

学術情報プラットフォーム 「核融合クラウド」(第2報) — 今後10年の戦略 —

大舘暁、中西秀哉

実験情報タスクグループが主としてLHD実験データを管理

- データ取得
- データの物理量への変換・仕組みの開発
- データの保存・提供・外部実験参加サポート
- 実験関連のWebサービスの開発・保守
- 国内の大学のデータ保存のサポート

情報通信システム部

部長
石黒 静児

副部長
渡邊清政 小林策治 大館暁

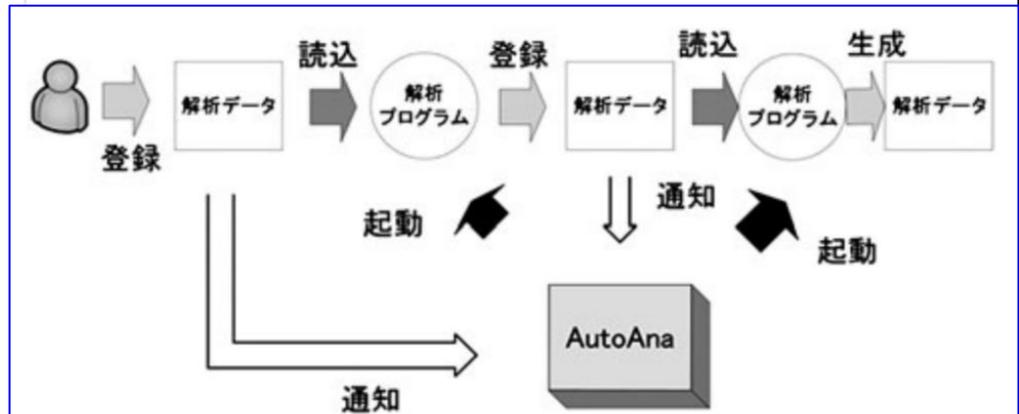
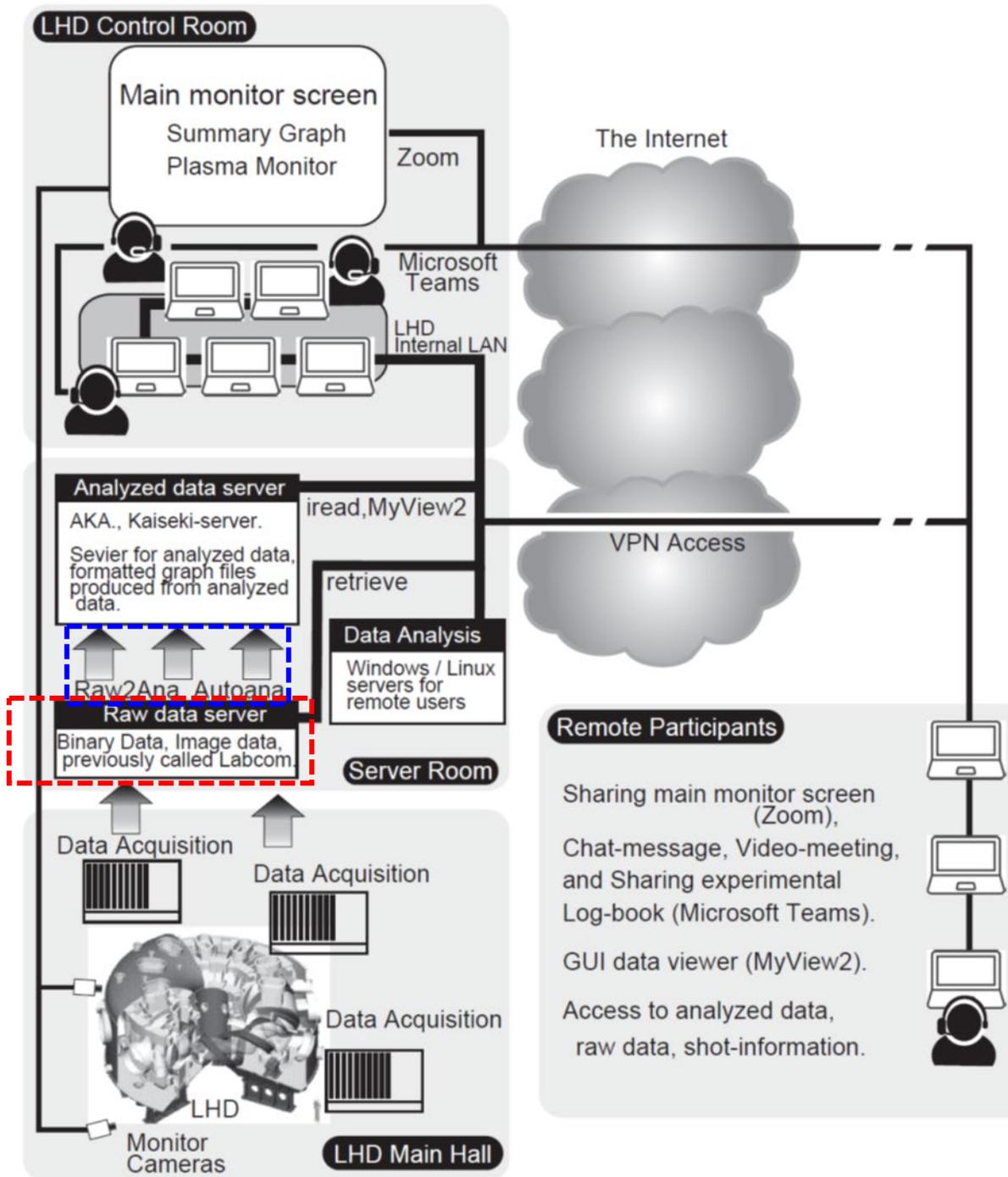


