

Japanese Genotype-phenotype Archive へのデータ登録

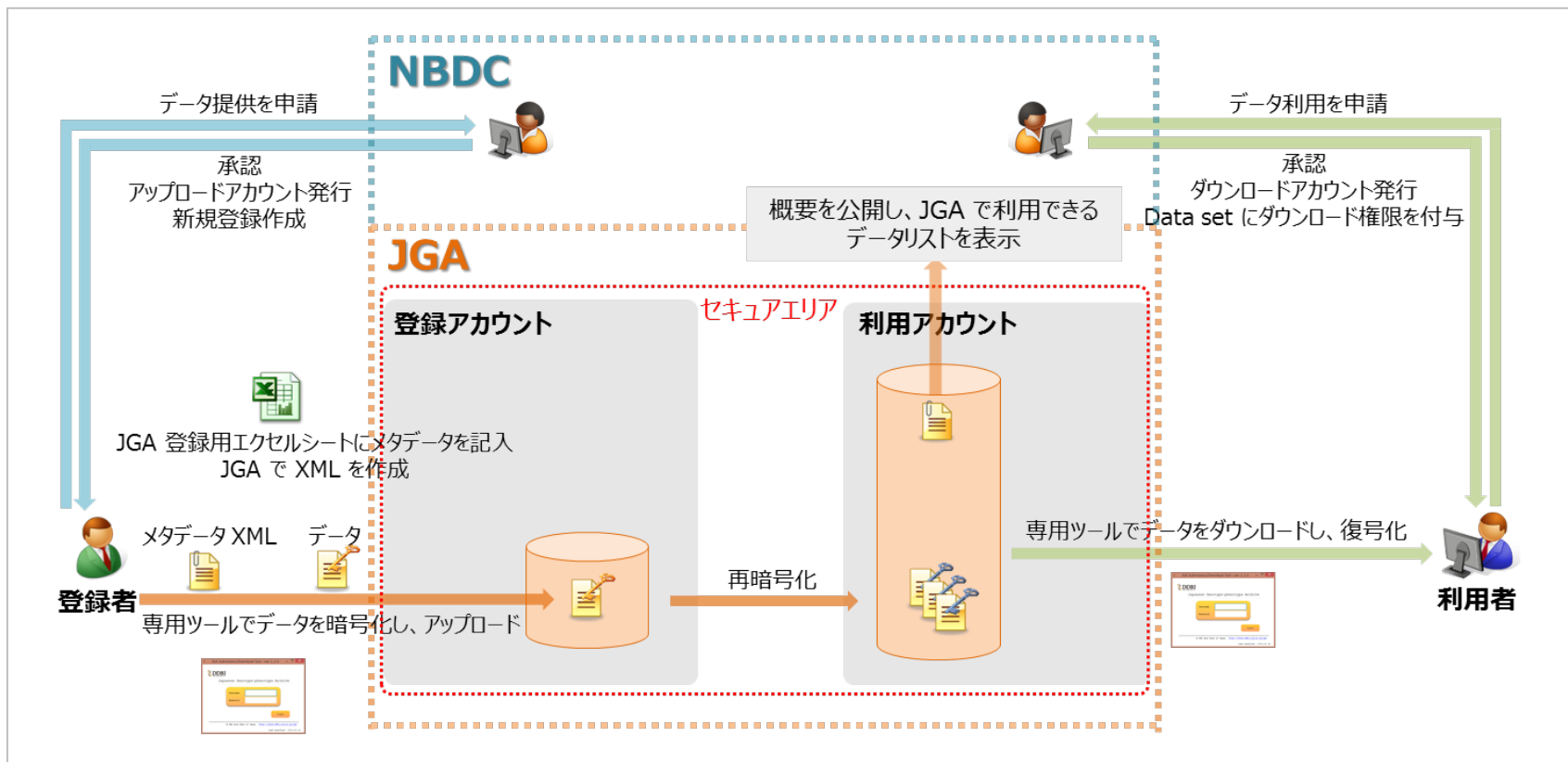
児玉 悠一

Kodama Yuichi, Ph.D

DDBJ センター、アノテータ

DDBJ center, annotator

DDBJ と NBDC の役割分担

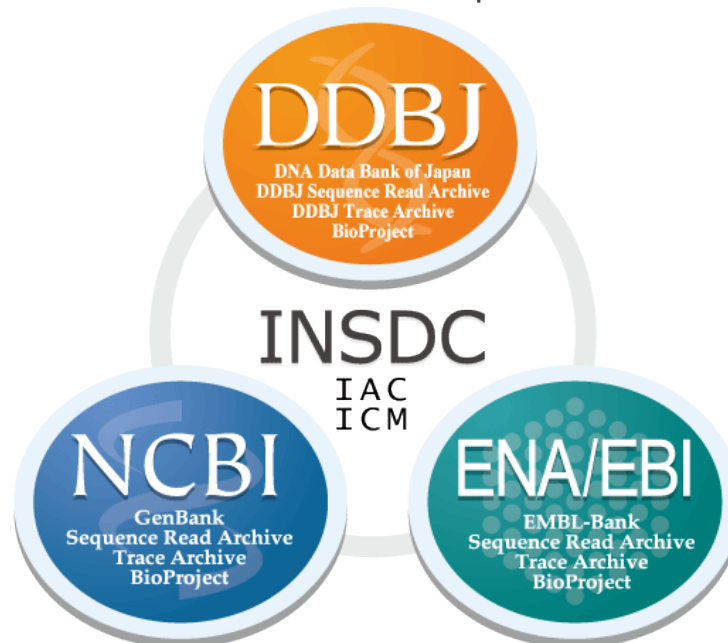


- ✓ 利用制限が必要な個人ゲノムデータを受入・保存・提供
- ✓ 匿名化されたメタデータのみ受付



