Minimal AIF Instructions

Daniel W. Siderius[1,*]

[1] *Chemical Sciences Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA*

(Dated: January 20, 2022)

[*] Corresponding author, email: daniel.siderius@nist.gov

DISCLAIMER

Certain commercially available items are identified in this document. This identification does not imply recommendation by the National Institute of Standards and Technology, nor does it imply that it is the best available for the purposes described.

I. INTRODUCTION

The purpose of this document is to provide a guide for formatting equilibrium adsorption data (e.g., an adsorption isotherm) in the "adsorption information format" (AIF) for inclusion as supplementary information in a special issue of J Chem Eng Data. Please note that the specifications here are not the final form of the AIF. Standardization of the AIF is a current project under the supervision of IUPAC. The original design of the AIF is to facilitate comparison of isotherm data from various instruments, but the broader purpose of AIF is standardize presentation of adsorption isotherms to avoid ambiguity in the presentation of adsorption isotherm data.

This document assumes understanding of basic adsorption concepts and terminology; field-specific jargon is used without definition. If there are questions regarding the instructions that follow, please contact Daniel Siderius `mailto:daniel.siderius@nist.gov`.

II. HELPFUL REFERENCES

The AIF was introduced by J. D. Evans, et al., in 2021 [1] to create a "universal" format for encoding adsorption isotherms, hence facilitating comparison of isotherms as measured by different adsorption instruments. As a broader project, the AIF aims to create a universal standard for presenting adsorption isotherms to facilitate collection of isotherms, curation of databases, and application of machine-learning approaches to isotherm data. Current information about the AIF and its ongoing development is available at `https://adsorptioninformationformat.com/`.

Additionally, software related to AIF is stored in a GitHub repository at `https://github.com/AIF-development-team/adsorptioninformationformat`. This repository includes a utility that can convert isotherm files from Micromeritics, Quantachrome, and BEL instruments to AIF.

Lastly, development of AIF has been accepted as an IUPAC Project. Details are available at `https://iupac.org/project/2021-016-1-024`.

III. DESCRIPTION OF AIF

AIF is a key-value dictionary that is encoded in the STAR file format[2]. Keys in an AIF dictionary are given by human-readable strings that begin with an underline ("_") sentinel and cannot contain spaces. Values are associated with a key and may exist as string, float, integer, or DATE/DATETIME variable types.

The easiest way to understand the AIF is to examine a sample isotherm encoded in AIF. In Append A, a sample AIF file is shown as pure text. This particular AIF-formatted file contains an adsorption-desorption isotherm of Nitrogen in the DUT-6 metal-organic framework, at 77.3 K. It was originally measured in a Quantachrome Autosorb instrument; the original text output from the instrument is is the AIF GitHub repository at `test/database/DUT-6/NK_DUT-6_LP_N2_114PKT (Raw Analysis Data).txt`

As can be seen in Appendix A AIF is composed of up to four parts:

1. AIF Dictionary Name (see Sec. III A)
2. AIF Metadata Header (see Sec. III B)
3. Adsorption Loop (see Sec. III C)
4. Desorption Loop [optional] (see Sec. III C)

The dictionary name, metadata header, and adsorption loop are mandatory. The desorption loop is only included if a desorption isotherm branch is measured. Authors may store both adsorption and desorption data in the adsorption loop if there is no hysteresis in the isotherm, although we note that doing so incurs a 'loss of information' about the measurement sequence.

We reiterate that keys in the AIF will change in its subsequent development. The present tutorial reflects the state of the AIF as of January 2022.

A. AIF Dictionary Name

The first line of an AIF file is a name for the AIF dictionary. In Appendix A, the dictionary name is `data_raw2aif`. It should begin with "`data_`" and follow with a user-selected string. Examples: `data_N2_CuBTC_77K`, `data_TrCO2_ZSM5_GCMC`, etc. The dictionary name is not used in further processing the AIF, so there is great freedom in the name choice.

