Journal of Surface Analysis, Vol. 29 No.3 (2023) pp. 146 - 154
柳生進二郎, 他　有機半導体材料における光電子収量分光(PYS)データベース構築
https://doi.org/10.1384/jsa.29.146

技術報告

有機半導体材料における光電子収量分光（PYS）データベース構築

柳生 進二郎[1,*], 吉武 道子[1], 長田 貴弘[1], 安田 剛[1], 桑島 功[1], 劉 雨彬[2], 中島 嘉之[2]
[1] *国立研究開発法人 物質・材料研究機構　〒305-0044 茨城県つくば市並木 1-1*
[2] *理研計器株式会社　〒174-8744　東京都板橋区小豆沢 2-7-6*
** YAGYU.Shinjiro@nims.go.jp*

（2022 年 12 月 1 日受付; 2023 年 1 月 22 日掲載決定）

有機・無機半導体材料のイオン化ポテンシャルの情報を与える光電子収量分光（PYS）のデジタルデータベース構築を行った．検索する際のキー（メタデータ）である，(1) 計測機から得られる計測メタデータ，(2) データ解析メタデータ，(3) 測定試料や実験手順などを記載した試料・実験メタデータ，(4) データの利用ライセンスに関するメタデータについて整理した．低分子有機材料については，材料名を一意に特定するために行列表記の構造ファイルから線形表記に変換した SMILES などをメタデータに含めた．これにより外部の有機物に関するデータベースと検索連携を取ることができる．

Create the database of the photoemission yield spectroscopy (PYS)

Shinjiro Yagyu[1,*], Michiko Yoshitake[1], Takahiro Nagata[1], Takeshi Yasuda[1], Isao Kuwajima[1], Yubin Liu[2], and Yoshiyuki Nakajima[2]
[1] *National Institute for Materials Science, Tsukuba, Ibaraki 305-0044*
[2] *RIKEN KEIKI Co., Ltd, Itabashi-ku, Tokyo 174-8744*
** YAGYU.Shinjiro@nims.go.jp*

（Received: December 1, 2022; Accepted: January 22, 2023）

A digital database of photoelectron yield spectroscopy (PYS) was constructed to provide information on ionization potentials of organic and inorganic semiconductor materials. The keys (metadata) for the search were organized as follows. (1) measurement metadata obtained from measurement instruments, (2) data analysis metadata, (3) sample and experiment metadata describing measurement samples and experimental procedures, and (4) metadata about data licenses. For low-molecular-weight organic materials, metadata such as SMILES converted from matrix notation structure files to linear notation was included in order to uniquely identify the material name. This enables linkage with external databases on organic materials.

1. はじめに

近年，データ駆動型材料開発（マテリアルインフォマティクス:MI）[1]が利用されるようになってきた．MI では，開発ターゲット材料周辺の様々な材料データ（例えば，手持ちの実験データ，構造データ，シミュレーションデータ，データベースから抽出したデータなど）を収集・整理・統合する必要がある．有機 EL，有機太陽電池などの有機デバイス材料の研究開発では，有機材料の様々な特性データのほかに，デバイス構造作成時に重要な，電子・ホール移動に関係するエネルギー準位データも必要になる．このエネルギー準位の情報（金属材料では仕事関数，半導体ではイオン化ポテンシャル，有機半導体材料のホール注入準位（HOMO: Highest Occupied Molecular Orbital）に相当する．今後，Ip と記載する．）は，光電子収量分光（PYS：Photo-emission Yield Spectroscopy）によって測定することができる．

中島らは，自社（理研計器）で開発した大気中光電子収量分光装置（AC シリーズ）測定装置を用いて代表的な有機 EL 材料についての測定を行い，2004 年に「有機電子デバイス研究のための有機薄膜仕事関数データ集」を出版した[2]．このデータ集には，有機材料の化学構造，その通称（名称），Ip，光電子放出率（放出光電子数の 1/n 乗の照射エネルギーに対する傾き），測定時の照射光量が記載されている．（残念ながら測定スペクトルは掲載されていない．）近年の MI 研究に対応するためには，このデータ集を参考に，PYS のデジタルデータベース化が必要である．この構築には，データ収集・保管方法の検討とともに，データの検索性や，PubChem[3]などの他のデータベースとのリンクも考えたメタデータ（検索する際のキー）の整理が必要である．

