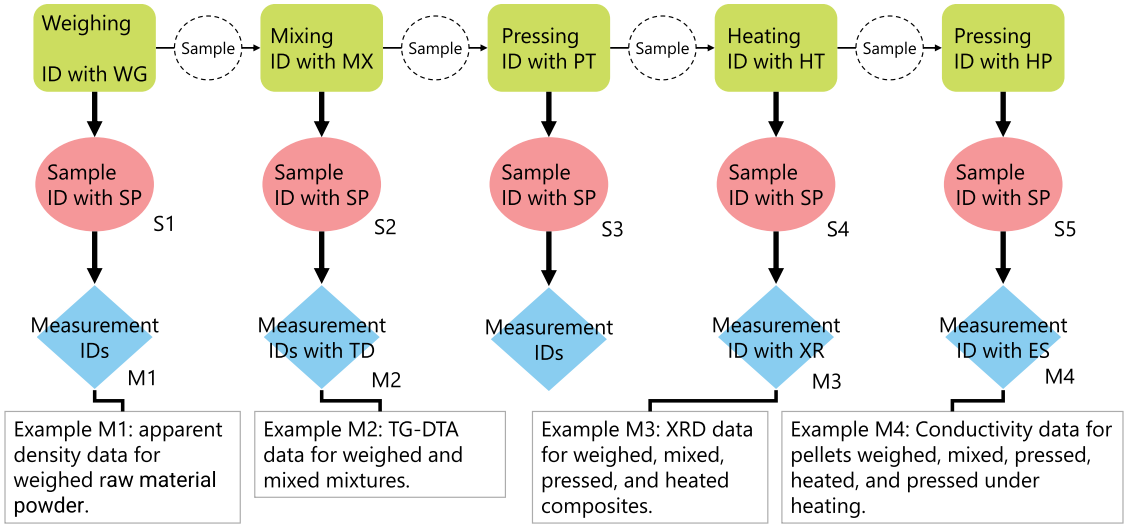


Figure 12. Data structure expresses the relationship between Process, Sample and Measurement..

(4) Sample: records the type of sample obtained through the process (mixture, sintered body, etc.) and its dimensions (diameter, thickness, mass). By entering the local material name and pasting the Process ID from another sheet, a Sample ID is automatically assigned.
(5) Property: allows entry of conductivity test start date, measurement conditions, and measured results. By entering the local material name and the Sample ID, a Property ID corresponding to the conductivity data is automatically assigned.
(6) Measurement (e.g. XRD): allows entry of other measurement information such as instrument name, data-file name, and measurement start date. By entering the local material name and Sample ID, a Measurement ID is automatically assigned.

4.2. Data format (2): RDE system

This section introduces RDE [24], a system capable of storing both process data and measurement data. RDE stores process-level metadata in a key – value scheme [31] together with the process data themselves and enables users to search and retrieve data on the basis of that metadata. By appending metadata suited to the intended use, any type of experimental data can be classified and accumulated [32]. Researchers record process and measurement data using a spreadsheet format that can be easily edited locally and then register the data in RDE via a conversion program. For tabular formats, various

Line 1 →	Item 1	record index		composition ID	weighting ID	process ID	PTa	
Line 2 →	Item 2						Green Compacting	
Line 3 →	Item 3						process starting date	pellet diameter
Line 4 →	Unit							mm
Line 5 →	Data type	string	string	string	string	string	date	number
Lines 6,7 →	Flag	ID						
	Prefix					PR		
Lines 8 →	Input type	Automatic	Manual	Automatic	Manual	Automatic	Manual	Manual
Lines 9- →		K003-CM0001[C1]-PR001[WG1-MX1]	K003	CM0001	WG1	PR001		
		K003-CM0001[C1]-PR002[WG1-MX1-PT1-HT1]	K003	CM0001	WG1	PR002	2021/6/14	10

