

○有馬 佳奈美¹、○岡本 瑠璃¹、小林 大樹²、吉沢 明康³、河野 信^{1,4}

1. 富山国際大学 現代社会学部、2. 新潟大学大学院 医歯学総合研究科、3. 京都大学大学院 薬学研究科、
4. 情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター(DBCLS)

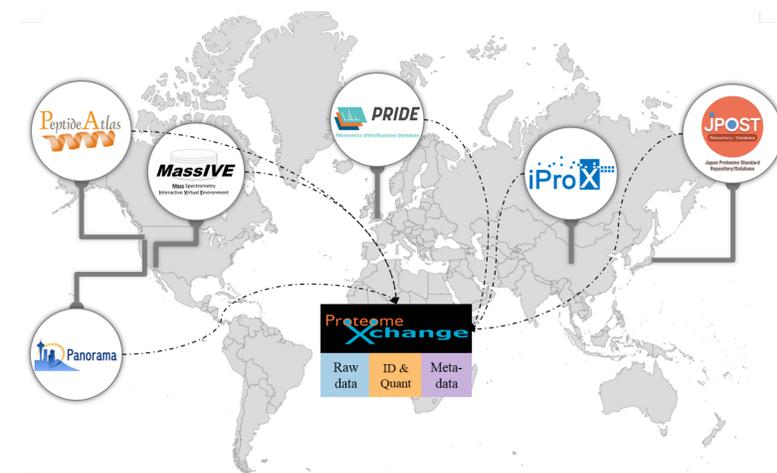
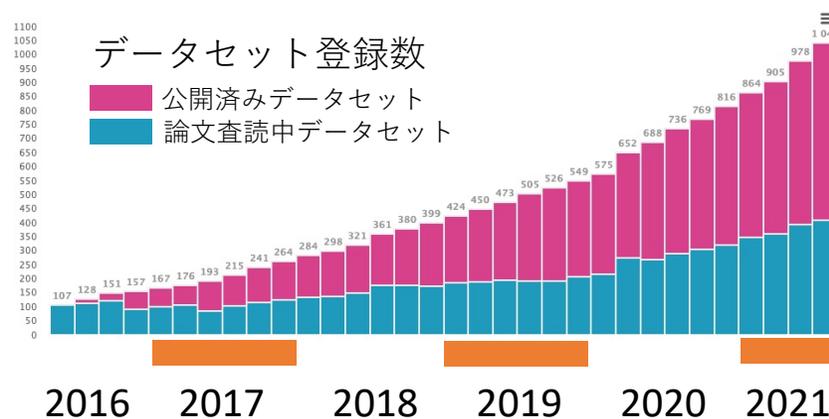


jPOST プロジェクト

jPOSTrepo¹⁾ProteomeXchange (PX)²⁾

プロテオーム実験データ
共有のためのリポジトリ

プロテオームデータ共有の世界連合



プロテオームデータの標準化: Proteomics Standards Initiative (PSI)

プロテオーム研究に関する国際的組織である Human Proteome Organization (HUPO) の PSI 分科会で標準を策定

- データの共有に必要なメタデータ項目

Minimum Information About a Proteomics Experiment (MIAPE)

Sample and Data Relationship Format (SDRF) 等

- ファイル/データフォーマット

mzML: スペクトルデータ

mzIdentML: ペプチド同定データ

Universal Spectrum Identifier (USI)³:

リポジトリ中のスペクトルを
識別するための識別子

mzSpecLib: スペクトラルライブラリの
データフォーマット

ProForma: タンパク質/ペプチドの
修飾を記述するデータ表記法

- 統制用語 (Controlled Vocabulary)

