

がん全ゲノム体制班の検討状況について

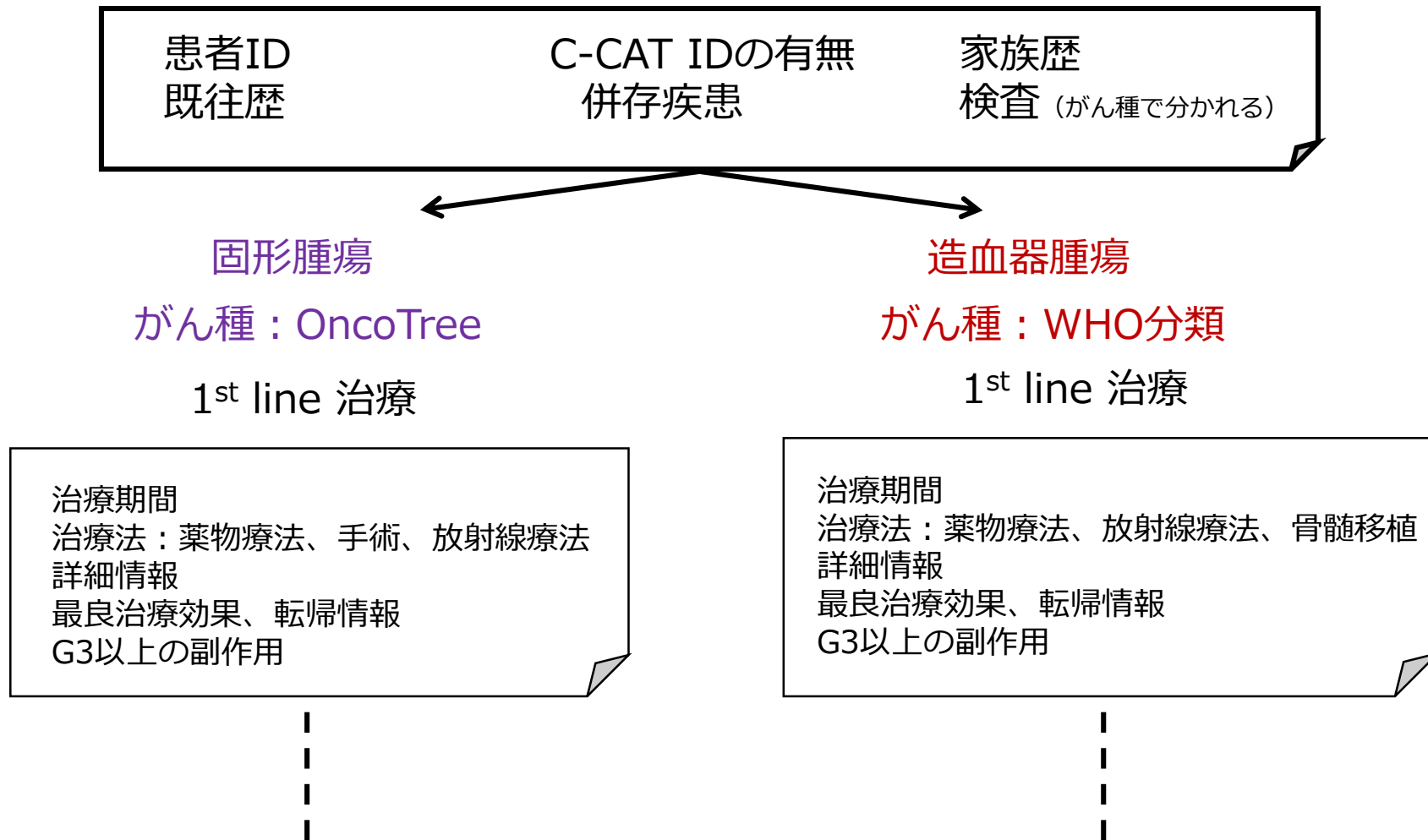
～がん全ゲノム体制班専門WG提出資料～

1. バイオバンクWG P2
2. 解析WG P6
3. データ共有WG P14
4. ELSI WG P20

バイオバンクWG

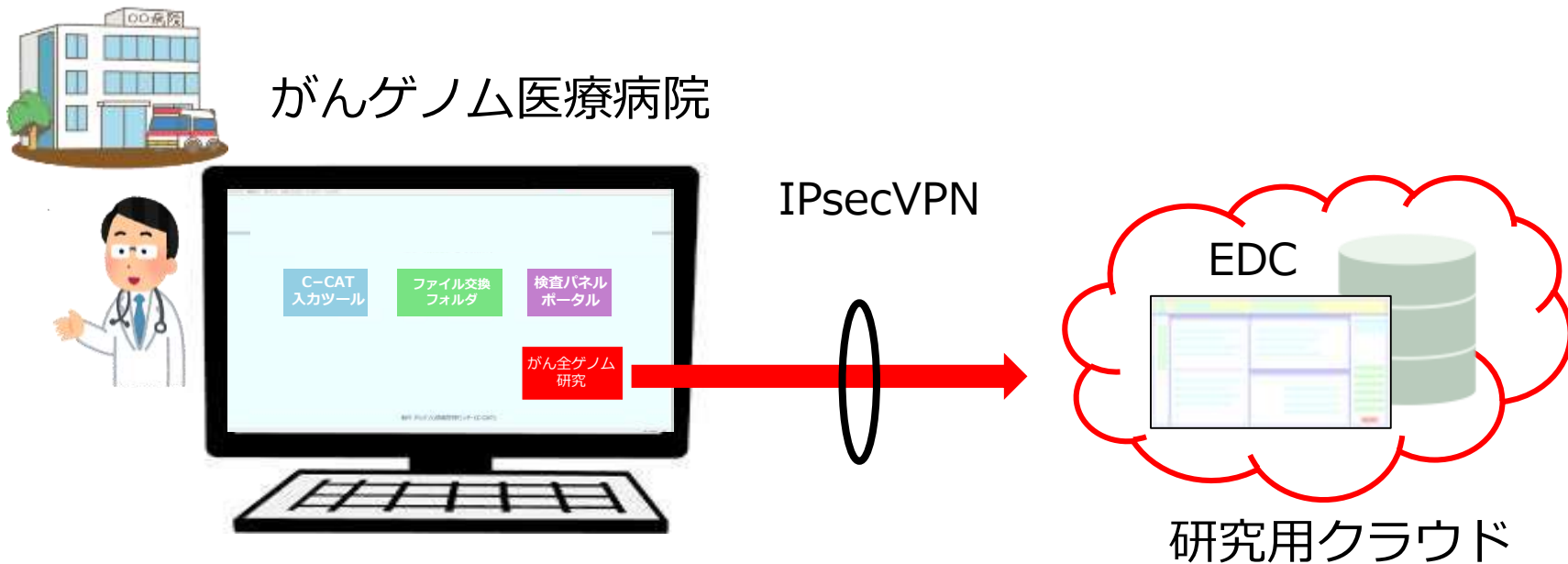
収集臨床情報の内容

固形腫瘍臨床情報v1のヒアリング + 造血器腫瘍臨床情報v1作成 →
収集臨床情報統合版v1.4作成 → 南谷先生ヒアリング → v2としてWGおよび製薬協に配布



臨床情報収集システム

情報セキュリティが確保された臨床情報収集システム案



EDC開発ベンダーと現在契約手続き

※がんゲノム医療で用いられている既存のセキュリティが確保された回線を活用

検体・臨床情報収集ロジスティクス

既存検体（先行解析）

- ・ 腫瘍部の新鮮凍結検体（-80℃以下）およびペア正常検体が既に保存され、網羅的解析・結果の公的データベースへの登録が可能な検体から解析をする
- ・ 検体保存方法は各バンクの既存インフラを用いる