Sheet name: Process
Line 1: Common index and process element
Line 2: Subcategories of Process
Line 3: Data items included in the subcategory of Process
Line 4: Unit of the data items of the third line
Line 5: Data types of data items (string type, date type, numeric type, etc.)
Lines 6,7: Lines for system use
Line 8: Data entry method for lines 9 and below (manual, automatic, etc.)
Lines 9-: Data lines

Figure 13. Example of data-record table format.

arrangements are possible – for example, subdividing multiple category columns on a single sheet, or entering one sample per row. All of these can be utilized by converting them into the RDE input format. In any case, data unicity is ensured through ID assignment, and the links among data are made explicit by the process chain, thereby allowing subsequent searches within RDE. This enables diverse data to be accumulated in a consistent manner. Figure 14 presents a conceptual diagram for registering solid-electrolyte experimental data using a single Excel sheet. In this sheet, columns corresponding to elementary process units (WG, MX, PT, HT, HP, etc.) are prepared in the order of execution, and labels predetermined for process order are provided. Users can rearrange the columns to match their own workflows—e.g. weighing → mixing → pressing → heating – and fill in only the necessary steps. Once a process step is completed, its contents are consolidated into one row; when the next process step is finished, the previous row is copied and pasted into a new row for continuation. When measurement data are acquired, users copy the Composition + Process ID from the Sample sheet and associate the measurement data files with that ID. RDE imports the Excel sheet of Figure 13, stores the process data rows containing the specified elementary units, and, at the same time, registers the measurement data files with the same IDs. Moreover, if the elementary process units are entered as a taxonomy (an item indicating the hierarchical relationship of data) in the designated field on the RDE screen, RDE automatically generates a tree structure that hierarchically classifies the processes and allows users to browse the data. Even when multiple samples and multiple elementary process units are listed side by side within a sheet, if the workflow for each sample is clearly described, the process chain can be traced to confirm raw-material information and experimental history. As described above, leveraging RDE allows experimental data on solid electrolytes obtained by multiple institutions to be recorded and accumulated in a unified format, thereby facilitating data sharing and reuse.

Thus, by accumulating data in the RDE according to the data structure proposed in this study, data management aligned with the FAIR principles [33] becomes possible.

Specifically, the RDE enables easy data discovery by utilizing user-specified metadata and taxonomies, ensuring data findability. Furthermore, RDE operates stably in the cloud, guaranteeing data accessibility. Moreover, standardizing the IDs – comprising composition ID, process ID, sample ID, measurement ID, etc.—enables interoperability between multiple laboratories. Finally, because data is stored in a consistent structure, machine learning datasets can be easily created, ensuring high reusability of the data.

Furthermore, metadata such as the device used to acquire the measurement data and the measurement conditions are essential. While this information is embedded within the measurement data, it is important to note that the metadata items and formats written to the measurement files can vary depending on the manufacturer and model. In RDE, structured programs to extract such metadata can be written separately according to format differences and pre-registered in what are called dataset templates. By selecting a dataset template corresponding to their own instrument, users can extract metadata appropriately in both machine-readable and human-readable formats. For XRD and TG-DTA, which are also the focus of this research, dataset templates have been developed for multiple instrument