B. AIF Metadata header

The second section of an AIF file is the metadata header, which includes required metadata key-value pairs and may include optional metadata key-value pairs. The mandatory metadata keys are given in Table A and are essential metadata for identifying an isotherm: identifiers for the adsorptive (adsorbate): `_exptl_adsorptive`, the adsorbent material: `_sample_id`, the isotherm temperature: `_exptl_temperature`, and units of the temperature, pressure, and loading. The table also gives data specifications for those keys along with example values. The key-value pairs in Table A **must be included using the key names exactly as shown**. Without this essential information, an isotherm is incomplete.

When possible, it is preferable to use existing systematic descriptors for identifying units, adsorptives, and (if possible) materials. An adsorptive can usually be identified by its InChIKey (e.g., Nitrogen is `IJGRMHOSHXDMSA-UHFFFAOYSA-N`, `https://webbook.nist.gov/cgi/cbook.cgi?Name=nitrogen&Units=SI`). Units may be identified using the QUDT framework, `https://qudt.org/`. (E.g., mmol/g or "millimoles adsorptive per gram adsorbent" may be represented in QUDT as `MilliMOL-PER-GM`; kPa or "Kilopascals" would be represented as `KiloPa`.) An (incomplete) list of QUDT units for adsorption is available in the included `QUDT_Units.xlsx` spreadsheet. Using systems like InChIKey or QUDT is not mandatory for isotherms included in the special issue of JCED.

The metadata header of an AIF file may contain other information deemed relevant by the author. The example AIF in Appendix refappendix:A includes the `_exptl_operator`, `_exptl_date`, `_instrument`, `_exptl_sample_mass`, and `_sample_id` keys. It is up to the author to decide which optional keys to include in an AIF file; the intention behind these keys is to provide useful or relevant metadata about the isotherm that aids interpretation and reproducibility. For example, an AIF file containing and adsorption isotherm from

molecular simulation could include keys that describe the simulation method or force field, e.g.:

```
_simulation_method 'Grand Canonical Monte Carlo'
_force_field_adsorptive 'TraPPE UA'
_force_field_adsorbent 'UFF with QEq point charges'
_force_field_mixingrule 'Lorentz-Berthelot'
```

An AIF file for experimental measurements could include information regarding adsorbent activation, equilibration time, measurement type, the physical properties set (e.g., a particular equation of state), etc.:

```
_activation 'Vacuum at 300K for 2 hours'
_measurement_type 'Manometric'
_equation_of_state 'from CoolProp'
```

Since the author has freedom in selecting the optional metadata keys, we only require that 1) the keys follow the STAR naming scheme (leading "_", no spaces) and 2) be interpretable based on information supplied in the associated manuscript or supplementary information. It is also important to consider whether including an optional key may in turn require and additional metadata key; for example, including `_exptl_sample_mass` (the mass of the adsorbent sample) would in turn require a key that clarifies the units of that quantity (e.g., `_units_mass`).

C. Adsorption and Desorption Loops

The final section(s) of an AIF file are the actual pressure-adsorption data, represented in a group of columns. In STAR terminology, these are termed 'loops.' An AIF file may contain one or two loops, one for **ad**sorption measurements and one for **de**sorption measurements. The adsorption loop is mandatory, while the desorption loop should be included if a desorption isotherm is measured. The inclusion of both adsorption and desorption loops does not automatically mean that the isotherm exhibits adsorption-desorption hysteresis; it

only means that measurements were done for adsorption (increasing pressure) and desorption (decreasing pressure) directions. However, if hysteresis is present, an author should use the adsorption and desorption loops to prevent confusion.

Each loop for AIF files submitted to JCED must contain at least three columns of data: measurements of pressure, adsorption, and uncertainty in the adsorption measurement. Additional columns could contain the saturation pressure of the adsorptive at the isotherm temperature or uncertainty in the pressure measurement. The saturation pressure of the adsorptive, p_0, should be included if the measurement is at subcritical conditions.