- ・ 核酸の調整は各バンクに依頼
- ・ 臨床情報入力は、各参加機関がCRC等、適切な人材を雇用して入力。研究費にその費用を計上

前向き収集検体（本格解析）

- ・ 統一ICFひな形に準じた倫理申請
- ・ 検体収集、保存のSOPを作成する必要あり
- ・ メタボローム、プロテオーム解析が出来るよう血清・血漿を保存？
- ・ 検体保存インフラは各施設に設置？
- ・ 核酸の調整は各施設？またはシークエンス拠点？
- ・ 臨床情報入力は、各参加機関がCRC等、適切な人材を雇用して入力。研究費にその費用を計上

解析WG

解析WG

◎小川誠司、油谷浩幸、上田宏生、浦上研一、岡田随象、片岡圭亮、柴田龍弘、白石友一、鈴木 穰、十時 泰、中村祐輔、南谷泰仁、藤本明洋、森下真一

解析	1	大規模解析を前提とした解析手法決定のため、パイロットプロジェクト支援、時系列解析、マルチオミックス、メタボローム解析可能な検体取り扱いガイドラインの決定、免疫ゲノム、エピゲノム
	2	標準パイプラインの開発、解析環境の決定 (CPU/GPU? Illumina/MGI?)
	3	データ解析拠点の構築
	4	1次データ共有クラウドの構築

