

Science ouverte dans le domaine des batteries

Eduardo Redondo-Iglesias¹, Marwan Hassini¹

¹Univ Eiffel, ENTPE, LICIT-ECO7, 69500 Bron, France

RESUME – Dans notre société de plus en plus électrifiée, les batteries lithium-ion sont un élément clé. La recherche sur ces systèmes de stockage évolue rapidement, avec une demande croissante pour des outils interopérables, des modèles de simulation robustes et accessibles, ainsi que des plateformes expérimentales collaboratives. Face au coût des expérimentations et aux nombreux domaines d'utilisation des données de batterie, la nécessité de tendre vers une démarche de science ouverte et reproductible s'impose. Pour cette transition, l'échelle de la communauté scientifique est la plus adaptée. Cet article propose une réflexion autour du cycle de vie des données sur les batteries et présente un ensemble d'actions pouvant être mis en place à l'échelle de la communauté. L'article se veut être un support afin d'initier une réflexion et structurer le besoin du point de vue des actions à mettre en place pour tendre vers une recherche reproductible et cumulative sur les batteries.

Mots-clés – batterie, données, science ouverte, recherche reproductible.

1. INTRODUCTION

Les batteries lithium-ion sont des systèmes de stockage d'énergie incontournables pour les objets du quotidien. À ce titre, une recherche florissante est menée afin de développer des batteries plus durables, plus sûres, moins chères et utilisables dans une large gamme d'usages. La collecte de données joue un rôle central dans la réalisation de cet objectif en éclairant les décisions à différents stades de la vie de la batterie. Ainsi, la quantité de données expérimentales produites croît en même temps que l'intérêt porté aux batteries se renforce. A titre d'exemple, le laboratoire LICIT-ECO7 de l'Université Gustave Eiffel génère 550 gigaoctets de données par an, tandis que le Réseau français de stockage d'énergie électrochimique (RS2E), avec ses 17 partenaires académiques, génère 1 pétaoctet par an [1].

Dans le domaine des batteries, une donnée expérimentale est généralement une mesure au fil du temps de la tension et la température d'une batterie sollicitée par un courant. Cette mesure peut être réalisée en conditions réelles sur un système ou en conditions contrôlées dans un laboratoire. Dans les deux cas, la production de ces données est un processus long et coûteux du point de vue économique et environnemental. L'achat de moyens de collecte des données peut être un frein pour de nombreuses équipes de recherche. La production de données fiables et reproductibles nécessite également un savoir faire [2].

La qualité des données produites est un enjeu majeur pour garantir la fiabilité des analyses sur les batteries. La diversité des travaux portant sur des systèmes alimentés par batteries est grande et croissante. La production de données de qualité est donc nécessaire pour garantir la pertinence des analyses sur ces systèmes. Les travaux d'optimisation de la charge de véhicule, les analyses du cycle de vie ou les analyses technico-économiques sont quelques exemples de travaux s'appuyant sur des données de batteries [3, 4, 5]. La communauté des batteries doit donc porter une attention particulière au partage de ses données vers d'autres communautés scientifiques. La reproductibilité de la recherche est également un enjeu majeur qui permet une validation des résultats par la communauté scientifique.

Cette validation est une condition sine qua non pour envisager une recherche cumulative basée sur un cumul des connaissances et non pas un agglomérat d'initiatives et résultats isolés.

Face à ces enjeux, plusieurs initiatives ont vu le jour à l'échelle mondiale au sein de la communauté scientifique des batteries. De nombreux jeux de données ont été partagés sous licence ouverte et plusieurs articles décrivant leurs contenus en ont assuré la visibilité [6, 7]. Quelques initiatives isolées ont également cherché à promouvoir des bonnes pratiques concernant la réalisation d'expérimentations [8], la définition de métadonnées associées aux données expérimentales ou l'utilisation de logiciels ouverts de traitement des données [2]. Plusieurs équipes de recherche se sont également organisées pour mettre en place des campagnes expérimentales communes au cours desquelles les outils et méthodes étaient mutualisés [9]. Ces initiatives louables sont souvent issues d'efforts isolés dont la mise en œuvre à une large échelle est rarement effective. Au sein de la communauté française s'intéressant aux batteries, l'intérêt pour la recherche reproductible et ouverte est grandissant. Cet article se veut être un support afin d'initier une réflexion et structurer le besoin du point de vue des actions à mettre en place pour tendre vers une recherche reproductible.