図2 AutoAna システムの動作の概要.

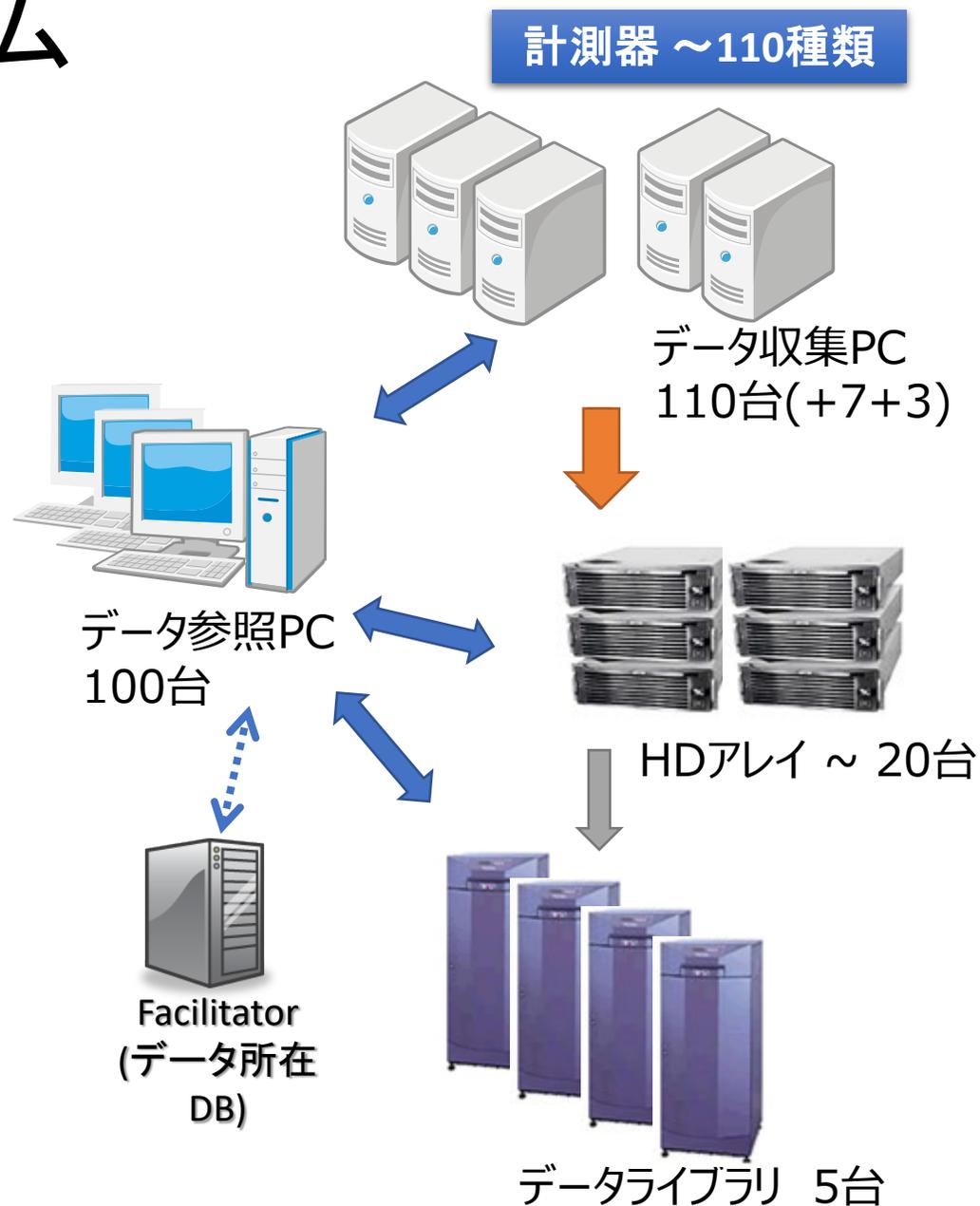


Remote Participants

- Sharing main monitor screen (Zoom),
- Chat-message, Video-meeting, and Sharing experimental Log-book (Microsoft Teams).
- GUI data viewer (MyView2).
- Access to analyzed data, raw data, shot-information.

LHD実験データシステム

- 大規模並行分散アーキテクチャ
 - ✓ 計測器1台に収集PC1台を用いて同時並行で収集・処理
 - ✓ 100超の収集ノードは、サイト毎にネットワーク上を流れる実験シーケンスに同期して運行
 - ✓ 制御コマンド／機器ステータスはIPマルチキャストで授受、マルチエージェント=マネージャ指向モデルで集中管理
- 中央ストレージも分散構成
 - ✓ 分散形態 ...> 拡張性+負荷分散
 - ✓ データは、**収集 → HDアレイ → BDライブラリ**と順次移送
 - ✓ 短パルス収集時はOK, but ...
 - ✓ **長パルス定常収集時、書込みラッシュで帯域が不足**