Each loop is preceded by a text block that identifies the loop columns, e.g.,

```
loop_
_adsorp_pressure
_adsorp_amount
_adsorp_amount_uncertainty
```

This block indicates that the loop contains three columns (pressure, adsorbed amount, and uncertainty in the adsorbed amount) and that the data is an adsorption loop (`_adsorb` prefix on the column names). The `loop_` line is absolutely necessary, and then the following lines must use the exact column names shown. If an author includes uncertainty in pressure, that would be indicated by the `_adsorp_pressure_uncertainty` line; similarly, saturation pressure would be identified by `_adsorp_p0`. The ordering of the column names following `loop_` must match the ordering of the data columns (see below). The units of the measurements given in the data loops are given in the metadata header. The desorption data loop would be preceded by

```
loop_
_desorp_pressure
_desorp_amount
_desorp_amount_uncertainty
```

where the column names have the `_desorp` prefix.

After the text block that identifies the contents of the adsorption or desorption data loop, the data are given in space-separated columns in the same order as the preceding text block.

Tabs should not be used to delimit the columns, as standard packages for interpreting STAR formatted files may fail to parse the columns.

It is easiest to think of these data in a table or spreadsheet, e.g., the following table:

Pressure (bar)	p_0 **(bar)**	**Amount Adsorbed (mmol/g)**	**Uncertainty (mmol/g)**
0.0001	0.0608	0.3400	0.0074
0.0049	0.0608	0.6022	0.0037
0.0108	0.0608	0.8470	0.0062
0.0181	0.0608	1.1079	0.0074
0.0203	0.0608	1.1820	0.0037

This table of isotherm data would be represented in an AIF file as

```
loop_
_adsorp_pressure
_adsorp_p0
_adsorp_amount
_adsorp_amount_uncertainty
0.0001  0.0608  0.3400  0.0074
0.0049  0.0608  0.6022  0.0037
0.0108  0.0608  0.8470  0.0062
0.0181  0.0608  1.1079  0.0074
0.0203  0.0608  1.1820  0.0037
```

Note regarding uncertainty: the meaning the uncertainty given in the adsorption and/or desorption loops should be clarified in the associated manuscript.

This completes the description of the contents of an AIF file. The most important items to keep in mind while composing an AIF file are 1) ensuring that the metadata header contains all required information and appropriate optional information and 2) that the loop columns match the order specified in the lines following the "`loop_`" sentinel.

IV. EXAMPLES

Several example AIF files are included with this guide:

`NK_DUT-6_LP_N2_114PKT.aif`: The AIF file included in this guide in Appendix A.

`CH4_RM8850_Exp.aif`: Isotherm of methane adsorbed in Zeolite Y measured using a manometric instrument. The isotherm contains only an adsorption branch of the isotherm.

`CO2_ZIF8_GCMC.aif`: Isotherm of CO2 adsorbed in ZIF-8 as measured using Grand Canonical Transition Matrix Monte Carlo simulation. It includes metadata that identify the simulation force fields, the simulation code, and the simulation cell size of the material.

`Xe_Vycor_Exp.aif`: Isotherm of Xenon adsorbed in Vycor Glass measured manometrically. This example contains both adsorption and desorption branches, the saturation pressure of the adsorptive, and information about the sample mass.

ACKNOWLEDGMENTS

Jack Evans
AIF Development Team

Appendix A: Example AIF File

```
data_raw2aif
_exptl_operator 'Nicole'
_exptl_date 2009-09-17T19:56:00
_exptl_instrument 'Autosorb Station 1'
_exptl_adsorptive Nitrogen
_exptl_temperature 77.3
_exptl_sample_mass 0.0339
_sample_id 'nk_DUT-6_LP_N2_114pkt'
_sample_material_id 'DUT-6'
_units_temperature K
_units_pressure Pa
_units_mass g
_units_loading mmol/g

loop_
_adsorp_pressure
_adsorp_p0
_adsorp_amount
0.269367243408 101860.98004799998 0.006484305926579284
0.27275763204 101860.98004799998 0.008217577878132952
0.273485572344 101860.98004799998 0.009928476539499517
0.27664664644799997 101860.98004799998 0.011659116339266468
0.279902379456 101860.98004799998 0.028835222823493427
0.293934562056 101860.98004799998 0.035156335339296206
0.315123491088 101860.98004799998 0.042143382257138595
0.31645404864 101860.98004799998 0.04724580849569103
0.32231623456799996 101860.98004799998 0.04894223032223065
0.32596660188 101860.98004799998 0.05063996822466363
0.332183425392 101860.98004799998 0.0523377061270966
0.333315332568 101860.98004799998 0.05403544402952958
```