- ✓ NBDC ヒトデータ共有ガイドラインに従い、データの登録と利用を審査

JGA Japanese Genotype-phenotype Archive



dbGaP

EGA

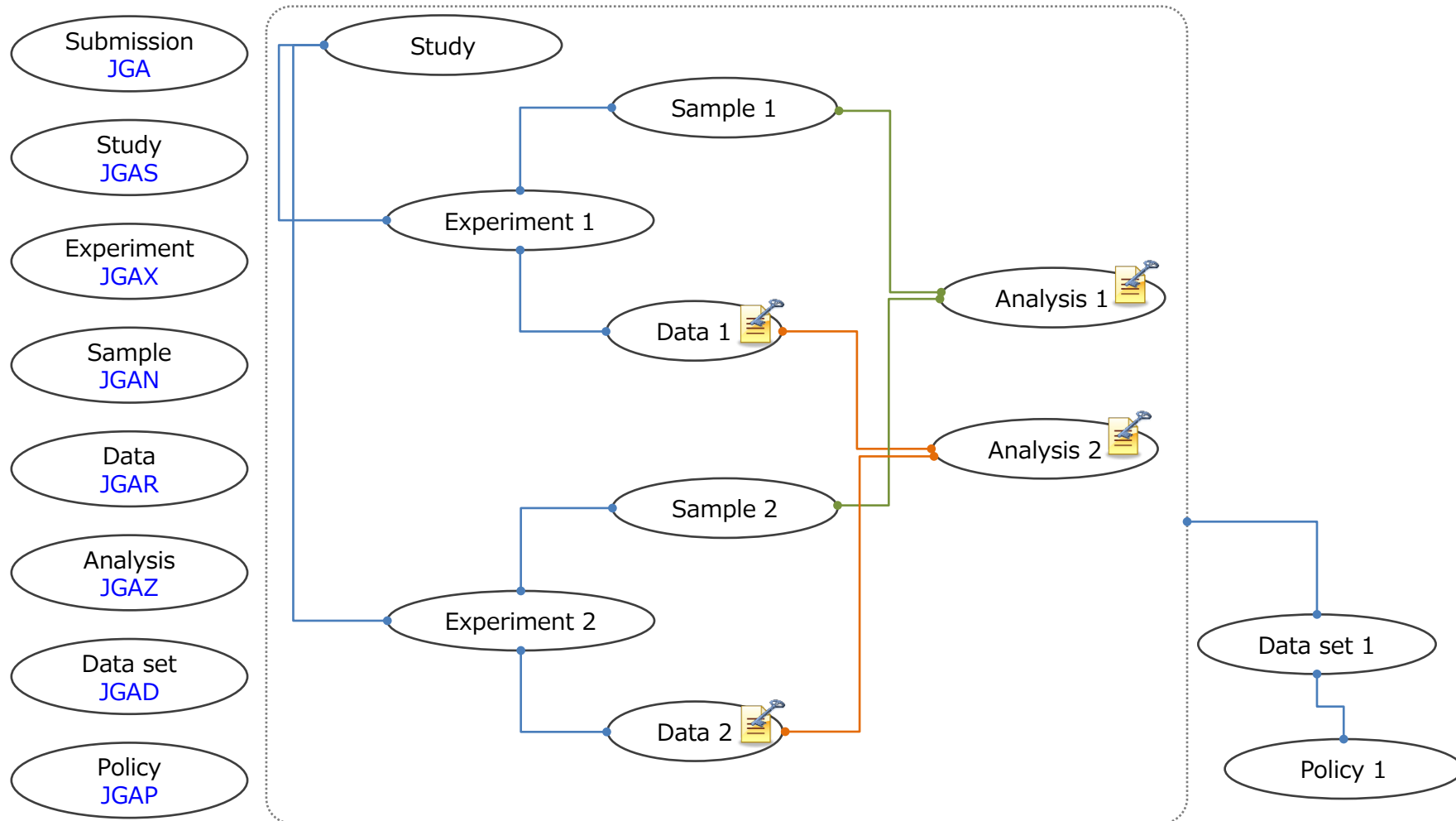
Database of Genotype and Phenotype

European Genome-phenome Archive

- ✓ JGA と EGA は SRA をベースにした同様のデータモデルを使用
- ✓ dbGaP と EGA はサマリー情報を交換
- ✓ INSDC ではない