PSI-MS 等



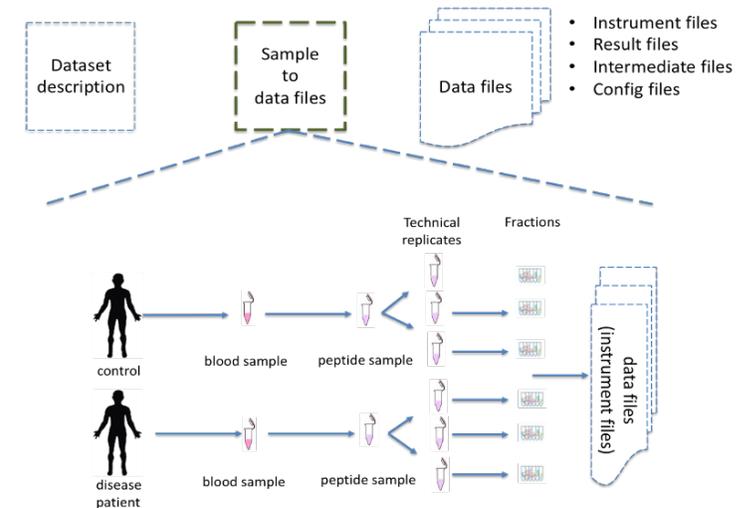
<https://www.psidev.info/>

HUPO-PSI Working groups and Outputs

Working Groups	Guidelines	v.	Formats	v.	Controlled Vocabularies	v.
Molecular Interactions	MIMix	1.1.2	PSI-MI XML	2.5.4	PSI-MI CV	2.5.0
	MIABE	1.0.0	PSI-MI XML	3.0.0		
Group charter	MIAPAR	1.0.0	MITAB	2.7, 2.8	PAR CV (deprecated)	n/a
Mass Spectrometry	Mass spectrometry (MIAPE-MS)	2.98	mzML	1.1.0	PSI-MS	4.0.15
			TraML	1.0.0		
Group charter	Identification (MIAPE-MSI)	1.1	mzData (deprecated)	1.05		
Proteomics Informatics	Mass spectrometry Quantification (MIAPE-Quant)	1.0	mzIdentML	1.2.0		
			mzQuantML	1.0.1		
Group charter	SDRF (Experimental Metadata) (in development)		mzTab	1.0.0		
			proBed	1.0.0		
Quality Control			proBAM	1.0.0		
			PEFF	1.0.0		
Group charter			USI	1.0.0		
			mzSpecLib (in development)			
Quality Control			ProForma (in development)			
			mzQC (PSI spec. under construction)			
Group charter						

SDRF (Sample and Data Relationship Format)⁴⁾ とは？

- リポジトリにはプロジェクト単位でデータが登録されるため、複数のサンプルのデータが、複数のファイルに分かれて登録されている。
- 機械可読かつ人間可読の形で、サンプルの情報と得られたデータファイルとを対応づけるためのメタデータフォーマット
- 遺伝子発現の分野で使われている MAGE-TAB-SDRF のプロテオーム版
- 多対多の情報を記述するので冗長になる



SDRF はサンプルとデータファイルの対応関係を記述する

sample properties					data file properties			study variables	
source name	characteristics[organism]	characteristics[disease]	characteristics[phenotype]	...	assay name	comment[fraction identifier]	comment[label]	comment[data file]	factor value[phenotype]
sample 1	homo sapiens	gastric carcinoma	control		Run 1	1	label free	fileRAW_Control_F1.raw	control
sample 2	homo sapiens	gastric carcinoma	primary tumor		Run 2	1	label free	fileRAW_Tumor_F1.raw	primary tumor
....									