SSDアレイによるストレージ2層化

- HDアレイ前段に高速SSDアレイ
増設 → 2層化

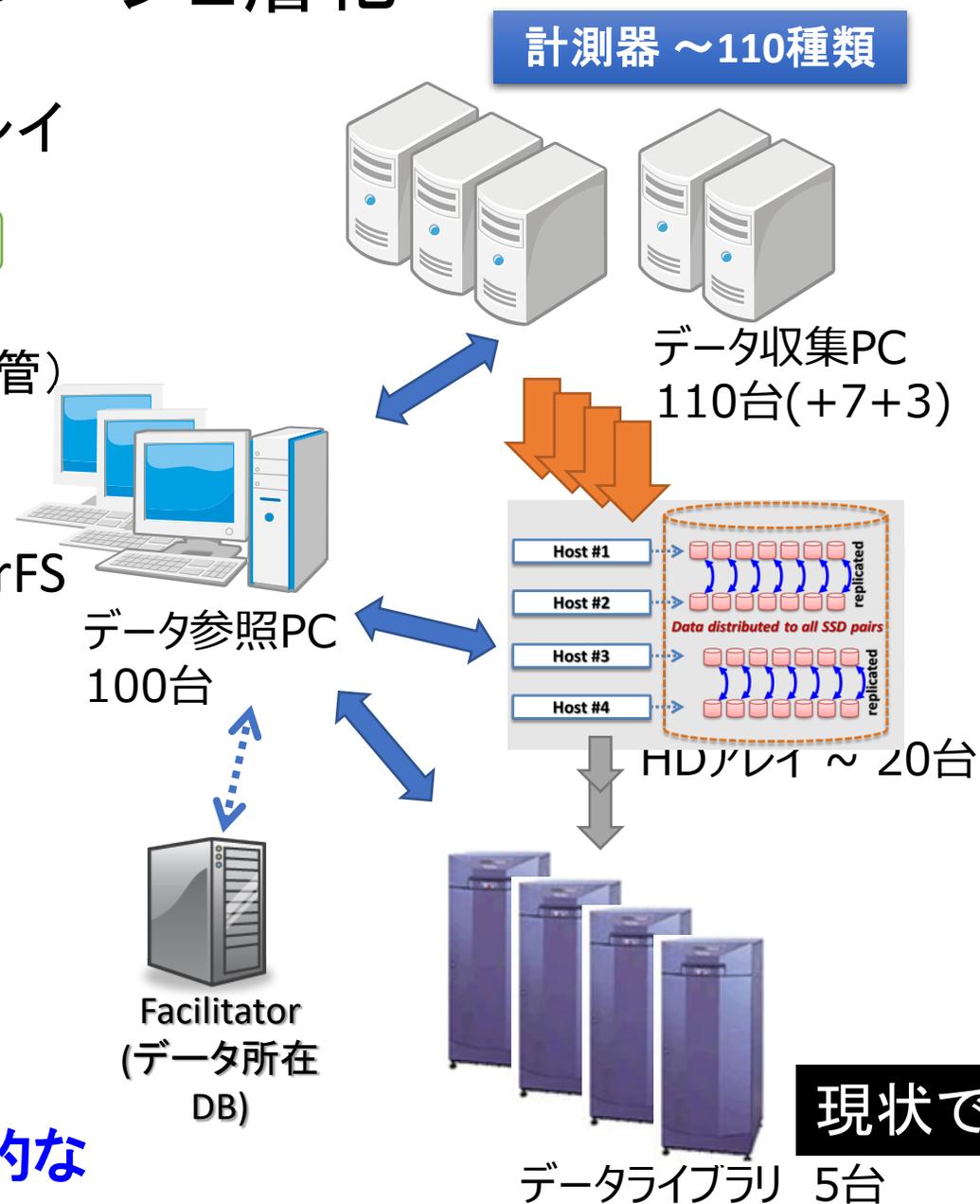
- ✓ SSD前段 ...> 高速I/O性、短期
- ✓ HD 後段 ...> 大容量、長期
- ✓ BDライブラリ ...> アーカイブ(保管)

- 分散ファイルシステムGlusterFS
を採用

- ✓ Distributed Replicated volume
(RAID-10 like) を提供
...> 分散書込み + 高速複製
- ✓ HD : rd 900 MB/s, wr 450 MB/s
SSD : rd 2 GB/s, wr ≒900 MB/s



