

# E3 コンテキストを説明する メタデータ ～核融合実験データの実例から～

Japan Open Science Summit 2023  
Session E3

自然科学研究機構 核融合科学研究所 中西秀哉



# 核融合実験装置の例

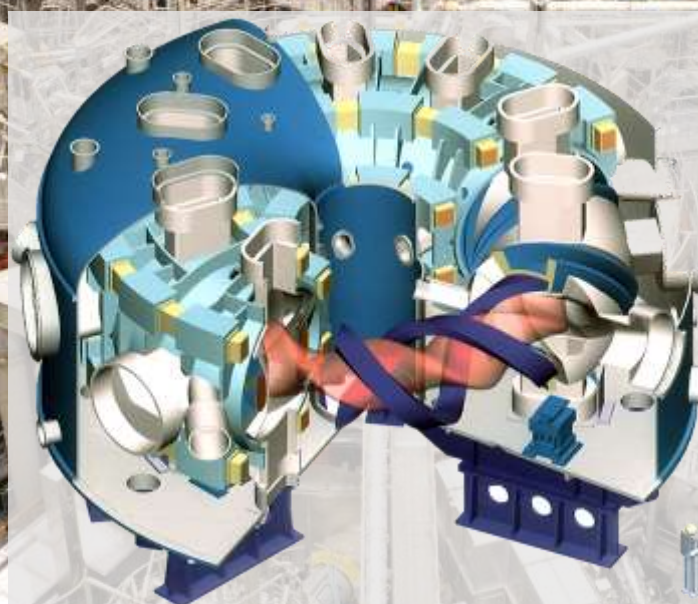
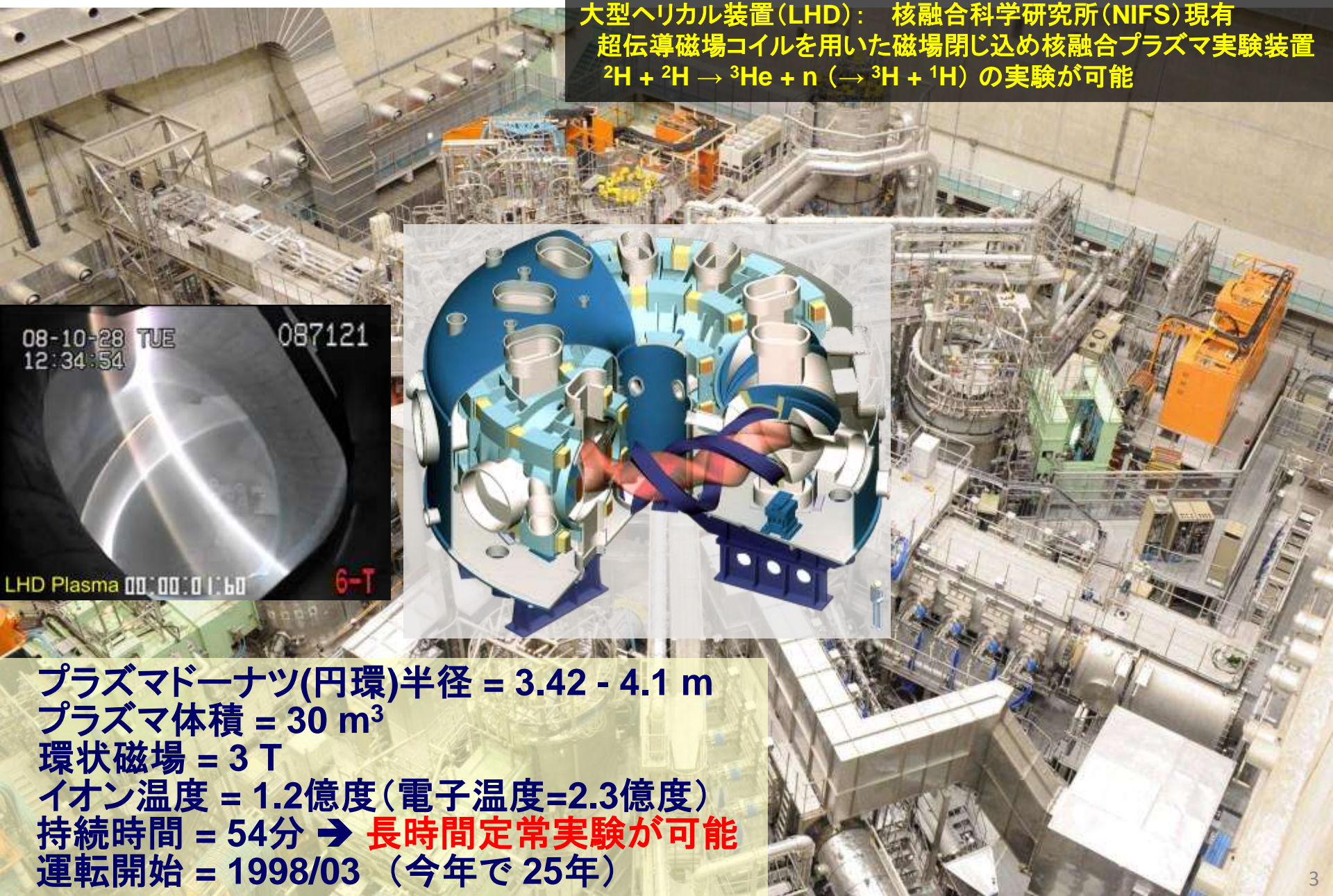
大型ヘリカル装置(LHD)： 核融合科学研究所(NIFS)現有  
超伝導磁場コイルを用いた磁場閉じ込め核融合プラズマ実験装置  
 $2\text{H} + 2\text{H} \rightarrow {}^3\text{He} + \text{n} (\rightarrow {}^3\text{H} + {}^1\text{H})$  の実験が可能