デジタルデータベースは，データ提供，データの格納，データからのメタデータ抽出，抽出メタデータのデータベース化，検索可能なデータベースの公開のプロセスから構成される．メタデータには，4 種類ありそれぞれ整理が必要である．（1）計測機から得られる計測メタデータ，（2）この計測データを使って解析した解析メタデータ，（3）測定試料や実験手順などを記載した試料・実験メタデータ，（4）データの使用許諾（ライセンス）に関するメタデータである．（1）および（2）については，PYS 測定装置は数社から発売されているが，計測・解析メタデータ構造を詳細に公開している理研計器の装置のメタデータを利用した．　理研計器では，MI 研究開発に対応するために，出力ファイルのフォーマットを統一した．そしてこれまでの計測・解析データのフォーマットを新形式のフォーマット（DAT）に変換するソフトを開発し，その詳細について公表した[4]．そして我々は，理研計器から公表された統一フォーマットから計測・解析メタデータを抽出する Python コードを作成し公表した[5]．（3）の試料・実験メタデータについては，検索者が望む検索結果を得るために，その要望を踏まえたメタデータを設定する必要がある．一方でデータ提供者側は，メタデータを入力しなければならないことから，多くのメタデータを設定することは現実的ではない．そこで，双方の折り合いがつくメタデータについて検討を行った．また，試料名については，表記ゆれが多く試料名による検索が機能しない場合がある．そこで今回，主に扱う低分子有機材料については，構造ファイル（MOL）[6]を添付することとした．Python のケミカルインフォマティクスライブラリーの RDKit[7] を用いることで MOL ファイルから SMILES[8]や InChI[9]といった線形表記法に変換することで一意に試料名を特定することができ，これを用いることで PubChem[3]などの外部の有機材料のデータベースとの検索の連携を取ることができる．（4）について，データベースを作成するにあたりデータのライセンスを明確にしておくことが，利用の観点で重要である．本作成のデータベースでは，データ作成者がみずからのデータの 2 次利用を許可す

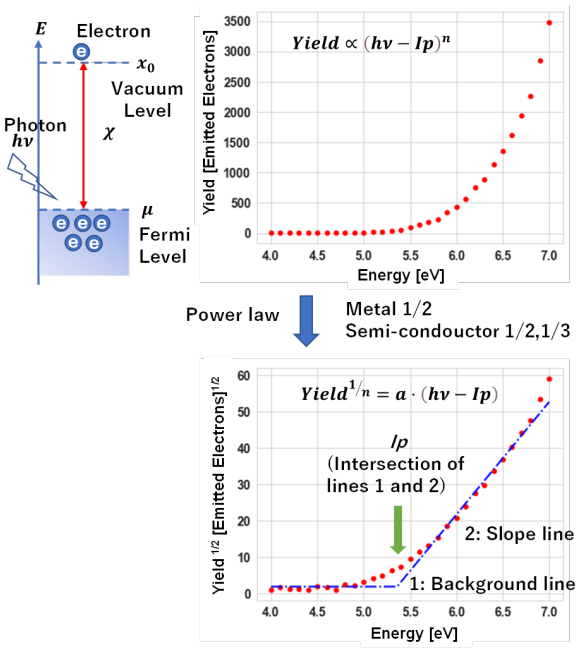

Fig. 1. Spectra obtained by PYS and their analysis (color online)

るという意思表示のクリエーティブコモンズ（CC）[10]の CC-BY または CC-BY-SA のライセンスを付与することとしている．

2. PYS 測定と解析値

PYS は，試料に照射する紫外光のエネルギーを走査して，放出される電子の数（1 光子当たりの放出電子数：Yield）を測定する．そして，電子が放出される閾値（Ip）の絶対値を求める．測定で得られるスペクトルとその解析法を Fig. 1 に示す．電子が放出される立ち上がり閾値を求める方法として，バックグラウンドと閾値以降の Yield のそれぞれを 1 次関数で近似してその交点を求める方法が用いられている．Yield は照射エネルギーに対してべき乗則に従って増加すると考えられている．閾値以降の Yield の直線化のために，Yield に対して金属ならば 1/2 乗，半導体ならば，1/1 乗，1/2 乗や 1/3 乗，有機物ならば 1/2 乗や 1/3 乗などが適用されている．AC 装置で測定したデータでは，基本的に 1/2 乗が適用され Ip が求められている．

3. メタデータ

3.1 計測・解析メタデータ

PYS 測定装置は数社から発売されているが，計測・解析データ構造を詳細に公開している理研計器のフォーマットにて計測・解析メタデータを構築している．理研計器のフォーマットの詳細および，AC の計測・解析ファイル（DAT）のメタデータ抽出ツールについては，すでに報告[5]しているが，利用者が興味ある情報は，Ip，光電子放出率，照射光量，解析に用いたべき乗数である．Ip の推定に利用したスペクトルデータの解析エネルギー範囲についての情報も解析記録としてメタデータに格納されている．データ提供においては，測定ファイルの DAT を添付することとした．