2. DONNÉES ET BATTERIES

Afin de mener une réflexion sur les pratiques de recherche au sein de la communauté de recherche sur les batteries, il est important d'avoir un certain nombre de définitions communes. La nature des données sur les batteries, leurs processus de génération et transformation en résultats doivent être formalisés.

Le cycle de vie des données sur les batteries peut être décomposé en plusieurs processus : la collecte, l'enrichissement, la structuration et le traitement. La figure 1 présente un schéma de cycle de vie de la donnée sous la forme d'une séquence linéaire de processus. Dans cet article, un cycle de vie en quatre étapes est proposé sous l'acronyme **CEST (Collecte, Enrichissement, Structuration, Traitement)**. Dans la figure 1, les étapes Enrichissement et Structuration sont représentées dans un même processus parce que, dans un grand nombre de cas, ces deux étapes du cycle de vie se superposent et ne peuvent pas être dissociées complètement car elles sont interdépendantes.

La **collecte** constitue le premier processus dans cette séquence. Ici, les données d'entrée sont sélectionnées et enregistrées pour générer les données brutes. Ce processus permet de sélectionner et mettre à disposition les données expérimentales d'intérêt pour l'étude réalisée. Ce travail est généralement réalisé manuellement et consiste à rassembler le jeu de données le plus riche possible.

Dans l'étape suivante, les données brutes sont **structurées** et **enrichies** avec des métadonnées. Les métadonnées, sont des informations qui permettent de qualifier les données brutes. Elle sont organisées souvent comme une série de couple clé-attribut. Par exemple : "date collecte" = "2024-08-15"; alors que les données brutes sont plus souvent vectorielles (p.ex. série temporelles) ou matricielles (p.ex. images, cartographie de paramètres). Pour garantir la pertinence de cette étape, les métadon-

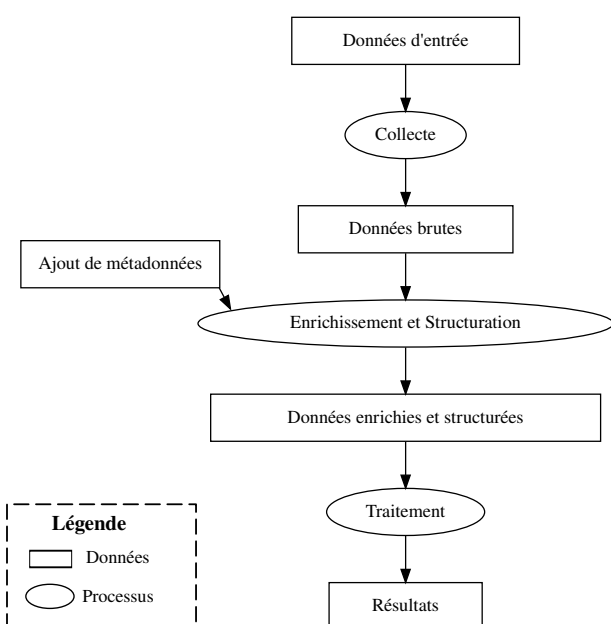


FIG. 1. Cycle de vie de la donnée sur les batteries.

nées doivent être structurées d'après des principes définis par le consensus de la communauté scientifique. Dans un article rédigé par plus de 20 auteurs internationaux de la communauté des batteries, Ward et al. ont proposé une structuration des données et métadonnées pouvant être utilisée [10].

La dernière étape du cycle de vie de la donnée est le **traitement** de données. Il s'agit de l'étape principale qui permet la génération de résultats pouvant avoir leurs propres cycles de vie par la suite. Les données traitées contiennent non seulement le résultat (grandeur physique, matrice de valeurs, etc.) mais aussi la spécification de la méthode de traitement de données. Ce dernier point reste un axe d'amélioration fort pour les travaux scientifiques sur les batteries. Dans la littérature, une grande variété de définitions et de méthodes de mesures des grandeurs caractéristiques existe. Les définitions et méthodes régulièrement proposées dans des normes ou réglementations internationales n'ont pas encore permis d'établir un ensemble de méthodes communes garantissant la comparabilité des résultats. La mise en place de logiciels d'analyse communs à la communauté scientifique et partagé sous licence ouverte est une voie d'amélioration qui permettrait de renforcer la qualité des résultats [2].