1. 大規模解析を前提とした予備的検討

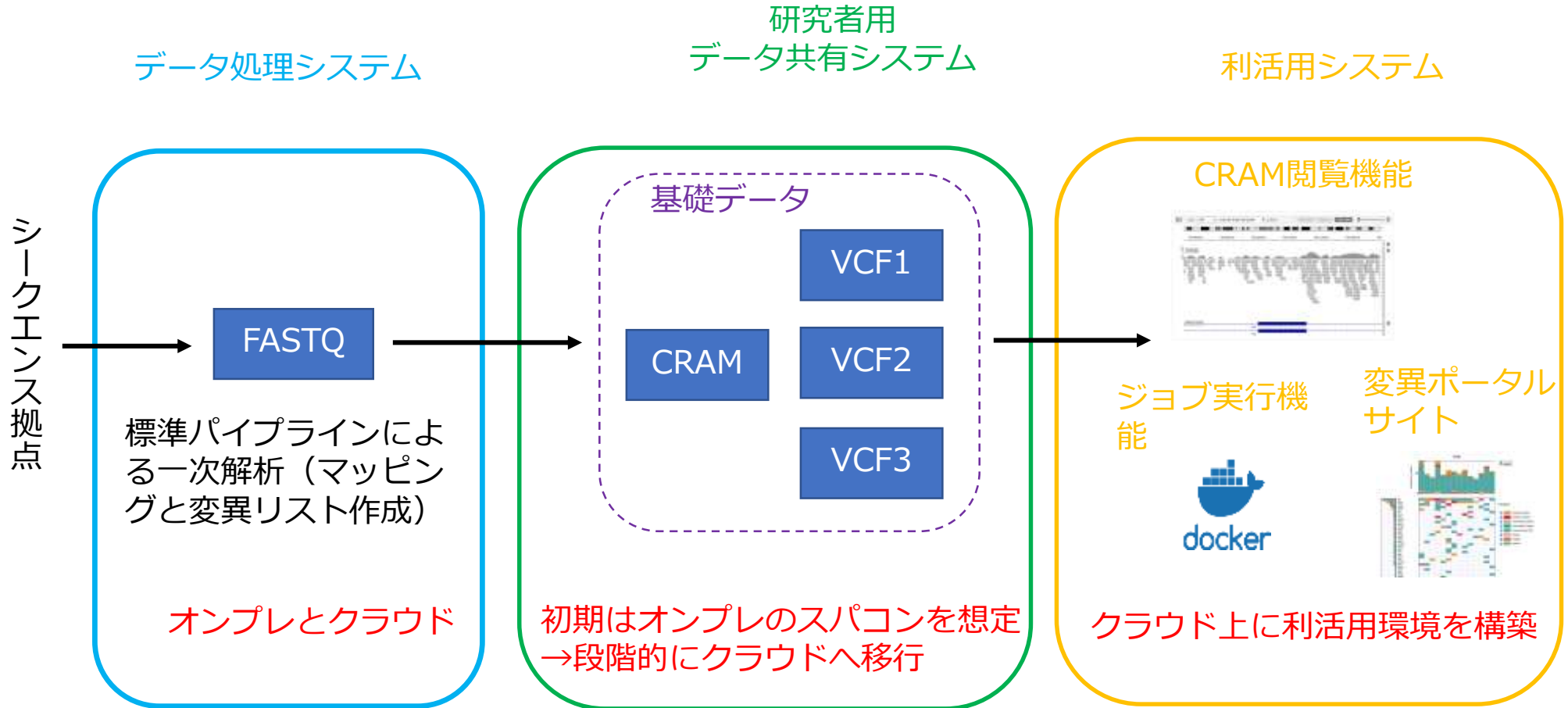
- シーケンスプラットフォームの検討 (進行中)
 - Long read seq
- 予備的解析 (進行中)
 - パイロットプロジェクト支援
 - 時系列解析
 - マルチオミックス解析(エピゲノム、免疫ゲノム、単一細胞)

2. 標準パイプラインの開発、解析環境の決定

3. データ解析拠点の構築

4. 一次データ共有クラウドの構築

データ解析の流れの概要

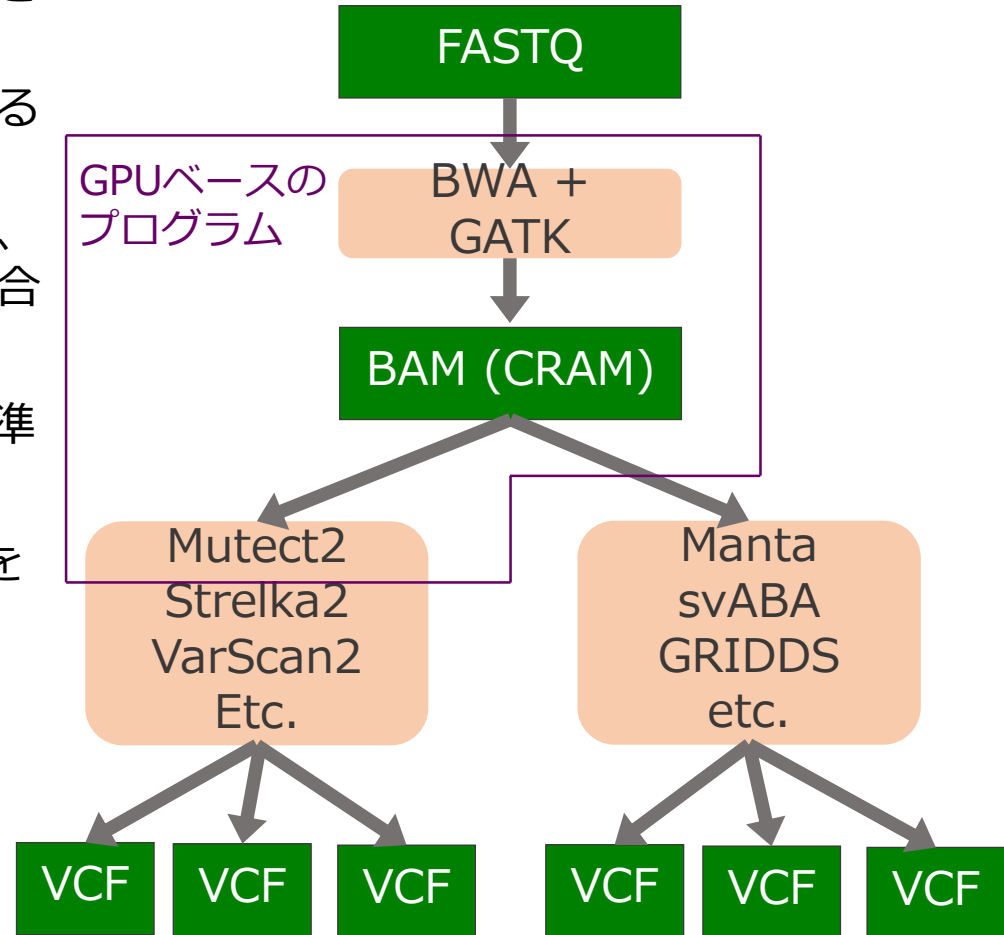


5年で10万人のdeep WGS、1万人のlong read sequence実行

GPUサーバー: 30台, CPUコア数: 10,000, ストレージ: 30PB。他に標準パイプラインなどのバイオインフォマティクス開発、各システムの開発・評価・運用のための人件費、クラウド利用料、企業委託費などが必要