3.2 試料名メタデータ

同じ有機化合物でも名称は様々あり，名称で検索できないことがある．（例えば，アラニンでは，L-Alanine, L-2-Aminopropionic acid, L-alpha-Alanine など数種類の呼び名がある．）そのため目的の有機化合物を検索する際には一意の名称が必要である．PubChem などのデータベースの検索では，化合物の構造は一意であることから構造から計算された表記法で検索が可能である．有機分子の構造を記述したファイルとして，MOL 形式（拡張子.mol）や SDF（Structure Data File，拡張子.sdf）形式[6]がある．これらは分子内の原子の座標を記録したものであり，行列表記法とも呼ばれる．（SDF 形式は，1 ファイルに複数の化合物や，物性値などの MOL 形式の構造以外の情報も保存されている．SDF ファイルは，試薬会社のウェブサイトでは材料情報として添付されていることが多い．）この行列表記法で書かれた構造ファイルから，アルゴリズムによりいくつかの代表的な線形表記法（SMILES, InChI, InChIKey など）へ変換することができる．Tabele1 に有機 EL 材料のホール輸送材として用いられる NPD（N,N'-Bis- (1-naphthalenyl)-N,N'-bis-phenyl-(1,1'-biphenyl)-4,4'-diamine）について，3 つの線形表記へ変換した結果を示す．

Table 1. Results of conversion to three commonly used linear notations. Example: NPD (N,N'-bis-(1-naphthalenyl)-N,N'-bis-phenyl-(1,1'-biphenyl)-4,4'-diamine). (color online)

表記法	説明	例（NPD）
SMILES	化学構造を少ないバイト長で表現できる。作成ルールが簡便で人間でも読み書きが容易。SMILES表記にはいくつか種類がある。	C1=CC=C(C=C1)N(C2=CC=C(C=C2)C3=CC=C(C=C3)N(C4=CC=CC=C4)C5=CC=CC6=CC=CC=C65)C7=CC=CC8=CC=CC=C87
InChI	SMILES 記法よりも多くの情報を表現できる。SMILES ほど容易ではないが構造情報を人間でも読むことが可能。	InChI=1S/C44H32N2/c1-3-17-37(18-4-1)45(43-23-11-15-35-13-7-9-21-41(35)43)39-29-25-33(26-30-39)34-27-31-40(32-28-34)46(38-19-5-2-6-20-38)44-24-12-16-36-14-8-10-22-42(36)44/h1-32H
InChIKey	27 文字のハッシュ化された固定長の文字列で、人間には構造を読めない。構造をユニークに同定すると共にデータベースへの格納・検索性能が高い。	IBHBKWKFFTZAHE-UHFFFAOYSA-N

SMILES 記法は，化学構造を少ないバイト長で表現でき，作成ルールが簡便で人間でも読み書きが容易である．SMILES にはいくつか種類があるが，PubChem では Isomeric SMILES が利用されている．InChI 記法は，SMILES 記法よりも多くの情報を表現でき，構造情報は（SMILES ほどには容易ではないが）人間でも読むことが可能である．InChIKey（hashed InChI とも呼ばれる）記法は，27 文字のハッシュ化された固定長の文字列で，人間では解読できないが，構造をユニークに同定すると共にデータベースへの格納・検索性能が高い．InChI とは異なり，まれに異なる分子から同じ値が生成されることがある（ハッシュ衝突）．なお，線形表記法について詳しくは「新規化学物質申出における構造を表すコードの記載のあり方に関する調査報告書」[11]にまとめられている．これらを踏まえ，MOL ファイルによる構造化・可視化とアルゴリズムによる線形

Table 2. Sample, experimental, and copyright metadata with descriptions and examples of entries (color online)