プラズマドーナツ(円環)半径 = 3.42 - 4.1 m  
プラズマ体積 = 30 m<sup>3</sup>  
環状磁場 = 3 T  
イオン温度 = 1.2億度 (電子温度=2.3億度)  
持続時間 = 54分 → **長時間定常実験が可能**  
運転開始 = 1998/03 (今年で 25年)



# 核融合実験装置の例

大型ヘリカル装置(LHD)： 核融合科学研究所(NIFS)現有  
超伝導磁場コイルを用いた磁場閉じ込め核融合プラズマ実験装置  
 ${}^2\text{H} + {}^2\text{H} \rightarrow {}^3\text{He} + \text{n} (\rightarrow {}^3\text{H} + {}^1\text{H})$  の実験が可能



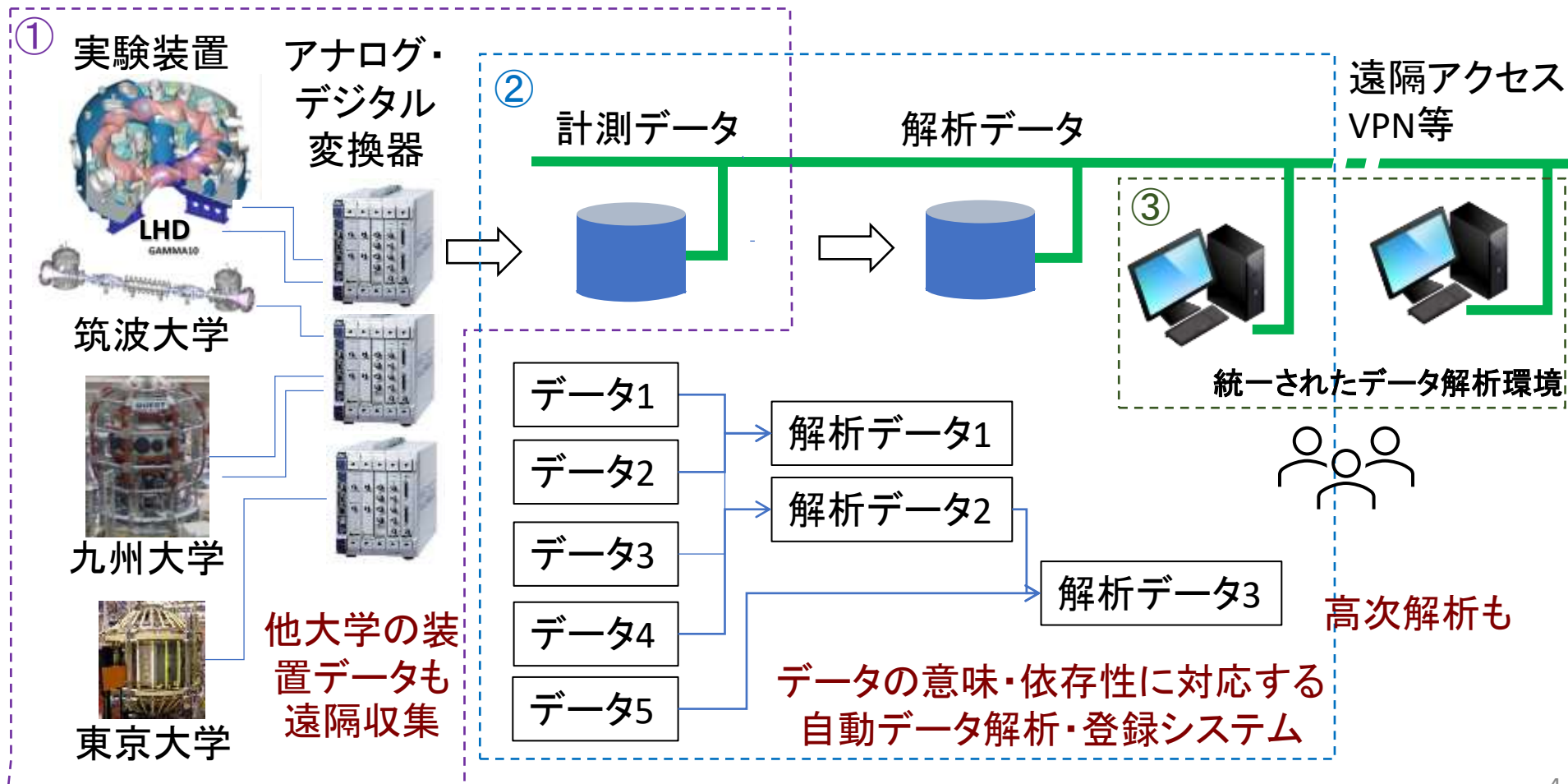
プラズマドーナツ(円環)半径 = 3.42 - 4.1 m  
 プラズマ体積 = 30 m<sup>3</sup>  
 環状磁場 = 3 T  
 イオン温度 = 1.2億度 (電子温度=2.3億度)  
 持続時間 = 54分 → 長時間定常実験が可能  
 運転開始 = 1998/03 (今年で 25年)



# 自動データ集録～解析～登録 → 参照

LHDデータ自動処理の流れ …… 1日に約180回、～8,000回／年

- ① 実験シーケンス(3分周期) → 計測データの収集(トリガ) → DB登録【130種】
- ② 計測データ登録を検知 → 自動データ解析 → DB登録【200～400個】
- ③ 計測・解析結果の自動可視化 + 対話的アクセスツール