標準パイプラインの案

- 予め主要なソフトウェアについては解析結果を配備しておく。
 - ユーザーが独自にプログラムを実行させる手間を省く。
 - 国際がんゲノムコンソーシアムなどでは、複数の変異検出ソフトウェアの手法を統合的に利用することが一般的。
- アラインメントの方法はできるだけ世界の標準に合わせてクセのない形を選定すべき。
- 解析WGなどで、アップデートを続ける体制を構築。
- パイプラインの対象
 - 全ゲノム解析 (germline, somatic)
 - トランスクリプトーム解析
 - バイサルファイトシーケンス解析
 - ロングリード?



* 基本的にパイプラインやその評価結果は、Githubなど全てオープンソースで公開し、広く閲覧可能とする(科学的透明性の担保)。

「データ処理システム」管理者の役割と体制

・ 役割案

- ・ 解析パイプライン開発
 - ・ 解析WGの要望に合わせて、種々のツールを評価する。
 - ・ 標準化解析パイプラインのメンテナンス。
 - ・ ドキュメントの整備。
- ・ 解析パイプラインの実行
 - ・ 上記で開発した解析パイプラインをオンプレの環境やクラウド上で実行して、所定の場所に格納する。

・ 体制案

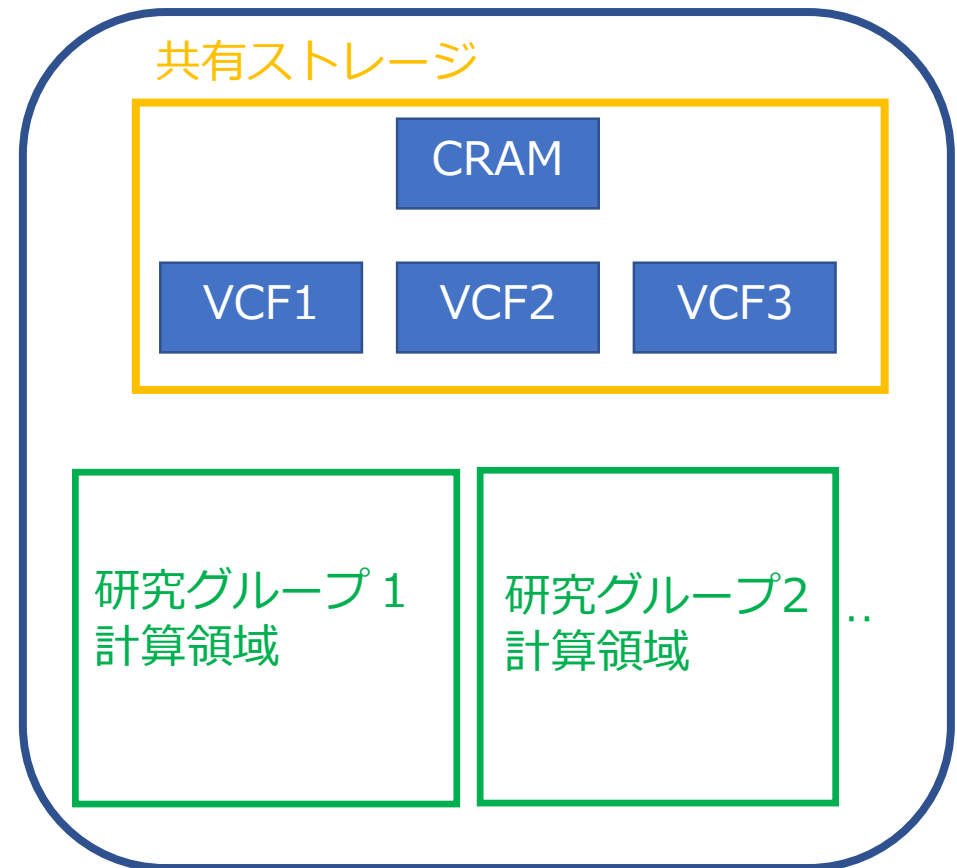
- ・ 基本的に外部委託が中心。
- ・ 恒久的な予算が確保されるのであれば、専門のITスタッフの雇用・教育を目指す。
- ・ 個別の研究グループの支援は行わない(今後の検討を要する)。ul>- ・ 解析パイプラインの決定などは解析WGなどの会議体を通じてなされる。
- ・ 人的リソースに著しい制限があることから、個別の研究グループの個々の要望に対応して解析をすること原則として行わない。
- ・ 解析を推進するための人材の育成はプロジェクトの成否に関わるもっとも重要な課題の一つ。

「データ処理システム」の進め方

- 解析パイプラインの開発
 - short readの全ゲノムにおいて、基本的な解析パイプラインの実行
- 解析環境の構築
 - オンプレの計算機器、クラウドのそれぞれでパイロット解析を実行してみる。
 - 解析能力、費用、コスト、運用上の問題点などをまとめるレポートを出す。
 - このレポートに応じて、選定を行う。