Afin d'illustrer le cycle de vie de la donnée sur les batteries, trois exemples peuvent être présentés : analyse de données expérimentales, modélisation, simulation.

2.1. Exemple 1 : analyse de données expérimentales

Expérimentalement, une batterie peut être testée en laboratoire ou au sein d'un système. Les données sont générées généralement à partir d'un ou de plusieurs équipements afin d'obtenir des séries temporelles de tension, courant, température, etc. Aujourd'hui, il existe un grand nombre de cycleurs commerciaux qui proposent chacun un logiciel de contrôle et un format de fichier différent.

La **collecte** de données se fait depuis le fichier brut rendu par le logiciel de cycleur de batteries. Ces données sont ensuite **enrichies** et **structurées** avec des métadonnées (horodatage, conditions de test, référence métrologique des équipements, etc.). Enfin, les données sont **traitées** pour obtenir différentes caractéristiques (capacité, impédance, analyse incrémentale de la capacité, etc.) selon le type d'essai et par différentes méthodes. DATTES est un exemple de logiciel où un effort a été fait dans sa conception pour mener à bien ce cycle de vie [11].

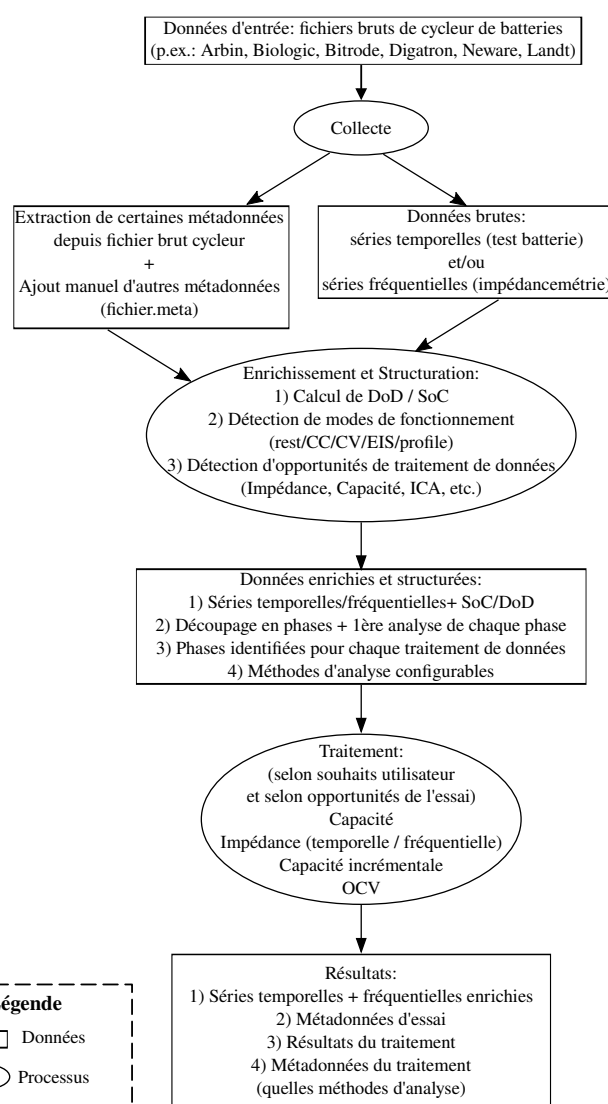


FIG. 2. Cycle de vie de la donnée dans le logiciel DATTES.

DATTES est un ensemble d'outils développés sous MATLAB et compatibles avec GNU Octave distribués sous licence libre [12]. Ce logiciel facilite la prise en main de données expérimentales sur les batteries. Son développement a été pensé pour rendre le traitement des données modulaire, configurable et traçable. En ce qui concerne la modularité, chacune des étapes du traitement pourrait être échangée par un utilisateur (développeur) en surchargeant la fonction correspondante à condition de respecter les interfaces (structuration des entrées / sorties) de la fonction remplacée. La configuration peut se faire pour une grande partie de traitements de données. Par exemple, le calcul de l'état de charge (SoC) qui dépend de la capacité de référence (nominale/initiale/actuelle) et de la condition de charge complète (CCCV à un certain niveau de tension paramétrable). Chacune des méthodes d'analyse est aussi configurable, par exemple pour une identification d'impédance nous pouvons spécifier la topologie du circuit équivalent (R+RC, R+CPE, etc.). Enfin, la traçabilité se fait à deux niveaux : origine de données expérimentales et méthodes d'analyse de données.