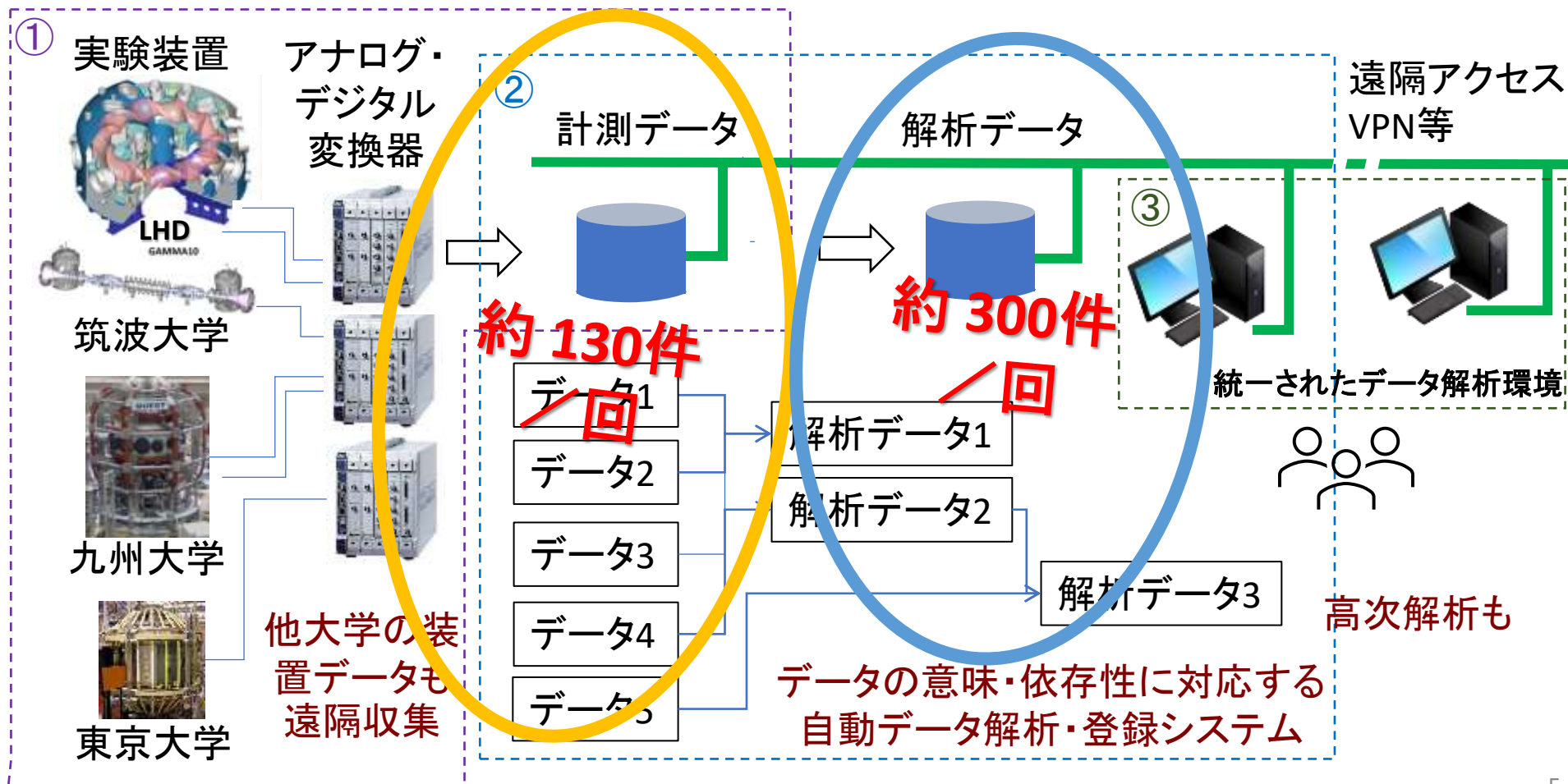




# 自動データ集録～解析～登録 → 参照

LHDデータ自動処理の流れ … 1日に約180回、～8,000回／年

- ① 実験シーケンス(3分周期) → 計測データの収集(トリガ) → DB登録【130種】
- ② 計測データ登録を検知 → 自動データ解析 → DB登録【200～400個】
- ③ 計測・解析結果の自動可視化 + 対話的アクセスツール

