J. Plasma Fusion Res. Vol.95, No.5 (2019) 206-207

小特集

プラズマ実験におけるデータ解析インフラストラクチャ

Data Analysis Infrastructures in Plasma Experiments

1．はじめに

1. Introduction

藤井恵介
FUJII Keisuke
京都大学工学研究科

（原稿受付：2019年2月26日）

規模の大きなプラズマ実験では，複数の研究者により多種多様な計測が実施される．得られたデータの解析方法も様々に開発されており，ある解析データを用いた別の解析など，データの利用方法が階層構造をなすこともしばしばである．このような解析を円滑に進めるためには，計測データ・解析データをデータベースに保存する仕組みはもちろん，各研究者が全てのデータにアクセスできること，基本的な解析データが自動的に生成されること，上流のデータの変更が追跡できることなどが重要である．このような多くの研究者が日常的に使うデータ解析に関わる仕組みをデータ解析インフラストラクチャと本小特集では呼ぶことにする．実際，各実験機関では上記機能を実装したインフラストラクチャを様々な工夫のもとに構築している．また国際熱核融合実験炉ITERでは国際協力実験となるため，その構築に多大な労力が割かれている．本小特集では，各機関でのデータ解析インフラストラクチャについて解説する．また研究室規模でも，多様な計測装置を組み込んだプラズマ実験を行うためには，小さな労力でいかに効率よくデータを生成・管理するかも重要となる．既存のソフトウェアを用いることで，これらのデータ生成・管理を簡単・効率的に行う方法についても述べる．

Keywords:
Data management, database, automatic analysis

プラズマ実験装置は一般に，様々な制御装置・計測装置からなる統合システムである．複数の研究者が多種多様な計測を行い，それらの結果を統合して新たな知見を生み出すことが求められる．各々が個人で行う実験であれば，制御データや計測データは，各人がわかるようコンピュータに保存したり，実験ノートに記載するだけで良かった．しかし実験が大規模になり，単独の研究者が担当できる範囲が小さくなるに従って，それぞれが提供するデータを相互に利用する必要が出てくる．そのためには，各々が提供するデータを共有し，皆が簡単にアクセスできる仕組みの構築が重要である．さらに近年では，生成されるデータ量が膨大になってきたこと，解析アルゴリズムの発展やコンピュータの処理速度の向上により，様々な解析方法も新たに確立されてきたことから，系統的なデータ管理が必要になってきた．ある解析で生成されたデータを別の解析で用いるというような階層的なデータのフローが生じることもしばしばである．

例えば環状プラズマ実験においては，平衡磁気面の座標は他の解析の基礎となる基本的な量である．磁気センサや電流センサからの計測データから再構成した磁気面位置の推定結果はそれ以降の解析で頻繁に用いられるため，皆に共有されることが必要となる．特にスムースに解析を進めるためには，各々が担当者に依頼せずとも，自動的に解析が行われてデータが共有されることが望ましい．一方で，磁気面位置を求めるための計測データが何らかの理由（例えばセンサの較正係数の修正など）で更新された場合，その結果は下流の解析にも伝播する必要がある．

また研究の捏造防止やその管理体制の確立のため，データ・解析の履歴管理を強化することも社会的に求められるようになってきた．そのため，生データの他，解析プログラムを追跡できることが必要である．

このように，データ解析インフラストラクチャには，

Graduate School of Engineering, Kyoto University, KYOTO 606-8501, Japan
author's e-mail:fujii@me.kyoto-u.ac.jp

ⓒ2019 The Japan Society of Plasma Science and Nuclear Fusion Research

(1)データの保管，(2)データへのアクセス確保のほか，(3)解析プログラムの実行・管理，(4)データ更新の対応など多岐に渡る．この問題自体は新しいものではないが，様々な解析方法が確立されてデータフローが階層性を増してくるに従って，(3)，(4)に系統的に対応することの重要性が認識され始めてきた．しかし，現在でも全ての要望に答える確立された方法論はなく，各機関でそれぞれの問題意識に応じた仕組みが検討され構築されている．このような活動はあまり公開されないため，どのようにデータが管理・運用されているかは外部からはよくわからないかもしれない．本解説では，各実験機関で構築されているデータ解析のためのインフラストラクチャの特徴について述べていただくことにした．

第2章では大型ヘリカル装置（Large Helical Device, LHD）を有する核融合科学研究所のAutoAnaシステムについて述べていただく．核融合科学研究所ではデータ解析の自動化による研究の加速・高度化が重視されており，担当者に依頼することなく広範な解析データを取得できる環境の整備に力が入れられている．そのため，「研究者が解析を行いその結果をサーバに登録するのではなく，解析プログラムそのものを登録する」というコンセプトで開発されたシステム＝自動解析システム（AutoAna）の構築が行われ，実際に運用されている．

第3章では，稼働が迫っている量子科学研究開発機構の大型トカマク装置JT-60SAにおけるデータ解析システムについて述べていただく．JT-60SAでは共用のデータ解析ソフトeDASを整備しており，それを用いた解析ルーチンを想定している．eDASを通すことで，様々なデータベースに格納された運転データ・計測データ・解析データのほか，シミュレーションデータにもそのフォーマットや保存されているデータベースの違いを意識することなくアクセスできるよう設計されている．

第4章では，大阪大学レーザー科学研究所における実験データベースSEDNAについて述べていただく．大型レーザー装置では，様々な目的に応じて個別に製作されるターゲットを用いた実験が行われる．各ターゲットの性質，レーザーの実験条件のほか，多種多様な計測データやその解析結果，シミュレーションを一元的に管理するシステムとして構築されている．また，将来的には天文学分野のように，データアーカイブスとして公開することも視野に入れて設計されていることが特徴である．

第5章では国際熱核融合実験炉ITERで開発が進められているデータ解析システムIMASについて述べていただく．各国の協同実験であるITERではインフラストラクチャはより重要となる．計算コード間のインターフェースや実験データを区別なく記述することをめざし，IMASデータモデルというファイル形式が定められている．それにより，例えば複数のコードを接続したシミュレーションの実行も可能になっている．また，ユーザ支援のため，実験データの検索システムや，プラズマプロファイルの可視化システムなども開発されていることが特徴である．

一方で，少し大きな実験装置を少数の研究室で運転する場合，ITERで採用されているようなシステムでは維持・運用にコストがかかりすぎる．データ管理に多大な労力を割くことができないため，既存の枠組みを使って効率よくインフラストラクチャを構築することが求められる．第6章では，核融合実験で広く使われているデータ管理ソフトウェアMDSplusというによる実験データ利用・管理について述べていただく．

このように，データ解析インフラストラクチャといっても優先度の違いにより様々な構造のものが開発・運用されている．おそらく読者のほとんどは大規模なインフラストラクチャを立ち上げるというよりは，大規模実験に携わる上でユーザとして利用するか，研究室規模での実験に際してインフラストラクチャを立ち上げるかであろう．ユーザとして利用する場合は，当該機関の記事を一読されると今後の共同研究をスムースに進める上で参考になるであろう．一方でインフラストラクチャを立ち上げる場合は第6章の記事に加え，データ管理のどの部分を重要視するかにより他の章も参考にされるとよいと思う．

J. Plasma Fusion Res. Vol.95, No.5 (2019) 208-212

小特集 プラズマ実験におけるデータ解析インフラストラクチャ

2．LHD における自動データ収集・解析インフラストラクチャ －AutoAna システム－

2. Automatic Data Compilation / Analysis Infrastructure in LHD - AutoAna -

鈴木千尋，江本雅彦，居田克巳
SUZUKI Chihiro, EMOTO Masahiko and IDA Katsumi
核融合科学研究所

（原稿受付：2018年12月20日）

核融合科学研究所の大型ヘリカル装置（LHD）における自動データ収集・解析システム「AutoAna」について解説する．核融合プラズマ研究では，多くの種類のデータ解析が必要である．これまでは，このような解析は各研究者が各々で実装していたため，各人で実施できる解析には限界があった．AutoAna は，このようなルーチン的な解析を自動化することを目的として開発されている．自動化によりデータ解析の効率が飛躍的に向上し，研究者の労力が減少するだけでなく，解析に専門性を有するデータも共有化されることになり，多くの研究者がより広範囲の計測・解析データにアクセスできるようになった．

Keywords:
AutoAna, LHD, raw data, analyzed data, data sharing, automation, mapping

2.1　AutoAna システム構築のコンセプト

2.1.1　データ解析スタイルの変化

規模の大きなプラズマ実験施設におけるデータ解析インフラストラクチャは，集中型システムから分散型システムへと変化してきた．集中型システムでは，ユーザは一つの計算機（または計算機クラスタ）にアクセスして，データ取得・解析を行ってきた．一方，分散型システムでは，基本的には計測器ごとに独立した計算機があり，その計算機でデータ収集と解析を行うようになった．分散型システムでは，ネットワークに繋がった計算機の数を増加させるだけで計測器の増加に対応できるという利点がある．一方，計算量が多い計測では処理できる量（ショット）が限られるのに対し，計算量が少ない計測では，計算機がアイドリング状態になる．CPU の稼働率という点において，計算機資源を有効活用できないという欠点があった．

プラズマ実験施設に設置される計測器の数が増えるにつれて，一人の研究者が解析できる範囲が狭くなり，複数の研究者が協力して行うスタイルに変化してきた．個人ベースの解析からグループベースの解析へと進化するにしたがい，A が解析した結果としての出力データが B の解析の入力データとなっている場合が増えてきた．そして，「B が A に依頼した解析の終了を待つ」という「作業の待ち時間」が解析全体のスピードを著しく遅くする結果となった．

2.1.2　グループベースの解析システムに必要な要件

「作業の待ち時間」をなくすために，大型ヘリカル装置（LHD）では自動解析システム（AutoAna）が導入され，解析時間の大幅なスピードアップが図られた．自動解析システムは，研究者が解析を行いその結果をサーバに登録するのではなく，解析プログラムそのものを登録するというものである．解析プログラムの登録により，計算自体は自動的に行われるので，「作業の待ち時間」が「計算機の計算時間」となり，大幅なスピードアップが実現できるという利点がある．一方，自動解析システムでは，データをチェックしながら計算ステップを進める対話型の解析が行われないので，各ステップの結果のチェックに問題が残る．計算ステップごとの結果をデータサーバに自動登録し，それを使って次の計算を行うという多層化された解析データ（n 次データ）から構成される自動計算システムが求められる．以下で解析システムに必要な要件について議論する．

１）多層化された解析データ（n 次データ）の共有

従来の対話型の解析ルーチンでは，各計算ステップの結果を担当者がチェックして計算を行うという方法が採られていたが，自動計算システムでは，このインタラクティブの過程がない．そこで，チェックすべき各計算ステップの結果を全てサーバに保存する必要がある．データベースに最終結果のみならず，中間結果のデータも残すことが重要となってくる．それらも全てのユーザがアクセスできるようにしたことで，計測担当者任せではなく全員がデータの確実性をチェックすることができるようなった．

２）解析プログラムの自動バッチ処理

解析プログラムの開発においても，対話型から自動型へ

National Institute for Fusion Science, Toki, GIFU 509-5292, Japan
corresponding author's e-mail: csuzuki@nifs.ac.jp

ⓒ2019 The Japan Society of Plasma Science and Nuclear Fusion Research

の変更にいかに対応するかが重要となってくる．解析は基本的には全てのショット・時刻で行われるので，通常は解析の必要がないデータ（例えば，非常にバラついた温度分布）に対しても計算が行われる．したがって，どんなデータでもクラッシュしない（または無限ループに入らない）プログラムが必要とされる．自動バッチ処理では，エラー処理がしっかりと組み込まれた解析プログラムの登録が重要となってくる．対話型処理でしばしば使われる，解析時のパラメータ設定（例えばFFTの時間幅設定）も固定または自動設定（イベント条件付平均におけるイベントトリガーのしきい値設定）する必要がある．このような，解析目的やショットによって変化するパラメータが入力できないという自動型の欠点を克服するために，以下に述べる手法を取り入れている．

(a) あらかじめ何種類かのパラメータについて全て計算を行い，その結果を異なる名前をつけて全て登録しておく．そうすることで，ユーザは解析目的やショットにあったパラメータに対応する解析結果を参照することができる．しかしながら，この方法を中間層の解析プログラムに適用すると，データの種類がネズミ算的に増えてしまうという問題が生じる．

(b) パラメータ設定だけを書いた入力解析データファイルを入力リストに加える．パラメータ設定を変更したファイルを再登録することで，ショット毎に最適な解析パラメータで計算することが可能となる．

(c) 解析時のパラメータをデータから決定するプログラムを開発してAutoAnaに組み込む．ユーザ側からすればパラメータの自動設定がなされるので理想的な方法であるが，その開発にはプログラム開発者の高度なスキルと時間が必要である．

3）解析プログラムの共有

解析プログラムが登録されることで，計測担当者のみならず，全てのユーザがその計算方法を知ることができる．これは，解析のブラックボックス化を防ぎ，解析ルーチンという知的財産のグループ内での共有を促す．そうすることで，「AとBの解析結果が異なる」という人による解析結果の違いという問題を，「Aが開発したプログラムとBが開発したプログラムの解析結果が異なる」というプログラム（解析手法）の違いに帰着させることができる．

2.1.3　LHDの自動データ収集・解析システムの概要と課題

図1に従来型の計測別分散システムと自動計算システムを組み込んだ役割別分散システムを示す．従来は，計測機器（担当者）毎にデータ収集・解析計算機がある分散型となっていた．このシステムは担当者が個人ベースの解析を行い，その結果だけをグループのメンバーが共有するという考えに基づいて構築されている．LHDで導入されている自動データ収集・解析システムは，データ収集の部分は分散型になっているが，解析システムの部分は，役割毎にグルーピングされた計算機群から構成されている．1台の自動計算管理サーバ，3台の平衡データベースサーバ，17台のデータ解析計算機群である．データ解析計算機群は12台のショット間解析計算機群（短時間計算）と，5台の輸送

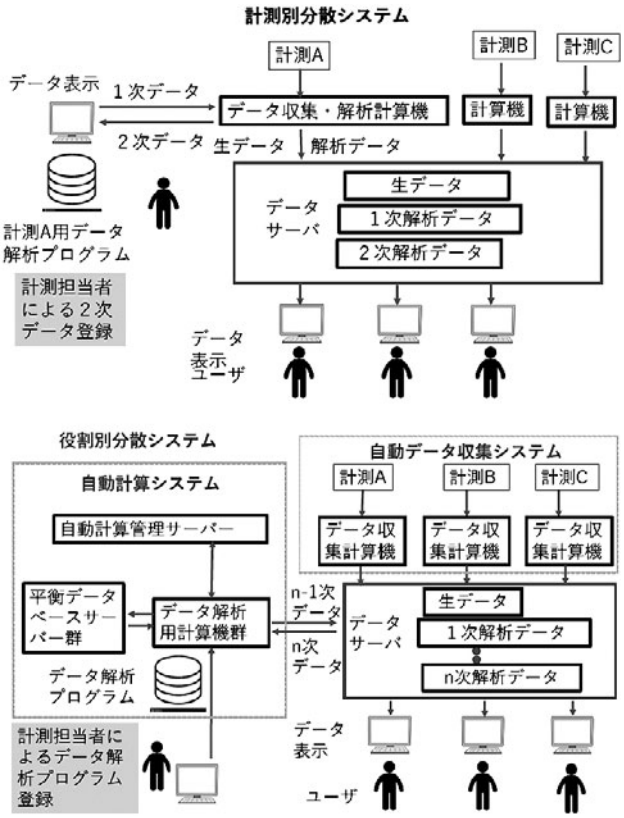

図1　従来型の計測別分散システムと自動計算システムを組み込んだ役割別分散システム．

解析用計算機群（長時間計算）から構成されている．計測担当者がシステムに解析プログラムを登録しておけば，基本的には計算はショットに同期して自動で行われる．各プログラムに対して計算を実行する計算機はあらかじめ決まっていない．自動計算管理サーバがアイドリング状態の計算機に次々とジョブを投入し，全ての計算機を使ってできるだけ短時間に全てのジョブが終了するように管理している．実験中のショット間に計算が終了する短時間計算を行う計算機群と，ショット間では計算が終了しない長時間計算に分離している．これは長時間計算のジョブが短時間計算のジョブを圧迫して，ショット間の表示が遅くなることを防ぐためである．

自動計算システムでは，プラズマ放電後に多層化された解析データ（n次データ）が大量に生成される．LHDではnが数十に達しているデータもある．ユーザは1次からn次までのデータを参照することができるので，最終結果のみならず，そのデータが作られた過程をたどることができる．この自動解析システムの課題は，今まで担当者に強く依存していた解析から，担当者に依存しない解析への変化にユーザがいかに対応していくかであろう．解析そのものは担当者任せではないので，ユーザは中間結果を見ながら自分でチェックするという心構えが必要になってくる．一方，担当者はルーチン的な解析作業から解放され，解析コードの改善や新しい解析手法の開発に専念できる．また，解析そのものは担当者に依頼せずに行えるがゆえに，解析プログラムを登録した担当者とのコミュニケーションが重要となってくる．

2.2　AutoAna システムの実際

2.2.1　システムの概要

LHD実験では，実験に関する物理データは解析サーバに登録されている．**図2**に示すように，AutoAna は解析サーバに登録されるデータを監視し，新たに登録されたデータを入力とするプログラムを自動的に起動するシステムである．具体的には下記のような動作を行い，プログラムの自動起動を行う．

1）解析サーバ上のデータが登録・更新されると，サーバはIPマルチキャストパケットを用いて，ネットワーク上の複数の計算機に，サーバ上のデータ構成に変更があったことを通知する．

2）AutoAna はこの通知を受けると，どのデータに変更があったかを確認し，そのデータを入力とするプログラムがあるかどうかをチェックする．実行するプログラムはモジュールという単位で管理されおり，全モジュールは，JSON（JavaScript Object Notation）ファイルに記述されている．プログラムを実行するために必要な入力データが全て揃っている場合，AutoAnaがプログラムの実行ジョブをキューに登録する．

3）プログラムの実行は複数の計算機で独立に行われており，それぞれのCPUコア数に応じた数の複数のプログラムが実行される．各計算機はキューから登録順にジョブを取り出してプログラムを実行し，計算結果を解析データとしてサーバに登録する．

2.2.2　最近の進展

AutoAna システムの構築に関しては，すでに本学会誌などでも報告されているが[1-3]，その後 AutoAna の重要度が高まり，いくつかの改良が施されている．2018年11月現在，AutoAna が管理するモジュール数は147，解析データ数は326まで増加した．増加したモジュールの管理を容易にするため，ジョブの実行状況・結果の確認や，モジュール間の依存関係の視覚化をWebで行えるように，ユーザインターフェイスの改良を行った．Web 上での表示結果の一例を**図3**に示す．これらの GUI の改善により，ある解析データがどの解析データに依存しているかや，障害が発生した場合の問題の所在を把握するのが容易になった．

AutoAnaによって実行されるジョブの一部には，非常に時間がかかるものもある．これらのジョブを実行すると他のジョブの実行が後回しになり，プラズマ実験中に直近のデータを用いて解析結果をグラフ表示させたい場合などに支障がでる．そのため，これらの長時間ジョブについては普段は無効にしておき，夜間・休日等に動かすといった運用を行えるように，Web 画面上からモジュールの有効・無効の切替ができるようにした．これらのジョブは手動で実行できるようにし，モジュール自体を無効にしたまま，ジョブの投入ができるようにした．また，非常に時間のかかるジョブについては，通常のジョブの実行に影響を与えないよう，特別なキューを別途用意することにより，対応を行った．

2.2.3　AutoAna の登録プログラムの条件

AutoAnaから起動されるプログラムには若干の制限がある．これらは，

1）ショット番号をコマンドライン引数で受け取り，該当するショット番号の解析データの登録を行う．
2）同一計算機で異なるショット番号の計算が同時に行われることがあり，中間ファイル等の衝突が起きないようにしなければならない．
3）データの登録が終了した時は，登録データを含め，実行中に作成したファイルは削除する必要がある．

である．使用するプログラムによっては，これらの条件を直接実装することが難しいものも存在するが，シェルスクリプトを用いて間接的に呼び出すことで，様々な処理系に対応することができる．現在，FORTRAN，Python，PV-Waveなどの言語で書かれたプログラムが実行されている．

実際に AutoAna から起動されるシェルスクリプトは B シェルで記述されており，その内部から，例えば PV-Wave などの外部処理系を呼び出す．AutoAna は，このシェルスクリプトを「スクリプト名＋ショット番号」という形で呼び出しており，対象となるショット番号は一番目の引数としてスクリプトに渡している．AutoAnaは人の手を介さないバッチ処理でプログラムを実行するため，シェルスクリプトのヒアドキュメントと呼ばれる機能を使い，PV-Wave では通常キーボードから行う入力をファイルから読み込ませて，PV-Wave 本体に渡している．

プログラムによっては，膨大な数の中間ファイルが作成されることもあり，ディスク容量を圧迫するため，不要となったファイルは直ちに削除することが必要である．シグナルを受けた時にコマンドを実行するtrapという機能を使い，正常終了および異常終了時に，中間ファイルを削除するようにしている．同時に異なるショット番号の計算が実

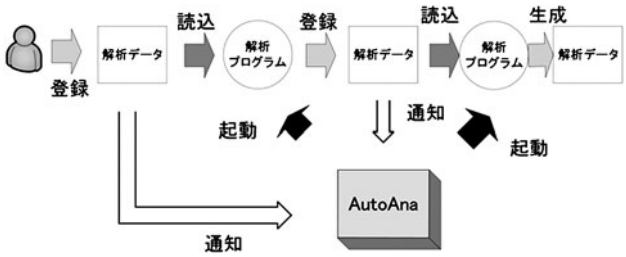

図2　AutoAna システムの動作の概要．

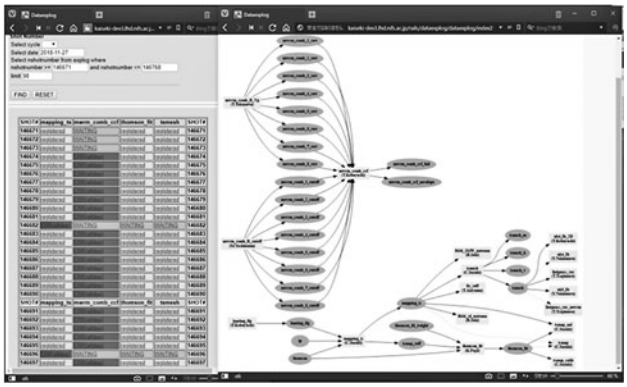

図3　ジョブの実行結果表示(左)とモジュールの依存関係図(右)．

行される可能性があるため，中間ファイルを作成する場合は，ショット番号毎に異なる名前を使用し，異なるショット番号の計算を行っているプロセスが同じ中間ファイルを参照しないようにしている．なお，同一のショット番号の計算が同時に実行されないことは，AutoAna 側で保証する．