研究者用データ共有システム（オンプレ）

- 既存のオンプレの計算リソースを利用する形。
- 共有のストレージにCRAM, VCFなどの基礎データを格納して、これに適切な権限を付与しつつ、各々のユーザーがアクセスする形。
- 共有ストレージの費用、管理代などは、「中央」で一括で支払う。



「研究者用データ共有システム」の進め方

- オンプレにおける環境の構築
 - 実際に国内のオンプレのスパコンで「共有ストレージ」のパイロット利用を行う。
- クラウドにおける環境の構築
 - クラウドの利用方針についての議論
 - クラウド上のデータ共有システムのパイロット版の構築

データ共有WG

データ共有WG

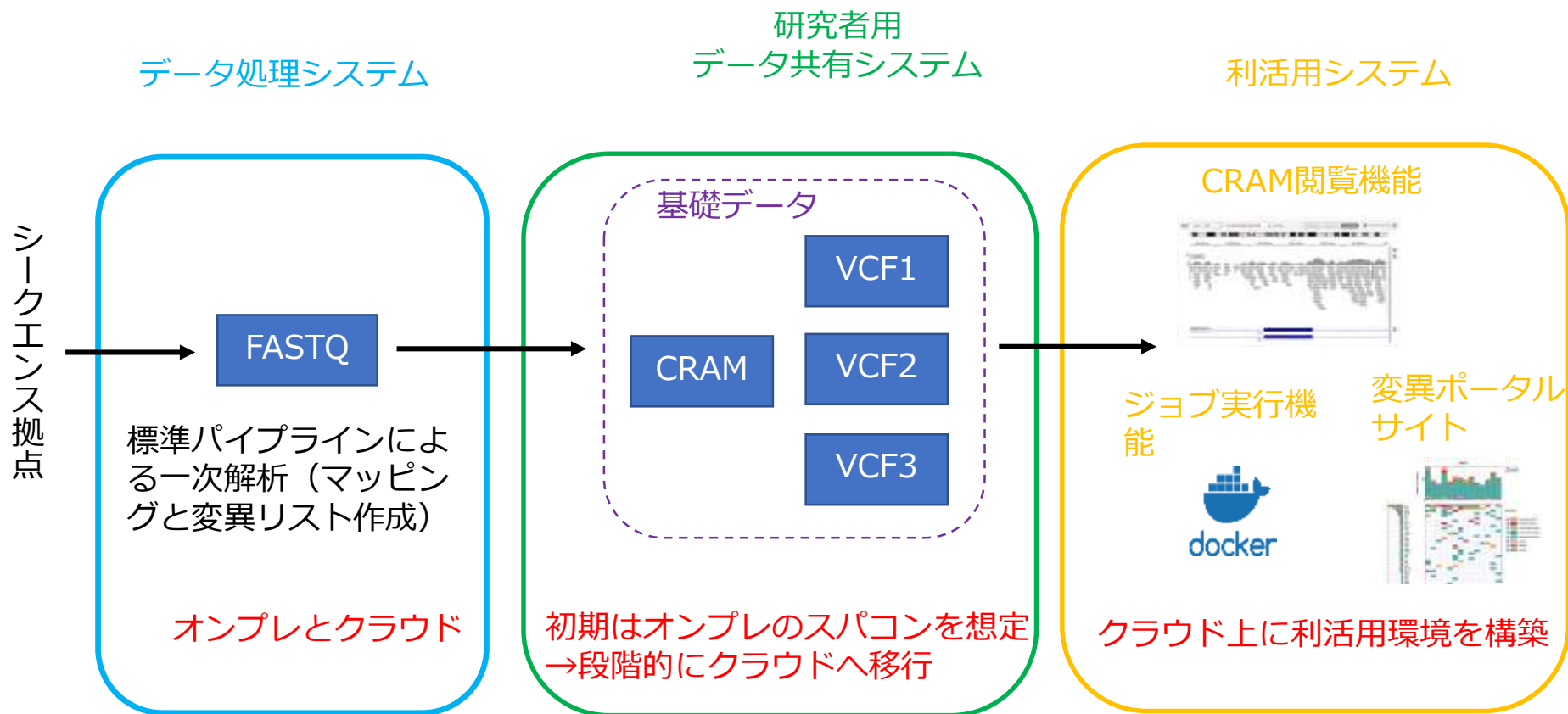
◎油谷浩幸、井元清哉、柴田龍弘、小川誠司、加藤護、浦上研一、宮野悟、森誠一、河野隆志、白石友一、白神昇平

検討事項	
1	○データ等の管理・運営方法の在り方について (進行中)
2	○データセキュリティについて世界的な標準化の動きと整合ルール
3	○データの二次利活用制度の整備・構築 (進行中)

WG内での検討内容

- がん全ゲノム情報共有に関する国際動向について
- がん全ゲノム情報共有に関する国内の取り組みについて
- がん全ゲノム情報共有に向けて創薬の観点から
- がん全ゲノム情報共有とスパコン連携について
- がん全ゲノム情報共有とクラウド環境について
- がん全ゲノム情報共有と臨床情報収集について