定常長パルス収集時の集中的な
書込みを受入れ

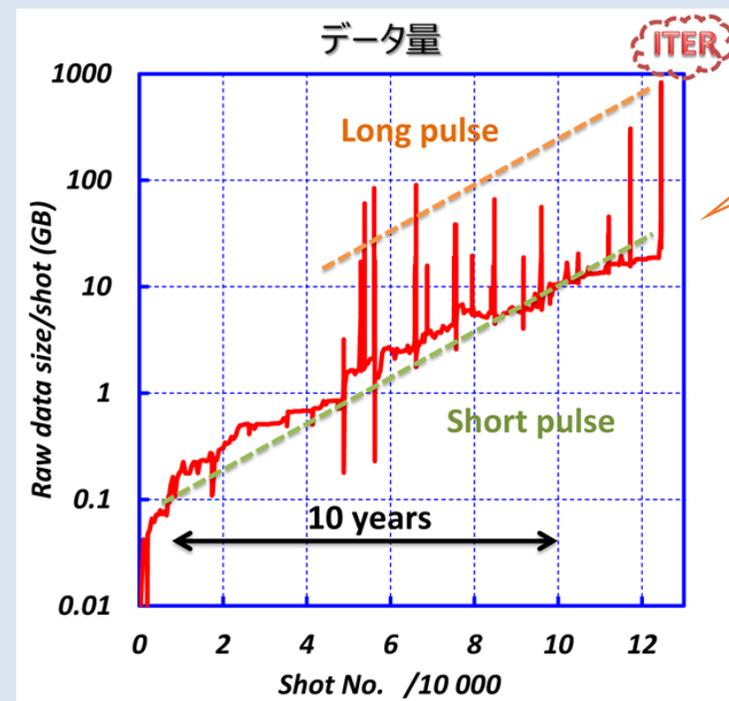
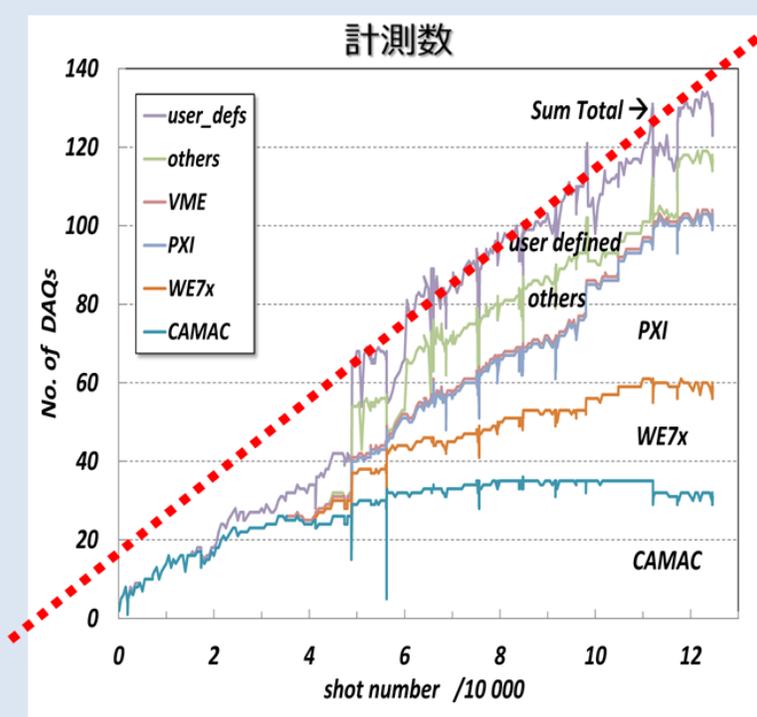


現状で1.1PB程度

(参考)

核融合実験データの実例 ～LHDの場合～

1. データ収集系の大規模分散化 ⇒ ノード数が一次関数的に増加
2. 計測データ量の恒常的な増加 ⇒ データ量は指数関数的に増加
3. 超伝導コイルによるプラズマ放電実験の定常・長パルス化
4. 計測技術の進化 (高速サンプリング、多チャンネル高分解能化)



ムーアの法則
(1.5年で2倍＝
5年で10倍)に
ほぼ沿った進化

保守運用にかかっている年間経費 (実績)

• システム保守等経費

1. データストレージ …… 1000万円
(BDライブラリ+HDDアレイ)
2. サイトライセンス …… 400万円
(PV-WAVE、LabVIEW ……レンタル)