Key	Value	記入例1（英語での記入を推奨）	記入例2	必須	キー（日本語）	属性	説明
dataLicense		CC-BY	CC-BY	○	ライセンス	ライセンス	データのライセンスCC-BYまたはCC-BY-SA
datasetTitle		Basic Organic EL Materials DB	metal natural oxidation process	○	データセット名	実験	データセット名または、実験名
dataProvider		Nims Taro	Nims Taro	○	データ提供者	実験	データ提供者(family given)
providerOrganization		NIMS	Rikenkeiki	○	データ提供者所属	実験	データ提供者所属
inputDate		2022/4/1	2022/9/1	○	記入年月日	実験	ファイル記入日(year/month/day)
aim		Basic Organic EL Materials	Al natural oxidation	○	計測の目的	実験	計測の目的や説明
sampleLavel		NPD	Al1	○	試料名	試料	任意に付けた試料名（試料略称と同じでもよい）
generalName		N,N'-Bis- (1-naphthalenyl)-N,N'-bis-phenyl-(1,1'-biphenyl)-4,4'-diamine	Al	○	試料一般名称	試料	論文等に記載するときの名前（著者が正式名称だと思っているもの）
sampleAbbreviation		NPD	Al	○	試料略称	試料	論文等で略称として使うときの名称
sampleDescription		hole transport material			試料の説明	試料	試料や計測上での説明
chemicalFormula			Al	△	化学式	試料	試料の化学式。molファイルが添付されているときは不要。
substrateName		ITO			基板名	試料	基板上に作成した場合の基板名。基板の効果の影響が考えられるときなど
sampleShape		Film	Bulk	○	試料形状	試料	（データ登録者の感性に一番合うもの）Film, Sheet, Powder, Pellet, Fiber, Block, Single crystal, Solution, Disk, Cylinder, Plate, Bar, Bulk,SAM
datFileName		NPD.dat	Al1.dat	○	datファイル名	試料	.datファイル名
molFileName		NPD.mol			molファイル名	試料	有機分子の場合、添付するmolファイルの名前
comment			Series from Al1 to Al5		コメント	試料	計測データに対するコメント（自由記述）
webReference		https://	https://doi.org/xxx		参考URL	実験	論文やWebなどのWebページアドレス、DOIなど
attachedReference		sample_set.pdf			添付文献ファイル名	実験	論文など添付するファイル名

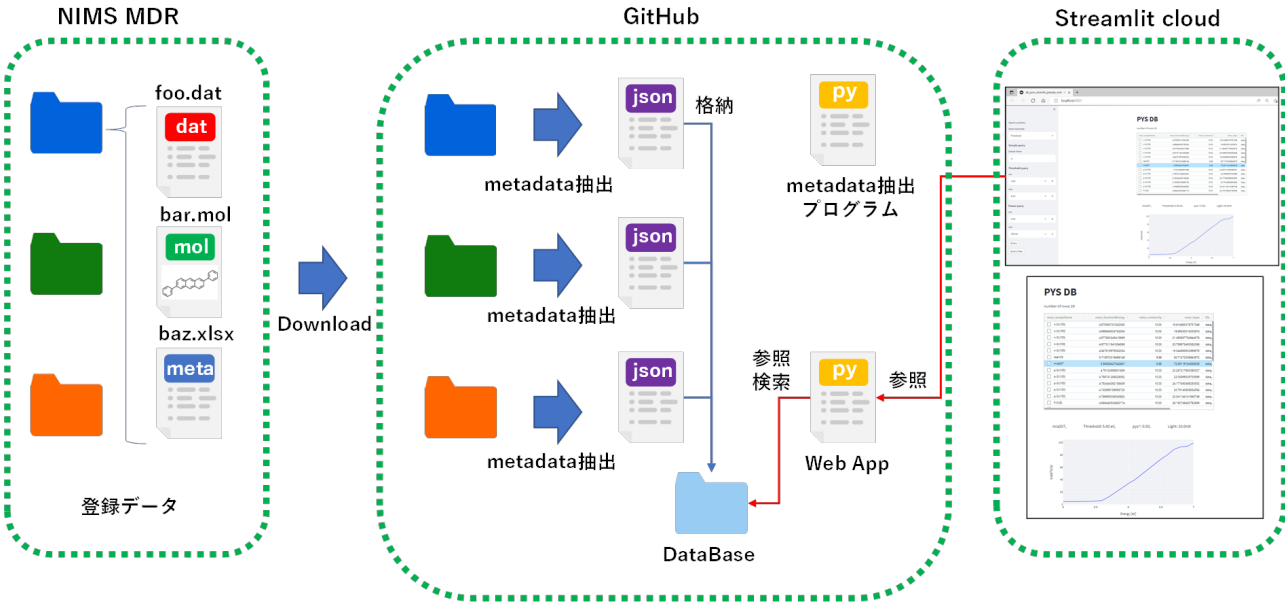

Fig. 2. Flow from data collection to database release (color online)

表記によって有機化合物を一意に特定することができる．有機材料を研究対象にしている研究者にヒアリングを行ったところ，発表資料を作成する際には，構造を可視化するために MOL ファイルを作成するため，有機化合物材料のデータ提供の際には，MOL ファイルを添付することとした． Python のモジュールの RDkit を用いて，MOL ファイルから，SMILES，InChl，InChlkey をそれぞれ計算し，先行している有機物のデータベースの PubChem でも複数の検索キーを提供しているためにそれに合わせた形で 3 種類をメタデータとしてデータベースに格納する．