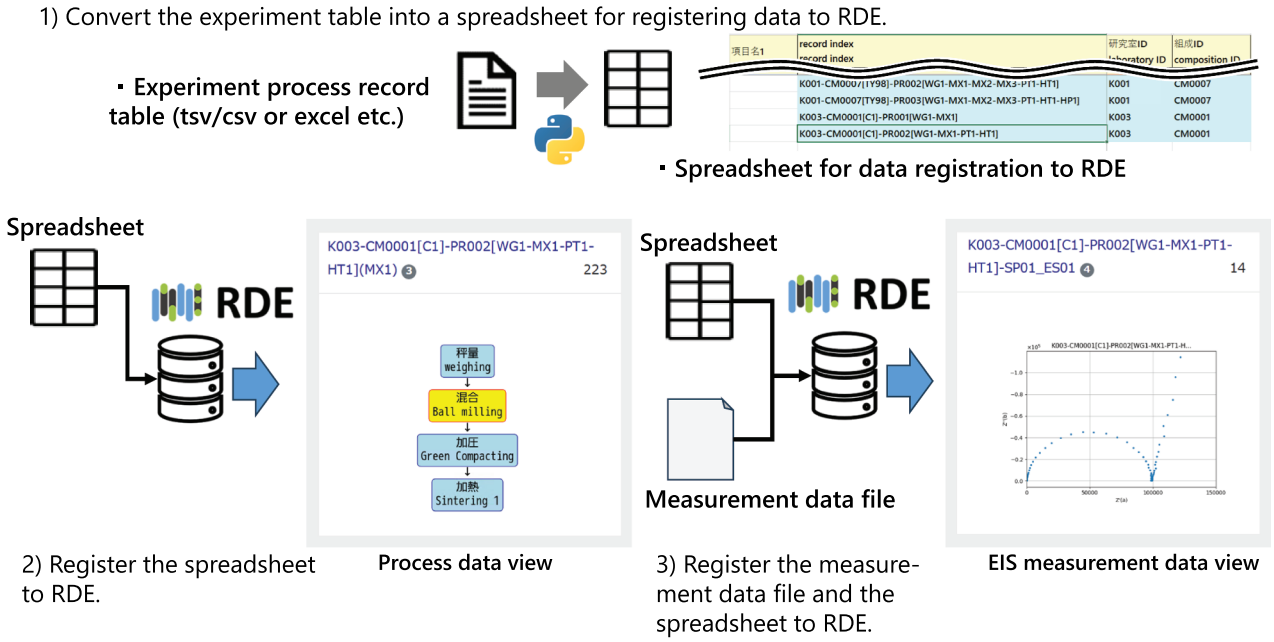


Figure 14. Schematic image of data registration to RDE.