3. サーバ更新 …… 200万円

• 開発要員経費

- ソフトウェア保守改修 …… ■■■■万円
(新計測対応を含む. 2名)

• その他、共通ユーティリティ経費

- 制御棟UPS保守 …… 270万円
 - 2026年にバッテリー交換で4300万円
- 電気代 …… ???万円

• SNET遠隔サイト保守経費

1. 遠隔複製ストレージ …… 400万円
(HDDアレイ【バックアップ】)
2. ネットワーク装置 …… 100万円

• DOI発行経費

- JaLC正会員年会費 …… 30万円～

全日本核融合データ収集・解析プラットフォーム(核融合クラウド)の構築

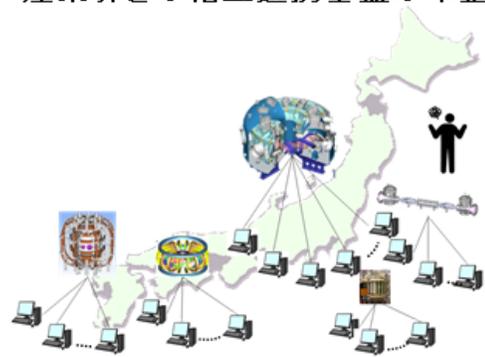
これまでの取組と解決すべき課題

取組:

- 核融合研主導で全日本核融合ネットワーク形成(他分野に先駆けた双方向型共同研究)
- アイデア、課題ベースの共同研究⇒「アイデアベースで研究に必要なデータを生み出す」取り組み