2.2.4 課題と今後の展望

現在の AutoAna に対する要望の一つに，各担当者自身によるプログラムの置き換えがある．上記のように，AutoAnaで動かす解析プログラムには，従わなければならないルールがあるため，現在では，解析プログラムを担当者から提供してもらった後，AutoAna 管理者（江本）が若干の修正を加えて実装している．ところが，解析プログラムは常に改良が加えられており，その影響をすぐに確認したい研究者からは，自身でプログラムの置き換えができることが望まれている．プログラムの制限の大きな要因は，複数のプログラムが同時に動いており，それらの干渉を避けるためであるが，この問題については，それぞれのプログラムを隔離された仮想環境下で実行させることで解決できる．そこで，実行環境の仮想化をめざした改良を進める予定である．

また，ジョブ管理の高度化も今後の検討課題である．現在 AutoAna では，基本的に早いもの順でジョブを処理している．実験データの校正や，解析手法の改良などにより，依存関係の上流に位置する何万ものデータが一斉に更新されることがある．このような場合，このデータに依存する何十万～百万ものデータが更新されることになり，実験中の直近のデータなどの，優先度の高い作業が中断されてしまう．現状，このような大量の更新作業を行う時は，運用で対処しており，キューにジョブが貯まりすぎないよう速度を調節しながら，ジョブを流し込んでいる．これらの問題に対処するためには，緊急度やリソースの利用状況に応じた高度なジョブ管理が必要である．このようなジョブ管理の高度化を目指した改良を続ける予定である．

2.3 AutoAna システムの活用例 –磁気座標マッピングと輸送解析–

2.3.1 磁気座標マッピングの概要

AutoAna システムでサーバに登録される多数の解析データの中でも，「磁気座標マッピング」に関連する解析データ群は，依存関係の上流に位置し，それに続くプラズマ輸送解析などの基盤として，LHD実験に関わる研究者に広く利用されるに至っている．本項では，AutoAna システムの具体的な活用例として，「磁気座標マッピング」から電子・イオン輸送解析に至る一連の流れを解説する．

LHDで得られる計測データには，時刻と場所の関数として表される二次元データ（電子温度分布など）が多い．1次データ（生データ）の段階では，各計測器の観測点の実座標（R，Z，Φ）における物理量が登録されることになるが，その後の解析のためには，対応するプラズマの平衡を決定して，実座標を磁気座標（r_{eff}，θ，ϕ）に変換（マッピング）する必要がある．ここにr_{eff}はトーラスプラズマの実効小半径であり，同じ放電の同じ実座標位置であっても，r_{eff} は時間的に変化することになる．

LHD のような非軸対称系の磁場配位では平衡の解析解が存在しないため，7つのパラメータで表される様々なプラズマ条件に対する平衡をあらかじめ VMEC コード[4]で計算して大規模データベースを構築しておき，そこから計測データに適合する平衡を選択することで磁気座標に変換している．LHD では原則として，トムソン散乱計測による電子温度分布を磁気座標の r_{eff} の関数で表した時に内外ができるだけ対称になるような平衡を選ぶ．トムソン散乱計測は常時稼働しており，空間分解能・時間分解能ともに高く，マッピングの精度が期待できるからである．また密度分布を使わず温度分布のみを使うのは，温度データの方が誤差が小さく，かつピークした分布形状になることが多いのでやはり精度が上がるからである．平衡データベースや磁気座標マッピングの詳細については文献[1,5]を参照されたい．

2.3.2 関連解析データの生成と依存関係

AutoAna システムにおける解析データ・解析プログラムの入出力の依存関係のうち，磁気座標マッピングから加熱分布の計算を経て輸送解析に至る一連の流れを抜粋して示したものが**図 4** である．解析データは楕円形で，解析プログラム（モジュール）は長方形で示されており，簡単のため一部の重要でない入出力は省略されている．平衡マッピング，NBI 加熱分布計算，ECH 加熱分布計算，イオン輸送計算，電子輸送計算のそれぞれに相当する部分を点線で囲んでおり，各部分のプログラムを各担当者 A, B, ……が開発するというイメージである．簡単のため，各部分内部の中間プログラム名やデータ名は省略し依存関係のみを表示しているが，1モジュールのみの簡単な場合もあれば，複雑な多段階構造の場合もあることがわかる．例えば平衡マッピングは，thomson（トムソン散乱），ip（プラズマ電流）の二つの1次データを入力として，いくつかの中間段階を経て tsmap_smooth などの解析データを出力する．thomson だけでなく ip も必要なのは，上述の平衡決定においてipの値を参照するからである．thomson，ip，lhdcxs7_cvi（荷電交換分光）などはその上流にデータが存在しない最上流の解析データ（1次データ）であり，各計測の担当者によって登録される．これに対し，tsmap_smooth などは1次データを解析することによって生まれた n 次データということが理解できよう．

LHDのトムソン散乱計測データには，計測限界や検出器の感度特性のばらつき，背景光の影響などの避けられない理由により，外れ値や異常値が含まれることがある．生データのままでは正しいマッピングが行われなかったり，その後の輸送解析において半径方向の勾配をとる際に問題となることがある．このため，外れ値の除去などの前処理を行った分布データに対して処理を行う必要がある．現在は，京都大学の藤井氏が開発した，統計学的なベイズ推定を応用した分布データの処理手法をモジュールとして組み込み，処理を行っている[6]．このような処理済みの密度分布形状と，干渉計により計測された線積分密度を用い

て，電子密度の絶対値較正も行っており，tsmap_smooth などには密度の絶対値が出力される．続いて行われる輸送解析の出発点となる解析データ群である．

NBI 加熱分布の計算では，平衡マッピング済みの解析データ（tsmap_smooth_a99）から出発し，いくつかの中間段階を経て，ビームの減速過程も考慮した加熱分布データ（fit3d_sd_autoana）を出力する．この部分には Boozer 座標への変換コードや fit3d コード[7]が組み込まれている．ここで，平衡マッピング部分の中間データから，NBI 加熱分布内部のプログラムに直接矢印が引かれていることからわかるように，前段階の最終出力データだけでなく，中間データも必要に応じて参照している．

ECH 加熱分布の計算では，ビームの光線追跡を行う LHDGauss コード[8]が組み込まれたモジュールにより，lhdgauss_deprof が出力される．これらの加熱分布データを入力として，イオン輸送や電子輸送の計算が行われる．電子輸送の計算では，時間分解能を上げるため電子サイクロトロン放射のデータ（ece_fast_calThom）も参照する．最終的にイオンや電子の粒子・熱輸送係数などのデータが dytrans_cxs 7，dytrans_ece に出力される．

2. 3. 3　AutoAna の採用による利点

このようにして，輸送解析に必要なイオンや電子の輸送係数データの生成が，全てのショットの全ての時刻について自動的に行われ，関係者全員で共有されるようになった．これまでのようにその都度担当者に依頼する必要がなくなり，データ解析の効率が飛躍的に向上した．また，**図 4** で平衡マッピング部分までの解析プログラムは，前述の「短時間計算」に相当する比較的軽い処理であるため，LHD 実験の放電間隔（最短で 3 分）の間に処理が完了し，実験の進行に追随して磁気座標に変換済みの解析データが確認できるようになった．例えば，磁気座標で表されたトムソン散乱データから計算した蓄積エネルギーと，反磁性ループによる蓄積エネルギーをすぐに比較してビーム圧力の影響を評価することもできる．

ここで重要な点は，ユーザは**図 4** の A～E の各部分の最終出力データだけでなく，各部分内部の中間データも参照できることである．全ての中間データが解析データサーバ上で関係研究者全員で共有されており，データ処理の過程を誰でも追跡できるようになっている．したがって，前項でも述べたように，例えばなんらかの原因で妥当でないフィッティング結果が登録された場合に，そのような不具合を発見しやすく，かつその原因（1 次データの問題か，解析プログラムの処理の問題か）を特定して改善するのも容易となる．また，中間データが全て参照できるということ

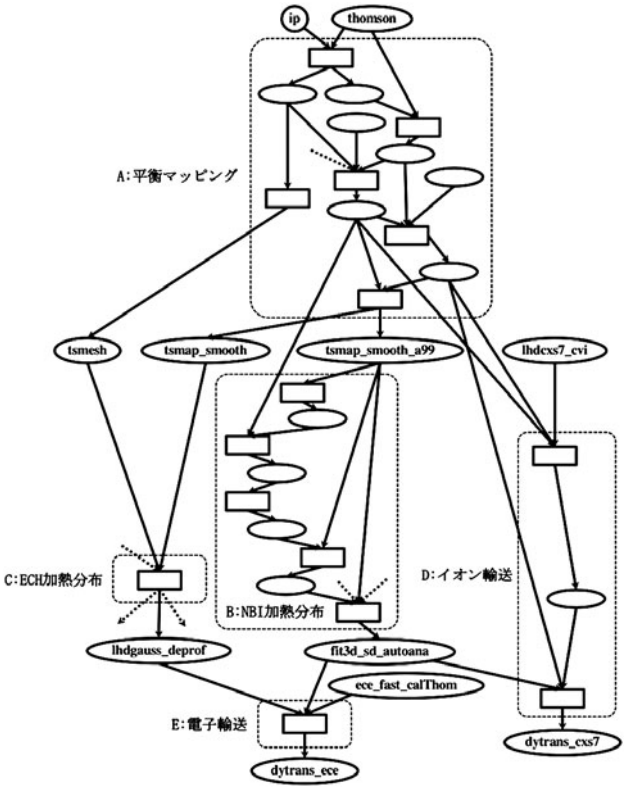

図 4　磁気座標マッピングから輸送解析に至る解析データと解析プログラムの入出力の流れ．解析データは楕円形で，解析プログラムは長方形で示されている．

とは，ユーザ側の処理の選択肢を広げることになる．例えば，電子温度ホールが生成された場合のような特殊なケースでは，前述のベイズ推定を用いた処理がうまくいかないこともある．その場合は，処理前の分布も合わせて中間データとして登録されているので，ユーザは処理前のデータを読み出して自身の解析を行うことができる．

参 考 文 献

［1］江本雅彦 他：プラズマ・核融合学会誌 **90**, 562 (2014).
［2］M. Emoto *et al.,* Fusion Eng. Des. **89**, 758 (2014).
［3］M. Emoto *et al.,* Fusion Sci. Technol. **74**, 161 (2018).
［4］S.P. Hirshman and J.C. Whitson, Phys. Fluids **26**, 3553 (1983).
［5］C. Suzuki *et al.,* Plasma Phys. Control. Fusion **55**, 014016 (2013).
［6］K. Fujii *et al.,* 45th EPS Conference, ECA, edited by S. Coda *et al.* (European Physical Society, Prague, 2018), Vol. 42A, p. 2.1007.
［7］S. Murakami *et al.,* Trans. Fusion Technol. **27**, 256 (1995).
［8］T. Ii Tsujimura *et al.,* Nucl. Fusion **55**, 123019 (2015).

J. Plasma Fusion Res. Vol.95, No.5 (2019) 213-215

小特集 プラズマ実験におけるデータ解析インフラストラクチャ

3．JT-60SA におけるデータ収集・解析インフラストラクチャ

3. Data Acquisition and Analysis Infrastructure in JT-60SA

3.1　JT-60SA におけるデータベースストラクチャ

3.1 Database Structure in JT-60SA

山崎莉穂，坂田信也，浦野　創
YAMAZAKI Riho, SAKATA Shinya and URANO Hajime
量子科学技術研究開発機構　那珂核融合研究所

（原稿受付：2019年 2 月22日）

大型トカマク装置 JT-60SA におけるデータ収集から解析までのデータの流れについて概説する．各計測機器から生成される生データは物理データに変換後，実験データサーバに格納されると共に，生データは元データサーバに格納される．一方，トカマク本体設備から生成される運転系データの一部は物理データに変換され，実験データサーバに格納される．JT-60SA の実験データベース・システムでは，これらのデータをすべて一元管理することにより，収集される多様なデータの管理および提供を容易にする．また，収集されたデータを用いてユーザが解析した平衡データや解析データについても同様に一元管理し，各ユーザが作成したデータを共通して利用するためのインフラ整備を進めている．

Keywords:
database structure, discharge data, plant monitoring data, raw data, equilibrium data, analysis data, JT-60SA

3.1.1　はじめに

量子科学技術研究開発機構が幅広いアプローチ活動（BA）協定の下で欧州と共同で設計・建設を進めている大型トカマク装置 JT-60SA[1]では，多種多様な計測機器・加熱設備が同期してプラズマ放電を行い，それらの装置からの大量のデータ収集が見込まれる．JT-60SA の実験参加は日本国内だけでなく，欧州からも想定されており，多数のユーザが円滑に実験・解析を行えるよう，大規模なデータ収集・解析システムを整備する必要がある．

本章では，JT-60SA におけるデータ収集から格納，解析までのデータの流れについて説明し，現在整備を進めている JT-60SA のデータベース（DB）・システムについて紹介する．**図 1** に，計測機器・トカマク本体設備側で取得されたデータがユーザに渡るまでの簡略的な流れを示す．本章では DB の構成，データ収集系，データアクセス等について各節に分けて述べる．3.1節では，JT-60SA における DB の構成と，各データの管理方法について説明する．3.2節では，JT-60SA の計測器及びトカマク本体設備からのデータ収集の流れとその利用方法について紹介する．3.3節では，複数の DB システムへのアクセス及びデータ解析のためのインフラ整備について紹介する．

3.1.2　JT-60SA の収集データと DB ストラクチャ

JT-60SA では，各計測器及びトカマク本体設備が放電に同期して収集するデータ，各本体設備が24時間連続的に収集するデータの他，各計測器が収集した物理量変換前の生データ，プラズマの平衡データ，さらにそれらのデータを用いて詳細な解析を行った解析データ等，様々なデータが DB に格納される．これらのデータは単に実験データとして参照されるだけでなく，格納された複数のデータを用いたデータ解析等にも利用される．そのため，大量の収集

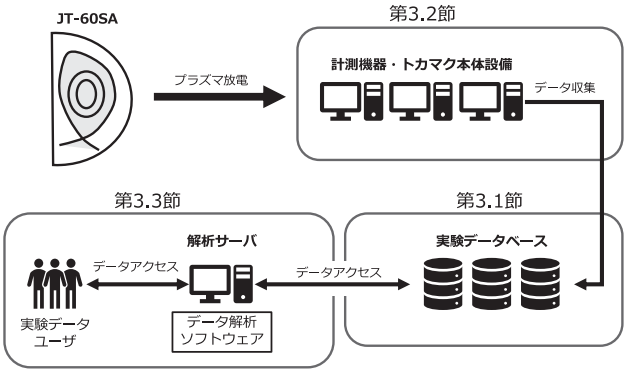

図 1　JT-60SA の実験 DB システムの概略．本章の各節で述べる領域を示す．

National Institutes for Quantum and Radiological Science and Technology, Naka, IBARAKI 311-0196, Japan
corresponding author's e-mail: yamazaki.riho@qst.go.jp

ⓒ2019 The Japan Society of Plasma Science and Nuclear Fusion Research

データを適切に管理し，かつユーザが各データの違いを意識せずに利用できるようなDBシステムを構築することが重要である．

上述の各データはデータフォーマットや利用方法が異なるため，すべてのデータに対し同一の格納方法を用いることは難しい．そこで，データの種類に応じ，複数のDBに分けることで，各データに適した管理を行う．JT-60SAにおける各種DBの構成を**図2**に示す．元データサーバは物理データの生成元となった生データ等を管理するサーバであり，実験データサーバは各設備が収集した物理データ及び平衡計算を行ったデータを管理する．解析サーバはこれらのDBに登録された各データをユーザに提供するための窓口としての役割を担うことに加え，ユーザが作成した解析データを管理する．解析サーバは各DBへのアクセスライブラリを持っているため，解析サーバを介して各データを利用することが可能である．これにより，ユーザは複数のDBへのアクセスを効率的に行うことができる．さらに，解析サーバ上にはJT-60SA用のデータ解析用ソフトウェア（eDAS）が整備されており，データのグラフ表示や基本的なデータ解析も行うことができる．本節ではJT-60SAで収集するデータとその管理方法について説明する．

3.1.3　生データとその関連データ

プラズマ放電によって各計測器が収集するデータは，一般にそのままの状態では物理データとしての解析に適さないため，変換や補正，間引き等のデータ加工処理（定型処理）を行い，物理量としての意味を持ったデータに変換する必要がある．元データサーバでは，物理データの生成元となった生データ及び定型処理に使用したデータ加工プログラム，計測器の較正データの管理を行う．

元データサーバに収集される生データ量は放電1ショット当たり最大170 GB（100秒放電を想定）と大量であるため，すべてのデータを一度に転送しようとすると，ネットワークに負荷がかかり，データ転送速度が遅くなる原因となる．そのため，生データの転送を実験終了後の夜間等，ネットワーク負荷の少ない時間帯に行う．転送する生データはヘッダ部と実データ部で構成される．ヘッダ部にはデータの種別やその構造といった，データを読み出す際に必要となる情報が記述されており，これらの情報から実データ部の構造を知ることができる．また，生データは放電ショット番号及び各計測量の識別番号（PID番号）をインデックスとして各フォルダに格納される．

図2の左上枠内に示すように，生データは放電後に定型処理が行われ，物理データに変換されるが，定型処理に使用したデータ加工プログラム，較正データは放電ごとに変更する必要がない．よって，データ加工プログラム，較正データはデータ転送年月日ごとに各フォルダに格納され，計算式等の修正が行われた場合に再度登録される．また，各ファイル名にはデータ転送年月日に加え，枝番を付けて登録するため，同日に複数のデータを転送することが可能である．これらのデータは生データと異なり，解析等の処理を行わないため，ヘッダ部を付加せずに転送する．

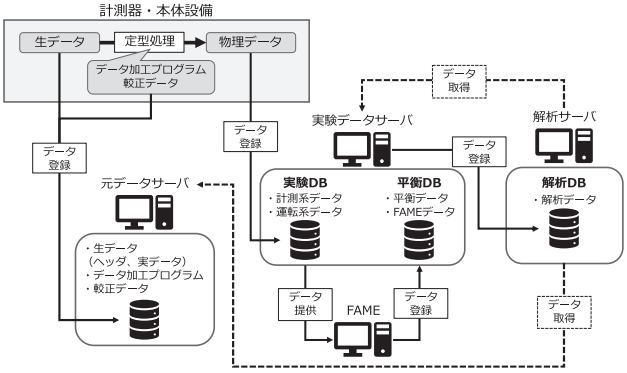

図2　JT-60SAにおける各種DBの構成と設備からのデータの流れ．

物理データの生成後においてもその生成過程を参照するためには，物理データとその生成元の生データ，使用した加工プログラム，較正データの関連性を明確に情報として残す必要がある．よって，各データの紐付けをデータベーステーブルにより行う．また，生データのヘッダ部に生成後の物理データのPID番号を記述することで，紐付けをより明確にする．これらの情報をDBシステムに残すことで，物理データの再計算等の際に，物理データの生成過程を参照することが可能である．

3.1.4　計測系データと運転系データ（実験DB）

各計測器が収集した物理データ（計測系データ）は放電に同期して収集されるが，トカマク本体設備が収集した物理データ（運転系データ）には放電に同期して収集されるデータ（設備放電データ）と，24時間連続的に収集されるデータ（設備監視データ）がある．計測系データ及び運転系データはどちらも実験データサーバに転送され，実験DBに格納される．また，両データはヘッダ部と実データ部で構成され，ヘッダ部と実データ部のデータフォーマットを統一している．これにより，ユーザは計測系，運転系の区別なく両データを扱うことができる．

計測系データ及び設備放電データはプラズマ放電の直後に全設備から転送され，生データと同様に放電ショット番号及びPID番号をインデックスとして各フォルダに格納される．物理データのヘッダ部には生成元となった生データのPID番号，使用した加工プログラムのファイル名及び較正データのファイル名が記述される．これらの情報から格納された物理データの生成過程を参照できる．

設備監視データは，トカマク本体設備が連続的に収集するデータであり，15分周期で実験データサーバに転送されたデータは，1日分ごとに分けて格納される．設備監視データはその時間軸となるデータ（時刻データ）と合わせて転送されることが必要であるが，同一の計算機から収集された設備監視データは同一の時刻データを持つため，1つの時刻データに対しそれを参照する設備監視データは複数存在する．例えば，超伝導コイル関連設備から転送される設備監視データは約2000種だが，その時間軸となる時刻データはわずか2種であるため，同一の時刻データが重複して登録されることになる．よって，時刻データと設備

監視データを別々に登録し，データの重複登録を防ぐ．また，各設備監視データのヘッダ部に，参照する時刻データのPID番号を記述することで，各設備監視データがどの時刻データに対応するかを明確にする．

定型処理の際に使用した計算式や較正データの見直しにより，生データに再計算等の処理を行う必要が生じた場合，物理データの更新が必要となる．データの更新履歴はRev.1，Rev.2のようにリビジョンによって管理される．これにより，格納されたデータを上書き，削除することなく，更新前後のデータを並べて比較することが可能である．また，複数の物理データを再加工して二次的な物理データを生成する場合，各物理データのリビジョンと生成した二次物理データのリビジョンを紐付ける必要があり，この機能の実装を進めている．