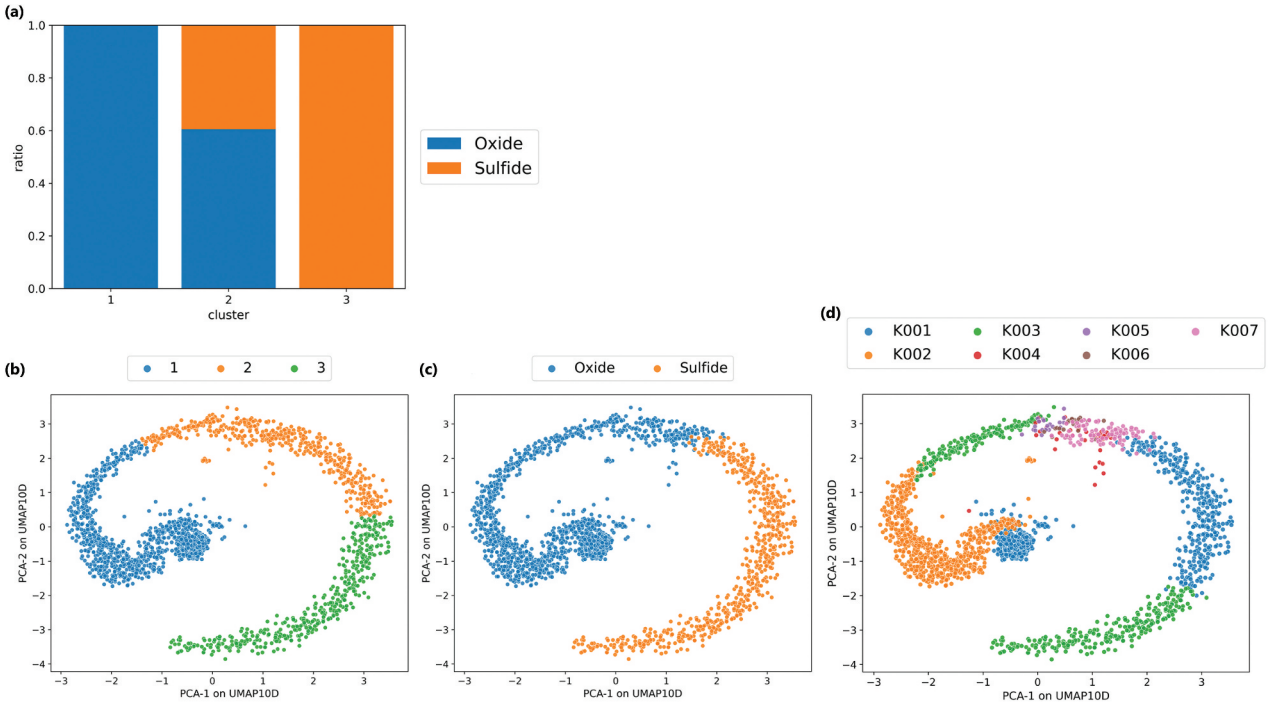


Figure 15. Cluster analysis of process-derived feature embeddings and comparison with external labels. (a) Stacked bar plot showing the proportion of oxide and sulfide labels in each cluster, normalized by cluster size; (b) three clusters identified by k-means in the 10-dimensional UMAP embedding space, visualized in two dimensions by applying PCA; (c) distribution of oxide and sulfide labels on the same PCA-projected space, illustrating that within cluster 2 the two classes occupy distinct regions; (d) distribution of laboratory identifiers (lab IDs) across the PCA-projected space, indicating that each cluster contains data from multiple laboratories.

manufacturers and models, some of which are publicly available on the NIMS website [34].

4.3. Use case

Data collected via RDE can be used to extract datasets suitable for machine learning. As an example, based on the data structure of this study, data from seven laboratories were aggregated and analyzed for characteristics.

Specifically, a table containing extracted process data was created and subjected to cluster analysis. Character strings within the process data were converted to numerical values using one-hot encoding. Dimension reduction to 10 dimensions was performed using the UMAP method [35], followed by K-means clustering [36]. For UMAP hyperparameters, we performed grid searches with n_neighbors set to [10, 20, 30, 50, 80] and min_dist set to [0.0, 0.1, 0.3, 0.5] to find the conditions yielding the highest Silhouette value in the K-means clustering. The results showed that a Silhouette value of 0.66 was obtained with n_neighbors = 10 and min_dist = 0.0, indicating that three clusters were optimal.

Figure 15 shows the results obtained from cluster analysis. When compared with the composition information, Cluster 1 corresponded almost 100% to oxides, Cluster 3 corresponded almost 100% to sulfides, and Cluster 2 contained a mixture of oxides and sulfides (Figure 15(a)). Furthermore, as evident from comparing Figure 15(b,c), within Cluster 2, oxides and sulfides each formed distinct groups, revealing that they occupy non-overlapping regions in the process space. Furthermore, as shown in Figure 15(d), each cluster contained multiple laboratories, indicating that this tendency for oxide and sulfide separation is a cross-cutting feature independent of specific laboratories. Thus, the data structure developed in this study enables the analysis of process data that was previously difficult to handle.

As a result of such data accumulation, data-driven research and materials development have already advanced, for example, the development of machine-learning models to predict the relationship between chemical composition and conductivity in solid electrolytes [37], and studies on autonomous experiments for battery materials [38]. In the future, the organized data are expected to be combined with diverse computational resources to enable more advanced analyses and data utilization [39–43].