3.3 試料・実験・ライセンスメタデータ

PYS は有機材料分野に限らず，様々な材料系でも利用されている．そこで，この装置を利用している研究者からヒアリングを行い，さらに，これまで NIMS 内のデータ整理で使われてきた分類法（主に

高分子データベース PoLyInfo[12]の分類法など）を組み合わせてメタデータの記載とした．Table 2 に試料・実験・ライセンスメタデータとその説明および記入例について示す．試料名については複数の記入欄（試料名，試料一般名称，試料略称）を設けた．これは今後，試料名の辞書を作成するためである．基板名についても自由記述欄を設けた．薄膜など基板の種類によって特性が変わる可能性があるためである．試料形状については PoLyInfo の試料形状の分類を参考に設けた．2 次利用が可能なデータベースを目指していることから，データ作成者がみずからのデータの 2 次利用を許可する意思表示のクリエーティブコモンズ（CC）の CC-BY または CC-BY-SA のライセンスの明示をお願いしている．

4. 収集からデータベースへの変換フロー

データ収集から公開までのフローを Fig.2 に示す．一つのデータセットは，計測に用いた DAT ファイルの他に，Tabel 2 に示した試料のメタデータ(EXCEL)，有機材料であれば MOL ファイルを一式のセットとしている．論文や発表などによって公開されたデータを物質・材料研究機構（NIMS）の公開材料データリポジトリ（MDR）[13]に登録する．登録されたデータのうち 2 次利用が可能なライセンスを持つデータについて，ダウンロードを行い GitHub リポジトリ[14]にデータを集約する．このデータを我々が開発したプログラムを用いて，メタデータの抽出を行う．メタデータは key と value で構成されており，その記述方式は Web アプリケーションなどで利用しやすい JSON 形式とし，JSON ファイルを集約しデータベースを作成する．メタデータの抽出ソースコードについても同じ GitHub リポジトリに集約・バージョン管理を行う[15]．（なお，抽出コードは BSD-3 ライセンスを適用する．データに関しては前述の通り CC-BY または，CC-BY-SA を適用する．これにより 2 次利用も可能である）また，外部から簡易に検索可能な Web アプリケーションも作成しそのコードも格納する．現状データ数が少なく，簡易な検索によるデモ運用の立場から，外部サービスの Stremlit Cloud[16]を利用する．これは，GitHub リポジトリにある Stremlit Web アプリケーションを参照して Web サービスを提供する仕組みである．（なお今後，様々な要望や仕様の変更により提供方法を変更することが考えられる．）

5. 収集データの例（有機 EL 材料とその分類）

著者らがこれまでに測定したデータから基本的な有機 EL 材料の一部を抽出したデータベースを作成した．登録データからその用途によって Ip を分類したグラフを Fig.3 に示す．それぞれの用途によるデータ数にばらつきがあるが，アクセプター材料では 5.8 から 6.5 eV，ホール輸送材料では 5.1 から 5.7 eV，発光材料では 5.5 から 5.7 eV，電子輸送材料と電子注入材料についてはデータが 1 点しかないが 6.5 eV と 5.6 eV，ホスト材料では 5.9 から 6.1 eV，ドナー材料では 5.1 から 5.6 eV に分布している．有機 EL 素子は，陽極から有機層の HOMO レベルにホールを注入し，陰極から LUMO（lowest unoccupied molecular orbital: 最低空軌道）レベルに電子を注入する．この注入された電子とホールがそれぞれの輸送層を通り発光層での結合により発光が起こる．ホール注入層では，陽極電極の仕事関数との差が小さい材料が望ましい．一般的に ITO 電極（およそ仕事関数は 5 eV 程度）が陽極として用いられることからその近辺に Ip を持つ材料が選ばれる．ホール輸送層では注入されたホールを発光層まで運ぶためにホール注入層より Ip が大きくそして発光層より小さくする必要がある．電子注入層および輸送層では，ホール注入・輸送と同様な考え方である[2]．電子注入レベルで重要なエネルギーレベルは，LUMO

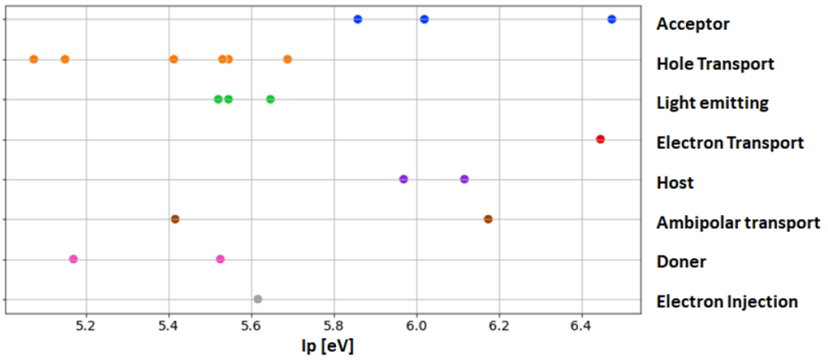

Fig. 3. Plot of Ip of basic organic EL materials classified by application. (color online)