データ解析の流れの概要

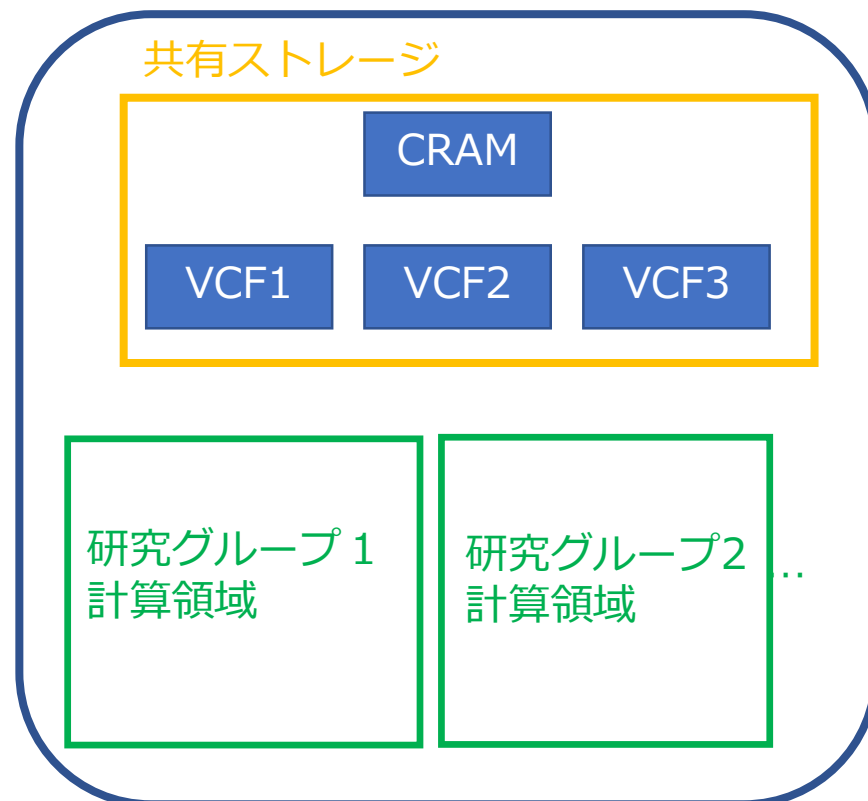


5年で10万人のdeep WGS、1万人のlong read sequence実行

GPUサーバー：30台、CPUコア数：10,000、ストレージ：30PB。他に標準パイプラインなどのバイオインフォマティクス開発、各システムの開発・評価・運用のための人件費、クラウド利用料、企業委託費などが必要

研究者用データ共有システム（オンプレ）

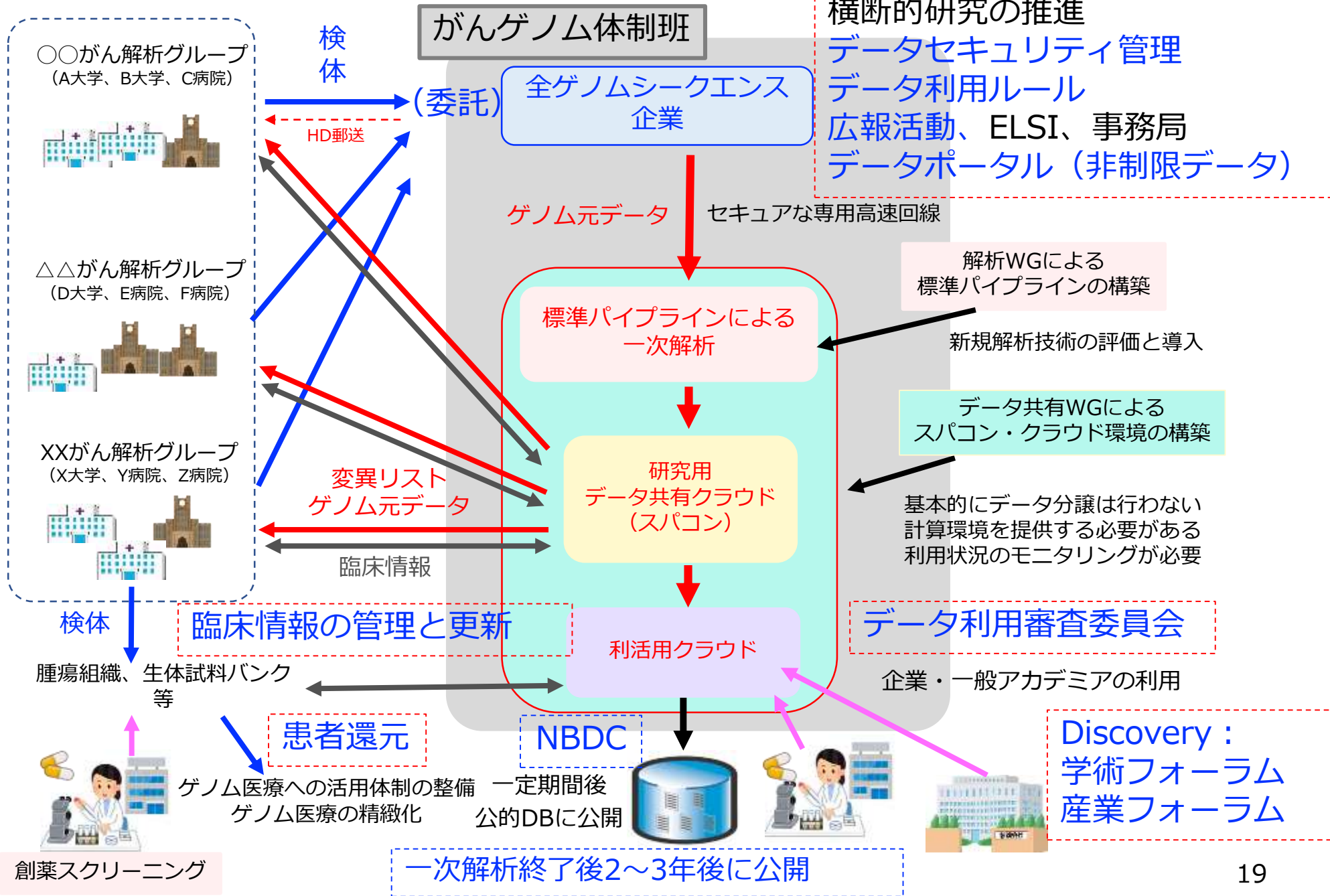
- 既存のオンプレの計算リソースを利用する形。
- 共有のストレージにCRAM, VCFなどの基礎データを格納して、これに適切な権限を付与しつつ、各々のユーザーがアクセスする形。
- 共有ストレージの費用、管理代などは、「中央」で一括で支払う。