解決すべき課題

- データ形式、データマネジメントは各機関独自⇒横の連携が困難
- データマネジメント人材育成・評価体制の不足
- コロナ禍による遠隔共同実験円滑化の必要性
- 産業界との相互連携基盤の未整備



本事業

5年後の達成目標、達成された時の姿

国内外の核融合中核装置を活用した最先端研究参画のためのプラットフォーム構築

⇒ 世界の核融合研究のイニシアティブ、国内外遠隔共同実験支援、汎用性担保による核融合大規模データの超分野利活用(アカデミア、産業界、DS人材育成)



ワンストップサービスの実現

- 核融合研への共同研究申請により、
- 国内機関の核融合関連実験装置の遠隔利用を可能に
 - 実験設備を持たない大学等からの参画も

- ITERから転送される大規模データを最大限活用するプラットフォームとの互換性
- ITER実験初期から日本の国際先導性を発揮、研究をリード

核融合大規模データの超分野利活用(実践の場、人材育成)

- データサイエンス分野
- 天体プラズマ分野 など
- 産業界⇒核融合産業興隆へ

各種政策等への貢献:第5次エネルギー基本計画(核融合エネルギー)、第6期科学技術・イノベーション基本計画(案)(Society 5.0)、大学共同利用機関法人等で構成する「連合体」、データサイエンス人材育成に係る多様な政策

『核融合クラウド』を取り巻く環境・動向

① 国内(核融合)実験データ取り扱いの共通化

- i. LHD/SNET遠隔データ収集保存プラットフォーム = LHD + 九大・筑大・東大 + 京大・工繊大・・・更に拡張?
- ii. データ利用規範に関する合意形成・・・R3年度より、プラ核学会専門委員会で検討始まる

② 実験データ解析+数値モデル計算 統合データインフラの構築

- i. QST「核融合情報科学センター構想」
= ITER全データ複製ストレージ(REC)+CSCスパコン → スパコンでオンラインデータ解析
【他分野に先駆け、世界初に挑戦。IT分野とも連携協力】
- ii. 統合データインフラ用のソフトウェア・プラットフォーム・・・QSTが弱い
→ 欧州が“Integrated Modelling & Analysis Suite (IMAS)”開発を主導 → 「世界標準」狙い
【NIFS/LHD成果のEG+AutoAna解析システムで参入・貢献・競争 → 分野を先導】

NIFS-QST連携
NIFS-ITER連携

③ オープンサイエンス&オープンデータの潮流・・・学術分野全般にまたがる

- i. 研究データ管理(RDM) ← 社会的要請 → データの永続公開
- ii. 欧州HORIZON2020プログラムがオープンサイエンス化を促進 → Fair4Fusion (EuroFusion/Euratom)

④ ITER/JT-60SA/DEMO時代に向けて

- i. ITER + JT-60SA + JA-DEMO → 六ヶ所の遠隔実験センター(REC)にデータ蓄積. LHD+SNETも?
【統合データインフラ(ハード)+プラットフォーム(ソフト)の研究開発が集眉の急】