0.33809094093599995 101860.98004799998 0.05573449800785591

0.34380913867199997 101860.98004799998 0.0574256555308221

0.34757283002399997 101860.98004799998 0.05912076128146835

0.35413362532799997 101860.98004799998 0.06081849918390134

0.42206005490399995 101860.98004799998 0.07288296689832234

0.525048942456 101860.98004799998 0.08901674127500907

0.6506679741839999 101860.98004799998 0.11353786732007201

0.814630528152 101860.98004799998 0.14562379760016192

1.026738467208 101860.98004799998 0.20794788760606583

1.2882250235519999 101860.98004799998 0.25209565344879004

1.61685407376 101860.98004799998 0.3202354828274469

2.0322066787199997 101860.98004799998 0.40431430735057333

2.5645496896799997 101860.98004799998 0.5130261243696983

3.22761531384 101860.98004799998 0.6335049759513451

4.056853977359999 101860.98004799998 0.7753766411795409

5.12407312488 101860.98004799998 0.9483366511606605

7.055114766479999 101860.98004799998 1.2170477627631064

8.65884991608 101860.98004799998 1.425415478579154

9.72756227448 101860.98004799998 1.5473235885809884

12.417968309759999 101860.98004799998 1.8373340724415548

15.7255104024 101860.98004799998 2.1391629178244846

19.8822361896 101860.98004799998 2.4524547842385696

25.122339799199995 101860.98004799998 2.7734851769056057

31.878719063999995 101860.98004799998 3.107505238640093

40.661332163999994 101860.98004799998 3.45725240730022

51.04341409679999 101860.98004799998 3.790667074123763

64.74682365839999 101860.98004799998 4.15361448399428

80.614055772 101860.98004799998 4.510981732076955

101.76498791999998 101860.98004799998 5.011103732312269

128.16855598319998 101860.98004799998 5.497301649595846

162.136037088 101860.98004799998 6.105631409783129

204.24725035199998 101860.98004799998 6.815088441616109

255.949677072 101860.98004799998 7.705084764500063

322.14824834399997 101860.98004799998 8.81774796778141

407.03195397599995 101860.98004799998 10.261062078510244

510.78077920799996 101860.98004799998 12.137312515118422

646.809623928 101860.98004799998 14.732548373027498

813.6106117919999 101860.98004799998 18.26581732392813

820.564708176 101860.98004799998 18.43414343068873

833.5369776959999 101860.98004799998 18.730918544641163

1052.5496838479999 101860.98004799998 24.060499482387346

1490.4644385599997 101860.98004799998 29.481547694721982

2065.17730824 101860.98004799998 32.44706150522758

2959.1306647199995 101860.98004799998 35.02946562317641

4053.44092392 101860.98004799998 37.30311833654324

5227.784619839999 101860.98004799998 39.40897137350681

6461.23013568 101860.98004799998 41.57312657254618

7561.806547679999 101860.98004799998 43.959961812741874

7915.0175819999995 101860.98004799998 45.06085929753659

10135.79546328 101860.98004799998 51.90774413273625

12857.092298639998 101860.98004799998 53.39596275294649

15287.279673599998 101860.98004799998 54.01715057461189

18043.986938399998 101860.98004799998 54.52383979355507

20803.627296 101860.98004799998 54.91905738433074

25597.100865599998 101860.98004799998 55.44259237470892

30844.6705296 101860.98004799998 55.88018760925075

36096.6398328 101860.98004799998 56.224209847774766

41302.74623039999 101860.98004799998 56.50163864609482

46463.656334399995 101860.98004799998 56.733926041272674

51603.6348216 101860.98004799998 56.93239028599118