JGA データモデル

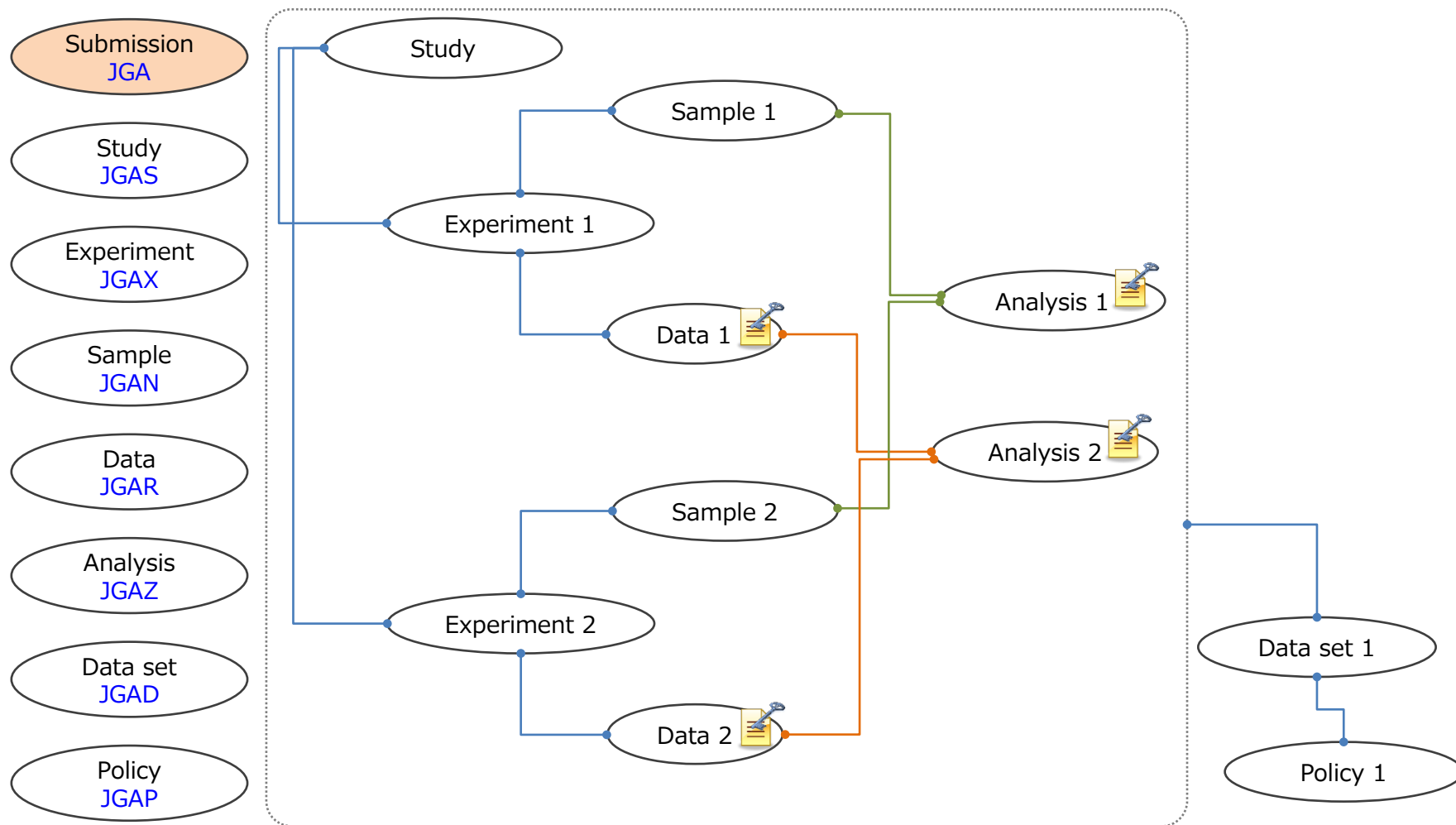
JGA データモデル



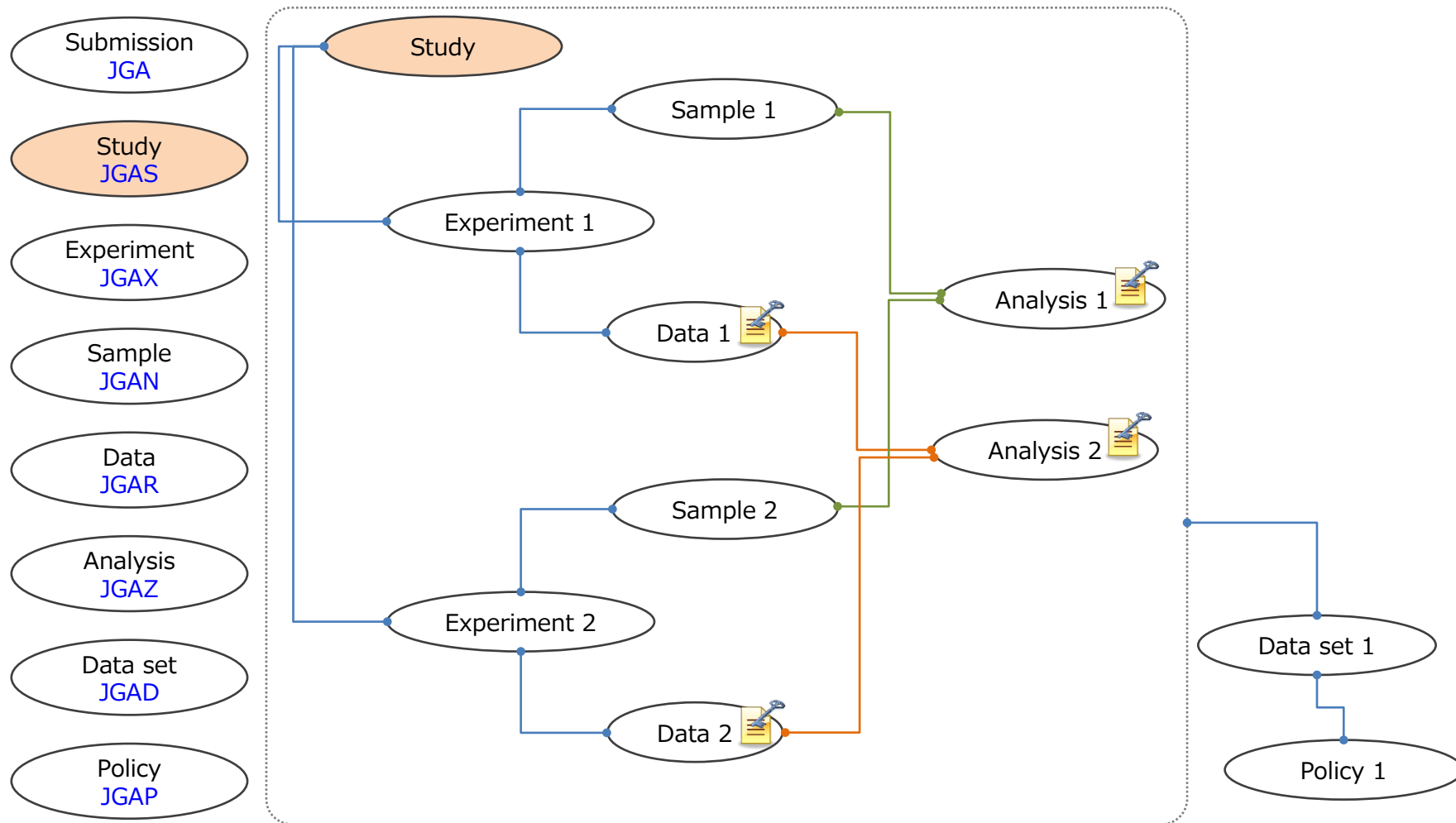