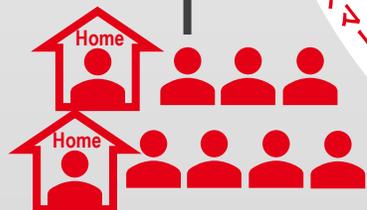
QST核融合科学情報センター構想

大規模実験データに基づくデータ駆動モデリング手法に関するNIFS、統数研、京大、九大との共同研究 2018~

HPCとAI開発両方に適した計算機アーキテクチャ/ストレージの検討 2020~

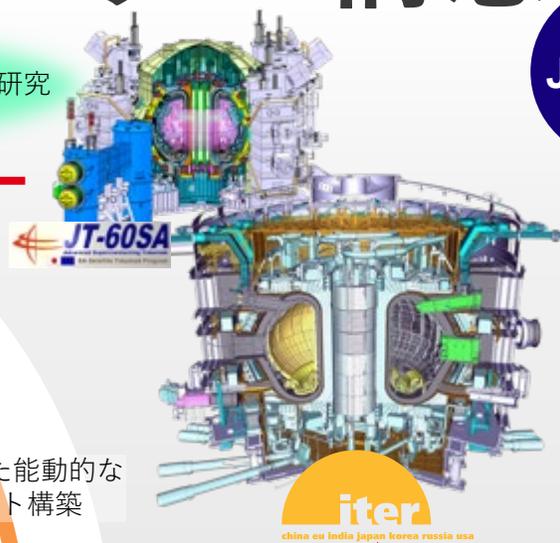
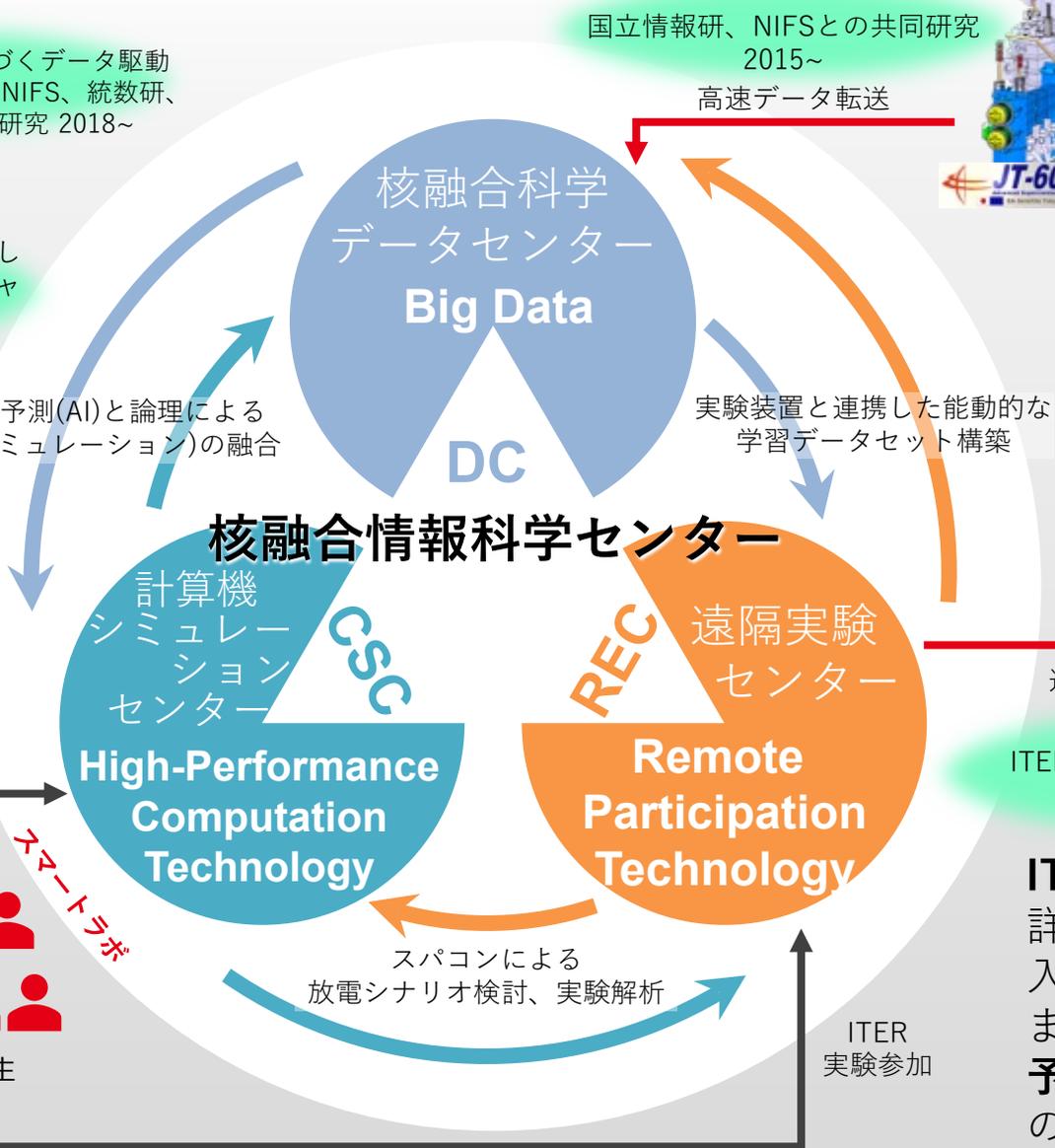
経験による予測(AI)と論理による予測(理論シミュレーション)の融合