3.1.5 平衡データ（平衡DB）

平衡DBには，MHD平衡計算コード（SELENE）[2]を用いて，Grad-Shafranov方程式を解いて得られた磁束分布の情報が格納される．JT-60SAでは，このデータに関して目的に応じて2種類のデータを配置する．1つは平衡データと呼ばれ，ユーザの要求に応じて特定のショット番号／時刻において随時計算を実行し，結果をDB上に格納するもので，これは主に放電終了後にユーザがデータ解析を行うための利用を想定している．したがって，平衡データはユーザが解析するたびに，その都度特定の時刻のデータが作成され，既にデータが登録されている場合は登録されたデータを読み出す仕組みである．一方で，JT-60SAにはFAMEと呼ばれる計算機が運用され，これは運転系設備から放電終了の信号を受信し，ある仮定の下で自動的に平衡計算を行い，結果をDB上に格納する．FAMEはJT-60U時代から存在したシステムであり，放電開始から終了まで100ミリ秒ごとに平衡計算を実行し，計算結果をFAMEデータとして格納するとともに，各種の平衡パラメータを時系列データとして，実験データサーバに転送する．格納されたFAMEデータは，磁束分布の時間変化をアニメーションとして見ることができるため，実験中にプラズマ位置形状制御が指令通りに動作しているか，計測器の視線や視点が実験に必要とされる磁気面上にあるかどうか等を確認する目的で利用される．

なお，平衡データ，FAMEデータともに磁束分布を計算するためには，磁気計測の値や装置の導体構造物に流れる渦電流を考慮する必要があるが，渦電流分布を計測することはできないため，導体モデルを仮定して放電開始からのコイル電流やプラズマ電流の時間変化から導出する．JT-60SAでは，放電中にプラズマ位置形状制御用に実時間で計算された渦電流値は，磁気計測データとともに実験DBに格納され，これらを用いて平衡計算が行われる仕組みになっている．

3.1.6 解析データ（解析DB）

解析データは実験DBや平衡DBに格納されている各データを用いてユーザが詳細な解析を行ったデータであり，各種解析コード類によるシミュレーション結果やユーザが独自に生成したデータ等を含み，解析サーバ上で管理が行われる．解析DBへのデータ登録は専用ツールで行われ，解析サーバの登録者であれば解析DBへのアクセスが可能である．この解析DBはFORTRANのネームリスト形式に準じたASCIIファイルをフォーマットとしているのが特徴で，ユーザは各自で作成した解析データを自分のホームディレクトリに登録し，データの編集を行うことができる．また，他のユーザが作成した解析データを読み出すことが可能である．解析DBへのデータ登録及び読み出しについては，一般ユーザであれば誰でも行うことができ，（パブリッシュされたデータ等で，ある程度ルール化することで）各自で作成した解析データを他の多くのユーザが共通して使用することができる．このように解析DBは様々なデータに対応する柔軟性を持つ一方で，所定のデータを規則的に格納するものとは限らないため，格納されているデータが何かという管理は原則ユーザに委ねられる．そのため，解析DBは別途Descriptionファイルというデータ定義の情報をまとめたファイルと紐づけられており，ユーザがこれを編集することができる．また，JT-60SAのデータ解析ソフトウェアや解析コードは，各データの説明，次元や単位等の情報を記載したDescriptionファイルを保持している．

3.1.7 まとめ

本節では，JT-60SAのDBストラクチャについて簡単に述べた．JT-60SAの実験運転では，フォーマットや利用方法の異なる複数のデータを大量に扱うため，各データに合った管理方法を採用することが重要である．JT-60SAのDBシステムでは，設備から収集された生データから，物理量への変換を行った計測系データ及び運転系データ，さらにそれらのデータを元に解析を行った平衡データ及び解析データに至るまで，すべてのデータを格納・管理し，ユーザに提供できるような環境を整備している．また，各データは元データサーバ，実験データサーバ，解析サーバでそれぞれ管理される．ユーザは複数あるDBに直接アクセスするのではなく，解析サーバというデータ提供の窓口を通して，複数のデータを取得することができる．

現在，既に登録された物理データから生成された二次的な物理データをDBに登録するための機能の実装を進めている．二次物理データを多くの研究者が共通して利用するために，二次物理データとその生成元となった物理データのリビジョンを管理し，各物理データの更新に伴い自動的に二次物理データを更新する機能を実装する．二次物理データについては，詳細を3.2節で述べる．

参考文献

[1] H. Shirai *et al.*, Nucl. Fusion **57**, 102002 (2017).
[2] T. Takeda *et al.*, JAERI-M 8042 (1979).

J. Plasma Fusion Res. Vol.95, No.5 (2019) 216-219

小特集 プラズマ実験におけるデータ解析インフラストラクチャ

3．JT-60SAにおけるデータ収集・解析インフラストラクチャ

3. Data Acquisition and Analysis Infrastructure in JT-60SA

3.2 JT-60SAにおけるデータ収集

3.2 Data Acquisition in JT-60SA

仲野友英，坂田信也，山崎莉穂，浦野　創，戸塚俊之
NAKANO Tomohide, SAKATA Shinya, YAMAZAKI Riho, URANO Hajime and TOTSUKA Toshiyuki
量子科学技術研究開発機構　那珂核融合研究所

（原稿受付：2019年2月22日）

JT-60SAにおけるデータの流れを概説する．分光器や干渉計などの計測器から生成されるデータは，物理データに変換され，実験データサーバに格納されるとともに，生データは元データサーバに格納される．他方，加熱装置やガスパフ装置などのトカマク本体設備から生成されるデータの一部も物理データに変換され，実験データサーバに格納される．これらの物理データのうち一部のデータはさらに他の物理データを用いて再加工され，より有用な物理データとして，ユーザに供される．JT-60SAデータベース・システムでは，このような再加工された物理データもデータベースに格納できる仕組みとリビジョン管理を実施する仕組みを実装する．また，運転系データのうち，プラズマ放電に同期せず24時間連続的に収集されるデータの時間軸をプラズマ放電と同期する計測系データと共通にするため，計測系データの時間を絶対時間に置換する仕組みを実装する．

Keywords:
Diagnostic, database system, data link, processed data, raw data, JT-60SA

3.2.1　データ収集（計測系，運転系）

JT-60SAでは数多くの計測器が設置され[1]，プラズマ放電に同期して種々のデータ（計測系データ）が生成される．他方，トカマク本体には加熱装置やガスパフ装置，真空排気装置などが付設され，オン・オフのタイミングや実績値（加熱パワーやガスパフ量，圧力）などのデータ（運転系データ）が生成される．運転系データは，必ずしもプラズマ放電中だけに生成されるのではなく，例えば，真空容器内部の圧力や熱電対によるダイバータ板の温度などは，プラズマ放電と次のプラズマ放電の間の時間にもデータとして記録される．運転系データは，計測系データと同様に研究に有用，場合によっては必須である．よって，ユーザには，運転系・計測系データの区別なく，シームレスなデータへのアクセスを提供するべきである．そのため，種々のデータの特徴を吸収して統一性を高め，効率よく管理することによって，データベースの利便性を高めるようにJT-60SAデータベース・システムの設計を進めている．

図1にはJT-60SAにおけるデータの流れを示す．分光器や干渉計など種々の計測器は生データを生成する．生データは，較正データなどを用いた物理データへの変換プロセス（定型処理）を経て，実験データサーバに転送される．この一連のデータの流れはプラズマ放電の終了後に直ちに実行されることが重要である．データ格納までに要する時間は，プラズマ実験実施者がデータを参照して，次のプラズマ実験の条件決定のための思考に許される時間を左右するからである．

一方で，生データは物理データほどに，緊急を要さない．むしろ，実験終了後に詳細解析のため，物理データに変換されなかったデータを抽出する，あるいは実験データベースに格納されている物理データより高い時間分解で物理データを再評価する，などの利用方法が想定される．また，物理データよりもデータ量が多いため，次のプラズマ放電までに元データサーバへの転送が完了しない可能性がある．これらの理由から，生データをプラズマ放電とプラズマ放電の間に転送するのではなく，実験終了後の夜間に，計測器毎に割り当てられた時間枠内で元データサーバへ転送する方針である．なお，元データサーバには，ヘッダを含む生データだけでなく，定型処理のために使用した較正データとデータ加工プログラムなど，生データから物理データを生成するのに必要な全ての情報が格納される．なお，運転系生データは各設備によって保管される．

National Institutes for Quantum and Radiological Science and Technology, Naka, IBARAKI 311-0196, Japan
corresponding author's e-mail: nakano.tomohide@qst.go.jp

©2019 The Japan Society of Plasma Science and Nuclear Fusion Research

3.2.2 一次および二次物理データの生成とデータベース・システムによる管理

多くの物理データは，**図1**の左下枠内に示すように，他の物理データを必要とせず，1種のみの計測器から生成される．たとえば，主プラズマ中のイオン温度は，荷電交換過程に由来して放射されるスペクトル線C VI（$n=7-8$：529 nm）を分光器によって測定し，波長較正されたスペクトル線のドップラー拡がりの幅から導出される．他方，1種のみの計測器で生成される物理データではプラズマ物理研究には不十分な場合があり，その場合には他の物理データを用いてより有用な物理データを生成する必要がある．たとえば，プラズマ実効電荷数 Z_{eff} を導出するには，以下の式で示されるように，制動放射光強度に加えて，電子温度と密度が必要である[2]（実際には，電子温度・密度分布のマッピングのため，磁場平衡も必要）．

$$P^{\mathrm{B}}(\nu)\mathrm{d}\nu = n_{\mathrm{e}} n_Z z J (R/T_{\mathrm{e}})^{0.5} \exp(-h\nu/T_{\mathrm{e}}) \frac{1}{2} g_{\mathrm{ff}} \mathrm{d}\nu,$$

$$\text{with } J = \frac{2^7}{3^{1.5}} (2\pi)^{0.5} \alpha \left(\frac{z^2 R}{mc^2}\right)^{1.5} hc(a_0/z)^2,$$

$$\text{and } Z_{\mathrm{eff}} = \frac{\Sigma_Z n_Z z^2}{n_{\mathrm{e}}}.$$

ここで，$P^{\mathrm{B}}(\nu)$ は周波数 ν の制動放射光放射率，n_Z は価数Zの不純物密度，R はリュードベリ定数（＝13.6 eV），h はプランク定数（$=6.6\times10^{-34}$ Js），g_{ff} はガウントファクター，α は微細構造定数（～1/137），m は電子の質量（$=9.1\times10^{-31}$ kg），c は光速（$=3.0\times10^{8}$ m/s），および a_0 はボーア半径（$=5.3\times10^{-11}$ m）である．制動放射光強度は多くのプラズマ物理研究者にとってあまり有用ではない．制動放射光強度よりもプラズマ実効電荷数の方が有用であるのは明らかである．そこで，データベースに登録済みの複数の物理データ（この例では制動放射光強度，電子温度・密度）を使用して，二次的な物理データ（同，プラズマ実効電荷数）を導出する仕組みが必要であり，導出された二次物理データも実験データサーバに登録できるようなデータベース・システムの柔軟性と二次物理データに特有の管理が必要である．二次物理データのリビジョン管理には，二次物理データと一次物理データの紐付けが必要になる．この紐付けによって，一次物理データが修正された場合に二次物理データが適切に修正される．上の例では，制動放射光強度の較正係数が変更された場合はもとより，電子温度や電子密度，さらには磁場平衡が修正された場合にも，プラズマ実効電荷数の修正が必要になる．この一連の修正作業を人間が解析手順を遡って構成データ一覧を調べるのではなく，プラズマ実効電荷数のリビジョンに対応する制動放射光強度，電子温度・密度のリビジョンを監視し，それらに不整合が認められた場合にはプラズマ実効電荷数の修正・再計算を自動的に促すようなリビジョン管理の仕組みが必要であり，JT-60SA データーベース・システムに実装を進めている．

このような管理機能を持たせたデータベース・システムを構築することによって，複数の計測器で生成された生データおよび一次物理データを用いて，二次物理データの高空間分解化や整合性の確認，冗長性の確保など，これまで持たなかった価値を生み出す利用方法に発展させることができる．**図2**には，制動放射光強度を測定可能な計測器を示す．X線イメージング結晶分光器（XICS）と荷電交換再結合分光（CXR）では，それぞれフォトンエネルギー5 keVおよび波長529 nmのスペクトル線とともに制動放射光が連続光として測定される．軟X線放射計測（SX）と可視分光（Visible）では，それぞれフォトンエネルギー1－2 keV付近および波長523 nmの制動放射光強度が測定

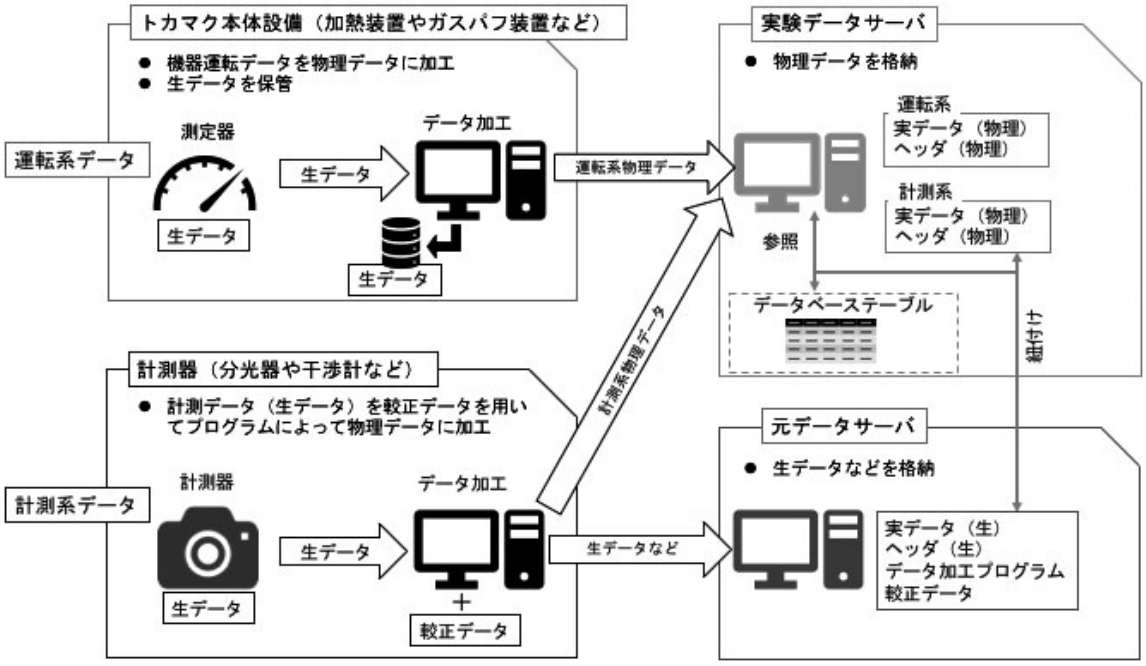

図1　JT-60SA におけるデータ収集とデータの流れ．

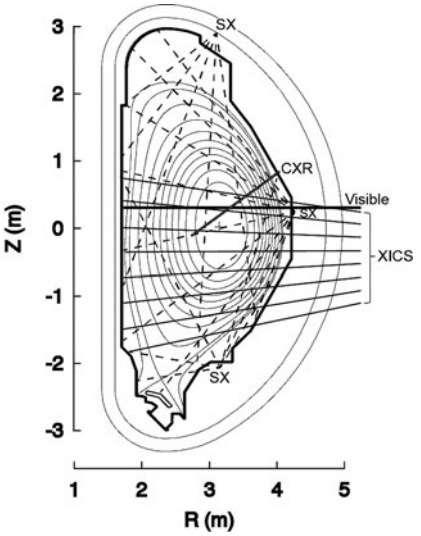

図２　JT-60SA 磁場平衡と分布計測の例．間引いた視線を示す．XICS：X 線イメージング結晶分光，CXR：荷電交換再結合分光（視線があるポロイダル断面を貫く位置），SX：軟 X 線放射計測および Visible：可視分光．(注)検討中の計測器の視線も含まれており，必ずしも全ての計測器が実装されるのではない．

される．これら４種の計測器から得られる制動放射光強度データのうち最初の３種は物理データとして実験データサーバに登録されず，生データとして元データサーバに保管され，一般には利用されない場合が多い．しかし，これらの生データを較正データなどを用いて適切に加工することによって，制動放射光強度の物理データを生成することが可能である．これによって，本来は制動放射光強度の計測は可視分光（Visible）の１視線のみであるため分布計測は不可能であるが，他の３種の計測器の生データを利用することによって分布計測へ発展する．これによって，アーベル変換により制動放射光強度の径方向分布を導出し，さらに電子温度・密度の径方向分布を用いて，プラズマ実効電荷数の径方向分布を導出することが可能である．導出されたプラズマ実効電荷数の径方向分布を比較することによって，これら計測器間の整合性を検証することにもつながり，これらの計測器の妥当性の確認という観点で成果が本来目的の物理データ（例えば，SX なら軟 X 放射強度の絶対値およびチャンネル間の相対強度の信頼性の向上，CXR なら C VI スペクトル線強度の絶対値の信頼性の向上，など）に還元される．JT-60SA データーベース・システムでは，このような解析を実行できるようにツール整備（3.3 節，eDAS）を進めている．さらには，全ての計測器で整合するプラズマ実効電荷数の二次元空間分布を再構築することも可能であろう（注：これら４種の計測器で測定される制動放射光のフォトンエネルギーまたは波長が異なるため，これらの制動放射光強度をそのまま使ったアーベル変換あるいはコンピュータ・トモグラフィーによる空間分布の再構築はできないので，プラズマ実効電荷数などを介する解析が必要）．

このような二次物理データやその空間分布の導出を自動的かつ省力化された保守のもとで行うには，径方向分布データのフィッティングルーチンや磁場平衡へのマッピングルーチンの共通化，磁場平衡の統一性など，標準ライブラリの整備といった技術的な基盤の構築も重要である．

3.2.3　実験データベースとして格納される運転系データ

運転系データは，プラズマ放電と同期して収集される設備放電データ，プラズマ放電とは同期せずに24時間連続的に収集される設備監視データから構成される．

設備放電データは，プラズマの生成・維持に直接的に必要となる装置の入出力データ，例えばガスパフ装置やポロイダルコイルの制御設定値やその実績データなどである．これらの生データは，既に，**図１**で示したように，物理データに変換後，実験データベースの一部として格納され，ユーザに提供される．なお，プラズマ放電の開始時刻や終了時刻などのプラズマ放電に関わる情報データも設備放電データとして格納される．

また，設備監視データは，実験運転中に機器状態の監視のため連続的に運転される装置，例えば真空容器内圧力やダイバータ板温度などのデータである．JT-60SA 実験運転より，その一部の生データが物理データに変換された後，15分周期で実験データサーバに転送され，実験データベースの一部として格納される．

3.2.4　設備監視データの利用

設備監視データを実験データベースとして格納するにあたり，粒子バランス研究等のプラズマ研究を目的としたデータ，例えば真空容器内圧力や容器内外の温度や排気系の質量分析器などのデータ，設備の機器状態の監視・開発等を目的としたデータ，例えば超伝導コイル関連のヘリウム冷凍系減圧機の回転数や減圧ポンプの出口オイル温度などのデータを選定した．ただし，2020年３月末より開始される統合コミッショニング期においては，真空容器内圧力，質量分析データおよびガスパフ装置などのトカマク本体設備が収集する全ての設備監視データ最大で約600点，加えて超伝導コイル関連のヘリウム冷凍機に関する全ての設備監視データ約2000点を実験データベースに格納できるように準備を進めている．

プラズマ研究を目的とした設備監視データの利用時には，プラズマ放電時に同期して収集される計測系データとの時間的な一致が必要不可欠である．しかし，計測系データがハードウェアのタイミング信号を介して厳密にデータ収集機器間の同期を図るのに対し，設備監視データのデータ収集機器間の同期は厳密に図られてはいない．また，計測系データの収集に関わる機器はプラズマ放電開始時刻を基準とした相対時間をデータ収集時の時間軸とするので設備監視データも同様の時間軸でも表示でできるような仕組みが必要である．この目的のため，設備放電データの一つであるプラズマ放電開始時刻及び終了時刻の絶対時刻を利用して，設備監視データの時間軸を計測系データ同様にプラズマ放電開始時刻からの相対時間に変換する仕組みを実装する．これに伴い，プラズマ運転に関わるデータ収集機

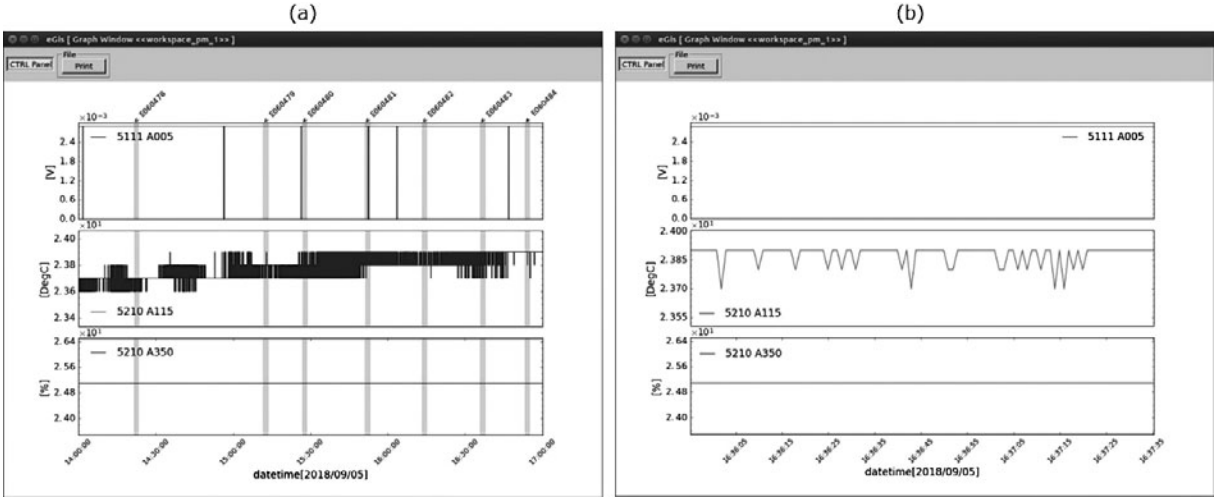

図 3　利用者への運転系データ提供例．(a)時刻に対する上段：TFコイル10-11クエンチ電圧，中段：減圧ポンプ出口オイル温度，下段：ヘリウム圧縮機回転数，および(b)ショット番号 E060483 放電中の拡大図．図(a)では，幅広い時刻（3時間）にわたって機器の運転状態を監視でき，図 (b) では，これらの機器とプラズマ放電に同期して収集されるデータ，例えば計測系データ，との比較を想定している．(注)これらのデータは通信試験に用いた模擬データの表示である．

器間の時刻の同期を目的とした全設備共通の NTP（Network Time Protocol）サーバの設置および設備監視データの時間軸として絶対時刻を採用するようデータフォーマットを共通化する．

また，これらの対策に併せ，設備監視データを利用者に提供するためのツールの整備を進めている．**図 3** に利用者への設備監視データ提供例として，超伝導コイル関連のヘリウム冷凍機データ（模擬データ）の表示を記す．**図 3** (a) は，機器の運転状態の監視を目的とした表示である．なお，グラフ上に時間軸の範囲内に行われたプラズマ放電をマーキングすることでプラズマ放電との関連を容易に調べられる．**図 3** (b)は，プラズマ放電に同期して収集される計測系データとの比較を目的とした表示である．プラズマ放電開始時刻と終了時刻を時間軸に設定することでプラズマ放電に同期した設備機器の状態変化をより詳細に把握することが可能である．

放電情報データの利用，NTPサーバの設置および運転系データのデータフォーマットの共通化により，設備監視データ利用における課題を解決し，より有効に活用できる環境を整備する．JT-60SA 実験運転より新たに実験データベースの一部となる設備監視データの活用がプラズマ研究の進展および設備機器の開発において大きく貢献することが期待される．

参 考 文 献

［1］ JT-60SA Research Plan v4.0, JT-60SA Research Unit; http://www.jt60sa.org/pdfs/JT-60SA_Res_Plan.pdf (current as of Feb. 1, 2019).