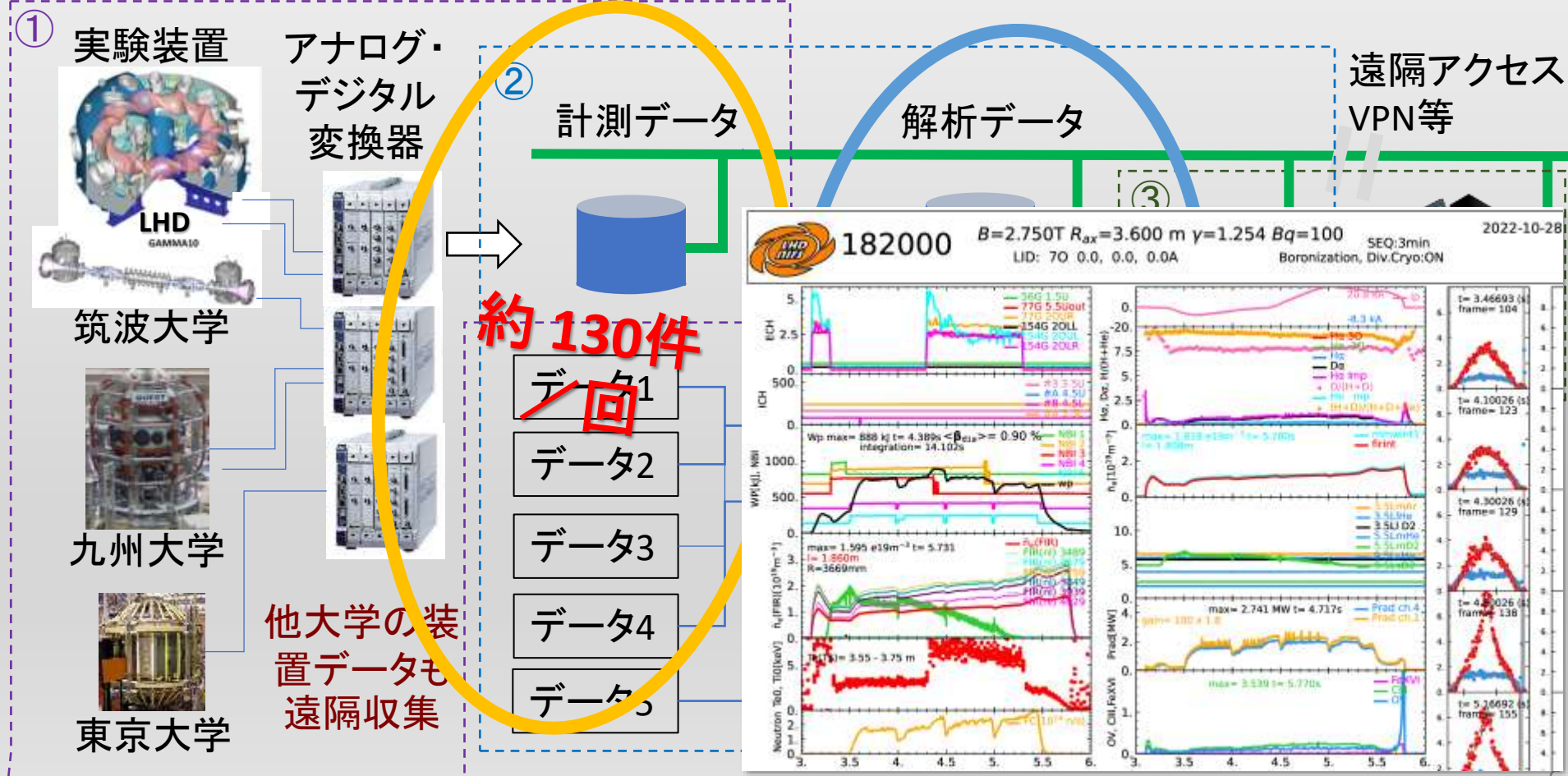




# 自動データ集録～解析～登録 → 参照

LHDデータ自動処理の流れ … 1日に約180回、～8,000回／年

- ① 実験シーケンス(3分周期) → 計測データの収集(トリガ) → DB登録【130種】
- ② 計測データ登録を検知 → 自動データ解析 → DB登録【200～400個】
- ③ 計測・解析結果の自動可視化 + 対話的アクセスツール