5. Conclusions

This study examined data structures and data formats that facilitate the straightforward collection, standardization, and reuse of experimental data generated by multiple laboratories in large-scale solid-electrolyte research projects. First, to structure the experimental data, we clarified the relationships among four items – processes, samples, measurements, and measurement results – and devised a mechanism for expressing the many variations

of experimental workflows as process chains composed of elementary process units. Leveraging this data structure, experimental data recorded in spreadsheet formats, together with assorted measurement files, can be aggregated and stored on a single platform, the RDE system. The RDE-based data-organization approach developed in this work is applicable not only to solid electrolytes but also to the experimental data of a wide range of solid materials. Moreover, because the data-collection methodology implemented in RDE is highly versatile, it can be broadly applied across different materials-research domains.

Note

1. Note: Multiple process units belonging to the same category may also be recorded uniquely. For example: Hand mill = MX_a, Ball mill = MX_b.

Acknowledgements

The authors gratefully acknowledge Yuki Tanaka, Nobuko Kubota, Michiyo Kamiya (Tokyo Institute of Science), and Kazuhiro Yamamoto (Gakushuin University) for their assistance in data collection. We are especially thankful to Michiyo Kamiya and Kazuhiro Yamamoto for their valuable comments on the details of the collected data.

We also thank Hideki Yoshikawa and Hiroko Nagao (National Institute for Materials Science) for fruitful discussions and guidance on operating the RDE system.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the GteX Program (Japan) under Grant Number [JPMJGX23S6]. This work also includes the research achievements supported by the Japan Science and Technology Agency (JST) [JPMJAL1301].

ORCID

Yasuhiro Fujiwara http://orcid.org/0000-0003-1005-6131
Kota Suzuki http://orcid.org/0000-0002-2473-0724
Satoshi Hori http://orcid.org/0000-0001-7845-3518
Yoshiyuki Inaguma http://orcid.org/0000-0002-1174-8407
Akitoshi Hayashi http://orcid.org/0000-0001-9503-5561
Masaki Azuma http://orcid.org/0000-0002-8378-321X
Yasutoshi Iriyama http://orcid.org/0000-0001-8639-0844
Hirotoshi Yamada http://orcid.org/0000-0003-0733-6992
Tetsuhiro Katsumata http://orcid.org/0000-0002-8430-935X
Kazunori Takada http://orcid.org/0000-0001-7568-1806
Ryoji Kanno http://orcid.org/0000-0002-0593-2515
Masahiko Demura http://orcid.org/0000-0002-7308-3041