［2］ T. Fujimoto, *Plasma Spectroscopy* (International Series of Monographs on Physics vol. 123) (Oxford, Oxford University Press 12004).

J. Plasma Fusion Res. Vol.95, No.5 (2019) 220-222

小特集 プラズマ実験におけるデータ解析インフラストラクチャ

3．JT-60SA におけるデータ収集・解析インフラストラクチャ

3. Data Acquisition and Analysis Infrastructure in JT-60SA

3.3 JT-60SA におけるデータベースアクセス

3.3 Database Access Infrastructure in JT-60SA

浦野　創，坂田信也，山崎莉穂
URANO Hajime, SAKATA Shinya and YAMAZAKI Riho
量子科学技術研究開発機構　那珂核融合研究所

（原稿受付：2019年 2 月22日）

JT-60SA ではデータの種類に応じた複数のデータベース（DB）を運用しており，これらの DB アクセスについては，C 言語や FORTRAN 等で記述された関数群をライブラリ化してユーザに提供する．JT-60SA では各々が構造の異なる全ての DB システムへのアクセスにおいて，ある程度共通化されたフォーマットを採用するように設計している．複数の DB を利用するような環境であっても，ユーザ側からは複雑なデータ構造を意識せずにデータ解析を行えるように，DB アクセスのユーザインターフェース上にはシンプルな DB 構造を仮想的に持ち込むなどの工夫が重要である．

Keywords:
data analysis infrastructure, database, data access library, eDAS, JT-60SA

3.3.1　概要

日欧の共同事業である JT-60SA 計画[1]では，核融合エネルギーの早期実現のために国際熱核融合実験炉 ITER の支援・補完研究および人材の育成を主要目的としており，日本国内からの実験参加の他，欧州からの実験参加がオンサイト並びにオフサイトで想定されている．また，大型実験装置でのデータ解析においては，単に大容量データを扱うというだけでなく，第3.1節で解説されたように，放電データ，運転データ，平衡データ，解析データ等の複数の異なる種類のデータを扱いながらデータ解析を行う必要があるため，可能な限りシンプルなデータアクセスのインフラを整備することが要求される．

JT-60SA におけるデータベース（DB）アクセスには，第一に JT-60SA 用のデータ解析ソフトウェア（eDAS）の利用を推奨している．eDAS については，過去にプラズマ・核融合学会誌における講座「今日からはじめる Python [2]」の中でまとめられているため，こちらも併せて参考にしていただきたい．eDAS は，大型装置におけるデータ解析ソフトウェアとして十分な機能の拡張性を有する設計の下で開発されており，特に DB アクセスの観点からは，仮想的な DB 構造を GUI 上に持ち込むことでユーザ側からは複雑なデータ構造を意識せずにデータ解析を行える環境を提供している．次に，独自にプログラミングする等，さらに高度なデータ解析を行うユーザのために，DB アクセスライブラリを用意している．これは C 言語や FORTRAN で記述された関数群をライブラリ化しており，Python で wrapper 処理したオブジェクトをユーザに提供している．本節では，JT-60SA の DB アクセスのためのインフラの整備について述べる．

3.3.2　ユーザによるデータ取得と解析

図 1 に JT-60SA の解析サーバからの各種データアクセス及び eDAS 周りのデータフローを示す．JT-60SA では上述の通りデータの種類に応じた複数の DB を運用しており，

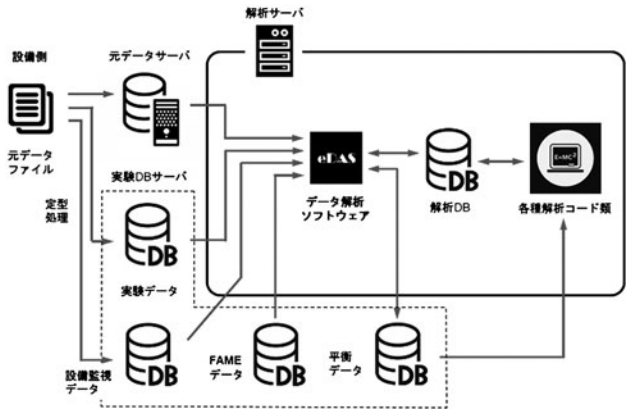

図 1　JT-60SA の解析サーバからの各種データアクセス及びデータ解析ソフトウェア周りのデータフロー．

National Institutes for Quantum and Radiological Science and Technology, Naka, IBARAKI 311-0196, Japan
corresponding author's e-mail: urano.hajime@qst.go.jp

©2019 The Japan Society of Plasma Science and Nuclear Fusion Research

DB アクセスライブラリ及び eDAS の各ツールは解析の目的に応じて DB との入出力を行う．JT-60SA の放電データ及び運転データの実体は，実験 DB 上に PID 番号と呼ばれる，任意に割り振られた英数字から成る 9 桁の識別子で一様に格納されており，解析サーバ上から DB アクセスの命令が来た場合には，この識別子によってデータを探索する．しかし，ユーザ側にとってはアクセスしたいデータに該当する PID 番号を知る必要がある他，階層を持たない一様なデータ構造であるため探索領域が広いという問題が生じる．そのため，JT-60SA での DB アクセスには，ユーザ側が効率的にデータ探索を行えるように「カテゴリ名／データ略称名」という見かけ上の階層構造を導入している．**図 2** に JT-60SA における DB 構造を GUI で可視化した実験データの例を示す．このような仮想的なインターフェースを用いることで，ユーザは PID 番号との関係はもとより，その存在を知る必要がなくなる．また，DB システムへのアクセスにおいては，異なる DB 間でもある程度共通化された関数形のインターフェースを採用するように設計している．

さらに，PID 番号を持たない平衡 DB や解析 DB にも同様に仮想的な階層構造を導入することで，DB の種別によらず同じフォーマットでのアクセスを可能にしている．これによって，実験DB上に存在する放電データと解析DB上に存在する解析コードで計算されたデータという，本来全く異なる構造の DB 上に存在するデータを容易に取得し，同レベルで解析することが可能となる．

3.3.3 ユーザによるデータの格納

JT-60SA では，放電データや運転データ等のルーチンによって自動的に格納されるデータの他，ユーザが解析したデータ等の人為的に生成されたデータが存在する．これは主に平衡 DB と解析 DB が該当する．

3.1節で解説されたように，平衡DBにはユーザが解析する特定のショット番号／時刻において，MHD 平衡計算コード（SELENE）[3]を用いて，Grad-Shafranov 方程式を解いて得られた磁束分布データが格納される．**図 1** で平衡 DB は解析サーバの外に配置されているが，この DB には，圧力分布と磁束分布の関数形を仮定し，さらに磁気軸における安全係数を 1 と仮定した条件の下で Grad-Shafranov 方程式を解いた平衡データが格納されており，基本的に誰がデータを作成しても同じ結果となるためにデータをユーザ間で共有している．一方で，厳密には平衡 DB の一部のデータは解析サーバ内にも配置されており，Grad-Shafranov 方程式を解く上で上記の仮定を変更する等した場合には，解析サーバ内のユーザディレクトリ上に平衡データが格納される．しかし，先に述べた通り，ユーザインターフェース上には仮想的な DB 構造を配置しているため，ユーザ側からは平衡データの実体が解析サーバの内側か外側かは全く意識する必要がない．

解析 DB には，eDAS や各種解析コード類で計算された結果等が格納されるが，このとき解析 DB 用のデータアクセスライブラリを用いることで指定のフォーマットに統一している．解析 DB は FORTRAN のネームリスト形式のフォーマットを採用しているが，解析 DB 用のデータアクセスライブラリは C 言語や Python 等の言語にも対応したデータアクセスを可能にしている．

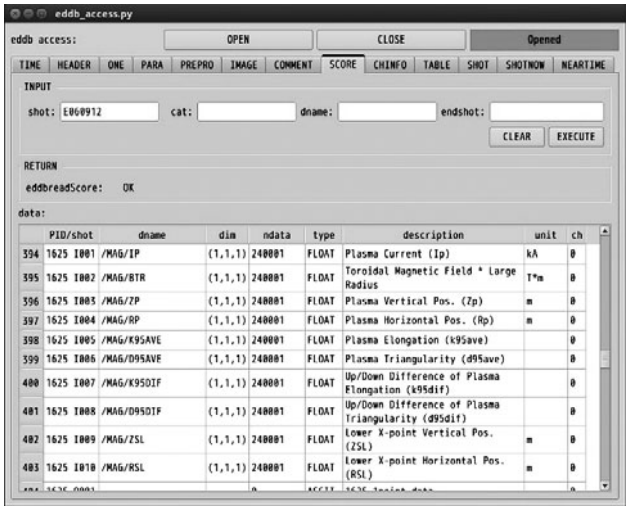

図 2　JT-60SA における DB 構造を GUI で可視化した実験データの例．各データが 9 桁の PID 番号（スペースを含む）で整理されており，それぞれに「カテゴリ名／データ略称名（= dname）」が割り当てられている．

3.3.4 eDAS を用いたデータ解析

JT-60SA では上述のように複数の異なる種類の大容量データを扱うため，データ解析の目的は多様にわたることは容易に想像できる．したがって，eDAS は十分な機能の拡張性を有する設計の下で開発されている．特に DB アクセス用のオブジェクトは各 DB へのアクセスライブラリを含むプラグイン形式で開発されているため，DB の仕様変更があった場合には，このオブジェクトをアップデートするだけで全てのツールが利用できる．プラグイン形式の最大の利点はソフトウェアの拡張性であり，将来的な機能拡張に加え，ユーザ定義のオブジェクトを容易に取り込むことができる．

図 3 に，eDAS の放電波形描画ツール（eGIS）のユーザインターフェースを示す．eGIS は，解析するショット番号，データ名等を GUI 上で指定し，DB からデータを取得してグラフ描画を行う基本的なツールである．単に取得したデータを表示するだけでなく，例えば複数の実験データからグリンワルド密度の時系列データを作成・プロットするなど，簡単な演算を行うこともできる設計になっている．**図 3** では，MHD 揺動解析のために磁気プローブ信号をスペクトル分析している様子を表している．

前述の通り eDAS はデータの種類に応じた様々な DB にアクセス可能なため，実験 DB 上の放電データと輸送解析コード等で計算した解析 DB 上のシミュレーション結果をグラフで比較することもできる仕様になっている．また，eDAS は，実験DBから呼び出した多チャンネル計測器の放電データ（各チャンネルの時系列データの他，ヘッダ部に各チャンネルの計測視野情報が格納されている）を平

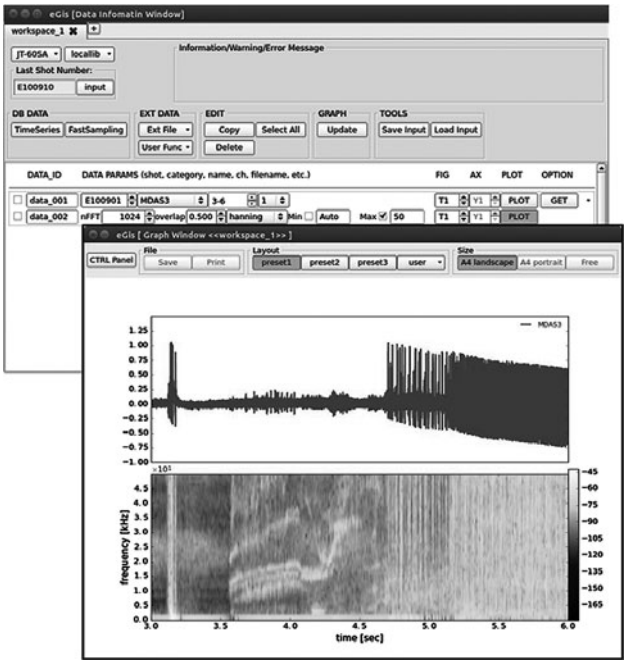

図 3　eDAS の放電波形描画ツールのユーザインターフェース．磁気プローブ信号から MHD 揺動解析を行っている例．

衡 DB から読み込んだプラズマの磁気面上にマッピング処理し，それを空間分布データとして解析 DB に格納した後，いくつかの解析コード用の入力ファイルを作成する機能を有する．

3.3.5　まとめ

本節では，JT-60SA の DB アクセスのためのインフラの整備について簡単に述べた．JT-60SA のようにデータの種類に応じた複数の DB を配置している場合，データ解析のために可能な限りシンプルなデータアクセスのインフラを整備することが要求される．

JT-60SA では，各 DB へのアクセスについては，C 言語や FORTRAN で記述された関数群をライブラリ化してユーザに提供している他，基本的なデータ解析を容易に行うことができるようにデータ解析ソフトウェア（eDAS）を整備している．各々が構造の異なる全ての DB システムへのアクセスにおいて，ある程度共通化されたフォーマットを採用するように設計しており，ユーザが独自にプログラミングする場合でも DB アクセスに対する敷居はできるだけ低く保つように心がけている．また，このように複数の DB を利用するような環境であっても，DB アクセスのユーザインターフェース上にはシンプルな DB 構造を仮想的に持ち込むことで，ユーザ側からは複雑なデータ構造を意識せずにデータ解析を行えるように工夫している．さらに将来的には，Windows Explorer に似た GUI を実装することで，データ名称の変更やユーザ独自のカテゴリ作成，ショートカットの利用等，この仮想的な階層構造を各ユーザが任意にカスタマイズできる仕様にすることを検討している．

JT-60SA におけるデータ解析の目的は多岐にわたるため，eDAS は機能の拡張性を想定して設計している．研究活動の進展に伴って，解析手法が変化する他，各種の DB や機器類，そして実験装置自体も改良されていくため，ソフトウェアはそれらの変化に対して柔軟に対応できる仕様が求められる．eDAS は，DB アクセスモジュールを含めて，このような機能拡張に関するオブジェクトをプラグイン形式で採用する設計になっている．また，一般ユーザが利用する汎用ソフトウェアの場合は，イベント管理を widget で構成された GUI として実装し，特にデータアクセスやグラフ描画をインタラクティブな操作で行えるように開発することが重要である．

最後に本章を執筆するにあたり，デジタルサーブ株式会社の阪本三姫氏には，JT-60SA のための eDAS 開発及び DB 開発における技術スタッフとして協力していただいた．感謝いたします．

参 考 文 献

［1］H. Shirai *et al.,* Nucl. Fusion **57**, 102002 (2017).
［2］浦野 創：プラズマ・核融合学会誌 **94**, 216 (2018).
［3］T. Takeda *et al.,* JAERI-M 8042 (1979).

J. Plasma Fusion Res. Vol.95, No.5 (2019) 223-228

小特集 プラズマ実験におけるデータ解析インフラストラクチャ

4．レーザーエネルギー学実験データベース SEDNA の整備と利用

4. Development and Usage of Experimental Database for High Power Laser System

藤岡慎介，石田正人，奥　浩行，久保田良典，阪本雅昭，谷口麻梨香，
永田みず穂，橋本賢子，山ノ井航平，長友英夫，白神宏之
FUJIOKA Shinsuke, ISHIDA Masato, OKU Hiroyuki, KUBOTA Yoshinori,
SAKAMOTO Masaaki, TANIGUCHI Marika, NAGATA Mizuho, HASHIMOTO Masako,
YAMANOI Kohei, NAGATOMO Hideo and SHIRAGA Hiroyuki
大阪大学レーザー科学研究所

（原稿受付：2019年１月15日）

大阪大学レーザー科学研究所は，大型レーザー装置を用いた高エネルギー密度科学研究の共同利用・共同研究拠点である．大型レーザー装置を利用した研究は，その限られた実験期間とショット数故に，作業の効率化の為，グループを横断したチームによって進められる．実験の準備，装置運転の指示，実験結果，発表資料，論文など，様々なタイミングで，グループ内及びチーム内での情報共有が必要になる．多種のデータを多様なレベルで共有できるシステムとして実験データベース SEDNA（System of Experimental Database for NAtional users' laser facility）を整備した．SEDNA は，研究成果への疑義が生じた場合のデータ追跡機能も有しており，捏造を予防する役割も担っている．

Keywords:
experimental database, high power Laser, data sharing

4.1　はじめに

大阪大学レーザー科学研究所（レーザー研）は，2006年に全国共同利用施設化して以降，実験データの保管と管理及び活用するためのシステムを，手探りながらも整備してきた．これには二つの動機がある，一つはポジティブな動機であり，レーザーエネルギー学研究センター（当時）の全国共同利用施設化によって，世界中の研究者がデータを保管し，活用し，そこから成果を発信するという理念が生まれ，その理念を実現するための仕組みとして，実験データベースを構築することが必要になったことである．

もう一つは相対的にネガティブな動機である．厳正な研究倫理に基づく研究遂行が求められるようになり，データ・解析のトレーサービリティーの管理を強化する必要に迫られた．国内では STAP 騒動，世界的には米国ベル研究所で起きたシェーン事件がデータ管理の強化に繋がったと考えられる．両者に共通するのは，「計測器の物理的な特性に起因する補正処理含む，一切のポストプロセス処理が行われていないデータ」，いわゆる生データの不在である．以来，生データの管理が厳しくなったことは，読者の多くが認識されていると思う．

レーザー研では，キャトルアイサイエンス社の研究フロー統合管理システム RCM（R&D Chain Management）System[1]をベースに共同利用レーザー施設実験データベースシステム SEDNA（System of Experimental Database for NAtional users' laser facility）を開発し，運用している．

4.2　実験データベース SEDNA 構築の背景

SEDNA を構築する以前の状況を最初にまとめたい．当時，データにアクセスできるのは，レーザー研内からのみであり，完全に部局内で閉じたシステムであった．また，USB スティックや光学記憶媒体を用いたデータ受け渡しが横行しており，データのやり取りを全くトレースできない状態にあった．その上，様々な指示・連絡がメールベースで行われ，指示書はメール添付で届き，最新のドキュメントが一体どれなのかを見通し難い状態であった．フォーマットの定まっていない重要な指示が複数あり，曖昧な指示による間違いや，再確認を要するなど非効率な場面もあった．また，大量の紙の配布物など，資源の無駄使いもあった．

SEDNA がめざしたものは，世界中の共同研究者が所属研究室で実験提案・データ解析ができる環境を作ることである．また実験に関わるプロセスを規格化することで，不慣れな共同研究者が犯す間違いを予防することも重要であると考えた．そして，オンライン化によって紙媒体での情報共有を減らすと共に，いつでも最新の重要情報にアクセスできるようにすることをめざした．その為にも，安心してデータ及び解析結果を保管できるセキュアな環境を構築

Institute of Laser Engineering, Osaka University, Suita, OSAKA 565-0871, Japan
corresponding author's e-mail: sfujioka@ile.osaka-u.ac.jp

ⓒ2019 The Japan Society of Plasma Science and Nuclear Fusion Research

すること．更にはデータ，データの流れ，解析の経緯を保管・管理できるようにして，漏洩や捏造の予防を行う必要があった．

4.3　研究フロー統合管理システム RCM System Software

SEDNA は，キャトルアイサイエンス社の RCM System Software 上に構築されている．RCM は，基幹部分に XML（eXtended Markup Langrage）を採用し，データベースからユーザインターフェースを XML をベースに構築することが可能である．RCM には簡易の XmlViewer が付属しており，XML で記述されたデータベースを直感的，フレキシブルに閲覧し編集することが可能である．XML を処理するための XSLT も利用可能で，データベースを目的に応じた形で再構築することが可能である．

権限設定の機能を有しており，公開範囲を任意に定めることで，個人の研究ノートからチーム内での情報共有，広く情報公開という形態をセキュアに実現することが可能である．全てのプロセスはログとして保管されており，過去の作業をトレースし，研究の再現性と信頼性を向上することができる．更に RCM を基盤に ASNARO[2]というプラットフォームも導入しており，ユーザフレンドリーな GUI による操作のみでデータベース処理が可能になっている．

4.4　大型レーザー装置実験の特徴

大型レーザー装置実験は，大きく四つのコンポーネントから構成される．一つ目は「レーザー」，二つ目はレーザーが照射される「ターゲット」，三つ目はターゲットから生成される「プラズマ」，四つ目は計測結果の背後にある物理を明らかにする「シミュレーション」である．

図 1 に示すように，複数のグループが組み合わさったチームによって実験は遂行される．グループは目的及び役割を単位として構成される．レーザーの運転を担うレーザー運転グループ，ターゲットの製作と検査を担当するターゲット製作グループ，プラズマ計測を支援するプラズマ計測支援グループ，研究代表者と研究協力者で構成された研究グループに大きく分類される．研究グループは研究課題毎に独立に組織されるため，チーム内には複数の研究グループが存在する．レーザー条件が似通った研究グループが集まり，そこにレーザー運転グループ，ターゲット製作グループ，プラズマ計測グループが加わって構成されるのがチームで，数ヶ月の単位で実験を遂行することで，作業の効率化を図っている．