「研究者用データ共有システム」の進め方

- オンプレにおける環境の構築
 - 実際に国内のオンプレのスパコンで「共有ストレージ」のパイロット利用を行う。
- クラウドにおける環境の構築
 - クラウドの利用方針についての議論
 - クラウド上のデータ共有システムのパイロット版の構築

データ管理・共有スキーム（案）



ELSI WG

ELSIワーキング検討状況

第2回がん全ゲノム解析等連絡調整会議（2020/10/27 オンライン）

ELSIワーキンググループ

◎武藤香織，横野 恵，安中良輔，井上悠輔，加藤和人，田代志門，丸 祐一，
森 誠一，山口 建

本WGの検討課題

1

新薬開発への活用や将来の追加解析に耐えうる包括的な同意取得の統一化に向けた検討

2

患者等へのリコンタクトも可能とする仕組みの構築に向けた検討

- いずれの課題についても具体的な取り組みは計画全体におけるシナリオの明確化が前提となるが、可能な範囲で検討を進める
- 厚労科研ELSI研究班（小杉班）、三菱総研委託調査をはじめとする関連研究・調査グループと連携し、既存の研究成果も有効活用しつつ検討を実施する
- 分析に際しては、今後予想されうる国内外の規制環境の変化も考慮する

ELSI
全般の情報
収集・分析

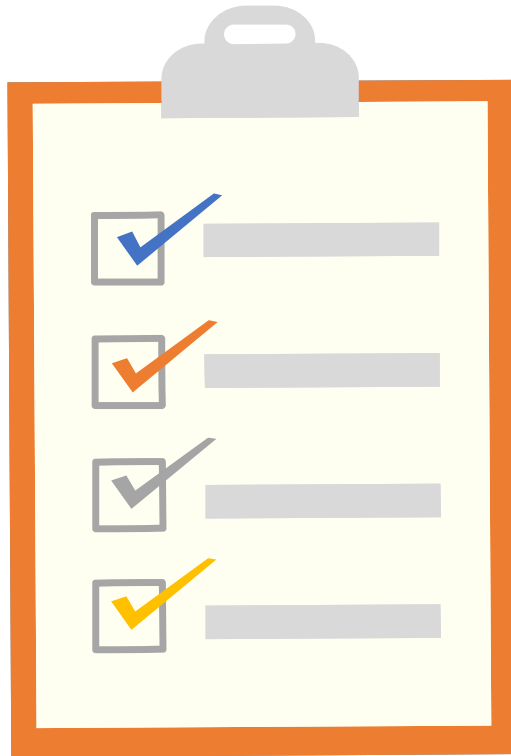
研究動向把握
課題予測

想定されう
るシナリオ
の分析

同意モデル
の提示

各モデルに
関わるEL
SIの整理

統一ICF



- 指針の定める説明事項を充足することが大前提
- 既存の類似の計画で用いられた I C F をベースに統一 I C F に盛り込む事項を整理する
- 必須事項とオプトアウト可能事項との区分けは難病の計画とも調整して決定する
- 動画をはじめとする説明補助ツールを作成して併用することを前提とする

I C が包括的になればなるほど、I C 以外の部分での倫理的配慮・ガバナンスの必要性・重要性が大きくなる。その部分は今後検討して計画に組み込む。その際、以下のもの等の国際的な議論・指針を参照する

- OECD, "Guidelines for Human Biobanks and Genetic Research Databases (HBGRDs)"
- GA4GH, "Global Alliance for Genomics and Health: Consent Policy"(v 2.0: September 2019)
- World Economic Forum, "Genomic Data Policy Framework and Ethical Tensions" (2020)

ICにおける説明事項

研究目的の範囲

実行計画が掲げる目的

- がん医療への活用
- 日本人のがん全ゲノム配列データベースの構築
- がんの本態解明
- 創薬等の産業利用



検討すべきポイント

- 保健医療全般やゲノム医療全般に関する目的の明示
- がん以外の疾患の研究への利用
- どの部分をオプトアウト対象とするか

新薬開発

- 新薬開発への活用は企業等がデータにアクセスして研究開発に利用する場合を想定（これで十分か？）
- 民間企業が商業目的で利用する可能性を明示
- 広範囲からのデータアクセスを想定した説明とし、データ利用に関するガバナンスの仕組みを示す
- 民間資金を活用する場合は適正な利益相反管理と研究参加者への説明等を通じた透明性の確保が重要