アクセッション番号プレフィックス

✓ SRA データモデルにオブジェクトを追加

JGA Submission

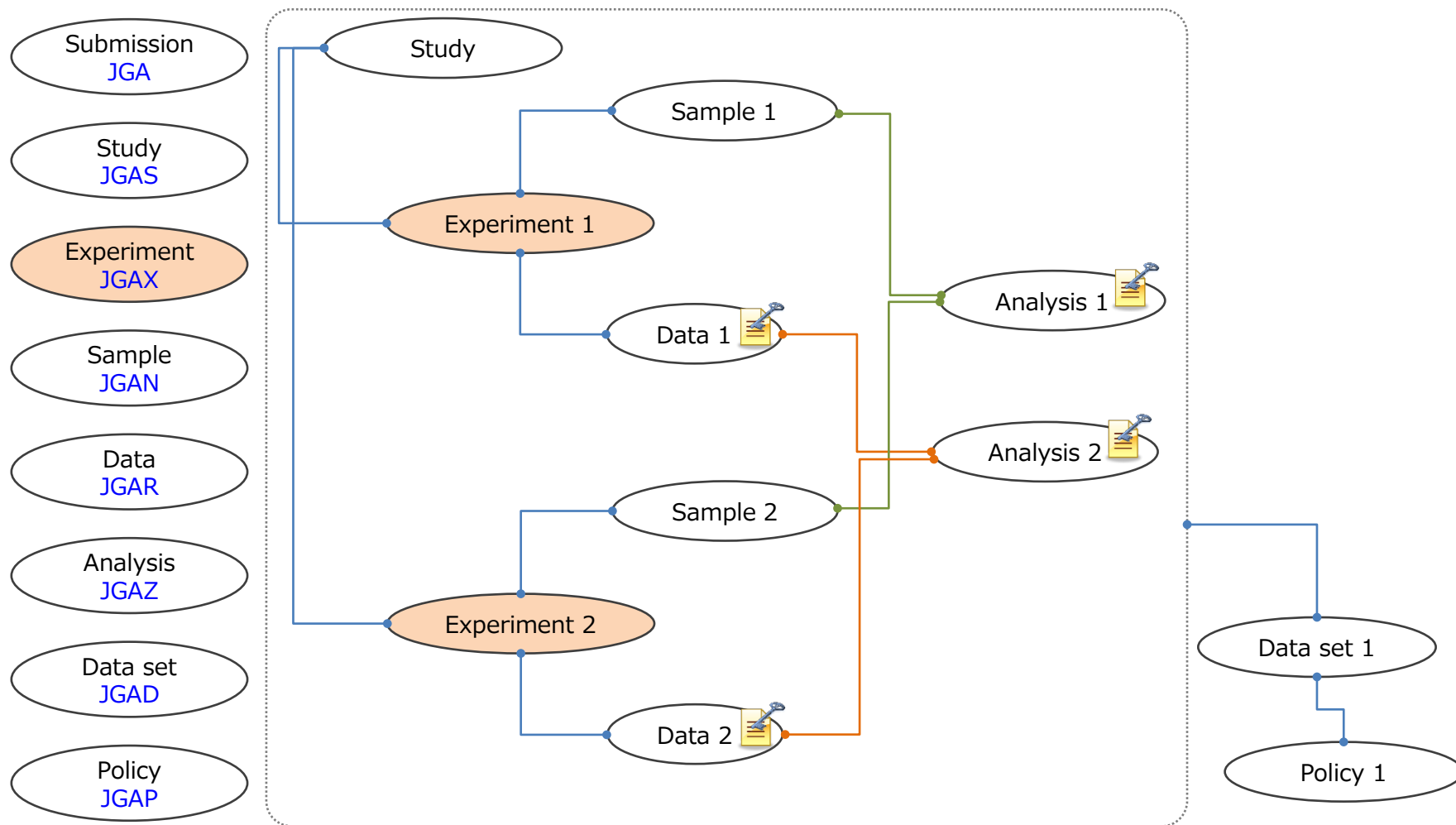


- ✓ 登録のためだけのオブジェクト
- ✓ 名前や連絡先などの登録者情報、提供開始日