4.5　大型レーザー装置実験におけるデータ共有

図 2 に大型レーザー装置実験のデータベースの概略を示す．「ターゲット」のデータは，レーザーショットの最短でも 1 ヶ月前から入力が開始される．「レーザー」のデータはショットと同時に生成され，「プラズマ」のデータは計測器から制御用のコンピュータにアップロードされることで保存される．「シミュレーション」はレーザーショット終了後，数ヶ月のタイムスケールで生成される．このように，各コンポーネントのデータはそれぞれ異なる時刻に生成され，データベースに蓄積される．そこで，レーザーを発射するたびに割り振られる「ショット番号」を共通の ID として利用し，四つのコンポーネントのデータを連携さ

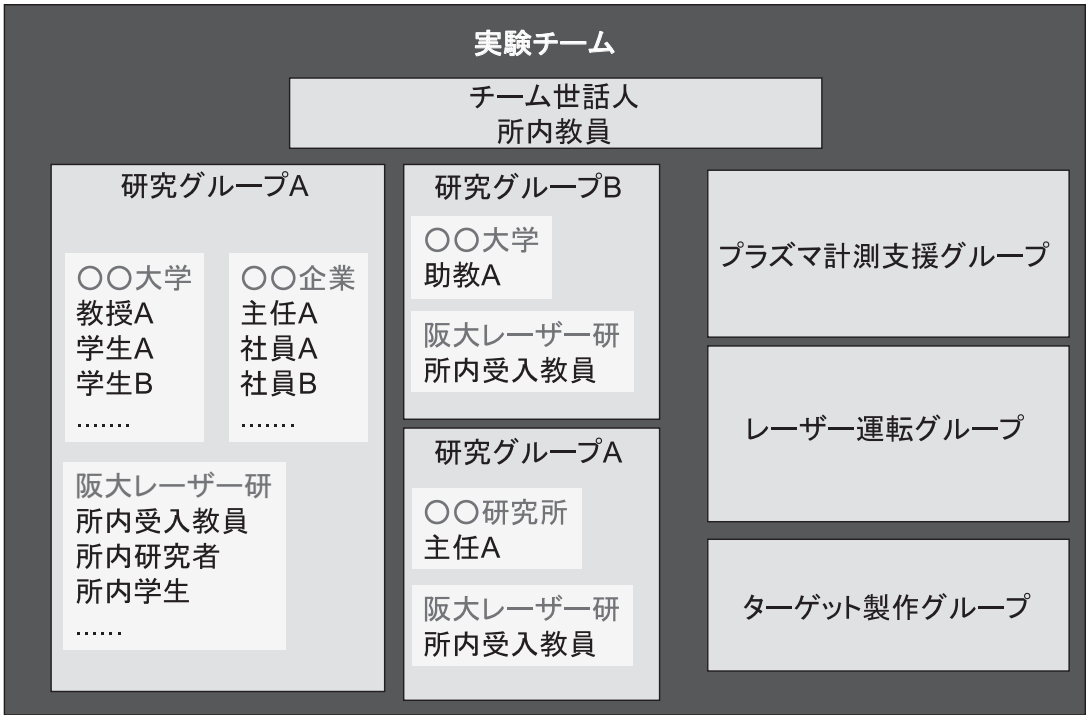

図 1　大型レーザー装置実験におけるチーム構成の概要図．チームは，世話人（所内教員）の下，数ヶ月の単位で活動する．レーザー条件や計測条件が近い複数の研究グループと，プラズマ計測支援グループ，レーザー運転グループ，ターゲット製作グループの連携で実験が遂行される．

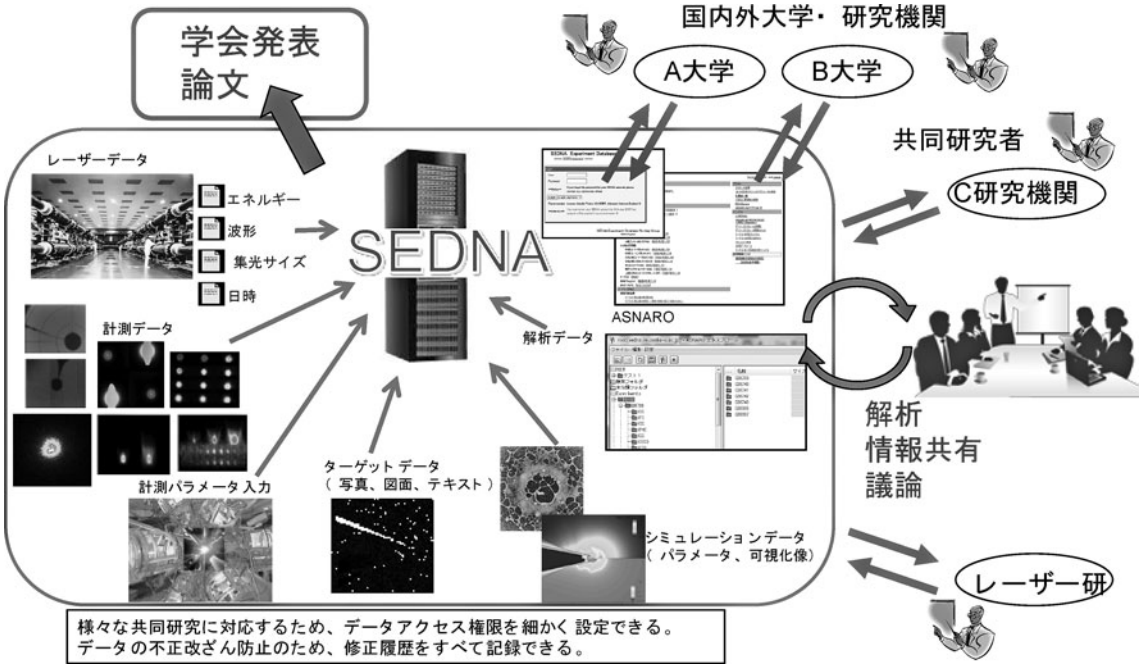

図 2　実験データベースの概略．レーザーデータ，計測データ，ターゲットデータ，シミュレーション，解析データ等を SEDNA に集約し，国内外の研究者らと共有する．

せる仕組みとしている．

「ターゲット」は少なくも実験の 1 ヶ月前に，実験グループからターゲット製作グループに発注される．ターゲット製作グループは，発注書に基づいてターゲットを製作するため，発注書には必要な情報が不足無く含まれなくてはならない．SEDNA ではターゲット発注書を入力するフォームを用意し，ターゲットに使用する部品の仕様，個数を明記し，部品の組み立て方法を不足なく入力できるようにしている．この発注書は実験の実務にあたる学生が作成することが多く，それ故に考慮不足や誤解による間違いは避けられない．ターゲット製作グループに提出される前に，所内の実験受入担当者がチェックし間違いが無いことを確認した後，ターゲット製作グループのグループリーダーに届くようにしている．ターゲット製作グループのリーダーが製作の可否を判断した後に，ターゲット製作グループの技術職員に送られる仕組みとしている．この時，ターゲットの種類毎に呼称が与えられ，この呼称によって発注書は管理される．

ターゲット製作を担当する技術職員は，部品を製作し，組立を行い，組み上がったターゲット毎にターゲット番号（6 桁の数字）を振っていく．ターゲットに使用した部品や組立後のターゲットの写真や寸法が，ターゲット番号と共にデータベースに保存される．またターゲット番号と発注書の呼称が関連づけられ，ターゲット番号さえわかれば，その発注書及び検査成績書がいつでも閲覧できるようになっている．

レーザーショットが始まると，実験グループとレーザー運転グループの連携が必要になる．レーザー運転グループは数十名の技術職員で構成されており，実験グループも多い時は数十人の規模になる．数十人が効率的に動くために，スケジュールの管理が必要である．一日のショットスケジュールを，使用するターゲットとレーザー条件と共に，デイリースケジュールという形で所内のネットワーク上で公開している．実験関係者は，デイリースケジュールを見ながら，実験の準備を進める．このデイリースケジュールは実験データベースの柱となるものである．デイリースケジュールはレーザー条件の要約のみが表示されており，その背後にはショットシートと呼ばれる，詳細なレーザー条件が記載されたデータが存在している．これは実験ノートに類するものである．また使用するターゲットは上記のターゲット番号と共に記載されている．ショット後にショットシートとショット ID が関連づけられ，更にターゲットとショット ID も関連づけられる．

ショット後に自動的に，プラズマ計測データを保存するフォルダが生成される．フォルダ名にはショットIDが使われており，このフォルダの中にはよく使用される計測器の略称が書かれたサブフォルダがある．実験者は，ショット後にショット ID が書かれたフォルダの担当する計測名が書かれたフォルダにデータをアップロードする．このようにすることで実験データとショット ID 及び計測器名が関連づけられる．実験後，特定の計測器で得たデータを一括でダウンロードできるようにもしている．

更に，研究チーム毎に学会発表資料や論文を共有できるようにもしている．研究チームには課題番号が与えられており，課題番号毎にデータを共有できるようになっている．またショットIDのフォルダに校正や解析を施したデータを保管することも自由に可能である．

4.6　データ解析手法の共有

大型レーザー装置実験は極めてショットレートが低い．激光 XII 号レーザーは一日で 5 ショット程度（使用するビームの組み合わせによって 2 倍以上増やすことも可能）

であり，LFEX レーザーでは一日 4 ショットが最大である．それ故に，単一の解析方法で大量のデータ処理をするという経験が少なく，なかなかデータ解析手法の共通化が図りにくいというのが実状である．MacとWindowsが混在しているというのも解析の共通化を妨げる要因である．

しかしながら，最近では Python，Matlab 等を用い，GUI 環境を備えた解析プログラムを書ける，またはそのような解析に興味がある学生が増加傾向にあり，OS 等のプラットフォームに依存しない解析法の確立が期待できる．RCMは，GUI環境で解析データとコードを呼び出し処理する機能を有しており，この機能を活用した研究が必要である．

パワーレーザーコミュニティとして，「パワーレーザーインテグレーションによる新共創システム J-EPoCH 計画」を提案している[3]．セラミックレーザー及びファイバーレーザーを活用したパワーレーザーの高繰り返し化技術と，AI を活用した多ビーム結合技術によって，キロジュール級のレーザーを繰り返し動作することが可能になった．そこに，プラズマデバイスによる多機能化が加わり，学術開発の加速，新学術の創生，そして産業イノベーションの創出が期待されている．レーザー学会「ハイパワーレーザーによる高エネルギー密度科学」調査研究委員会のもとで様々な分野からの要求仕様をまとめ，装置設計にフィードバックをかけている．

このような高繰り返しレーザーが生み出す大量のデータを処理するためには，各種データの自動取得，データの予備解析の自動化，ディープラーニングを活用したデータの一次選別が欠かせない．RCM は大量データ処理にも対応可能なシステムであり，今後は SEDNA をシングルショットベースのシステムから，繰り返しショットへの対応可能なシステムへの進化を遂げる必要がある．

4.7　データ共有駆動型の共同研究の重要性

我が国の科学技術分野における国際競争力の衰えが危惧され，例えば発表論文数は主要国内で唯一減少傾向にあることがマスメディア等でも盛んに指摘されている．このような憂慮すべき事態の打破には，研究のスタイルの変革が求められる．

実験分野においては現場至上主義，別の言い方をすると，現場での様々な辛酸を共になめた仲間内での協力が過剰に重視されていると感じている．著者自身，データの産みの苦しみは何度も経験しているし，大型装置実験は連携で支えられていることは重々承知している．しかしながら，研究・教育を取り巻く環境は急速に変化（悪化？）し，長期間に亘って現場に研究者又は学生を配置することができる，人的，金銭的，時間的余裕を持つ研究者はごく一部に限られているのは，ご承知の通りである．現場至上主義の下で，取得データを延々に限られた仲間内で占有することは，最善の策ではないと考えられる．データ占有期間を明確に定め，期間終了後は解析能力を有しながらも，リソースが限られる研究者に参加を促し，研究ネットワークの広がりを図る必要があるのではないか？という問題提起をさせていただきたい．

例えば，高エネルギー物理学や天体観測の分野では，取得したデータを共有することで，実験現場に参加できなかった研究者や，実験を専門としない研究者に対しても，

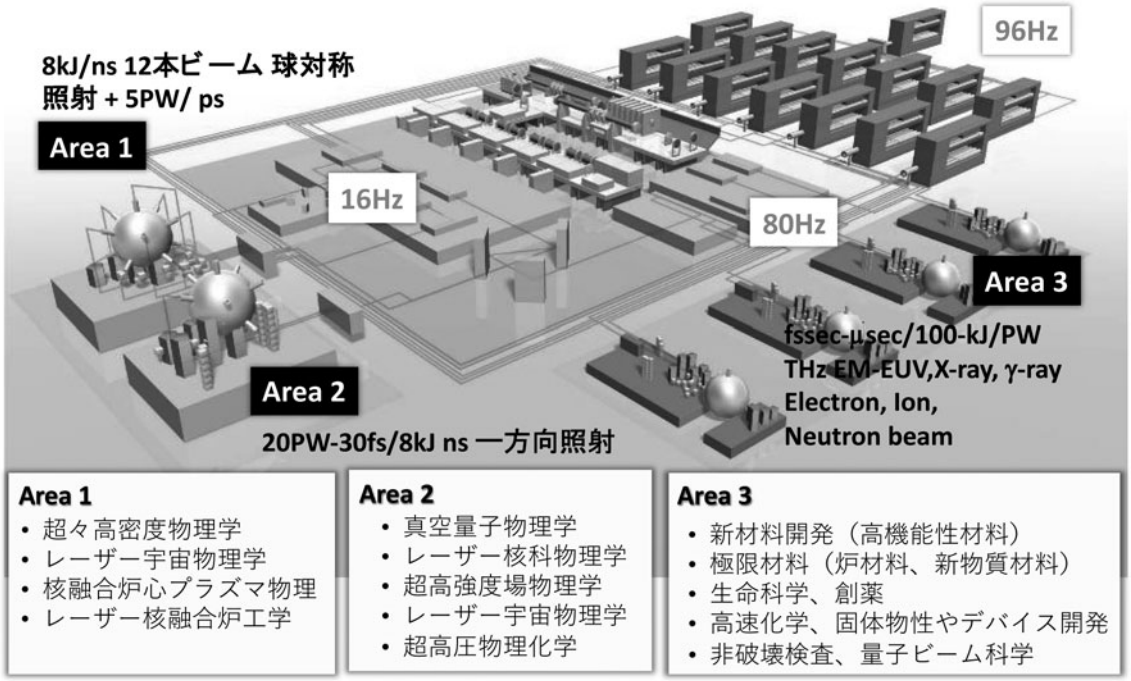

図 3　パワーレーザーインテグレーションによる新共創システム　J-EPoCH 計画．繰り返しレーザーを中心に，学術開発，新学術創出，産業イノベーションの創出をめざす．高繰り返しレーザーが生む大量のデータを処理するためには，それに対応したデータベースシステムの構築が不可欠である．

研究の機会を提供している．プラズマ分野においても，データ共有を中心におき，データ駆動型の共同研究の促進が求められていると考えている．我々も予算は限られているが，データ共有の理念を掲げながら，システムを構築していきたいと思っている．

データ公開を前提にせずとも，データの適正な管理は，取得したデータを活用するために不可欠である．個人管理のコンピュータの中で，データを保管し，処理し，解析するような方法では，どんなに貴重なデータであっても，数年後には全く素性のわからないゴミデータとなってしまいかねない．データの適正な管理のためにも，データベースの整備と活用は必須である．

図 4 はデータ公開の過程を図示したものである[4]．実験データベースに蓄積されるデータの大半は生データである．これを解析可能なレベルに加工したデータを高レベルデータと呼ぶ．実験提案者は，高レベルデータを優先的に利用する権利を一定期間有する．一定期間終了後は，データアーカイブスとして公開し，アーカイブスユーザとして登録された研究者にも解析する権利を付与する．これは，天文学分野では既に実現しているスタイルである．

琉球大学の野崎真也准教授による，データ共有による成果を一つ紹介させていただきたい[5]．この研究では，大量のプラズマ X 線画像を主成分分析法を用いて統計処理した．ノイズは非主成分であるため，この処理によって，ノイズのみを選択的に除去することが可能である．**図 5** (a) は，信号対ノイズ比の悪い信号である．**図 5** (b) はノイズ除去に一般的に用いられているウィナーフィルター（Low Pass フィルターの一種）を用いて，ノイズ除去を行った．ウィナーフィルターでは，高周波を多く含むノイズを除去できているが，信号とノイズを弁別できないため，信号値の高周波成分も同時に失われ，信号が鈍っているのがわかる．**図 5** (c) は主成分のみを使って再構成した結果である．ごま塩ノイズが選択的に除去され，シャープな信号が得られているのがわかる．

4.8　最後に

教育・研究分野においては，コンプライアンスの強化が推し進められ，ファカルティー・デベロップメントや，研

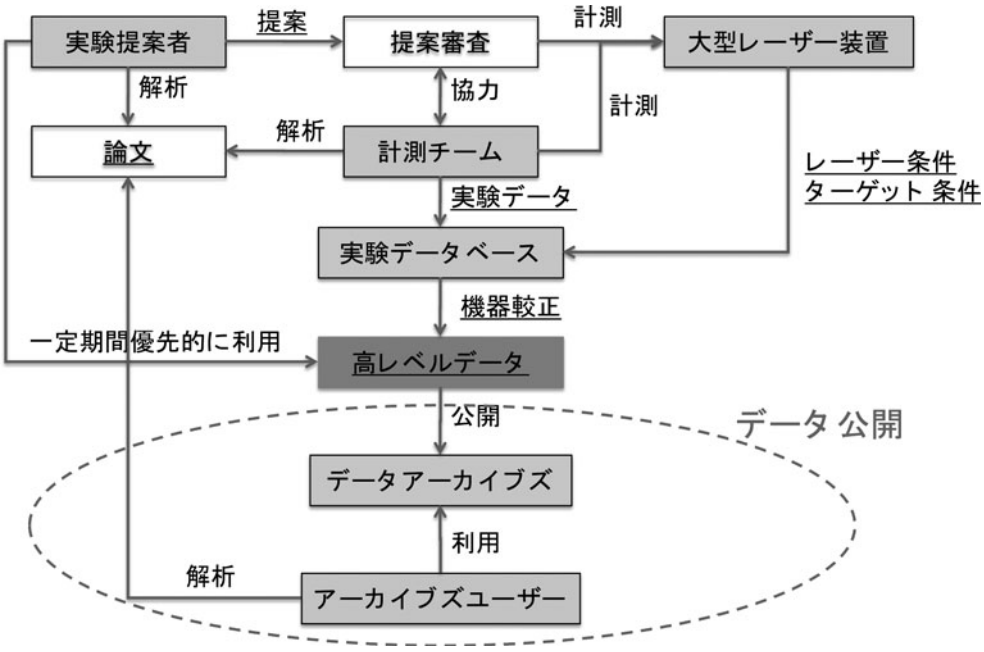

図 4　データ公開に基づくデータ共有駆動型共同研究．

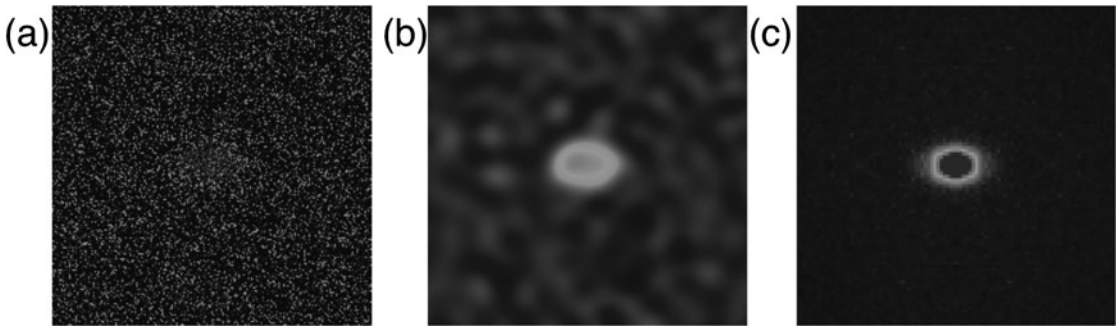

図 5　主成分分析法を用いたデータ処理．大量データを統計処理することで，プラズマ X 線画像を構成する成分を抽出し，非主成分であるノイズを除去することに成功した．(a) 処理に利用した信号対ノイズ比の低い画像．(b) ウィナーフィルター（いわゆるスムージング）処理を施した実験データ．(c) 主成分のみを抽出した画像．

究倫理教育，情報セキュリティ講習，競争的資金に関する教育等々が必須となっている．いかに優れたデータベースを構築しても，それを利用するのは人間である．データ保存の重要性，データ共有の重要性，取得したデータの二次活用を意識した保存方法などの教育も必要であると考えられる．

謝　辞

本開発の一部は大阪大学レーザー科学研究所の共同利用・共同研究拠点事業の支援の下で実施された．SEDNAの開発にあたっては，森尾登氏と退職された西原功修名誉教授，福田優子氏，島田京子氏及び山田研二氏の協力が不可欠であった．システム開発に当たっては，キャトルアイサイエンス社の上島豊氏，前田茂樹氏から多くのサポート及びアドバイスをいただいた．最後に利用者からのフィードバックにも感謝いたします．

参 考 文 献

［1］ https://www.i4s.co.jp/rcm/rcmabs.html
［2］ https://www.i4s.co.jp/asnaro/asnaroinfo.html
［3］ http://www.ile.osaka-u.ac.jp/ja/research/mp2018/index.html
［4］ 海老沢 研：日本物理学会誌 **63**, 670 (2008).
［5］ S. Nozaki *et al.*, Plasma Fus. Res. **7**, 2404120 (2012).