SDRF の記述形式

サンプル情報、データファイル情報、実験のキーとなる情報の3項目が必要

<https://github.com/bigbio/proteomics-metadata-standard>



SDRF 化された jPOSTrepo に登録されているデータセット

PXD005173 (JPST000210) の SDRF 例 (一部省略)

Tomonaga *et al.*, Phosphoproteomics data of colon tissues (tumor and non-tumor).

jPOSTrepo に登録されている
SDRF 化された全11データセット

source name	characteristics [organism]	characteristics [organism part]	characteristics [disease]	assay name	comment[data file]	comment [fraction identifier]	comment[label]	comment [technical replicate]	comment[instrument]
CRC_N	Homo sapiens	Colorectum	Normal Colorectum	run 1	121026_CRC_N_Fr01.raw	1	label free sample	1	NT=LQ Orbitrap Velos;AC=MS:1001742
CRC_N	Homo sapiens	Colorectum	Normal Colorectum	run 2	121026_CRC_N_Fr02.raw	2	label free sample	1	NT=LQ Orbitrap Velos;AC=MS:1001742
CRC_N	Homo sapiens	Colorectum	Normal Colorectum	run 3	121026_CRC_N_Fr03.raw	3	label free sample	1	NT=LQ Orbitrap Velos;AC=MS:1001742
CRC_T	Homo sapiens	Colorectum	Colorectal cancer	run 30	121102_CRC_T_Fr01.raw	1	label free sample	1	NT=LQ Orbitrap Velos;AC=MS:1001742
CRC_T	Homo sapiens	Colorectum	Colorectal cancer	run 31	121102_CRC_T_Fr02.raw	2	label free sample	1	NT=LQ Orbitrap Velos;AC=MS:1001742
CRC_T	Homo sapiens	Colorectum	Colorectal cancer	run 32	121102_CRC_T_Fr03.raw	3	label free sample	1	NT=LQ Orbitrap Velos;AC=MS:1001742

comment[proteomics data acquisition method]	comment [cleavage agent details]	comment[modification parameters]	comment[modification parameters]
NT=Data-Dependent Acquisition;AC=NCIT:C161785	AC=MS:1001251;NT=Trypsin	NT=Carbamidomethyl;AC=UNIMOD:4;TA=C;MT=Fixed	NT=Oxidation;MT=Variable;TA=M;AC=UNIMOD:35
NT=Data-Dependent Acquisition;AC=NCIT:C161785	AC=MS:1001251;NT=Trypsin	NT=Carbamidomethyl;AC=UNIMOD:4;TA=C;MT=Fixed	NT=Oxidation;MT=Variable;TA=M;AC=UNIMOD:35
NT=Data-Dependent Acquisition;AC=NCIT:C161785	AC=MS:1001251;NT=Trypsin	NT=Carbamidomethyl;AC=UNIMOD:4;TA=C;MT=Fixed	NT=Oxidation;MT=Variable;TA=M;AC=UNIMOD:35
NT=Data-Dependent Acquisition;AC=NCIT:C161785	AC=MS:1001251;NT=Trypsin	NT=Carbamidomethyl;AC=UNIMOD:4;TA=C;MT=Fixed	NT=Oxidation;MT=Variable;TA=M;AC=UNIMOD:35
NT=Data-Dependent Acquisition;AC=NCIT:C161785	AC=MS:1001251;NT=Trypsin	NT=Carbamidomethyl;AC=UNIMOD:4;TA=C;MT=Fixed	NT=Oxidation;MT=Variable;TA=M;AC=UNIMOD:35