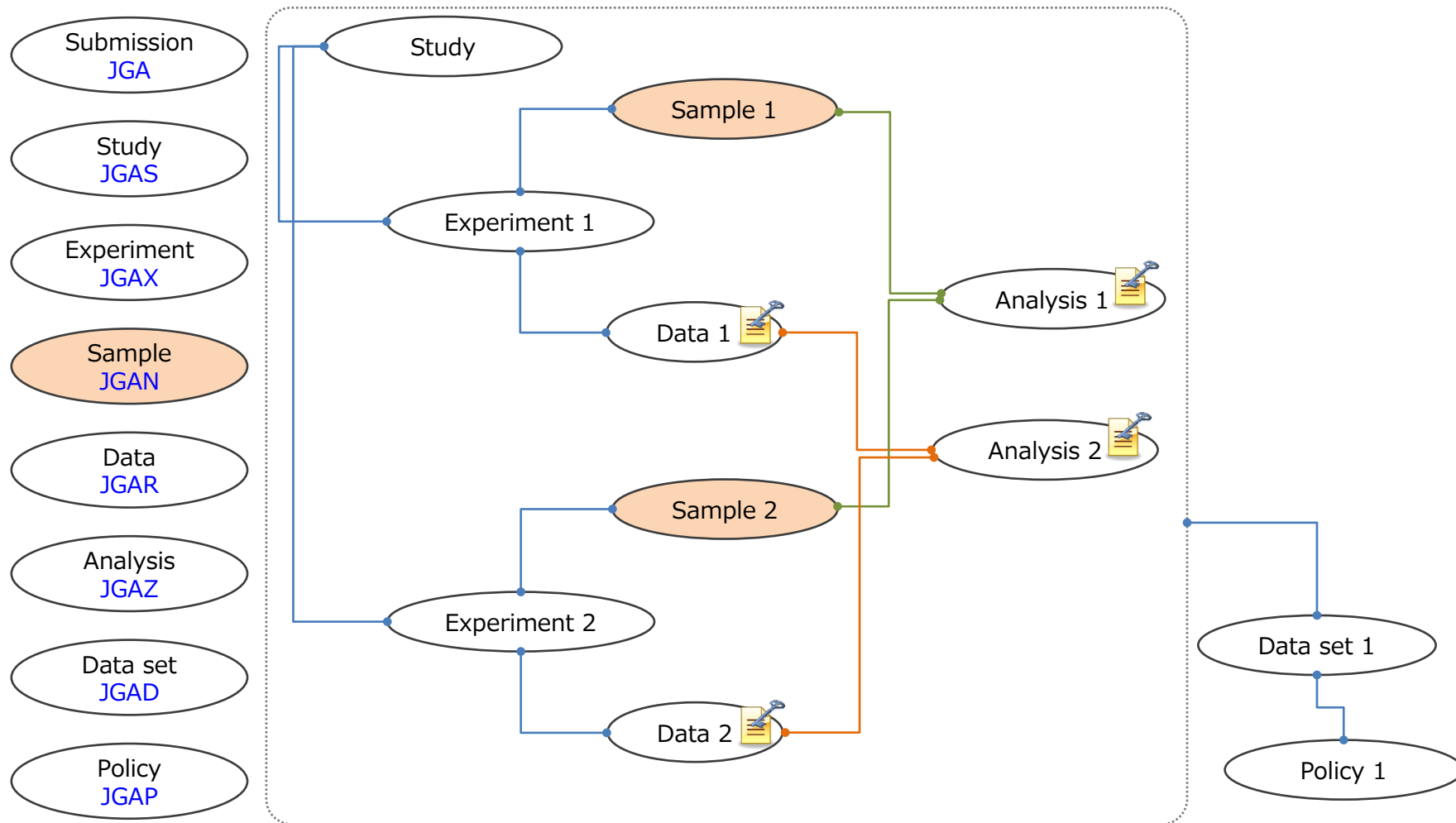
- ✓ トップレベルのオブジェクトで研究内容・研究費・論文情報を含む
- ✓ データ提供開始後、研究概要を示すために一般に公開される

