

NIMS における分析データの”Scheme on Read“型収集の試み

Trial of Data Collection of Instrument Analysis in “Scheme-on-Read” Type

物材機構¹ ◦鈴木 峰晴¹, 長尾 浩子¹, 吉川 英樹¹

NIMS-MaDIS-DPFC¹ ◦Mineharu Suzuki¹, Hiroko Nagao¹, and Hideki Yoshikawa¹

E-mail: SUZUKI.Mineharu@nims.go.jp

新規材料の開拓・探索を短期間で進めることを目的の一つに掲げている材料データ科学は、できるだけ多くのデータを先入観なく、また分野を横断して収集し、最新の数理統計学的手法により解析しようとするものであり、計測インフォマティクスはその一部をなしている。そこでは、デジタル化され、かつ第三者可読性の高い計測データ群が必要とされる。物質・材料研究機構 (NIMS) 材料データプラットフォームセンター (DPFC) では、実用に向けた材料研究データベースの構築に着手し、我々は分析データの可読化・収集を開始している。

分析装置からは通常、バイナリー形式で書かれた生データが生成される。異なる機種が生成するデータを統計的に扱ったり相互比較したりするには、データ形式の統一が必要とされる。そこで我々は生データ記述側にデータ形式に関する要求を与えることのない”Scheme on Read”方式の立場で、できる限り拘束条件を排したデータ形式を表面分析に限らず広範囲の分析 (計測) データの表現に活用できるよう検討を開始したので、状況を報告する。

変換後のデータ構造の概念図を Fig. 1 に示すが、上位から効率的な検索を可能にする「計測メタ情報群 (Characterization metadata)」, その下位にはパラメータ・セット (Parameter set) として、再現測定・繰り返し測定を可能にすると共に第三者可読性を高める「主要パラメータ群 (Primary parameters)」, 装置出力データから抽出した「測定条件パラメータ群 (Raw parameters)」を置き、最下層に「条件付き数値データ (Formatted numerical data)」を配した。

「測定条件パラメータ群」は装置に依存した記述であり、装置を使用している熟練担当者でも理解しづらい用語で書かれているが、その中から抜き出した主要な条件である「主要パラメータ群」は、ISO 等で定義されている共通用語にできるだけ置き換えている。「条件付き数値データ」のヘッダーには変数と測定量に関する最低限の情報を与えている。

たとえ数値部分の記述形式が異なっても、ヘッダー部分に解読情報が書かれていることで、第三者可読性のあるデータとして取り扱うことができる。また、このデータセットを解読ソフトウェアとリンクさせて保管することで、計算機やプログラム言語の変化に耐性を持たせ、データの利活用寿命を延ばすことができる。

講演では、この可読化プロセスを実際にいくつかの分析に適用した例を紹介する。

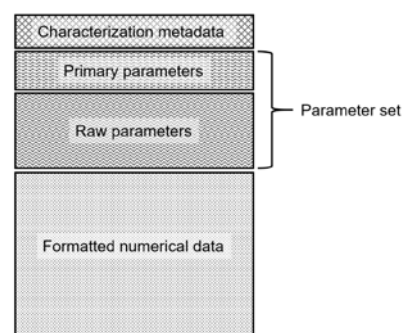


Fig. 1 Structure of converted characterization data package.