La figure 2 montre le cycle de vie de la données sur DATTES. Le résultat de DATTES est ainsi une structure de données contenant les séries temporelles et fréquentielles enrichies (SoC, DoD, mode de fonctionnement du cycleur), leurs métadonnées, les résultats du traitement de données et les métadonnées associées à ce traitement de données (information sur les méthodes d'analyse utilisées pour obtenir ce résultat). Ces résultats sont

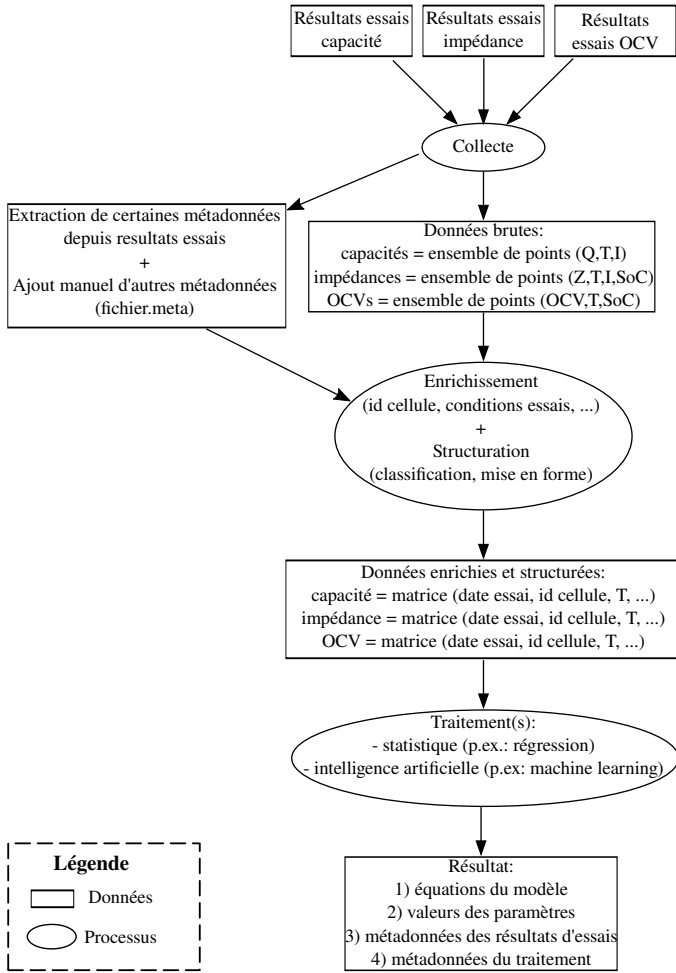


FIG. 3. Cycle de vie de la donnée pour la modélisation.

stockés par défaut sous forme de fichier.mat. Ce fichier peut être ouvert sous MATLAB, GNU Octave et Python (avec la bibliothèque scipy). DATTES propose aussi d'autres formats de fichiers (json, csv) pour rendre l'interopérabilité plus facile.

2.2. Exemple 2 : modélisation d'une batterie

Un autre exemple est celui de la modélisation (électrique, thermique, vieillissement, etc.) d'une batterie. La figure 3 montre une représentation du cycle de vie des données pour la modélisation.

Ici, les données d'entrée peuvent être les résultats d'essais. Par exemple, les valeurs de capacité obtenues à fur et mesure que la batterie vieillit (modèle de vieillissement), les valeurs d'impédance et de tension (OCV) à différentes températures (modèle électrique), ou encore les caractéristiques thermiques (p.ex. coefficient d'entropie, capacité thermique, etc.) nécessaires pour le paramétrage d'un modèle thermique. La **collecte** d'un grand nombre de ces grandeurs peut être faite depuis les résultats fournis par DATTES. L'**enrichissement** avec des métadonnées consisterait à définir dans quelles conditions ces paramètres ont été obtenus et sur quelles cellules (identifiant unique de cellule). Ces données devront également être **structurées** pour faciliter leur **traitement**. Le résultat fourni à l'issue de ce cycle de vie est composé des équations du modèle, les valeurs des paramètres du modèle ainsi que des métadonnées : d'une part les métadonnées à partir des résultats d'essais (identifiant cellule, conditions d'essais, origine de données) et d'autre part les métadonnées qui décrivent les traitements réalisés.