追加解析

- 追加解析の指針上の位置付け
 - ① 当初の同意の範囲
 - ② 既存試料・情報の利用
 - ③ 既存試料・情報の提供
- ①でどこまで対応できるようにするか
- 診療情報以外の情報の（追加）収集とデータリンケージは想定するか

リコンタクトの仕組みの検討

1

電子的 I C を採用

- ・ 説明の統一性確保
- ・ 同意内容管理
- ・ オンラインによるリコンタクト

2

オンラインプラットフォームの活用

- ・ 追加データ収集
- ・ データ連携
- ・ P P I や情報発信



先行研究が指摘する電子的ICの利点

効率の向上

ICFの内容の修正やトラッキングが容易

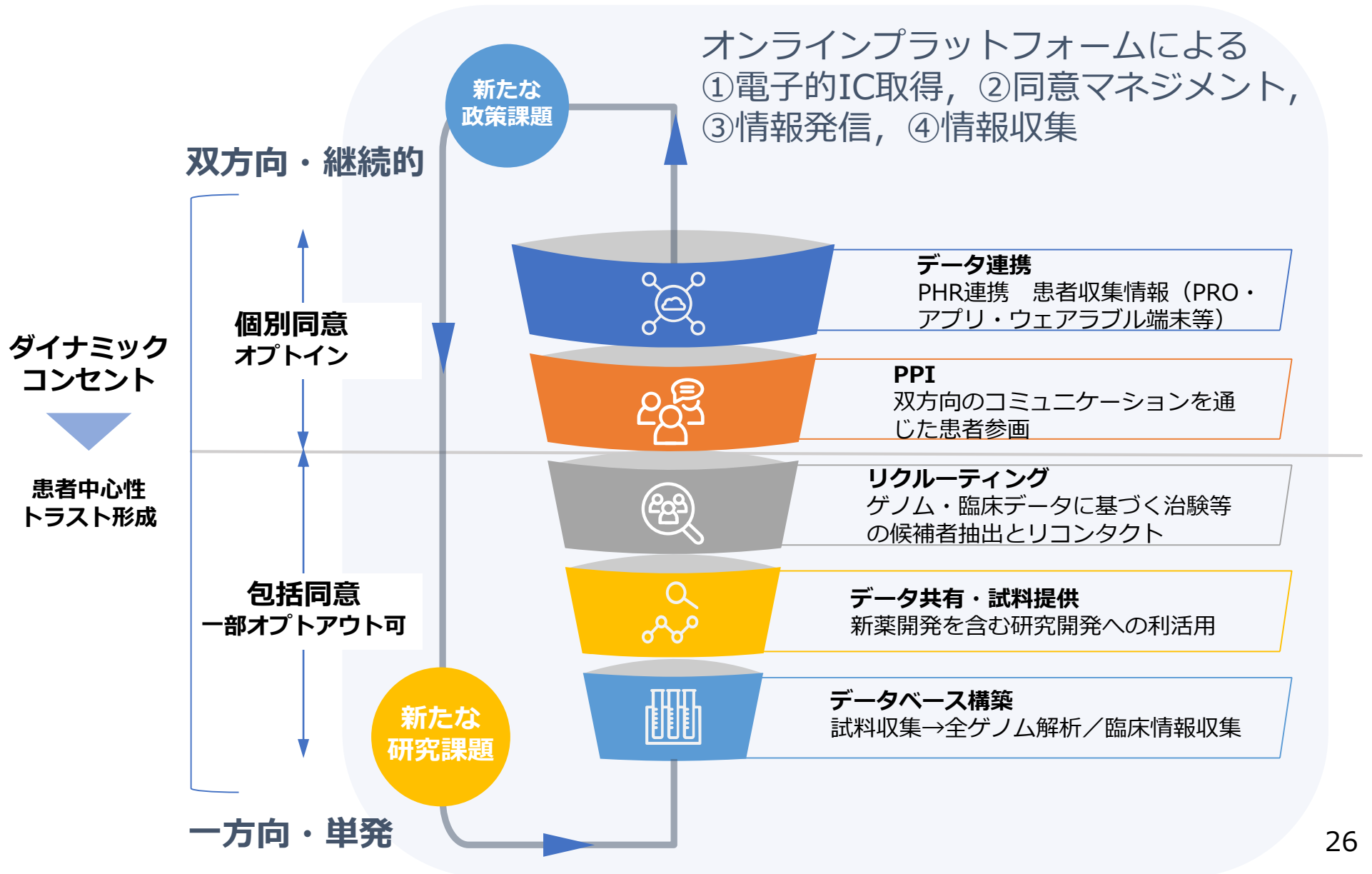
参加者の理解の向上

個々のペースに合わせて説明コンテンツ
を利用できる

同意内容の振り返り・確認が容易になる

参加率の向上

オンラインプラットフォームによる展開



産業利用についての検討



課題 1・2 に共通する論点であるため取り出して検討する



本計画で想定される産業利用のシナリオを具体化した上で各シナリオに関わる E L S I を分析して提示する

- 新薬開発への活用については製薬協からも協力を得ながらシナリオ分析を実施
- E L S I 分析では以下の点に着目する
 - レギュレーション面での整理
 - 利益相反管理のあり方
 - 社会的・倫理的課題の整理（国内外の先行調査では、学術利用に比べ産業利用に否定的な態度をもつ一般市民・研究参加者が多いとされる）
 - 国内外の先行事例における対応の検討