# FAIR原則の実装状況

現行LHD(+国内)システムで機能提供中  
今後に向け要件定義・追加実装が必要

## To be **F**indable: (見つけられるために)

- ▶ F1. (メタ) データが、グローバルに一意で永続的な識別子 (ID) を有すること。 ← PID
- ▶ F2. データがメタデータによって十分に記述されていること。
- ▶ F3. (メタ) データが検索可能なリソースとして、登録もしくはインデックス化されていること。
- ▶ F4. メタデータが、データの識別子 (ID) を明記していること。

## To be **A**ccessible: (アクセスできるために)

- ▶ A1. 標準化された通信プロトコルを使って、(メタ) データを識別子 (ID) により入手できること。
- ▶ A1.1 そのプロトコルは公開されており、無料で、実装に制限が無いこと。
- ▶ A1.2 そのプロトコルは必要な場合は、認証や権限付与の方法を提供できること。 ← 認証・権限
- ▶ A2. データが利用不可能となったとしても、メタデータにはアクセスできること。 ← データ履歴

## To be **I**nteroperable: (相互運用できるために)

- ▶ I1. (メタ) データの知識表現のため、形式が定まっていて、到達可能であり、共有されていて、広く適用可能な記述言語を使うこと。
- ▶ I2. (メタ) データがFAIR原則に従う語彙を使っていること。
- ▶ I3. (メタ) データは、他の(メタ) データへの特定可能な参照情報を含んでいること。

## To be **R**e-usable: (再利用できるために)

- ▶ R1. メタ(データ)が、正確な関連属性を豊富に持つこと。
- ▶ R1.1 (メタ) データが、明確でアクセス可能なデータ利用ライセンスと共に公開されていること。
- ▶ R1.2 (メタ) データが、その来歴と繋がっていること。 ← データ履歴
- ▶ R1.3 (メタ) データが、分野ごとのコミュニティの標準を満たすこと。 → 標準的形式が無い!?

→ 参照情報 (Data References)

# NIFS 実験データDOI登録の諸課題

- データ総件数・・・計測・解析データ・解析コード、全て公開済
  - LHDのデータセット総件数は、凡そ、3,000万件  
 $Total = (130 + 300) \times 180,000 \div 2 \approx 30,000,000$ 
    - ✓ (これに、大学装置分が追加されますが、件数は少ない)
    - ✓ **データ粒度** = 1ファイル(30~40波形)で**平均500 MB**・・・十分大きい
    - ✓ (実験中)データ生成率 = 500個/回 × 180回/日 = **9~10万個/日**
  - DOIランディングページの**動的生成**プログラムを開発・公開 (右図→)
- 登録速度・期間・経費
  - 年1,000万件登録 × 3年で、全LHD計測・解析データ登録完了する!?  
→ メタDB連動のDOI登録自動化ツール開発に着手。
  - JaLCシステム残容量切迫 → 現状、3,000万件は登録NG、要拡張。
  - JSTによって、R4年度に1,000万件の登録テスト ← NIFSデータ提供  
→ 処理可能と判明。但し、他ユーザの検索処理に悪影響の恐れ  
→ 30万件/日あたりが実効的上限。
- データ検索・参照方法と、DOI登録の「粒度」
  - ✓ 分野内では、2つの主キー【計測(解析)データ名】、【実験番号】で検索・参照する・・・実験データシステムは数あれど、検索形態は同じ
  - ✓ この「粒度」を変えると、分野内のデータ(再)利用性が大きく低下する?
    - 【データ名】でDOI登録 → 25年、18万回に及ぶデータ・・・不連続、同一性?
    - 【実験番号】でDOI登録 → 利用方法の異なる500種データセット・・・参照困難

Fusion Diagnostics Data Repository

LHD FPellet #182000.1

**Data description**

Reaching Helium Filaments (FPellet): Inject hydrogen isotopes pellets into core plasma to control plasma density. The in situ probe can type 2D-lenslet injector and the arrow extruder type repetitive pellet injector are implemented.