References

[1] Tan D-H, Banerjee A, Chen Z, et al. From nanoscale interface characterization to sustainable energy storage using all-solid-state batteries. Nat Nanotechnol. 2020;15(3):170–180. doi: 10.1038/s41565-020-0657-x
[2] Ma M, Zhang M, Jiang B, et al. A review of all-solid-state electrolytes for lithium batteries: high-voltage cathode materials, solid-state electrolytes and interfaces. Mater Chem Front. 2023;7(7):1268–1297. doi: 10.1039/D2QM01071B
[3] Medina J, Ziaullah A-W, Park H, et al. Accelerating the adoption of research data management strategies. Matter. 2022;5(11):3614–3642. doi: 10.1016/j.matt.2022.10.007
[4] Gao Z, Sun H, Fu L, et al. Promises, challenges, and recent progress of inorganic solid-state electrolytes for all-solid-state lithium batteries. Adv Mater. 2018;30(17):1705702. doi: 10.1002/adma.201705702
[5] Muzakir M, Yang F, Wang Z, et al. Inorganic solid electrolytes for all-solid-state Li/Na batteries: recent developments and applications. J Mater Chem A. 2025;13(1):73–135. doi: 10.1039/D4TA06117A
[6] Hayashi A, Sakuda A, Tatsumisago M. Development of sulfide solid electrolytes and interface formation processes for bulk-type all-solid-state Li and Na batteries. Front Energy Res. 2016;4(25):1–13. doi: 10.3389/fenrg.2016.00025
[7] Miura A, Rosero-Navarro N-C, Sakuda A, et al. Liquid-phase syntheses of sulfide electrolytes for all-solid-state lithium battery. Nat Rev Chem. 2019;3(3):189–198. doi: 10.1038/s41570-019-0078-2
[8] Japan Science and Technology Agency (JST). Alca-spring: specially promoted research for innovative next generation batteries. 2013. Available from: https://www.jst.go.jp/alca/alca-spring/en/index.html
[9] Japan Science and Technology Agency (JST). Gtex: innovative GX technology creation project. 2023. Available from: https://www.jst.go.jp/gtex/en/index.html
[10] Dima A, Bhaskarla S, Becker C, et al. Informatics infrastructure for the materials genome initiative. JOM. 2016;68(8):2053–2064. doi: 10.1007/s11837-016-2000-4
[11] Draxl C, Scheffler M. NOMAD: the FAIR concept for big data-driven materials science. MRS Bull. 2018;43(9):676–682. doi: 10.1557/mrs.2018.208
[12] Stier S-P, Kreisbeck C, Ihssen H, et al. Materials acceleration platforms (MAPs): accelerating materials research and development to meet urgent societal challenges. Adv Mater. 2024;36(45):2407791. doi: 10.1002/adma.202407791
[13] Brandt N, Griem L, Herrmann C, et al. Kadi4Mat: a research data infrastructure for materials science. Data Sci J. 2021;20:8. doi: 10.5334/dsj-2021-008
[14] Wang R, Xu C, Dong R, et al. A secured big-data sharing platform for materials genome engineering: state-of-the-art, challenges and architecture. Future Gener Comput Syst. 2023;142:59–74. doi: 10.1016/j.future.2022.12.026
[15] Pizzi G, Cepellotti A, Sabatini R, et al. AiiDA: automated Interactive Infrastructure and Database for Computational Science. Comput Mater Sci. 2016;111:218–230. doi: 10.1016/j.commatsci.2015.09.013
[16] Puchala B, Tarcea G, Marquis EA, et al. The materials commons: a collaboration platform and information