J. Plasma Fusion Res. Vol.95, No.5 (2019) 229-235

小特集 プラズマ実験におけるデータ解析インフラストラクチャ

5．ITER のデータ解析インフラストラクチャ IMAS

5. Data Analysis Infrastructure for ITER ～ IMAS ～

細 川 哲 成，PINCHES Simon D.，ABADIE Lana
ITER Organization

（原稿受付：2019年 2 月23日）

本章では ITER の建設・運転期間を通じその統合モデリング計画（ITER Integrated Modelling Programme）を支援するための解析インフラストラクチャ IMAS（The ITER Integrated Modelling & Analysis Suite）の概要と最近の開発動向について解説する．IMAS は汎用性・拡張性が高い ITER データモデルを中心として解析コードのモジュール化とワークフローの定義による統合シミュレーションの開発，及び共通なインターフェースを通じて実験・解析データベースの利用を可能とするソフトウェア・インフラストラクチャであり，建設・運転期間で必要となる計測，加熱，燃料供給及び制御システムなどの開発のための運転シナリオや解析ツールを提供する．また，ITER 実験制御・データシステム CODAC における実験データアーカイブシステムの概要と IMAS との関連について紹介する．

Keywords:
ITER, integrated modelling, data structure, physics workflow, database, CODAC

5.1　はじめに

ITER 統合モデリング計画とその実装である IMAS[1,2] の主要な目的は，30年に渡る ITER の建設・運転期間において ITER Research Plan（IRP）[3] を遂行するための解析インフラストラクチャを提供することにある．IMAS インフラストラクチャは，主として ITER を構成する各極（ITER メンバー）から提供される要素コード，炉心プラズマや装置を記述するデータモデル，データアクセスインターフェース，データベース，ワークフローと，それらを ITER メンバーの研究施設で利用可能とするインストーラや各種ソフトウェアツールから成る．ITER 統合モデリング計画では，ITER の建設期における，より正確なプラズマの性能予測や制御システムの開発及び運転計画の作成を支援し，運転期においてはプラズマ制御やモデリングなどに資することが期待されている．ITER プラズマの物理的な理解を促進し理論・実験的な知見を集積していくために ITER メンバー各極との国際協力を通じて IMAS インフラストラクチャの開発が行われているが，同時にそのための広範囲な調整も求められている．

本章では IMAS の概要について，その中心的な構成要素の説明を5.2節で，IMASの統合シミュレーションと運転計画データベースを5.3節で概説する．5.4節では IMAS の開発を進める上で利用されているソフトウェア開発支援ツールについて，また5.5節ではITER実験制御・データシステム CODAC（Control Data Access and Communication）の実験データアーカイブシステムと IMAS との関連について紹介する．

5.1.1　ITER メンバーとの連携

ITER 統合モデリングに関する ITER 機構と ITER メンバーの各極内組織との関係を**図 1** に示す．統合モデリング専門家グループ（Integrated Modelling Expert Group: IMEG）は ITER 統合モデリングの諮問に応じ，ITER メンバーで開発された要素コードの提供，ITER メンバーにおける IMAS のインストールなど，ITER 統合モデリング活動と各極との調整を行っている．ITER 統合モデリング計

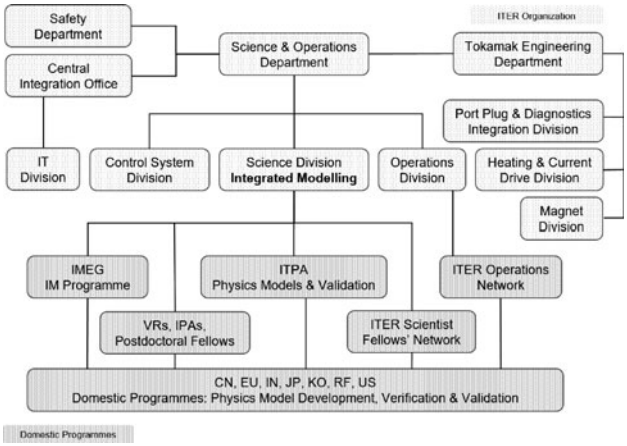

図 1　ITER 統合モデリングは ITER メンバーとの広範囲に渡る協力・調整を要するが，特に IMEG,ITPA 及び ISF ネットワークが主要な役割を担っている．IMEG は ITER 統合モデリング活動とインフラストラクチャ開発を，ITPA は基盤となる物理モデルの開発と実証を，ISF ネットワークは個別の物理モデリングの開発を支援する．

ITER Organization, Route de Vinon-sur-Verdon, CS 90 046, 13067 St. Paul-lez-Durance Cedex, France
corresponding author's e-mail: Masanari.Hosokawa@iter.org

©2019 The Japan Society of Plasma Science and Nuclear Fusion Research

画はITER協定によって規定されており，IMEGは国際トカマク物理活動（ITPA）による基礎研究・物理モデリングの成果と共にそれを補完する形でITERの炉心プラズマ研究を支えている．

ITER統合モデリングはこの他にも客員研究員（VR），ITERプロジェクトアソシエイツ（IPA）やITER博士研究員が参加し，さらにITERサイエンティストフェロー制度（ISFN），ITERオペレーションネットワーク（ION）及びITERメンバーの各研究機関と協調して活動している．ITPAの「高エネルギー粒子物理」グループでは，IMASを利用して複数の加熱装置から発生する高速イオン分布のベンチマークを行い，将来的にはITER運転シナリオにおける高速イオンの安定性評価を計画している．また同「統合運転シナリオ」グループ及び「輸送と閉じ込め物理」グループでは，ITERの運転キャンペーンPFPO（Pre-Fusion Power Operation）初期における定格の1/3の運転シナリオ（5 MA/1.8 T）についてHモード遷移の評価のためにIMASを利用したベンチマーク活動を行っている．

5.2 IMASの構成要素

5.2.1 概要

2011年に開発が始まったIMASインフラストラクチャは現在ITER機構において実装され，ITERメンバーの複数の研究機関で既に利用されている．その設計はthe European Integrated Modelling Task Force（EU-ITM）[4]に由来し，IMASの構成要素の多くがこれを基礎としている．IMASインフラストラクチャで中心的な役割を担うのがITERデータモデルであり，開発者・利用者が別個に定義された要素コードのデータ形式を結合毎に変換する労を軽減し，計算モデルや装置に依存しない標準的なデータ形式を持つデータベースを作成する際に効果的である．要素コードは互いに疎結合モジュールとしてワークフローを構成するが，これは想定する30年の運用期間中の個別技術の進展に対応するとともに，特定の物理モデルや解析コードの変更が他へ与える影響を最小限にするためでもある．

5.2.2 IMASデータモデル

データモデルは一般的にどのようなデータが存在しているかを定義し，そのデータ構造についての情報をデータの作成者・利用者に示すものである．記憶装置上でのIMASデータはMDSplusファイル形式（またはHDF5）を取るが，利用者は炉心プラズマや装置を抽象化したデータモデルを意識すればよく，物理的なデータ形式を気にする必要はない．この抽象化により，ITERデータモデルの適用範囲は解析計算だけでなく，ITERなどの実験データにも広げることが可能となる．ITERデータモデルは解析コード間のインターフェースや実験データを区別なく記述するという目的をもっており，磁場閉じ込め装置を一般化することも視野に入れることができる．

ITERデータモデルは2つの主要な要素に分けられ，まず利用者が直接意識することになる物理的対象を抽象化したData Dictionary（DD），次にDDにマッピングされた解析・計測データをショット・ラン番号に関連付ける機能である．例としては，トムソン散乱計測の解析計算結果と実験計測結果をDDにあるIDS "thomson_scattering"に格納することで，両者を同じデータ構造・名称で比較できることになる．ITERデータモデルは炉心プラズマ研究に資することを目的としており，本質的にその進展に従って変更・拡張されることを前提としている．このため詳細なライフサイクル管理基準が決められていて，複数のグループ共同でデータモデルを修正することなどにも対応している．また，多岐に渡るITERデータモデルはDDをXML言語で記述するが，その変更にあたっては全体的に矛盾が発生しないよう設計ガイドラインが定められている．

Interface Data Structure（IDS）はDDにあって最も粒度が高く，単体として利用できる構造体（複数の異なる型を格納できるデータ型）であり，例としては装置を構成する各設備（コイル，計測，加熱装置）や物理的概念（平衡，プラズマプロファイル）などがある．IDSにより個々のデータの由来を追跡する労が減り，異なるデータ通信を要する様々なモジュール間の結合を容易にする．**表1**にIMASバージョン3.21.0で定義されているIDS名の一覧を，**表2**にIDS: thomson_scatteringのデータモデルの例を示す．

5.2.3 データアクセス

IMASインフラストラクチャではUALのAPI（Application Programming Interface）として複数の言語（FORTRAN，C++，Python，Java，Matlab）をサポートしており，異なるプログラミング言語で記述された物理モデル間の通信や，遠隔データへのアクセスを容易にする．

ITERの運転開始に先立ち，より正確な放電予測を行うためにはIMAS上の解析コードと既存装置の実験結果について比較し妥当性検証を行う必要から，IMAS-UALはIDAM[5]システムに由来する統合データアクセス（Unified Data Access: UDA）技術によって拡張された．UDAは個別のデータ形式をITERのデータモデルに対応付ける（マッピングする）プラグイン形式を採用することで遠隔

表1 IMAS Physics Data Model（ITERデータモデル）v3.21.0で定義されているIDS名の一覧表．DDの拡張・更新は主に新規のユース・ケースやユーザからの要望に従って行われる．

amns_data	edge_transport	pf_passive
barometry	em_coupling	polarimeter
bolometer	equilibrium	pulse_schedule
charge_exchange	gas_injection	radiation
coils_non_axisymmetric	ic_antennas	reflectometer_profile
controllers	interferometer	sawteeth
core_instant_changes	iron_core	sdn
core_profiles	lh_antennas	soft_x_rays
core_sources	magnetics	spectrometer_visible
core_transport	mhd	summary
dataset_description	mhd_linear	temporary
distribution_sources	mse	thomson_scattering
distributions	nbi	tf
ec_antennas	neutron_diagnostic	transport_solver_numerics
ece	ntms	turbulence
edge_profiles	pellets	wall
edge_sources	pf_active	waves

表 2　IDS をウェブブラウザで表示するコマンド“dd_doc”を利用した IDS: thomson_scattering のデータモデル表示例.

Full path name	Description	Data Type	Coordinates
▸ ids_properties	Interface Data Structure properties. This element identifies the node above as an IDS	structure	
▾ channel(i1)	Set of channels (lines-of-sight)	struct_array [max_size=100]	1-1...N
name	Name of the channel {static}	STR_0D	
identifier	ID of the channel {static}	STR_0D	
▸ position	Position of the measurements (intersection between laser beam and line of sight)	structure	
▾ t_e	Electron temperature	structure	
data(:)	Electron temperature {dynamic} [eV]	FLT_1D	1-channel(i1)/t_e/time
time(:)	Generic time [s] {dynamic}	flt_1d_type	1-1...N
▾ n_e	Electron density	structure	
data(:)	Electron density {dynamic} [m^-3]	FLT_1D	1-channel(i1)/n_e/time
time(:)	Generic time [s] {dynamic}	flt_1d_type	1-1...N
▸ code	Generic decription of the code-specific parameters for the code that has produced this IDS	structure	
time(:)	Generic time [s] {dynamic}	flt_1d_type	1-1...N

データにアクセスすることを実現しており，ASDEX Upgrade，JET，MAST，TCV 及び WEST などの実験データベースなどのためのプラグインが用意されている．UDA には FAT または non-FAT の 2 種類の実装モードがあり，non-FAT モードでは UDA クライアントとサーバが互いに遠隔の別計算機上のプロセスとして動作し，FAT モードではクライアントとサーバプロセスが同じ計算機上で実行され，UDA サーバーは実装されない．

UDA では，サーバプログラムとして Unix 系オペレーティングシステムで動作する XINETD（eXtended Internet services Daemon）を利用している．XINETD はインターネット上でのクライアント・サーバ型のデータベース管理を可能とし，inetd に比べてよりセキュリティが確保されている．加えて UDA では X.509証明書によるアクセス認証にも対応している．**図 2** は UDA を利用する Python クライアントプログラムの例で，WEST の遠隔実験データを IDS "magnetics"として取得している．

ITER 実験制御システム CODAC では，一日当たり最大で数 PB の実験データを蓄積することが求められるが，構成の異なるデータ記憶装置への統一的，かつ柔軟なアクセス方法として UDA を利用する開発が進んでいる（5.5節）．

5.2.4　ワークフロー

ITER データモデルによるインターフェースを持った要素コードはワークフローツール上で統合することができるようになるが，現在 IMAS インフラストラクチャはワークフローツールとして Kepler[6]と Python をサポートしている．同じプログラミング言語で記述される要素コードの場合，例えば主プログラム内で関数副プログラムの呼び出しをするように比較的容易に結合することができるが，異なる言語で記述されていたり，プログラム間で疎結合が要求される場合にはこれらを共通のワークフローツール上で動作するモジュールに変換する必要がある．IMAS インフラストラクチャではこのための変換プログラム"FC 2 K"を用意しており，要素コードをワークフローツール上で動作する 3 種類のファイル(1)単独実行ファイル，(2)Kepler Actor，(3)Python関数を自動生成する．**図 3** に Kepler を用いた IMAS ワークフローのスクリーンショットを示す．

```
>>> import imas                          <- IMAS モジュールのインポート
>>> imas_entry=imas.ids(53223,0)         <- Shot/Run 番号設定
>>> imas_entry.open_public('WEST')       <- UDA プラグイン指定
>>> imas_entry.magnetics.get()           <- IDS 取得
imas::get(expName='WEST', idx=0, group='magnetics', variable=
'ids_properties/comment', type=string, rank=0, shot=53223, )
...
```

図 2　Unified Data Access（UDA）を利用する Python クライアントコード例．ITER クラスタ上で WEST の実験データを IDS “maganetics”として取得している．

5.3　IMAS アプリケーション

5.3.1　統合シミュレーション

鍵となる支配的な物理現象や装置を包括的に含んだ統合シミュレーションを行うことは，より正確な運転計画を作成する上で重要である．自由境界平衡コード DINA[7,8]と統合モデリングコード JINTRAC[9]は，データ交換にそれぞれ IMAS-DD を採用しているが，この目的のために IMAS 上で DINA+JINTRAC の統合シミュレーションが行われた[10,11]．

本ワークフローでは，中心部からプラズマ対抗壁までを含む空間スケールで，ポロイダル磁場コイルやダイバータ板への熱流束といった各装置の工学的な許容条件を満たしながら，運転シナリオの最適化（プラズマ位置形状制御法，電流駆動，加熱や燃料供給など）による定格運転シナリオ（15 MA/5.3 T）を実証した．X 点形成のランプアップ初期から X 点がリミター配位へ遷移するランプダウンまでの全放電時間について，自由境界条件の下に自己無撞着なペデ

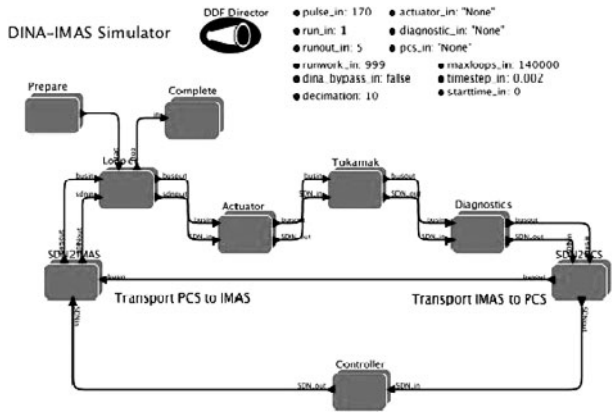

図 3　プラズマ制御システム開発支援のために DINA コードを利用したトカマクシミュレーター（Kepler ワークフロー）の作成例．

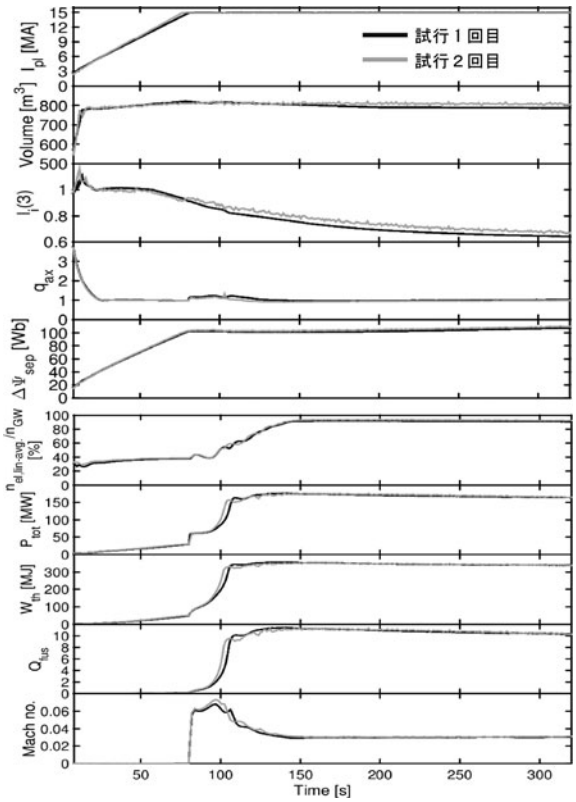

図 4　ランプアップからフラットトップ初期までの DINA+JINTRAC IMAS ワークフローによる DT 定格運転シナリオ（15 MA/5.3 T）の計算結果比較，試行回の違いは，放電立上げ間隔調整（80→75秒）と，立上げ後の H モード遷移までの時間変更による．

スタル MHD 安定性解析による圧力を考慮して検証したのは初めての試みだったと考えられる．**図 4** に DINA+JINTRAC ワークフローによる Q＝10定格運転シナリオ（15 MA/5.3 T）のうち，ランプアップからフラットトップ初期までの時系列図を 2 ケースの施行回毎に示す．本計算では両コードの計算結果は 1 %以内で一致し，シナリオが異なる 2 ケースの計算については約 1 %の誤差で解が早期に収束している（計算時間は ITER クラスターで約 2 週間）．

本統合計算についてはさらに，HELENA[9]，MISHIKA[10] コードを用いてペデスタル圧力勾配の MHD 限界を考慮する安定性解析を行う改良が計画されている．一方，中心，周辺および SOL プラズマの輸送コード，JETTO[11]，SANCO[12]，EDGE2D-EIRENE[13-15] を統合したワークフローについても，DINA コードからの平衡データを利用する改良が行われている．DINA+JINTRAC IMAS ワークフローではさらに，燃料補給により生じる中 Z 不純物の中心プラズマにおける蓄積が，フラットトップにおけるタングステンの蓄積を防ぐ可能性について報告されている[16]．

5. 3. 2　シナリオデータベース

ITER の定格放電予測の統合シミュレーションの進捗について前述したが，ITER メンバー各極の Domestic Agency（DA）や研究機関と解析結果を共有する場合，IMAS-DD を共通のファイル形式とすると効果的と考えられる．2016年の ITER Council（IC）において ITER の段階的アプローチの採用が決まったことを受けて IRP が改定され，各運転フェーズにおける複数の運転シナリオ（燃料の種類，閉じ込めモードなど）の解析結果を共有・保存するために IMAS データベースが開発されている．このデータベースでは，既存の計算結果を IMAS データフォーマットに変換すること，及び新規にワークフローを実行して直接 IMAS データファイルを作成することなどでデータベース化を進めており，ASTRA[17]，CORSICA[18]，DINA[19]，JINTRAC[9]及び METIS[20]コードが利用されている．IMAS データモデルとして存在しないデータ構造などが明らかとなった場合（例としてはコレクティブトムソン散乱計測で必要となる高速イオン分布関数など）には，DD の改定が必要に応じて行われている．

IMAS データベースには現時点で IRP にあるすべての運転フェーズのシナリオである定格の 1/3(5 MA/1.8 T)，定格の 1/2（7.5 MA/2.65 T）及び定格運転シナリオ（15 MA/5.3 T）が用意され，プラズマの平衡データと中心プラズマのプロファイルが少なくとも共通に利用できるようになっている．それらの間となるシナリオデータに加えて，プラズマ電流を増大する 17 MA，長時間定常放電（9 MA/5.3 T）及びハイブリッド運転シナリオ（12 MA/5.3 T）もデータベース化された．

IMAS データベースではユーザを支援するためのツールが複数用意されており，データの一覧を取得してフィルタリング結果を表示する Python スクリプト "scenario_summary"（**図 5**），2 次元平衡データや中心プラズマプロファイルを可視化する Python スクリプト "equiplot"，"kinplot"（**図 6**）などがある．また，個別のデータに変更があった場合に電子メールで知らせるアラート機能も追加された．IMAS データベースは ITER に資するすべての方がアクセスし，シナリオデータの可視化や IMAS ワークフローへの入力などに利用可能となっている．利用者によって作成されたデータを IMAS データベースへ追加する場合には，受領に際してデータの検証やその目的などが審査される．

5.4　IMAS インフラストラクチャの開発

各極に地理的に分散するIMASユーザ・開発者が課題を共有し問題解決していくために，Webベースの課題管理システム JIRA が利用されている．提起された課題はチケットとして共有され，その種類に応じてタスクとして担当者へ自動的に振り分けられる．現時点では，総計で約2000件の課題が報告され，約75%が ITER と ITER メンバーの担当者の協力で解決されている．**図 7** では過去 4 年間で提起された JIRA の課題数の推移と，問題の種類を示す．プロジェクトの管理という観点からは，JIRA は各タスクの進捗状況や消化率などをWeb上で可視化し，タスクを適切に配分することなどを容易にする．

IMAS インフラストラクチャでは，それ自体の移植性を高めてITERメンバーの研究機関におけるUnix系計算機で動作するようインストール過程を簡素化しており，主要なコンポーネントについて"configure - build - install"などの標準的な方法でインストールが可能である．ITER クラスターでは，動的環境切替を行う"Environment Modules"（Module コマンド）がコンパイラ環境や特定のバージョンのソフトウェアへの切替・選択などで利用されており，最新の IMAS インフラストラクチャ環境は，

```
$ module load IMAS
```

というコマンド入力で設定できる．一方，IMAS インフラストラクチャのインストール自体も Module コマンドによって必要なソフトウェア環境を設定するので，最初に IMAS インフラストラクチャのインストールを試みる場合，前提条件となるソフトウェアが Module コマンドを通して利用できる必要がある．

IMAS インフラストラクチャでは分散型バージョン管理システム Git[21]をベースとする Web ツールを利用して要素コード，ツール，インフラストラクチャ自身のソースプログラムとその履歴を集中管理しており，開発ブランチ毎にアクセス権限をコントロールしたり，ソースプログラム編集時（プルリクエスト）のレビューなどに利用されている．また，このシステムと連動する継続インテグレーション（Continuous Integration: CI）システムを併用し，前述の変更があった場合に自動的にコンパイル，ビルド，単体・回帰テストを行ってその品質の保全を図っている．

5.5　ITER 実験データの保存と利用

CODAC は全部で220を数える計測装置と制御システムから成る全系制御システムで，駆動装置，センサー及びプラズマパラメータ計測制御に関わる全計測・制御機能を統括する．CODAC の主要な目的の一つは，大規模な ITER の実験データの集積と，データアクセス及び可視化方法を提供することであるが，これはそれぞれ異なる目的と実装