**Creator(s)**

YASUNORI Iwano<sup>1,2</sup>, YOSHIOHISA Iwano<sup>2,3</sup>  
 national institute for fusion science (NIFS)

**Acquisition status**

Completed with no errors

**Primary info.**

Sheet	182000
SubChnl	1
Management	50.0.0
Acquisition Date	Fri, 18 Dec 2022 17:05:11 +0900
Acquisition Date	Fri, 18 Dec 2022 17:05:11 +0900
ModuleGroup	FOC
Position	1
Channels	16
AcquiredChannels	16
Comment	
Rights	Coverage by National Institute for Fusion Science (NIFS)
Access Rights	Rights and Status of LHD Data
AIRC	Not charged
Publisher	National Institute for Fusion Science (NIFS)
Identifier	10.17461/nifs.data.182000
Identifier Registration	
Relation	
Funding Reference	LHD-Project-Diagnos (LHD project, WPI)
Resource Type	Dataset
Version	1.0
File	Name: /Project/182000/Exp... Size: 4866271 Type: CSV

**Channel parameters**

Name	Value	Type
Sheet	182000	INT
SubChnl	1	INT
ChannelNumber	1	INT
ModuleGroup	Component/FOC	STRING
ModuleType	FOC133	STRING
ConnectTo	/Dev/FPel_Trig	STRING
connectPort	/Dev/NET	STRING
DeviceID	/Dev/FPel_Trig	STRING
NumOfCoefficients	2	INT
VCoefficient	0	DOUBLE
VCoefficient	0.0003517518125	DOUBLE
RawDataResolution	1bit	STRING

**Timing summary**

Delay (sec)	Interval (sec)	Number of samples
0.000000	0.000001	1300000

**Trigger**

Channel	Delay time(sec)	Pulse width(sec)	Repetition count	Repetition time(sec)
TRIG1	0.000000	0.000000	1	0.000000

**Clock**

Channel	Delay time(sec)	Pulse width(sec)	Repetition count	Repetition time(sec)	Magnification	Range	Frequency
TRIG1	0.000000	0.000000	1	0.000000	1	1bit	1MHz

**Signal Plot**

LHD FPellet #182000.1





# OS化に向けた対応と分野的な背景

## 1. 今後のデータシステムOS化

1. プライベート/グループ研究環境を、**実験データストレージ**と透過的に接続。  
← NII RDC フレームワークを最大限活用 (CiNii, JAIRO Cloudは機関導入済)
2. データシステムの**認証・権限管理機能強化**・・・研究データエコシステム構築事業で実施
3. 大規模実験・・・**運転値・設定パラメータ**など、複数メタデータ媒体に分散保管。  
→ DOIランディングページに纏め、一般公開を進める → **(メタ)データ散逸防止**。

## 2. 研究データとしての特徴、課題

1. 総データ容量は ~1.5ペタバイト(圧縮後)・・・圧縮でサイズ1/3 (→ コスト削減目的)  
総データセット数は ~4,000万件、参照数は 年間3,000万ダウンロード以上。
2. 大規模実験・・・所属・業務指示が異なる、**多数多様な関係者が装置運転・保守に従事**  
(研究者、技術者以外に、メーカー保守員、運転監視員)  
→ 不統一なメタデータ(**RDBMS/オフライン/紙手書き**)が日々生成
3. 「**データ標準化**」の文化がない・・・「**標準化による(異分野)データ再利用**」=OSと衝突!?
4. 学術研究基盤事業による**LHD実験プロジェクトは2025年度に終了予定**。その後の体制?



# OS化に向けた対応と分野的な背景



## 1. 今後のデータシステムOS化

1. プライベート/グループ研究環境を、**実験データストレージ**と透過的に接続。  
 ← NII RDC フレームワークを最大限活用 (CiNii, JAIRO Cloudは機関導入済)
2. データシステムの**認証・権限管理機能**強化・・・研究データエコシステム構築事業で実施
3. 大規模実験・・・**運転値・設定パラメータ**など、複数メタデータ媒体に分散保管。  
 → DOIランディングページに纏め、一般公開を進める → **(メタ)データ散逸防止**。

## 2. 研究データとしての特徴、課題

1. 総データ容量は ~1.5ペタバイト(圧縮後)・・・圧縮でサイズ1/3 (→コスト削減目的)  
 総データセット数は ~4,000万件、参照数は 年間3,000万ダウンロード以上。
2. 大規模実験・・・所属・業務指示が異なる、**多数多様な関係者が装置運転・保守に従事**  
 (研究者、技術者以外に、メーカー保守員、運転監視員)  
 → 不統一なメタデータ(**RDBMS/オフライン/紙手書き**)が日々生成
3. 「**データ標準化**」の文化がない・・・「標準化による(異分野)データ再利用」=OSと衝突!?
4. 学術研究基盤事業による**LHD実験プロジェクトは2025年度に終了予定**。その後の体制?