であるためにこの測定データでは判断することができない．簡易的な見積もりとして PYS によって測定した Ip から光学バンドギャップを引いて LUMO を見積もることができる．光学バンドギャップは，電子的なバンドギャップよりもエキシトン束縛エネルギーのため 0.2 - 1.0 eV 程度狭いと考えられている[17]．一方で，逆光電子分光法によって直接電子注入準位の LUMO を測定することができる [18]．逆光電子分光法による LUMO と PYS と光学バンドギャップから求めた値では，0.1 から 1 eV 程度異なっているとの報告がある[17]．実際の開発では，Ip の他に，それぞれのキャリアの移動度，結晶性，それぞれの界面層での相互作用など様々なファクターを考慮に入れて材料選択が行われている．

6. まとめ

有機・無機半導体材料イオン化ポテンシャルについての情報を与える光電子収量分光(PYS)のデジタルデータベースの構築を行った．検索する際のキー（メタデータ）について，（1）計測機から得られる計測メタデータ，(2) データ解析メタデータ，(3) 測定試料や実験手順などを記載した試料・実験メタデータ，(4) データのライセンスに関するメタデータを整理した．（1）及び（2）のメタデータについては，理研計器が測定出力ファイル仕様の公開によりその情報を利用した．(3) については，この測定装置を利用している研究者やこれまで NIMS 内で構築されたデータベースの知見を利用し整理した．特に低分子有機材料については，材料名を一意に特定するために行列表記の構造ファイルから線形表記に変換した SMILES などをメタデータに含めた．これにより外部の有機物に関するデータベースと連携を取ることができる．(4) のライセンスについては，2 次利用が可能なライセンス表示をデータ提供者にお願いしている．検索可能なデータベースのデータの登録から公開までのフローについて構築した．データ登録は，NIMS のデータリポジトリである MDR を利用する．MDR には，計測ファイル（DAT），低分子有機材料であれば MOL ファイル，試料情報・プロセス・ライセンスなどが記載されている EXCEL ファイルを登録する．MDR に登録された 2 次利用可能なライセンスを持つデータをダウンロードし GitHub を利用して，集約・メタデータ抽出・Web アプリケーション化を行い，検索可能なデータベースの構築を行った．基本的な有機 EL 材料を登録し，その用途による Ip の分類を行ったデータ分析の結果を示した．

7. 文献

[1] 吉田亮: *統計数理* **69**, 35 (2021).
[2] 安達千波矢，小山田崇人，中島嘉之，*有機薄膜仕事関数データ集*，シーエムシー出版 (2004).
[3] PubChem https://pubchem.ncbi.nlm.nih.gov/
[4] 理研計器 M285-22010「AC-2S データ変換ソフト取扱説明書」
[5] S. Yagyu: *Journal of Surface Analysis* **29**, 97 (2022).
[6] MOL，SDF ファイルの説明 https://funatsu-lab.github.io/open-course-ware/molecular-design/file-format
[7] RDKit https://www.rdkit.org/docs/index.html
[8] SMILES https://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system
[9] InChl https://en.wikipedia.org/wiki/International_Chemical_Identifier
[10] クリエーティブコモンズライセンス https://creativecommons.jp/licenses/
CC-BY：原作者のクレジット（氏名，作品タイトルなど）を表示することを主な条件とし，改変はもちろん，営利目的での二次利用も許可する CC ライセンス．
CC-BY-SA：原作者のクレジット（氏名，作品タイトルなど）を表示し，改変した場合には元の作品と同じ CC ライセンス（このライセンス）で公開することを主な条件に，営利目的での 2 次利用も許可される CC ライセンス．
[11] 平成 29 年度化学物質安全対策（新規化学物質申出における構造を表すコードの記載のあり方に関する調査）報告書 https://www.meti.go.jp/meti_lib/report/H29FY/000629.pdf
[12] PoLyInfo　https://polymer.nims.go.jp/
[13] MDR（Materials Data Repository） https://mdr.nims.go.jp/
[14] GitHub https://github.co.jp/
[15] https://github.com/s-yagyu/PYS-DB
[16] Streamlit Community Cloud https://streamlit.io/cloud
[17] 吉田弘幸，*応用物理* **83**, 245 (2015).
[18] H. Yoshida, *Chem. Phys. Lett.* **539-540**, 180 (2012).