Pulse	Run	Machine	Reference	Ip[MA]	BO[T]	Fuelling	Confinement	Workflow
100001	1	ITER	ITER-full-field-H	-15.0	-5.3	H	L-mode	METIS
100002	1	ITER	ITER-half-field-H	-7.5	-2.65	H	L-mode	METIS
100003	1	ITER	ITER-third-field-H	-5.0	-1.8	H	L-H-L	METIS
100007	1	ITER	ITER-intermediate-3T-H	-8.5	-3.0	H	L-H-L	METIS
100008	1	ITER	ITER-intermediate-3.3T-H	-9.5	-3.3	H	L-H-L	METIS
100009	1	ITER	ITER-intermediate-4.5T-H	-12.5	-4.5	H	L-mode	METIS
100013	1	ITER	ITER-PFPO1-1.8T-H	-5.0	-1.8	H	L-H-L	METIS
100501	1	ITER	ITER-nonactive-H	-7.5	-2.65	H	L-H-L	CORSICA
100502	1	ITER	ITER-nonactive-H	-7.5	-2.65	H	L-H dithering	CORSICA
100503	1	ITER	ITER-nonactive-H	-7.5	-2.65	H	L	CORSICA
100504	1	ITER	ITER-nonactive-H	-9.6	-3.25	H	L	CORSICA
100505	1	ITER	ITER-nonactive-H	-12.7	-4.7	H	L	CORSICA
100506	1	ITER	ITER-nonactive-H	-15.0	-5.3	H	L	CORSICA
101000	0	ITER	ITER_I-7.5_B-2.6_H	-7.5	-2.65	H	H-Mode	ASTRA
101001	0	ITER	ITER_I-7.5_B-2.6_H	-7.5	-2.65	H	H-Mode	ASTRA
101002	0	ITER	ITER_I-7.5_B-2.6_H	-7.5	-2.65	H	H-Mode	ASTRA
101003	0	ITER	ITER_I-7.5_B-2.6_H	-7.5	-2.65	H	H-Mode	ASTRA
101004	0	ITER	ITER_I-7.5_B-2.6_H	-7.5	-2.65	H	H-Mode	ASTRA
101005	0	ITER	ITER_I-7.5_B-2.6_H	-7.5	-2.65	H	H-Mode	ASTRA
101006	0	ITER	ITER_I-7.5_B-2.6_H	-7.5	-2.65	H	H-Mode	ASTRA
101007	0	ITER	ITER_I-5.0_B-1.8_H	-5.0	-1.8	H	H-Mode	ASTRA
101007	1	ITER	ITER_I-5.0_B-1.8_H	-5.0	-1.8	H	H-Mode	ASTRA
101008	0	ITER	ITER_I-5.0_B-1.8_H	-5.0	-1.8	H	Ohmic	ASTRA
101007	2	ITER	ITER_I-5.0_B-1.8_H	-5.0	-1.8	H	H-Mode	ASTRA

図 5　ITER シナリオデータベースにおけるデータ一覧の例．現時点で115ケースのシナリオが利用可能．

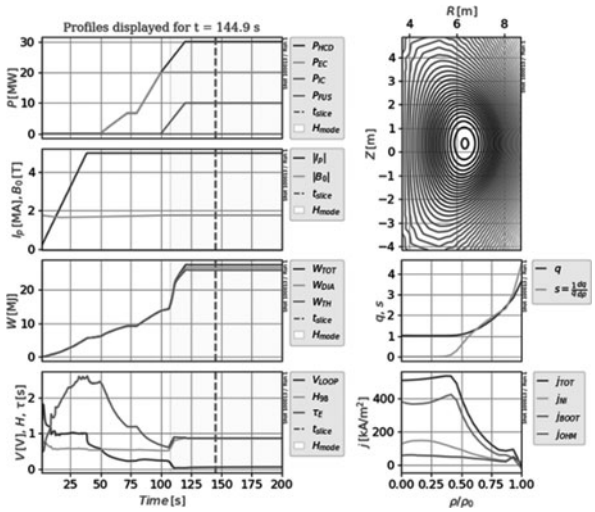

図 6　ITER シナリオデータベースにある定格の 1/3（5 MA/1.8 T）の運転シナリオをデータベースツールで可視化した例．

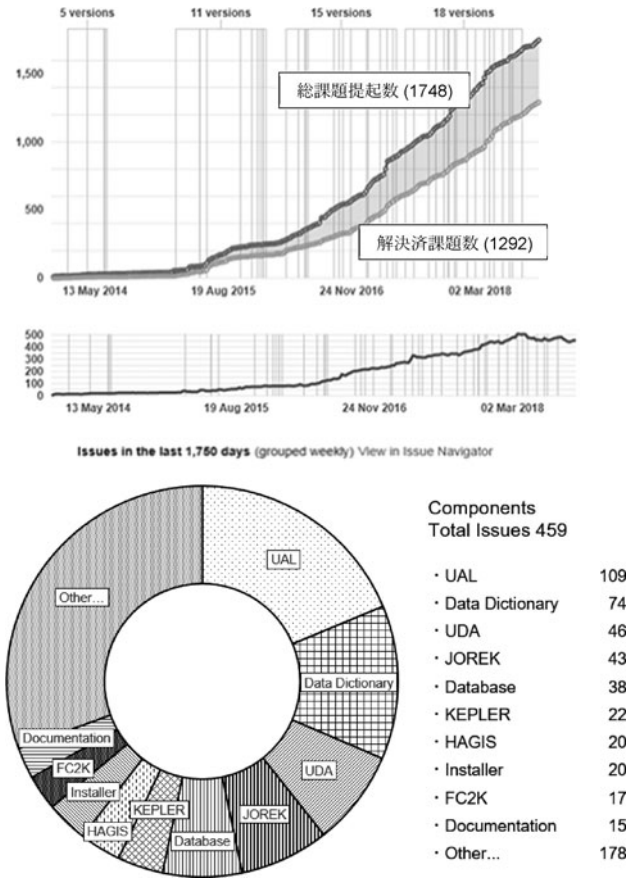

図 7　（上段）IMAS の JIRA システムにおいて作成された総課題数の時系列図；縦線＝IMAS バージョン更新日．（中段）未解決課題数の時系列図．（下図）未解決課題の IMAS コンポーネント毎の内訳．

を持つ4つのCODACの制御ネットワークを通じて行われる．

❖Timing Control Network（TCN）
IEEE-1588プロトコルによる同期信号用ネットワーク．
❖Plant Operation Network（PON）
EPICS用のネットワーク．
❖Synchronous Databus Network（SDN）
プラズマ制御システム用のUDPマルチキャストネットワーク．
❖Data Archiving Network（DAN）
TCPプロトコルによる大容量データ伝送ネットワーク．

EPICS（Experimental Physics and Industrial Control System）は実験装置の分散制御システム開発のためのソフトウェア環境[22]でありCODACの基礎となっている．実験制御はPlant Operating Zone（POZ）で行われ，CODACシステムの多くがここで動作するが，ITERはフランス政府の認可を受けた原子力基本施設（INB-174）としてその法令に従いPOZでの厳密な品質保証（Quality Assurance: QA）が求められること，及び空間的制約などから，解析コードをPOZ内の計算機に自由にインストールすることはできず，またすべての実験データをPOZ内では恒久的に保存することはできない（POZには10日間保存可能）．このため実験データはXPOZと呼ばれるPOZの外のエリアに設置される記憶装置に，リアルタイムに（ほぼ遅延なく）転送され，解析コードへの入力となったり，各極のITERメンバーに転送される．**表3**に4つのCODACネットワークの内，実験データ保存に主に関係するPONとDANの主要なスペックを示す．

DANとSDNではデータ形式が共通にHDF5であるものの，それぞれ異なるネットワークアーキテクチャを持つことから互いの物理的データ構造に互換性がない．一方，PONネットワークではPOZとXPOZ共通でRDBMS（Relational Database Management System）であるPostgreSQLに保存する（XPOZへのデータ転送時にHDF5フォーマットへ変換可能）．**図8**にCODACのXPOZにおけるデータアクセスの流れを示す．同じCODACシステム内でありながら設計上の要請から異なるデータ構造を持つ各記憶装置に対して共通のアクセス手段を与える目的から，多様なAPIを持ち，実験データだけではなく平衡解析などの計算結果やメタデータへのアクセスも可能とするUDAの利用が検討され，実装が始まっている．すでにDANネットワークからUDAを介してメタデータを取得でき，実験データに関しては，その大きいデータ容量のダウンサンプリングによりUDAを通じて計測データの利用ができるようになった．IMAS-UDAを利用して実験放電波形や平衡解析結果をリアルタイムに表示するライブディスプレイの開発も進んでいる．

5.6　まとめ

IMASインフラストラクチャは，ITERの実験・データ解析で想定されるソフトウェアツールへの需要を満たすためITER及び各極の研究機関と共同で開発が進んでいるが，同時に既存実験装置の需要に答えることも指向していることが特徴である．この共同開発を通じて成果物をITERメンバー間で共有し，ITERの解析ワークフローの妥当性を各極の実験データを下に検証することなどを可能とするだけではなく，ITERメンバーにとっては事前にITER

表3　ITER/CODACの実験データアーカイブシステムの主なスペック．

項目	値
計算機数（PON）	1000
総シグナル数（wires）	100000
PONデータ転送速度	25[MB/s]
PON最大データ転送速度	50[MB/s]
PONデータ記憶量	2.2[TB/日]
DANデータ転送速度	2-50[GB/s]
DAN最大データ転送速度	50[GB/s]
DANデータ記憶量	90－2200[TB/日]

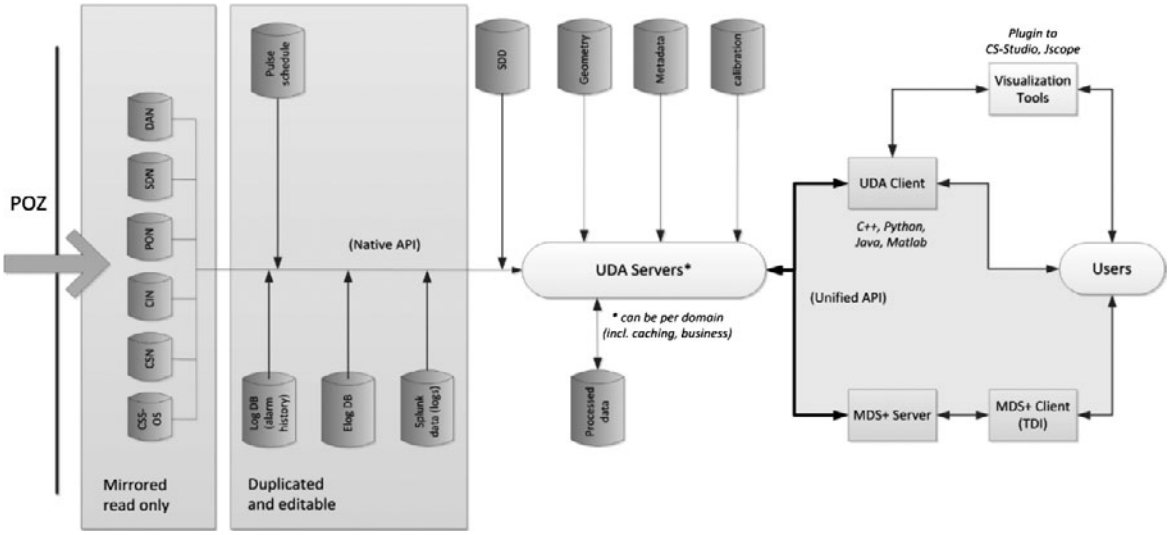

図8　ITER/CODACのXPOZにおけるUDA利用時のデータフロー．

の実験データへのアクセス・解析ツールの利用経験が得られる良い機会となっている．汎用性・拡張性が高いITERデータモデルを中心としたIMASインフラストラクチャの運用が既に始まり，ファーストプラズマに向けて数年後にはITERメンバーとの共同開発による解析ワークフローもすべて出揃う見込みとなっている．大変良い時機ですので，是非IMASの導入をご検討なさってはいかがでしょうか．

Acknowledgements

ITER is the Nuclear Facility INB no. 174. The views and opinions expressed herein do not necessarily reflect those of the ITER Organization.

参 考 文 献

［1］S.D. Pinches *et al.,* Proc. 27th IAEA, FEC TH/P6-7 (2018).
［2］F. Imbeaux *et al.,* Nucl. Fusion **55**, 123006 (2015).
［3］ITER Organization, ITR-18-003 (2018).
［4］G.L. Falchetto *et al.,* Nucl. Fusion **54**, 043018 (2014).
［5］D.G. Muir *et al.,* Fusion Eng. Des. **83**, 406 (2008).
［6］http://kepler-project.org
［7］V.E. Lukash *et al.,* Plasma Devices Oper. **13**, 143 (2005).
［8］V.E. Lukash *et al.,* 38th EPS Conference on Plasma Physics P2.109 (2011).
［9］G.T.A. Huysmans *et al.,* Proc CP90 Conference on Comp. Phys., World Scientific Publ. Co. 371 (1991).
［10］A.B. Mikhailovskii *et al.,* Plasma Phys. Report **23**, 844 (1997).
［11］M. Romanelli *et al.,* Plasma Fusion Res. **9**, 3403023 (2014).
［12］L. Lauro-Taroni *et al.,* Proc. 21st EPS vol 1 (1994).
［13］R. Simonini *et al.,* Contrib. Plasma Phys. **34**, 368 (1994).
［14］A. Taroni *et al.,* Contrib. Plasma Phys. **32**, 438 (1992).
［15］S. Wiesen *et al.,* FZ-J?lich, EDGE2D-EIRENE coupling at JET ITC (2005/2006) Project Report.
［16］F. Koechel *et al., Proc. 27th IAEA FEC,* EX/P7-25 (2018).
［17］G.V. Pereverzev and P.N. Yushmanov, IPP-Report IPP 5/98 (2002).
［18］J.A. Crotinger *et al.,* LLNL Report UCRL-ID-126284; NTIS #PB2005-102154 (1997).
［19］R.R. Khayrutdinov and V.E. Lukash, J. Comp. Phys. **107**, 106 (1993).
［20］J.F. Artaud *et al.,* Nucl. Fusion **58**, 105001 (2018).
［21］https://git-scm.com
［22］https://epics.anl.gov

J. Plasma Fusion Res. Vol.95, No.5 (2019) 236-240

小特集 プラズマ実験におけるデータ解析インフラストラクチャ

6．研究室におけるデータ解析インフラストラクチャ

6. Data Analysis Infrastructure for a Small Laboratory

辻 井 直 人
TSUJII Naoto
東京大学大学院新領域創成科学研究科

（原稿受付：2018年12月10日）

計測データ・解析プログラムの共有化は，大学の研究室規模の実験においても重要である．小規模な実験施設では研究者同士のコミュニケーションの敷居は低い一方で，共有化するための仕組みを構築・管理することが難しい．核融合実験で広く使われているデータ管理ソフトウェアMDSplusを利用することで，基本的なデータ共有の仕組みであれば，ごく簡単に構築することができる．また，簡単なデータプロット・解析だけであれば，MDSplus付属のソフトウェアでも行える．解析プログラムを共有する特別な枠組みはないが，MDSplusをベースにデータ収集・解析インフラストラクチャを組むことで，システム構築の労力を相当軽減することが可能となる．本章では，MDSplusによる実験データ利用・管理を，東京大学のTST-2球状トカマク装置での実装例を元に紹介する．

Keywords:
nuclear fusion, tokamak, data acquisition, MDSplus

6.1 MDSplusとは

MDSplus[1]はパルス実験データを収集，管理及び解析するためのソフトウェアである．従来型のシステムでは，デジタイザーのパラメータや生データ，解析データはそれぞれ別々に管理されていた．これら実験データの処理を，**図1**のように全てMDSplusを介して行うことで，データの管理や解析を大幅に簡略化することができる．時系列データのプロッター等，汎用の解析プログラムは，MDSplusデータベースにアクセスする形で実装することで共通化することができる．MDSplusはLinux，Windows，Mac上で動作し，データベースへアクセスするためのインターフェースはC，C++，FORTRAN，Java，MATLAB，IDL，Python，Visual Basic，LabVIEW，PHPで実装されている．

MDSplusはマサチューセッツ工科大学等で共同開発されている．MDSplusをベースに高度なデータ管理インフラストラクチャを構築することが可能であり，Alcator C-ModではMDSplusが主要なデータ解析インフラストラクチャとして運用されていた．また，DIII-D，JET等，世界中の磁場核融合実験にも実装されている．基本的なデータ共有だけであれば比較的簡単に新規導入して実装できるので，ITスタッフのいない大学研究室での利用にも適している．

本章では，まずデータユーザによるMDSplusデータベースへのアクセス方法，データプロバイダ（計測器の管理者）によるMDSplusの利用方法を紹介する．大学研究室においては，このようなシステムの構築が一番のハードルになるので，MDSplusのセットアップや実験装置の制御系との統合等，実際の導入方法についても，全体の流れがわかるよう，ごく簡単にではあるが説明する．

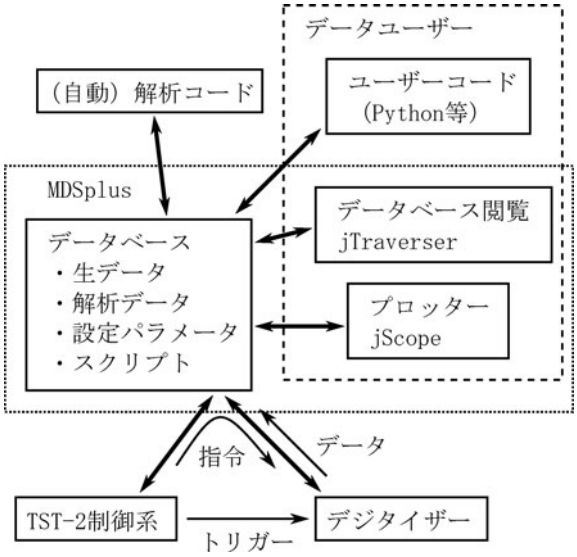

図1　東京大学TST-2装置におけるMDSplusの実装．MDSplusソフトウェアはツリーデータベースとデータベース閲覧・プロット等のためのツール群から成る（図点線内部）．データ収集やデータ解析は全てMDSplusを介して行われる．データユーザはjTraverser，jScopeといったMDSplus付属のツールを用いたり，Python等の言語によりデータベースに直接アクセスする（図破線内部）．

Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, CHIBA 277-8561, Japan
corresponding author's e-mail: tsujii@k.u-tokyo.ac.jp

©2019 The Japan Society of Plasma Science and Nuclear Fusion Research

6.2 データユーザによる利用

6.2.1 jTraverser によるデータベース閲覧

MDSplus データベースは，ショット毎のツリー構造になっている．データにアクセスするにはツリーの名前とショット番号が必要となる．ツリー構造を閲覧するために GUI ツール jTraverser（java プログラム）が用意されている．**図 2** に東京大学 TST-2装置の実験データのルートツリー（TST2ツリー）を示す．木のマークのあるノードは別のツリーへの参照になっている．RF や計測に関連するデータはそれぞれ独立なツリー（RF，ELECTRONS，MAGNETICS 他）にまとめられ，全て TST2 ツリーの下に配置されている．ツリー構造とノード名を工夫して作っておくことで，データユーザは，ある程度直感的に，自分の必要とするデータのパスを見つけることができる．例えば「MAGNETICS」の下にある「BT」はトロイダル磁場の時系列データである．

6.2.2 jScope による放電データのプロット

パラメータ等スカラーデータは，jTraverser 上で対応するノードをクリックすることで値を確認することができる（**図 3**）．放電の時系列データの表示には jScope と呼ばれるプロッターが用意されている．**図 4** に jScope のプロット画面を示す．各パネルには，プロットしたい信号のツリー，ノード名を設定する（**図 5**）．ノード名はデータプロバイダー（計測器の管理者）に確認するか，整理されたツリーであれば jTraverser から自分で見つけても良い．ノード名の正式なシンタックスは

```
\(ツリー名)::TOP[.(ノード名)] ... :(信号名)
```

である．大文字，小文字は区別しない．例えば，上記の「BT」であれば，

```
\TST2::TOP.MAGNETICS.TF_COIL:BT
```

となる．GUI 下部にあるテキストボックスにショット番号を入れる．TST2 ツリーの最新ショットをプロットしたい場合は，

```
CURRENT_SHOT('TST2')
```

とすれば良い．これらのシンタックスはデータプロバイダーによる利用の項で説明するが，TDI と呼ばれる言語である．放電が終了し，データ収集が完了したタイミングで，表示を自動的にアップデートする設定も可能である．

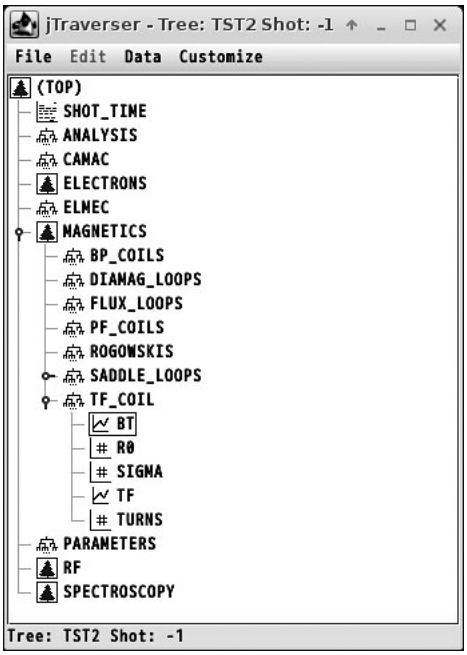

図 2　TST-2 における MDSplus ツリーデータベース（TST-2 ツリー）．「MAGNETICS」（磁気計測），「RF」等のツリーが TST-2 ツリーの下に配置されている．データは各ツリー，ショット毎に別々のファイルに保存される．