56737.48047839999 101860.98004799998 57.10545426596788

61848.661327199996 101860.98004799998 57.254170841917436

66975.174252 101860.98004799998 57.38512039330665

72095.421024 101860.98004799998 57.50225114781559

77192.2030536 101860.98004799998 57.60872168758832

82298.31765119999 101860.98004799998 57.698478063515395

87379.63428239999 101860.98004799998 57.782443705511696

92476.94960159999 101860.98004799998 57.85719681625448

97575.86478959999 101860.98004799998 57.926948838602506

100284.8426352 101860.98004799998 57.96643111540327

loop_

_desorp_pressure

_desorp_p0

_desorp_amount

97113.50270639999 101860.98004799998 57.919842028778376

90734.559156 101860.98004799998 57.83008565285129

85586.58132479999 101860.98004799998 57.74993663094573

80459.1351432 101860.98004799998 57.66294401439472

75389.2842384 101860.98004799998 57.56660725900084

70279.56993599999 101860.98004799998 57.45592527636935

65213.71870319999 101860.98004799998 57.33431986382299

60101.604597599995 101860.98004799998 57.19784279368168

55018.2881304 101860.98004799998 57.04188780031864

49960.569564 101860.98004799998 56.86316469400051

44881.1194464 101860.98004799998 56.656672386332495

39824.3341368 101860.98004799998 56.41675175097317

34758.8828712 101860.98004799998 56.13076845934628

29716.629703199997 101860.98004799998 55.780955486891486

24722.639243999998 101860.98004799998 55.34164935368829

19778.244717599995 101860.98004799998 54.7678402641838

17352.576972 101860.98004799998 54.391837381451175

14968.905782399997 101860.98004799998 53.93055278082887

12692.892430799999 101860.98004799998 53.313313186843565

10183.5648792 101860.98004799998 51.94143567560625

7976.852511119999 101860.98004799998 45.48529377314484

5352.00109992 101860.98004799998 39.73167318255841

2877.2707111199998 101860.98004799998 34.896936780715166

1413.8307230399998 101860.98004799998 29.463122632214958

---

[1] J. D. Evans, V. Bon, I. Senkovska, and S. Kaskel, Langmuir **37**, 4222 (2021), URL https://doi.org/10.1021/acs.langmuir.1c00122.

[2] S. R. Hall, Journal of Chemical Information and Computer Sciences **31**, 326 (1991), URL https://doi.org/10.1021/ci00002a020.

TABLES

Key	**Description**	**Data Type**	**Example**
_exptl_adsorptive	Adsorptive Identifier	String	IJGRMHOSHXDMSA-UHFFFAOYSA-N
_exptl_temperature	Temperature of Isotherm	Float	293.2
_sample_material_id	Adsorbent Identifier	String	DUT-6
_units_temperature	Units of Temperature	String	K
_units_pressure	Units of Pressure	String	Bar
_units_loading	Units of Adsorbate Loading	String	MilliMOL_PER_GM

TABLE I. Mandatory Metadata in AIF Header.

Key	**Description**	**Data Type**	**Example**
_exptl_operator	Name of Operator	String	'Daniel Siderius'
_exptl_date	Date of Measurement	See note	2009-09-17T19:56:00
_exptl_instrument	Model of Instrument	String	'Self-built Seiverts Apparatus'
_exptl_sample_mass	Mass of Adsorbent Sample	Float	0.0339
_sample_id	Sample Identifier	String	'nk_DUT-6_LP_N2_114pkt'

TABLE II. Optional Metadata in AIF Header.

Note: The date (and time) of the measurement is best represented by a DATETIME data type, but a simpler type such as the date by itself (e.g., 2022-01-19) is sufficient. We do strongly recommend that dates use the YYYY-MM-DD format as this simplifies sorting and has wider international usage than alternatives.