# 標準スキーマ、標準データ形式

## □ 標準スキーマから逸脱するコンテキスト

1. 装置プロジェクトのサイズ感 → 関係者の総数、存続年数(核融合は30年超)
2. 運転形態 → 連続何日・何時間、何回、2~3運転シフト … Creatorの意味が変わる  
→ 運転間隔(何秒、何分、何時間)

## □ 標準データ形式

- 存在しない分野もある  
→ 核融合:非可読なボーン・デジタル・データ → API関数呼出し(ファイル化の利点がない)
- 核融合実験(計測)では、プラズマ(イオン・電子)温度・密度、磁場、発光(電波~赤外~X/γ線)、粒子、不純物元素量(水素以外)、流速、それらの揺動など、多様な計測を同時に扱う。  
(0次元量 or 時間変化 or 2次元(断面)空間分布 or 3次元量 なのか、、、)
- 小型装置で~10計測、LHDのような大型実験だと100計測超~
- データ形式が多種多様で共通化が難しい → 標準化は行われぬ  
(反面、メジャーな実験データ集録“システム”はあります) … 次世代実験ITERが標準になる!?

## □ 研究ノート・運転日誌から、電子ログ・ブック(e-Logbook)へ

- 核融合の大型実験では、運転や研究データに関して、紙の研究ノートは無くなって久しい。  
… システムによる(半)自動ロギングや、Webポータル(実験予定表・実験報告書)に移行
- 反面、運転監視者による業務日報など、オフラインのメタデータは存在する。(全容把握が困難)  
… 装置の状態(情報)を内包している場合もある。

# NIFS 大型実験のコンテキストを残す

- 建設に10年、運転期間は30年が一般的。
- 装置サイズに応じ、仕様書・設計検討書・設計図等々が数多存在
- コンテキストを残す ≡ アーカイブズ(科学的史料として)
- 保管・再利用・公開方法の検討が必要。看過すれば散逸霧消。



- Web掲載されたPDF等も、必ずしも検索可能になっておらず、FAIR原則に沿ったアクセス性の確保が必要。

実験予定表・報告一覧 Japanese English

Cycle 24

実験日誌(まとめ)

実験日	日	種別	報告書
第1週	2022/09/29 (Thu)	[multi-ion]	
2022/09/30 (Fri)	[multi-ion]	[multi-ion]	
2022/10/03 (Mon)	[multi-ion]	[multi-ion]	
2022/10/04 (Tue)	[multi-ion]	[multi-ion]	
第2週	2022/10/05 (Wed)	[multi-ion]	[multi-ion]
2022/10/06 (Thu)	[multi-ion]	[multi-ion]	
2022/10/07 (Fri)	[multi-ion]	[multi-ion]	
2022/10/10 (Mon)	[multi-ion]	[multi-ion]	
2022/10/11 (Tue)	[multi-ion]	[multi-ion]	
2022/10/12 (Wed)	[multi-ion]	[multi-ion]	
2022/10/13 (Thu)	[multi-ion]	[multi-ion]	
2022/10/14 (Fri)	[multi-ion]	[multi-ion]	
2022/10/17 (Mon)	[multi-ion]	[multi-ion]	
2022/10/18 (Tue)	[multi-ion]	[multi-ion]	
第4週	2022/10/19 (Wed)	[multi-ion]	[multi-ion]
2022/10/20 (Thu)	[multi-ion]	[multi-ion]	
2022/10/21 (Fri)	[multi-ion]	[multi-ion]	