repository for the global materials community. JOM. 2016;68(8):2035–2044. doi: 10.1007/s11837-016-1998-7
[17] Statt MJ, Rohr BA, Guevarra D, et al. The materials experiment knowledge graph. Digit Discover. 2023;2(4):909–914. doi: 10.1039/D3DD00067B
[18] Butt AS, Fitch P. ProvONE+: a provenance model for scientific workflows. In: Huang Z, Beek W, Wang H, Zhou R, Zhang Y, editors. Web Information Systems Engineering – WISE 2020. Lecture Notes in Computer Science. Vol. 12343. Cham: Springer; 2020. p. 405–418. doi: 10.1007/978-3-030-62008-0_30
[19] Statt MJ, Rohr BA, Brown K, et al. ESAMP: event-sourced architecture for materials provenance management and application to accelerated materials discovery. Digit Discover. 2023;2(4):1078–1088. doi: 10.1039/D3DD00054K
[20] Lin TS, Rebello NJ, Beech HK, et al. PolyDAT: a generic data schema for polymer characterization. J Chem Inf Model. 2021;61(3):1150–1163. doi: 10.1021/acs.jcim.1c00028
[21] Matsuda A, Demura M, Yamazaki M, et al. A unified model for metallic materials reliability data and its application in NIMS metallic materials database (Kinzoku). Sci Technol Adv Mater Methods. 2025;5(1):2518745. doi: 10.1080/27660400.2025.2518745
[22] Christianna L-N, Tony T, Taylor J-R, et al. Voltaiq data format—a standard for battery data across the lifecycle. Front Energy Res. 2022;10:1059154. doi: 10.3389/fenrg.2022.1059154
[23] Demura M. Design of material data structures: balancing management and utilisation. J Soc Inorg Mater Jpn. 2023;30(427):304–309. doi: 10.48505/nims.4744
[24] National Institute for Materials Science (NIMS). DICE portal - Research Data Express (RDE). 2025. Available from: https://dice.nims.go.jp/services/RDE/
[25] Iwamizu Y, Suzuki K, Matsui N, et al. Search for lithium-ion conducting oxides using predicted ionic conductivity by machine learning. Mater Trans. 2023;64(1):287–295. doi: 10.2320/matertrans.MT-Y2022004
[26] Hori S, Nakamura T. A database of experiments for the development of sulfide-type solid electrolytes. Electrochemistry. 2024;92(4):325–329. doi: 10.5796/denkikagaku.24-FE0401
[27] Mittal D, Mease R, Kuner T, et al. Data management strategy for a collaborative research center. Gigascience. 2023;12:giad049. doi: 10.1093/gigascience/giad049
[28] Voigt S-P, Kalidindi S-R. Materials graph ontology. Mater Lett. 2021;15:129836. doi: 10.1016/j.matlet.2021.129836
[29] Ojala T, Över H-H. Approaches in using MatML as a common language for materials data exchange. Data Sci J. 2008;7:179–195. doi: 10.2481/dsj.7.179
[30] Tremouilhac P, Nguyen A, Huang Y-C, et al. Chemotion ELN: an open-source electronic lab notebook for chemists in academia. J Cheminform. 2017;9(1):54. doi: 10.1186/s13321-017-0240-0
[31] Michel K, Meredig B. Beyond bulk single crystals: the Physical Information File (PIF) for structure–property–processing relationships. MRS Bull. 2016;41(8):617–623. doi: 10.1557/mrs.2016.166
[32] Suzuki M, Nagao H, Harada Y, et al. Raw-to-repository characterization data conversion for repeatable, replicable, and reproducible measurements. J Vac Sci Technol A. 2020;38(2):023204. doi: 10.1116/1.5128408
[33] Wilkinson MD, Dumontier M, Aalbersberg I-J, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3(1):160018. doi: 10.1038/sdata.2016.18
[34] National Institute for Materials Science (NIMS). DICE portal - Research Data Express (RDE) service menu. 2025. Available from: https://dice.nims.go.jp/services/RDE/service_menu/catalog/
[35] Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;37(1):38–44. doi: 10.1038/nbt.4314
[36] Kanungo T, Mount DM, Netanyahu NS, et al. An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans Pattern Anal Mach Intell. 2002;24(7):881–892. doi: 10.1109/TPAMI.2002.1017616
[37] Iwamizu Y, Suzuki K, Kamiya M, et al. Chemical composition-driven machine-learning models for predicting ionic conductivity in lithium-containing oxides. Electrochemistry. 2025;93(2):1–8. doi: 10.5796/electrochemistry.25-71007
[38] Kraus P, Bainglass E, Ramirez F-F, et al. A bridge between trust and control: computational workflows meet automated battery cycling. J Mater Chem A. 2024;12(18):10773–10783. doi: 10.1039/D3TA06889G
[39] Sbailò L, Fekete Á, Ghiringhelli L-M, et al. The NOMAD artificial-intelligence toolkit: turning materials-science data into knowledge. Npj Comput Mater. 2022;8(1):250. doi: 10.1038/s41524-022-00935-z
[40] Bekemeier S, Rêgo C-R-C, Mai H-L, et al. Advancing digital transformation in material science: the role of workflows within the MaterialDigital initiative. Adv Eng Mater. 2025;27(8):2402149. doi: 10.1002/adem.202402149
[41] Roters F, Aslam A, Bai Y, et al. Stahldigital: ontology-based workflows for the steel industry. Adv Eng Mater. 2025;27(8):2402148. doi: 10.1002/adem.202402148
[42] Yakutovich A-V, Eimre K, Schütt O, et al. AiiDAlab – an ecosystem for developing, executing, and sharing scientific workflows. Comput Mater Sci. 2021;188:110165. doi: 10.1016/j.commatsci.2020.110165
[43] Kalinin S-V, Ziatdinov M, Ahmadi M, et al. Designing workflows for materials characterization. Appl Phys Rev. 2024;11(1):11314. doi: 10.1063/5.0169961