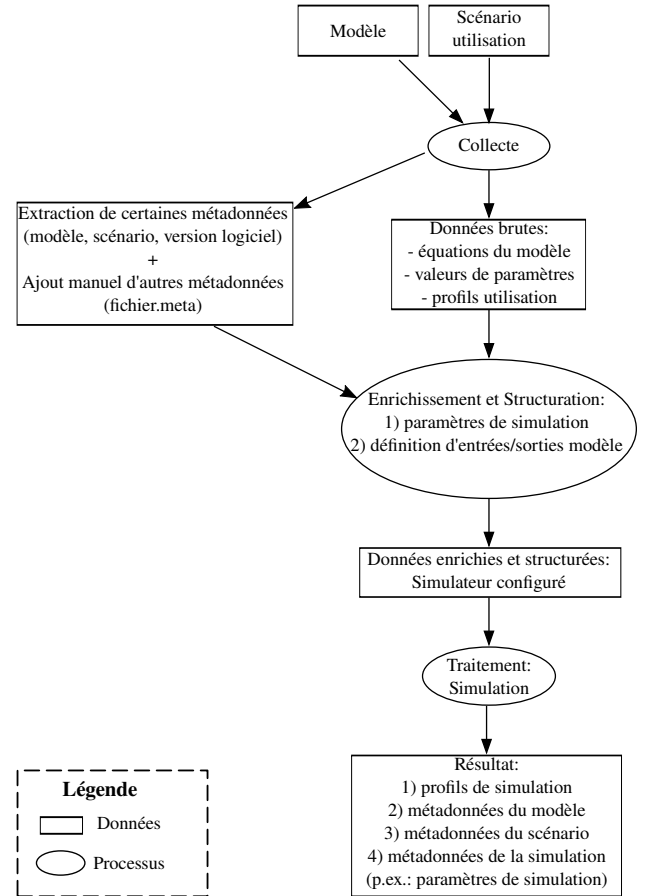


FIG. 4. Cycle de vie de la donnée pour la simulation.

Certains projets en développement, tels que le projet BANANES [13] pour la modélisation du vieillissement ou le projet CERISES pour l'analyse conjointe de plusieurs jeux de données, suivront à terme ce schéma de cycle de vie.

2.3. Exemple 3 : simulation d'une batterie

L'utilisation d'un modèle de batterie permet de générer de nouvelles données. Un nouveau cycle de vie peut donc être défini. D'une manière générale, un logiciel de simulation a besoin d'une définition de modèle (équation et paramètres) et d'un scénario d'utilisation (p.ex. : sollicitation d'une batterie). Ensuite, la simulation doit être paramétrée en choisissant le solveur, le niveau de précision souhaitée, etc. Enfin, la simulation réalisée constitue le résultat de ce cycle de vie.

La figure 4 montre un cycle de vie des données pour la simulation selon le modèle **CEST**. Les données d'entrée sont cette fois un modèle de batterie (électrique, thermique et/ou vieillissement) et un scénario d'utilisation. La **collecte** de données représente le renseignement du modèle (équations et paramètres) ainsi que les vecteurs qui définissent l'utilisation de la batterie (p.ex. : séries temporelles de courant et de température). L'**enrichissement** consiste à collecter différentes informations qui conditionneront la simulation réalisée (paramètres de la simulation). Ces données seront **structurées** dans le logiciel de simulation, par exemple : définition des entrées (profil d'utilisation) et de sorties (réponse de la batterie). Le cycle de vie s'achève avec la simulation, ce qui constitue le **traitement** de ces données structurées. À l'issue de cette simulation, les résultats fournis sont des séries temporelles (entrées et sorties du modèle), les métadonnées du modèle (équations, paramètres, origine du modèle), les métadonnées du scénario (origine de ce cas d'utilisation) et les métadonnées de la simulation (paramètres de simulation, version logiciel, etc.).