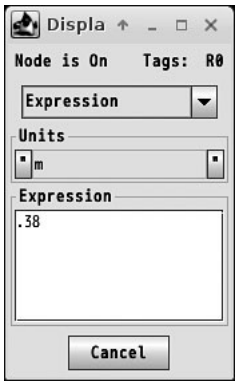

図 3　図 2 のパラメータ「R0」の設定画面．

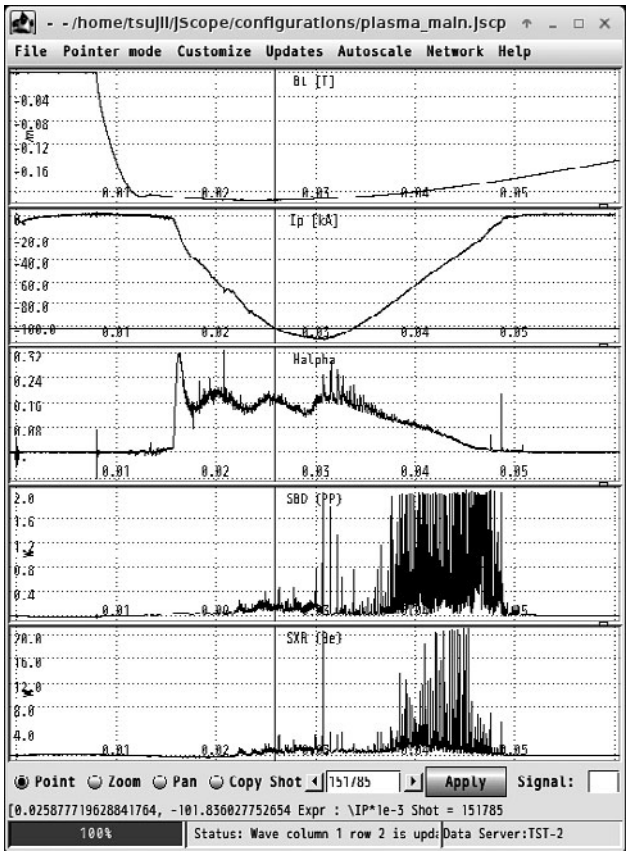

図 4　jScope による時系列データ表示画面．拡大（Zoom），パン（Pan），データ値の読み取りカーソル（Point）等プロッターに必要な基本的な機能が一通り実装されている．

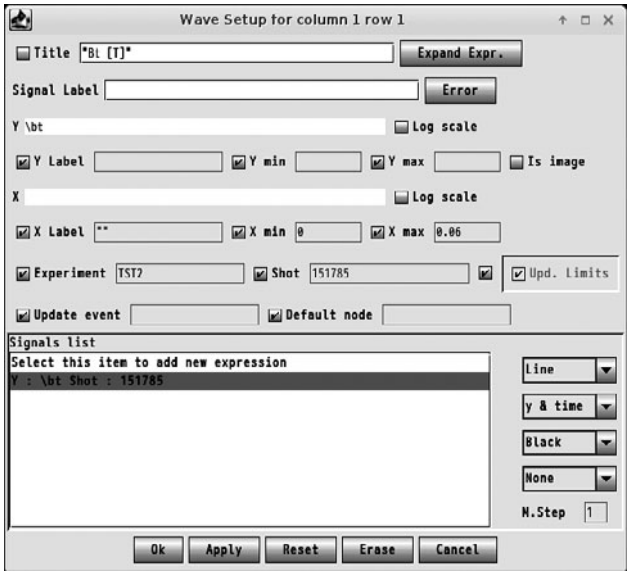

図5　jScopeの信号設定画面．信号型であれば，y軸だけ登録すれば良い．

jScopeはプロッターとしての基本的な機能を一通り備えており，直感的な操作で，プロットの拡大・縮小やパンができる．また，全てのプロットのx軸を特定のパネルに揃えてy軸を自動的にリスケールするボタンや，データカーソル等，解析に役立つ機能が実装されている．

jScopeを用いて，ツリーに登録されている信号型データをプロットすることができる．信号の種類が生データであるか，または校正されたデータや平衡計算結果等の解析データであるかを気にする必要はない．

6.2.3　プログラムによるデータの読み込み

データを表示するだけでなく，より複雑な解析を行いたい場合には，何かしらのプログラミング言語でデータを読み込む必要がある．MDSplusにおいては，上に述べたように，様々なプログラミング言語でデータにアクセスするインターフェースが実装されている．ここではPythonを使ったMDSplusデータベースの利用例を示す．前節の「BT」（トロイダル磁場の信号）を，ショット番号10000について読み込む場合は以下のようになる：

```
>>> import MDSplus
>>> tree = MDSplus.Tree('TST2', 10000)
>>> node = tree.getNode('MAGNETICS.TF_COIL:
BT')
>>> y = node.getData().data()
>>> x = node.getDimensionAt(0).data()
```

y，xはそれぞれトロイダル磁場の信号と対応する時間軸である．また，ノード名は上に書いた絶対パスではなく，TST2下の相対パスで指定した．ここで，data()メソッドで返されるオブジェクトはnumpy配列になっているので，大きなデータでも効率的に処理できる．

上記の数個のコマンドを使うだけで，全ての実験データ，解析データの利用が可能となる．全てのデータをMDSplusデータベースで管理することで，同じような構造を持ったデータであれば，共通のプログラムにより解析できるようになるので，プログラムの共有も促進される．

6.3　データプロバイダーによる利用

6.3.1　プラズマ放電時のMDSplusの動作

MDSplusツリーのショット番号-1はテンプレートツリーになっている．プラズマ放電時には，まずテンプレートツリーを複製することで新しいショット番号のツリーが生成される．ツリーにはアクションと呼ばれる，データ収集時に実行されるスクリプトを登録するノードがある．アクションには，初期化，放電，データ収集といった放電シーケンスのどのタイミングで実行されるかが設定されており，これらアクションが実行されることで，デジタイザーの初期化やデータ収集がショットサイクルと同期して行われる．各アクションには依存関係を定義できるので，必要なデータが揃った後に，平衡計算を実行し，分布計測データをマッピングするというような動作を自動化できる．また，アクションの終了時にイベントを生成する設定をすることで，6.2.2節で説明した，jScope上での対応する信号の自動アップデートが行われる．

6.3.2　パラメータの設定

放電前には，デジタイザーのサンプリング周波数等のパラメータ設定を行う必要がある．これは新規ツリーのテンプレートであるショット番号-1のツリーに対して行う．全てのパラメータはjTraverserから変更することができる．デジタイザーの運用のためにデバイス型が定義されている．現在ではあまり使われなくなっているが，CAMACモジュールは既に実装されているので，単にメニューから選んで追加するだけでMDSplusシステムに組み込むことができる．ノード名を右クリックし，メニューから「Setup Device」をクリックすると，GUIの設定画面でパラメータを設定することができるので便利である．デジタイザーのデータ収集や，解析スクリプトの入・切はノードの右クリックメニューの「Toggle On/Off」で行う．

6.3.3　プログラムによるデータの書き込み

パラメータの設定をスクリプトにより自動的に行う場合や大量のデータを入力する場合等，プログラムからデータを書き込むには以下のようにすれば良い：

```
>>> import MDSplus
>>> tree = MDSplus.Tree('TST2', 10000)
>>> node = tree.getNode('MAGNETICS.TF_COIL:
R0')
>>> node.putData(0.35)
```

上記の操作により，ショット番号10000のツリーの「R0」というノードに値0.35が代入される．

6.3.4　ツリーへのパラメータ・信号の追加

ツリーの変更にはjTraverserを用いるのが便利である．よく使うデータタイプは「NUMERIC」（数値），「TEXT」（文字列），「SIGNAL」（信号）である．数値と文字列はそれぞれパラメータとコメントに使える．信号型は，信号と軸座標からなる．信号は一般に多次元でも良く，

対応する軸座標も次元の数だけ登録できる．特に軸座標がなければ，登録しなくても良い．

通常は，過去のデータ構造は変えずに，今後の放電のデータ構造を変更するので，テンプレートツリー（ショット番号-1）を編集することが多い．デジタイザーの生データやシミュレーション結果等，数値データの保存には，放電後にデータを格納できるように中身が空の信号ノードを定義しておく．生データから加減乗除等の簡単な計算で得られる信号（校正信号等）を登録したい場合は，TDI（Tree Definition Language）で信号を定義すると良い．例えば，信号型 RAW_DATA から数値パラメータ OFFSET を引いて SENSITIVITY で割った校正信号が欲しければ，信号型のデータとして

```
(RAW_DATA－OFFSET) / SENSITIVITY
```

という表式を登録しておく．数値の読み込み・演算はノードが評価された時（Python であれば data() メソッドが呼ばれた時）に行われる．TDI に関する補足として，時間軸が等間隔の数列であれば，直接全ての数値を保存するのではなく，

```
(開始値):(終了値):(ステップ値)
```

のように定義しておくと，ノードが評価された時にデータが生成されるので，保存データ量を節約できる．TDI は比較的高機能のプログラミング言語になっており，TDI 命令だけで，ある程度複雑な解析が可能である．より複雑な計算の場合は，C や Python の外部関数を呼び出すことも可能である．ちなみに，小規模なグループではアーカイブデータの総量を抑えることも重要であるが，MDSplus の数値データは（デフォルトの設定で）自動的に圧縮されるので便利である．

最終的にデータユーザが使うべき校正信号や解析データを，直感的にわかりやすいノード名の信号型で登録しておくことで，データの共有が容易になる．

6.3.5 データの管理について

MDSplus を用いることで，デジタイザーの設定値や校正用パラメータ，取得データ等計測器に関する情報を全て一元的に管理することができるようになる．バックアップについても，次節で述べるが，対応するツリーファイルを保存するだけである．小規模なグループでは自分なりのルールで実験ノートやテキストファイル等にデータをまとめてあるだけのことが多いが，MDSplus を積極的に活用することで，データの共有化が容易になるだけでなく，データの一括処理等解析の効率化も可能となる．

6.4 MDSplus の導入

6.4.1 MDSplus のインストール

MDSplus のウェブページ[1] から各プラットフォーム用のバイナリをダウンロードできる．必要であれば，ソースからのコンパイルも特に難しいことはない．一般的な Linux，Windows 環境であれば問題なくインストールできるはずである．

6.4.2 ツリーパスの設定

導入の際には，データベースの実体をどこに置くかを定義する必要がある．ツリーの実体はショット毎に 3 つのファイルに保存される．例えば，TST2 ツリーのショット番号10000のデータであれば，

```
tst2_10000.tree,
tst2_10000.characteristics,
tst2_10000.datafile
```

の 3 つのファイルである．テンプレートツリーはショット番号の-1ではなく「model」がファイル名につく：

```
tst2_model.tree,
tst2_model.characteristics,
tst2_model.datafile .
```

これらのファイルが保存されるディレクトリ（ツリーパス）は，環境変数

```
(ツリー名)_path
```

で設定する．例えば TST2 ツリーであれば，tst2_path にファイルを参照するディレクトリを優先順位の高い順にセミコロン（;）で区切って設定する．新しいツリーは最も優先順位の高いディレクトリに作成される．ツリーパスには，ツリー毎に，最新のショット番号を格納するファイルも保存される．このショット番号は任意に設定でき，jScope による最新ショット表示や，放電時の新ツリー生成の際に参照される．

データベースアクセスのパーミッションはツリーの実体であるファイルのアクセス権限と同じであるので，例えば，計測器毎にツリーを分け，それぞれパーミッションを設定することで，データユーザが誤操作により，自分が管理しないデータを変更してしまうのを防ぐことができる．デフォルトの設定では，権限のあるユーザが誤ってデータを上書きしてしまうのを防ぐことはできない．データのバージョン管理をオンにすることで，新しく保存したデータが以前のデータを上書きすることはなくなる．データを読み込むと，デフォルトでは最も新しいデータが返されるので，データユーザの利用法は変わらない．

MDSplus はショット毎にファイルが作られるので，単一のディレクトリに大量のファイルが生成されてしまい，管理の効率が落ちる場合がある．この問題については，ツリーパスを

```
tst2_path=/tst2/new/~t;/tst2/models/~t;/
tst2/archives/~g~f~e/~d~c/~t
```

のように設定して置くことで解決できる．ここで，「~t」はツリー名（ここでは tst2）に展開される．「~c」，「~d」,...,「~g」はショット番号の 3 桁目から 7 桁目に展開される．例えばショット番号10000にアクセスした場合は/tst2/archives/100/00/tst2/が検索されるので，各階層でのファイル・ディレクトリ数を抑えられる．新しいショットは/tst2/new/tst2 に生成されていくの

で，実験終了後に/tst2/archives/...に移動するスクリプトを OS にスケジュールしておくと良い．

6.4.3 データ収集と自動解析

データの収集・解析はツリーに登録されているアクションが実行されることで行われる．実際にアクションを実行したり，その実行を制御したりするサーバの設定はここでは割愛するが，[1]に詳しく説明されている．MDSplus のサーバに対して，制御系から放電シーケンスに合わせて TCL コマンドを送ることでデータ収集・解析が行われる．最も簡単な設定では，まず TCL シェルを起動して，現在のショット番号を加算し，テンプレートツリーから新しいツリーを生成する：

```
$ mdstcl
TCL> set current tst2 /increment
TCL> set tree tst2 /shot=-1
TCL> create pulse 0
```

この時，TST2 ツリー下にある子ツリーも全て一緒に複製される．次にアクションの実行テーブルを準備し，INIT フェーズに実行するものとして登録されているアクションを実行する．

```
TCL> set tree tst2 /shot=0
TCL> dispatch /build
TCL> dispatch /phase INIT
```

このフェーズでデジタイザーの設定，アームが行われる．放電開始を MDSplus で制御する場合は

```
TCL> dispatch /phase PULSE_ON
```

とする．最後に，STORE フェーズのアクションを実行する．

```
TCL> dispatch /phase STORE
```

このフェーズでは取得したデータのデータベースへの保存を行う．平衡計算等のデータ解析を行うのも，このフェーズである．

6.4.4 データのリモートアクセス

MDSplus はクライアント・サーバモデルで実装されているため，一度設定してしまえば，ユーザはローカルかリモートかを意識せず，データベースにアクセスすることができる．サーバ側で mdsip サービスが動いていれば，ユーザ側としては，ツリーパスの環境変数に，サーバ名とポート番号を入れて

(サーバ名)：(ポート番号)：：(ディレクトリ)

と設定するだけである．mdsip でデフォルトの8000番ポートを使っている場合はポート番号の指定は省略できる．

6.5 まとめ

MDSplus は核融合実験データの収集・管理・解析に必要な一通りの機能を備えており，大型装置を動かすシステムを構築するのに十分な汎用性を持っている．ただし，高い信頼性を持ったデータベースとして運用するためには，そのためのハードウェア等のインフラが必要であり，それらを構築・運用するスタッフは結局必要である．また，MDSplus に実装されていない機能については，外部に付加的に実装するしかなく，そのような部分を共有する手間は従来型のシステムと同様である．そういう意味では，MDSplus の導入により，データの共有はすぐに実現できるが，解析プログラムの共有についてはデータ共有の結果として促進される程度である．とはいえ，MDSplus をベースにデータ収集系を構築することで，ゼロからシステムを構築する場合に比べて労力は相当軽減されるはずである．

参 考 文 献

［1］MDSplus wiki (www.mdsplus.org)

小特集 プラズマ実験におけるデータ解析インフラストラクチャ

7．おわりに

7. Conclusion

藤井恵介
FUJII Keisuke
京都大学工学研究科

（原稿受付：2019年２月26日）

本小特集では，複数人で行うプラズマ実験に際して必要となるデータ解析インフラストラクチャについて述べた．特に，大型実験装置ではデータの管理・共有のほか，基本的な解析データが自動的に生成されること，上流のデータの変更が追跡できることなどが重要である．大型装置を有する研究機関では，それぞれの優先度に従ってデータ解析のための環境を整備している．

データの保管・共有については標準的なデータベースの構築法がある程度確立されており，多くの機関・装置で現在は大きな問題なく行われているようである．また，研究室規模の実験においても，MDSPlusなどの既存の枠組みを使うことで，効率的にインフラストラクチャを運用していくことも可能になってきている．

一方で課題となるのは，解析の自動化やデータ変更の追跡をどのように行うかである．一般的に行われる対話型の解析ルーチンでは，各計算ステップの結果を担当者がチェックしつつ，適切なパラメータ（平滑化フィルタの窓の大きさなど）を決めて計算を行うという方法が（無意識のうちに）採られている．これらを自動化し，ロバストなプログラムとして開発するのは単純ではない．開発にはそれなりの時間と労力が必要である．また，自動化により誤った結果を出力しそれが後段に伝播してしまうというのではないかという恐れを解析担当者が抱いてしまうことも考えられる．（実際にはプログラム開発者が全てのデータの結果を人力でチェックすることができないことから，ユーザが自身の使うデータを中間ファイルも含めてチェックするべきという運用になるであろう．）

さらにITERのように超大規模なシステムになると，解析プログラムの性能が及ぼす影響が広大になるため，その開発にはより一層慎重にならざるを得ない．システムにプログラムを搭載するための審査も行われることになっている．解析ソフトウェア開発に対する敷居はますます高くなってしまうだろう．

ソフトウェア開発の潮流の一つに，オープン・ソースという概念がある．例えばPythonエコシステムを形成するNumpyなどの多くのパッケージや，Pythonというプログラム言語自体オープン・ソースである．オープン・ソースという言葉自体はソフトウェアのソースコードが公開されているという意味であるが，それ以上に誰がどの部分を開発・修正したかもオープンであることが特徴である．どのソフトウェアにどの程度開発・維持に貢献したかがクリアであり，それがソフトウェア・エンジニアの実績の一つとして評価される土壌もある．

プラズマ・核融合コミュニティにおいても，解析プログラムの開発・維持による他の研究者への貢献が実績として評価されるようになってくると筆者は考えている．

Graduate School of Engineering, Kyoto University, KYOTO 606-8501, Japan
author's e-mail: fujii@me.kyoto-u.ac.jp, fujii@me.kyoto-u.ac.jp

©2019 The Japan Society of Plasma Science and Nuclear Fusion Research

小特集執筆者紹介

藤井恵介（ふじい けいすけ）

京都大学工学研究科．専門は光計測，機械学習．2012年3月京都大学工学研究科博士後期課程修了．豊富に蓄積されているプラズマ実験データの有効利用をめざして，統計的手法・機械学習の導入に奮闘しています．趣味ではオープンソースソフトウェアの開発にも参加しています（https://github.com/fujiisoup）．プラズマ・核融合分野でもオープンソース・ソフトウェアの文化が広まればいいですね．

鈴木千尋（すずき ちひろ）

核融合科学研究所ヘリカル研究部助教．主な研究分野は，磁場閉じ込め高温プラズマにおける，不純物分光計測を中心とした実験研究．最近は，各種の高原子番号元素をプラズマ中に導入した際の，多価イオンからの極端紫外・軟エックス線分光スペクトルの計測とその解析に携わっている．そのせいで，以前はあまりなじみのなかった，周期表の下の方にある元素の並びをほぼ覚えてしまった．

江本雅彦（えもと まさひこ）

核融合科学研究所ヘリカル研究部助教．主な研究分野はデータ処理．LHDで収集されたデータの管理システムや自動解析システムの開発を行っています．好きなプログラミング言語はRuby．現在の趣味は未確認生物（Pokémon）の採集および生態調査のフィールドワーク．

居田克巳（いだ かつみ）

所属　核融合科学研究所ヘリカル研究部教授．主な研究分野はプラズマ物理，磁場閉じ込めプラズマの熱・運動量の輸送及び磁気流体不安定性．趣味は若い時はロッククライミング，アイスクライミング．今はパラグライダー，バックカントリースキー，マウンテンバイク．所属学会は日本物理学会，プラズマ・核融合学会，アメリカ物理学会．

山崎莉穂（やまざき りほ）

量子科学技術研究開発機構那珂核融合研究所技術員．現在は核融合研究のためのデータ処理設備の開発を行っています．核融合分野に来て3年目ですが，ようやく核融合用語が分かるようになってきました．趣味は音楽鑑賞，特にJAZZ，インストバンドをよく聴きます．最近はライブにも足を運ぶようにしています．

仲野友英（なかの ともひで）

量子科学技術研究開発機構那珂核融合研究所上席研究員．博士（工学）．1998年京都大学大学院工学研究科修士課程修了．専門はプラズマ分光．2017年よりITPAスクレイプオフ層とダイバータ物理トピカルグループ副議長．最近は娘と一緒に受験勉強しています．

浦野　創（うらの はじめ）

量子科学技術研究開発機構那珂核融合研究所上席研究員．工学博士．2002年北海道大学大学院工学研究科博士課程修了．独国Max Planck研究所，日本原子力研究開発機構を経て現職．専門は周辺プラズマ物理．2017年よりITPA周辺ペデスタル物理TG議長．趣味はルービックキューブ速解きで，自己ベストは11．9秒．最近は子どもたちにマジックを見せて笑顔にさせるのが楽しい．

細川哲成（ほそかわ まさなり）

ITER Organization, Science & Operations Department, Science Division, Confinement & Modelling Section, Computational Specialist Integrated Modelling.

Simon D. Pinches（サイモン・ピンチ）

ITER Organization, Science and Operations Department, Science Division, Confinement and Modelling Section, Section Leader (acting) / Scientific Coordinator and R.O. for ITER Integrated Modelling Programme. ITPA Energetic Particle Physics Topical Group Co-Chair.

Abadie Lana（アバディ・ラナ）

ITER Organization, Science and Operations Department, Control System Division, CODAC Section, Database Engineer, Responsible for scientific data archiving and data access.

辻井直人（つじい なおと）

東京大学新領域創成科学研究科講師．2012年アメリカ合衆国マサチューセッツ工科大学博士課程修了．同年ドイツマックス・プランク研究所ポスドク研究員．2013年東京大学新領域創成科学研究科助教を経て2018年より現職．専門はイオンサイクロトロン周波数帯から低域混成波帯の波によるトカマク・ヘリカルプラズマの加熱・電流駆動解析とレーザー・マイクロ波によるプラズマ計測．趣味はピアノを弾くこと．

藤岡慎介（ふじおか しんすけ）

大阪大学レーザー科学研究所教授・副所長．大阪大学大学院理学研究科物理学専攻の協力講座を担当しています．高強度レーザーで生成される強磁場を利用し，レーザー核融合プラズマの加熱効率の向上，相対論的磁化プラズマ科学等を研究しています．レーザー科学研究所では100 Hzの高繰り返しパワーレーザー装置の開発を進めており，最近は，このレーザーを多様な科学研究に活用する方法について，想像をめぐらしています．本小特集で扱っているデータベースインフラストラクチャーは，高繰り返しパワーレーザーと対になるべきものと考えています．