査読コメント，質疑応答

査読者 1. 眞田 則明（アルバック・ファイ）
[査読者]
表面分析のデータベースの構築について参考となる考え方が多く示されており，JSA 誌が掲載すべき技術報告と思います．なお，光電子収量分光（PYS）について詳しくない JSA 読者も多いと思います．初心者向けに多少の改訂のお願いと質問をお許しいただければと思います．

[査読者 1-1]
abstract および６．まとめで「無機」半導体と書かれておりますが，記事の内容には無機半導体は含まれていないように思われます．将来，無機材料への適用も考慮している，ということでしょうか？

[著者]
この測定手法が，有機 EL 材料などの有機半導体分野で多く使われてきたことから本論文の内容は有機半導体を中心に記載されています．しかし，無機半導体分野でも利用されており，メタデータについても無機半導体も考慮に入れて作成しておりますので，今後，無機半導体材料についてもデータベースにデータを乗せることを考えております．

[査読者 1-2]
１．はじめにで Ip という用語の説明がされておりますが，「Ip の情報」について（）内で定義されているようにも読めます．Ip が何かを明確にしていただければ幸いです．

[著者]
Ip の説明が不明瞭でしたので本文中に Ip について記載しました．金属では仕事関数，無機半導体ではイオン化ポテンシャル，有機半導体では HOMO レベルと呼び方が異なるために，統一的に Ip に統一しました．

[査読者 1-3]
Table1 で MOL 形式データの変換結果が示されていますが，実際に試料名メタデータとして使われているのは MOL 形式ということですので，変換前の MOL データも示していただけますでしょうか？

[著者]
MOL ファイルは少し大きい TSV 形式のファイルになりますので，論文中では記載いたしません．MOL 形式の詳細は，論文中のリファレンスにもあるこちらのサイト（[ファイル形式(MOL,SDF)] (https://funatsu-lab.github.io/open-course-ware/molecular-design/file-format)）を参照していただけると助かります．

[査読者 1-4]
データベースに実際にデータを登録する場合は MOL 形式で登録することになると思います．3.2 節の内容ですと，SDF ファイルを試薬会社から提供していただき，それに含まれる MOL 形式の構造データを取り出す，と読めますがそうなのでしょうか？初心者が誤解しないように MOL 形式の作り方をご教示いただけますと幸いです．

[著者]
低分子有機物の場合 MOL 形式の構造データを添付していただけると，検索にヒットしやすくなります．試薬メーカの試薬の説明の SDF ファイルの中に MOL 形式の 2 次元行列（TSV）が含まれています．（こちらも同様に先ほどのリファレンスに詳しく書かれています．）
SDF ファイルを Text などで開いて，最初の行から“M END”までをコピーし別の text にペーストして，拡張子を.mol にすると SDF ファイルから Mol ファイルを作成することができます．また，MOL ファイル作成ソフトは，フリーで利用することができるものが公開されています．
（https://www.acdlabs.com/products/chemsketch/）．
また，経済産業省から MOL ファイル作成の Web サイトが公開されており，そこで構造を描画してダウンロードすると，Mol ファイルを作成することができます．

MOL，SDF ファイルの説明
[ファイル形式(MOL,SDF)]
(https://funatsu-lab.github.io/open-course-ware/molecular-design/file-format)
MOL ファイル作成 Web サービス（経済産業省）
[少量新規化学物質の申出に必要な MOL ファイルの作成]
(https://www.nite.go.jp/chem/kasinn/syouryou/mol/)

[査読者 1-5]

今回構築されたデータベースを利用する場合は, GitHub や MDR になじみのない方が把握できる形で, 利用法についての現状をご教示いただけますでしょうか？

[著者]

論文が受理されたのちに DataBase の URL を公開設定にしますので, 公開後の URL を入力していただければ検索画面にたどり着くことができます.（DB 利用者（検索）は，Github，MDR を意識することはありません.）また，Github（このデータとプログラムが入っているリポジトリ）に README を付けますのでそれを参考に利用していただければと思います.

[査読者 1-6]

測定のエネルギー分解能は重要なファクターと思われますが，光源や装置の型式，チャージアップの有無などがデータベースから読み取れるものなのでしょうか？

[著者]

計測メタデータには，装置名など測定装置を特定するメタデータは記載されていますが，分解能などについては，装置メーカの装置マニュアルを参照していただく必要があると思います．なおチャージアップなどの情報については，測定者が気にかけていればコメントで記述している場合もありますが，データベースではそれらの情報で検索することはできません．同様の試料が集まってくれば，値のばらつきから判断することも可能になると思います.