2.4. Interactions entre cycles de vie de données

Les trois exemples précédents illustrent des cycles de vie de données batterie (expérimentation, modélisation, simulation). Lorsque chaque cycle de vie individuel et suffisamment renseigné, il facilite l'interopérabilité et la traçabilité.

La figure 5 montre l'interaction des cycles de vie des exemples précédents depuis les données expérimentales jusqu'à la simulation. Ces cycles de vie pourront interagir d'une manière efficace lorsque les méthodes de collecte, enrichissement, structuration et traitement sont renseignées et fournies avec le résultat.

La bonne compréhension de l'origine des données et des différents processus de transformation de ces données ouvre la voie à l'interopérabilité, et cela à deux échelles : interne et externe. L'interopérabilité interne consiste à permettre l'échange des processus (par exemple : différentes méthodes d'acquisition de données ou différentes méthodes d'analyse) à l'intérieur d'un cycle de vie. Par exemple, pour un même jeu de données d'entrée les processus d'analyse de différentes équipes de recherche doivent produire des résultats comparables. L'interopérabilité externe consiste à permettre la récupération de données issues d'un cycle de vie pour les utiliser en entrée d'un autre cycle de vie. Par exemple, le logiciel de simulation utilisé pour la gestion de l'énergie par un laboratoire doit pouvoir utiliser les modèles expérimentaux produits par un autre laboratoire.

Le deuxième avantage de cette organisation est celle de la traçabilité. Dans l'approche proposée (CEST), les données permettent de remonter à chaque étape de chaque cycle de vie et de connaître ainsi comment les données ont été obtenues et transformées. Avec cette connaissance, les résultats peuvent in fine être comparés minutieusement.

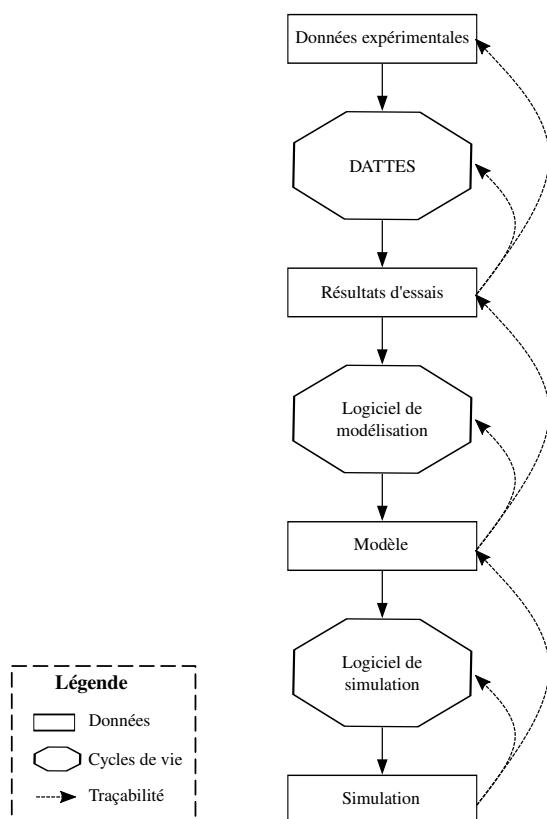


FIG. 5. Exemple d'interaction entre les cycles de vie.

La figure 5 montre une interaction séquentielle entre des cycles de vie. D'autres formes d'interactions peuvent exister : par exemple, les résultats obtenus en simulation peuvent consti-

tuer l'entrée de nouvelles expérimentations pour réaliser une validation des modèles obtenus.

Dans d'autres cas de figure, comme par exemple, un problème d'optimisation ou une analyse de cycle de vie, des multiples simulations peuvent être générées en utilisant un modèle en faisant varier ses paramètres et/ou son scénario d'utilisation. Ici, les cycles de vie seraient imbriqués.

Enfin, cette approche constitue un premier pas vers la recherche reproductible. Les actions à considérer pour y parvenir seront décrites dans le paragraphe suivant.

3. VERS UNE RECHERCHE REPRODUCTIBLE

Face aux coûts de production des données et à l'impératif de reproductibilité de la recherche, des actions doivent être mises en place à l'échelle de la communauté scientifique française des batteries. Cette section présente un certain nombre de bonnes pratiques visant à maximiser l'ouverture des données, à garantir l'interopérabilité des analyses et la reproductibilité des résultats.