国内研究者のためのシミュレーション・解析研究インフラ



国内研究者・学生

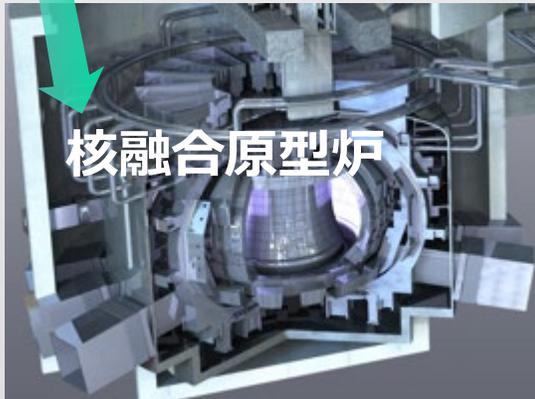
2021/9/1



JT-60SA



ITER



核融合原型炉

ITER, JT-60SA両方の詳細データセットが手に入る日本の優位を生かし、また理論と組み合わせたハイブリッドモデリングにより、**予測外挿性に優れたデータ駆動型モデル**を構築、信頼性の高い**原型炉**の性能・挙動予測につなげる。

NIFS/LHDで得られた卓越技術とデジタル資産の運用

- 世界に先駆けて、広帯域定常データ収集系を開発【学会賞】
↓
- プラズマ実験1回あたりの計測データ量世界一・・・記録更新中
 - 既に、ITER初期データ量見積り(1 TB/回)にほぼ到達
- 大陸間高速・長時間データ転送実験 (ITER → 日本)
 - ... 世界記録樹立するも、約3ヶ月で素粒子分野(欧→豪)に塗り替えられた
- 自律的データ解析プラットフォーム (EG+AutoAna)を開発【学会賞】
 - (恐らく)データ駆動型多段自動解析の先駆け的システム
 - ... 現在は、毎ショット200~300データが自動登録

IMASへの技術移植 & 国内核融合のデータ統合 & データ科学への提供

+

共通プラットフォーム+大量データ資産 ⇒ 人材教育の場 & 共同開発

『核融合クラウド』が今後取組むプラットフォームテーマ

- ① 国内(核融合)実験データ取り扱いの共通化
 - i. LHD/SNET遠隔データ収集保存プラットフォームのサービス範囲拡大 → 京大・工繊大・他
 - ii. コミュニティとの共通データ利用規範(案)の検討・作成

- ② 実験データ解析＋数値モデル計算 … 統合データインフラ＋(ソフト)プラットフォームの実現
 - i. 統合データインフラの構築 = 実験データ・ストレージ＋スパコン… QSTが先行(IT業界とも連携)
 - ii. 統合プラットフォーム・ソフトウェアの研究開発… QSTが弱い
→ NIFS/LHD成果の遠隔実験、EG+AutoAna 技術をIMASと統融合して「世界標準」を狙う

 **戦略的パートナーシップ:
QST連携、ITER連携**

- ③ オープンサイエンス&オープンデータの潮流
 - i. LHD資産の実験データに永続識別子(DOI)を付与 → 永続公開(オープンデータ化) → データ科学
 - ii. 他分野に先駆け、「Data Citation/Reference」本格運用を開始…計測(データ)名×ショット番号
→ (JSTの全面協力を得て)LHDデータのDOI登録件数でも世界一に!

- ④ ITER/JT-60SA/DEMO時代に
 - i. ITER + JT-60SA → 六ヶ所を国内核融合データ集積拠点に ← LHD/NIFS + 大学等SNET遠隔収集データ

RAW DATA FLOW & STORAGE

- Distributed DAQs, 1st/2nd storages & data archiving through 10/40 Gbps high-speed network
 - I/O multiplicity, bi-directional r/w
 - Data retrieval from any stages
- ✓ HDD : r/w = 900 / 450 MB/s
SSD : r/w = 11 / 9 GB/s