comment [fractionation method]	comment[fractionation method]	comment[file uri]	factor value [disease]
NT=IMAC;AC=PRIDE:0000587	NT=Strong cation-exchange chromatography (SCX);AC=PRIDE:0000561	https://repository.jpostdb.org/data/JPST000210.0/121026_CRC_N_Fr01.raw	Normal Colorectum
NT=IMAC;AC=PRIDE:0000587	NT=Strong cation-exchange chromatography (SCX);AC=PRIDE:0000561	https://repository.jpostdb.org/data/JPST000210.0/121026_CRC_N_Fr02.raw	Normal Colorectum
NT=IMAC;AC=PRIDE:0000587	NT=Strong cation-exchange chromatography (SCX);AC=PRIDE:0000561	https://repository.jpostdb.org/data/JPST000210.0/121026_CRC_N_Fr03.raw	Normal Colorectum
NT=IMAC;AC=PRIDE:0000587	NT=Strong cation-exchange chromatography (SCX);AC=PRIDE:0000561	https://repository.jpostdb.org/data/JPST000210.0/121102_CRC_T_Fr01.raw	Colorectal cancer
NT=IMAC;AC=PRIDE:0000587	NT=Strong cation-exchange chromatography (SCX);AC=PRIDE:0000561	https://repository.jpostdb.org/data/JPST000210.0/121102_CRC_T_Fr02.raw	Colorectal cancer
NT=IMAC;AC=PRIDE:0000587	NT=Strong cation-exchange chromatography (SCX);AC=PRIDE:0000561	https://repository.jpostdb.org/data/JPST000210.0/121102_CRC_T_Fr03.raw	Colorectal cancer

PXD004528
PXD004603
PXD004616
PXD005163
PXD005171
PXD005172
PXD005173
PXD005174
PXD005175
PXD005176
PXD005177



<https://github.com/bigbio/proteomics-metadata-standard/tree/master/annotated-projects>

SDRF化するにあたっての問題点：詳細なメタデータが無い！！



Journal of Proteome Data and Methods (JPDM)

プロテオームデータの詳細なメタデータを記述するためのデータジャーナル⁵⁾
詳細なメタデータ提供に対して論文出版のインセンティブを付与



<https://www.jhupo.org/jpdm/>

	A	B	C	D	E	F	G
1	Raw File	Species	Species_Ontology	Sample Type	Sample Type_Ontology	Cell line	Cell line_Ontology
2	Data110719_PC12_2050V-E15.wiff	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
3	Data110719_PC12_2050V-E15.wiff.scan	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
4	Data110719_BSA10fmol_2050V_5.wiff	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
5	Data110719_BSA10fmol_2050V_5.wiff.scan	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
6	Data110720_PC12_E14_2050V-5.wiff	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
7	Data110720_PC12_E14_2050V-5.wiff.scan	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
8	Data110720_PC12_E15_2050V-4.wiff	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
9	Data110720_PC12_E15_2050V-4.wiff.scan	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
10	Data110721_PC12_E11_2050V-4.wiff	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392
11	Data110721_PC12_E11_2050V-4.wiff.scan	Rattus norvegicus	10116	Cell Line	C16403	PC-12 cell	CLO:0008392

JPDM Data Description 投稿時に必要となる Supplementary ファイル (EXCEL)

参考文献

1. [jPOSTrepo: an international standard data repository for proteomes](#), Shujiro Okuda, *et al.*, *Nucleic Acids Res.*, **45**, D1107-D1111 (2017)
2. [The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics](#), Eric W Deutsch, *et al.*, *Nucleic Acids Res.*, **48**, D1145-D1152 (2020)
3. [Universal Spectrum Identifier for mass spectra](#), Eric W Deutsch, *et al.*, *Nat. Methods*, **18**, 768-770 (2021)
4. [A proteomics sample metadata representation for multiomics integration, and big data analysis](#), Chengxin Dai, *et al.*, *bioRxiv*, 10.1101/2021.05.21.445143 (2021)
5. [From Bench to Internet: Sharing Proteomics Data and Methods through the Open Access Journal](#), Yasushi Ishihama, *J. Proteome Data and Methods*, **1**, 1 (2019)

謝辞

研究データを共有してくださるすべての研究者、ならびに今回追加でメタデータを提供いただいた研究者の方々に感謝致します。また、SDRFの仕様策定ならびに実際にSDRF化を担当されている [メンバー](#) に感謝致します。本成果は、統合化推進プログラム「プロテオームデータベースの機能深化と連携基盤強化」(18063028)の支援のもと得られたものです。また、JPDMは科研費 研究公開促進費 国際発信強化B (21HP2004)の支援のもと運営されています。