La figure 6 présente une synthèse non exhaustive des recommandations visant à promouvoir la science ouverte, comparable et reproductible dans le domaine des batteries. Ces recommandations sont fortement inspirées des principes FAIR [14] et orientées à la promotion de la recherche reproductible dans la communauté scientifique.

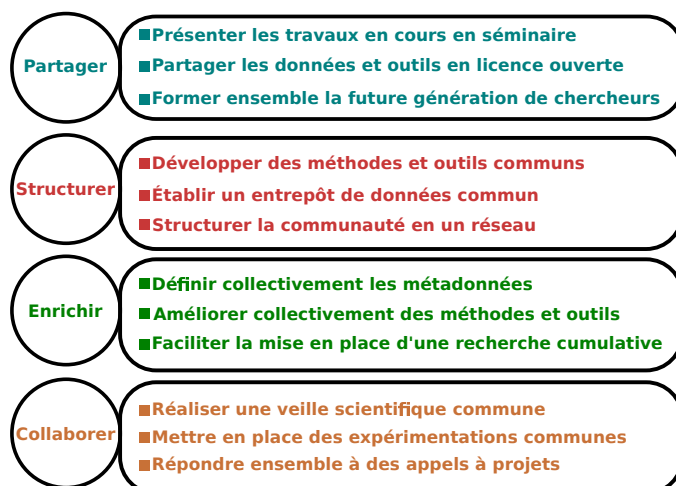


FIG. 6. Bonnes pratiques de promotion de la science ouverte dans le domaine des batteries.

Les bonnes pratiques listées dans la figure ci-dessus sont structurées en quatre piliers :

- **Partager** : dans ce pilier peuvent être incluses les différentes forme de partage de données de logiciels et, en général, de connaissances. Il s'agit ici de rendre les résultats des recherches accessibles à la communauté.
- **Structurer** : ce pilier recense les actions visant à uniformiser les connaissances pour faciliter les interactions.
- **Enrichir** : les actions à mettre en place dans ce pilier ont pour objectif de construire collectivement un écosystème de méthodes et outils pouvant aboutir sur une recherche cumulative.
- **Collaborer** : Ce dernier pilier regroupe des actions ayant besoin d'une coordination forte de la part de la communauté.

Ces quatre piliers impliquent des niveaux de coordination différents allant du plus faible (Partager) au plus fort (Collaborer), mais ils ne suivent pas forcément comme une séquence d'étapes. Dans ce sens, depuis notre point de vue, un certain nombre d'actions sont déjà à un état relativement acquis au sein de notre communauté. Nous pouvons identifier certaines actions du pi-

liar **Partage** : comme par exemple les séminaires (webinaires) du GT Stockage du GDR SEEDS, ou les efforts faits pour l'ouverture de données et de logiciels dans différents laboratoires. En ce qui concerne le pilier **Structuration** le réseau RS2E est un exemple remarquable d'une communauté scientifique ayant structuré son activité de recherche. Le GT Stockage du GDR SEEDS travaille à sa structuration en ayant démarré le processus de rédaction d'une feuille de route commune et en ayant établi un certain nombre d'événements réguliers permettant de rassembler la communauté scientifique.

Les deux derniers piliers **Enrichir** et **Collaborer** sont peut-être ceux qui sont à un état d'avancement moins développé en raison de leur besoin en coordination plus important. À ce jour, les collaborations n'impliquent en général qu'un faible nombre d'acteurs et ont une durée limitée.

On peut néanmoins citer le consortium COMUTES², proposé par 6 partenaires de recherche français : IFPen, CEA, EIGSI La Rochelle, IMS (Univ. Bordeaux), UTC, LICIT-ECO7 (Univ. Eiffel), qui est le fruit d'une suite de projets de recherche ayant impliqué ces partenaires depuis 2007 (SIMSTOCK, SIMCAL, MOBICUS). Cela a permis à ce groupe d'équipes de recherche de développer conjointement des protocoles d'essais de vieillissement, des bases données expérimentales ainsi que des pratiques communes de traitement de données.