査読者 2. 鈴木 峰晴（SA コンサルティング）

[査読者]

PYS を対象にしたデータ利活用の内容について,《データ収集・メタ情報を含めた可読的変換・データ利用例》一連を報告しており，JSA 誌への掲載を勧めます．読者の理解のために数点コメントさせていただきます.

[査読者 2-1]

他社装置も理研計器社 PYS の記述法にそろえるといった記述がされています．メタ情報の抽出を含めて，どのような変換ツールを用意されているのかを短くて良いので紹介していただくと良いと思います.

[著者]

各装置メーカが測定ファイルのフォーマットについてのドキュメントを公開してくだされればそれに合わせて変換ツールを Python で作成し Github リポジトリに公開します．（必要があれば Exe 化して配布することも可能です．）その際，すでに公開されている理研計器のフォーマットまたは，こちらの論文（S. Yagyu: *Journal of Surface Analysis.* **29** [2] (2022) 97-110.）または，サイト（https://github.com/s-yagyu/ACdataConverter）を参照してドキュメントを整理してくださると助かります.

[査読者 2-2]

「はじめに」第 2 段落中央当たりに「光電子放出率（Ip を推定する際の傾き）」とあります．PYS に不慣れな方にも理解できるように『何のどのような傾き』なのか説明を加えていただくとよいと思います.

[著者]

Yield は，単位フォトンあたりの光電子数となります.

放出光電子量は以下の式で表すことができ

$$Yield \propto (h\nu - Ip)^n$$

両辺を 1/n 乗をすると

$$Yield^{1/n} = a \cdot (h\nu - Ip)$$

となります.

この時の係数aが傾きに相当します.

本文中には，「放出光電子数の$1/n$乗の照射エネルギーに対する傾き」と記載いたしました.

[査読者 2-3]

3 章に DAT が紹介されていますが，ここには数値として Ip が書かれているのでしょうか，それともグラフから読み取るのでしょうか.

[著者]

DAT ファイルには，Ip の値は記載されていません．Ip を手動で求めた際のべき乗数，利用したエネルギー領域のデータが解析記録として保存されています．データベースでは，その解析記録をもとに Ip を求め，メタデータとして Ip を格納しています.

なお，DAT ファイルから計測メタデータおよび,

Ip を抽出するソフトはこちらに公開しています（https://github.com/s-yagyu/ACdataConverter）.

[査読者 2-4]

3.2 章の最後に「Python のモジュールの RDkit を用いて，MOL ファイルから，SMILES，InChl，InChlkey をそれぞれ計算し，メタデータとしてデータベースに格納する．」と述べられています．SMILES，InChl，InChlkey の 1 つに絞らず，3 種類として扱う理由を簡単に加えていただくと良いと思います.

[著者]

先行している有機物のデータベースの PubChem でも複数の検索キーを提供しているためにそれに合わせた形で 3 種類を採用しました.

本文中にも上記内容を記述しました.

[査読者 2-5]

Table 2 について．「Table 2 に試料・実験・権利メタデータとその説明・・・」と書かれています．各々どの行が試料・実験・権利メタデータに対応しているか示すことは可能でしょうか．また，例なので全てのメタ情報を列挙することの意味はないと思いますが，繰返し性・再現性の観点からは，装置名，測定日時は重要なパラメータです．是非この例の中に入れ込んでいただける良いと思います.

[著者]

Table2 については，実験者が記入するものなので最低限の記入を考えて作られています．Table2 試料，実験，ライセンスの属性の欄を設けました．最終的なデータベースのメタデータ（JSON ファイル）には，装置名，測定日時なども DAT ファイルに含まれている情報も統合されています.

[査読者 2-6]

Fig. 3 について．プロット図の右軸はデバイスの利用分野かと思います．そうすると，この図は，各分野で物性特性が優れているデータのみがプロットされているのでしょうか．言い換えると，例えば acceptor として検討していたが不適当だというデータが含まれていたら識別できるのでしょうか．MI としてどのようにデータを活用するかということにつながるかと思いますが.

[著者]

有機 EL や太陽電池分野で基本的と思われるデータをプロットしたものです．また，特性についても，試料の説明は必須ではありませんが，論文での分類を参考に我々がつけたものです．ご指摘の例については，説明がなければ識別でできません．恐らく利用法方法としては，設計する材料の Accepter に合致する Ip の値を検索して候補となる材料系を探すということになると思います.

[著者付記]

「データの権利」と記述していましたが，正確には，データの利用許諾（ライセンス）ですので，かかる場所について修正を行いました.

この論文がアクセプトされましたら，PYS-DB の確定した公開アドレスをリファレンスに記載いたします.