Enfin, il est nécessaire que les réflexions et les actions pour la construction d'une recherche reproductible à l'échelle de la communauté implique un grand nombre d'acteurs à tous les niveaux : dès l'expérimentation jusqu'à l'application en passant par la modélisation.

4. CONCLUSIONS

Cet article a présenté les enjeux de reproductibilité et ouverture de la recherche sur les batteries. Après avoir introduit ces concepts, un cycle de vie des données sur les batteries a été présenté. Ce schéma a permis de structurer la réflexion autour des bonnes pratiques de production des données. Dans une seconde partie, les actions pouvant être mises en place par la communauté française travaillant avec les batteries ont été décrites.

Enfin, cet article met en évidence que, lorsqu'il s'agit d'ouvrir la science de batteries, des efforts individuels peuvent être mis en œuvre (ouverture de données, logiciels, méthodes, etc.). Par contre, des actions collectives sont nécessaires pour que cette ouverture soit efficace, vers une recherche reproductible et cumulative.

lysis", *Frontiers in Energy Research*, vol. 10, p. 1023555, 2022.

- [9] M. Lecompte, J. Bernard, E. Calas, L. Richardet, A. Guignard, F. Duclaud, D. Voyer, M. Montaru, B. Crouzevialle, L. Lonardoni, *et al.*, "Experimental assessment of high-energy high nickel-content NMC lithium-ion battery cycle life at cold temperatures", *Journal of Energy Storage*, vol. 94, p. 112443, 2024.
- [10] L. Ward, S. Babinec, E. J. Dufek, D. A. Howey, V. Viswanathan, M. Aykol, D. A. Beck, B. Blaiszik, B.-R. Chen, G. Crabtree, *et al.*, "Principles of the battery data genome", *Joule*, vol. 6, no. 10, pp. 2253–2271, 2022.
- [11] E. Redondo-Iglesias, M. Hassini, P. Venet, and S. Pelissier, "DATTES : Data analysis tools for tests on energy storage", *SoftwareX*, vol. 24, p. 101584, 2023.
- [12] DATTES (Data Analysis Tools for Tests on Energy Storage), 2024. <https://gitlab.com/dattes/dattes>.
- [13] M. Gillet, E. Redondo-Iglesias, H. Helbling, and A. Sari, "Prise en compte du vieillissement des batteries li-ion dans un modèle d'analyse sur cycle de vie", *Symposium de Génie Electrique (SGE 2025)*, 2025.
- [14] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship", *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

- [1] T. Lombardo, M. Duquesnoy, H. El-Bouysidy, F. Áren, A. Gallo-Bueno, P. B. Jørgensen, A. Bhowmik, A. Demortière, E. Ayerbe, F. Alcaide, *et al.*, "Artificial intelligence applied to battery research : hype or reality?", *Chemical reviews*, vol. 122, no. 12, pp. 10899–10969, 2021.
- [2] M. Hassini, E. Redondo-Iglesias, and P. Venet, "Lithium-ion battery data : from production to prediction", *Batteries*, vol. 9, no. 7, p. 385, 2023.
- [3] A. Houbbadi, R. Trigui, S. Pelissier, E. Redondo-Iglesias, and T. Bouton, "Optimal scheduling to manage an electric bus fleet overnight charging", *Energies*, vol. 12, no. 14, p. 2727, 2019.
- [4] T.-D. Patil, E. Vinot, S. Ehrenberger, R. Trigui, and E. Redondo-Iglesias, "Sensitivity analysis of battery aging for model-based PHEV use scenarios", *Energies*, vol. 16, no. 4, p. 1749, 2023.
- [5] M. Gaetani-Liseo and H. Helbling, "Modélisation de micro-réseaux pour l'optimisation technico-économique et environnementale de la gestion de l'énergie", *Symposium de Génie Electrique (SGE 2023)*, 2023.
- [6] G. Dos Reis, C. Strange, M. Yadav, and S. Li, "Lithium-ion battery data and where to find it", *Energy and AI*, vol. 5, p. 100081, 2021.
- [7] Q. Mayemba, R. Mingant, A. Li, G. Ducret, and P. Venet, "Aging datasets of commercial lithium-ion batteries : A review", *Journal of Energy Storage*, vol. 83, p. 110560, 2024.
- [8] M. Dubarry and D. Anseán, "Best practices for incremental capacity ana-