JGA Experiment



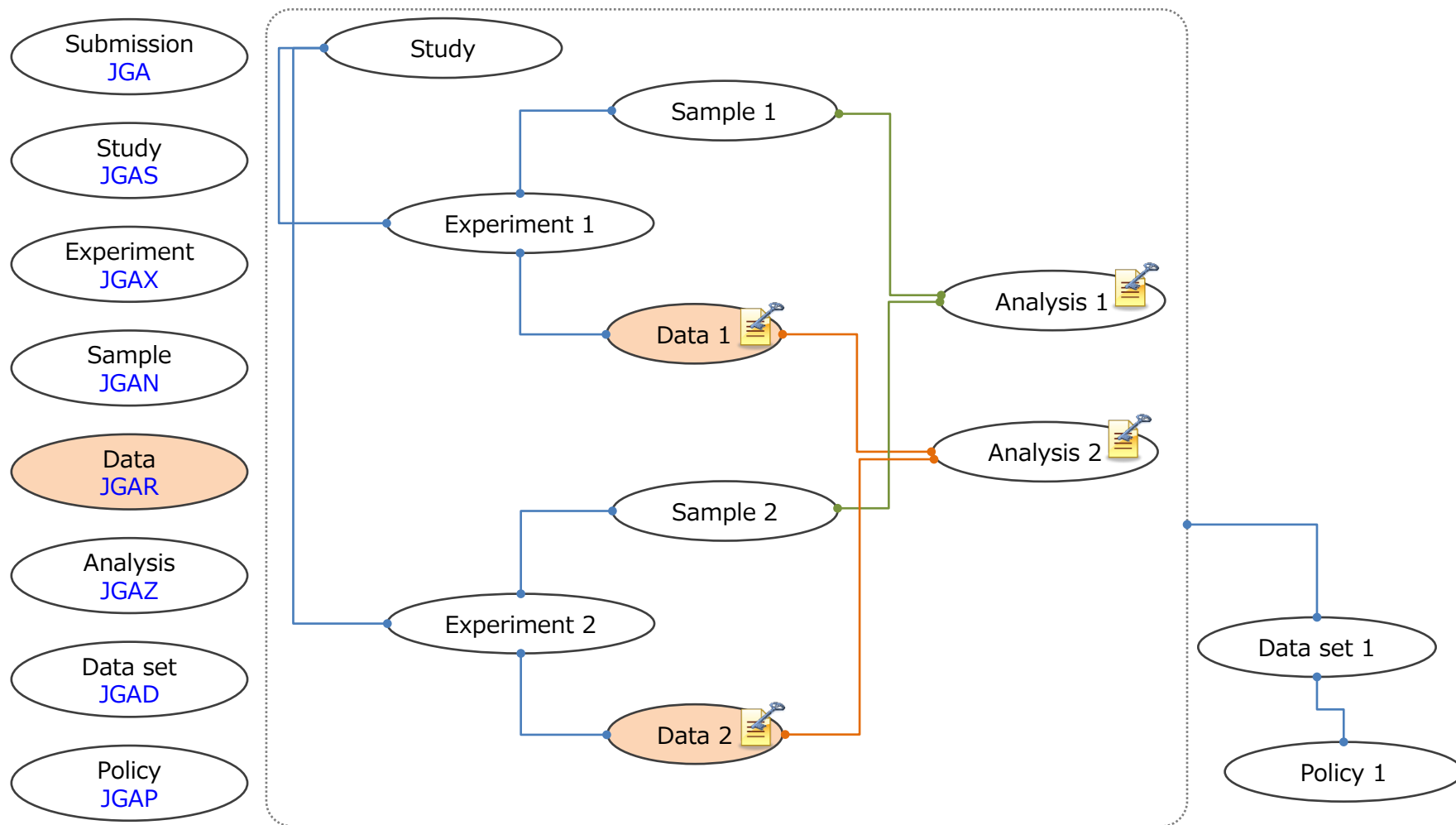
- ✓ 実験手順、質問票、ライブラリー情報やシークエンサなど
- ✓ 1つの Sample と複数の Data オブジェクトを連結

JGA Sample

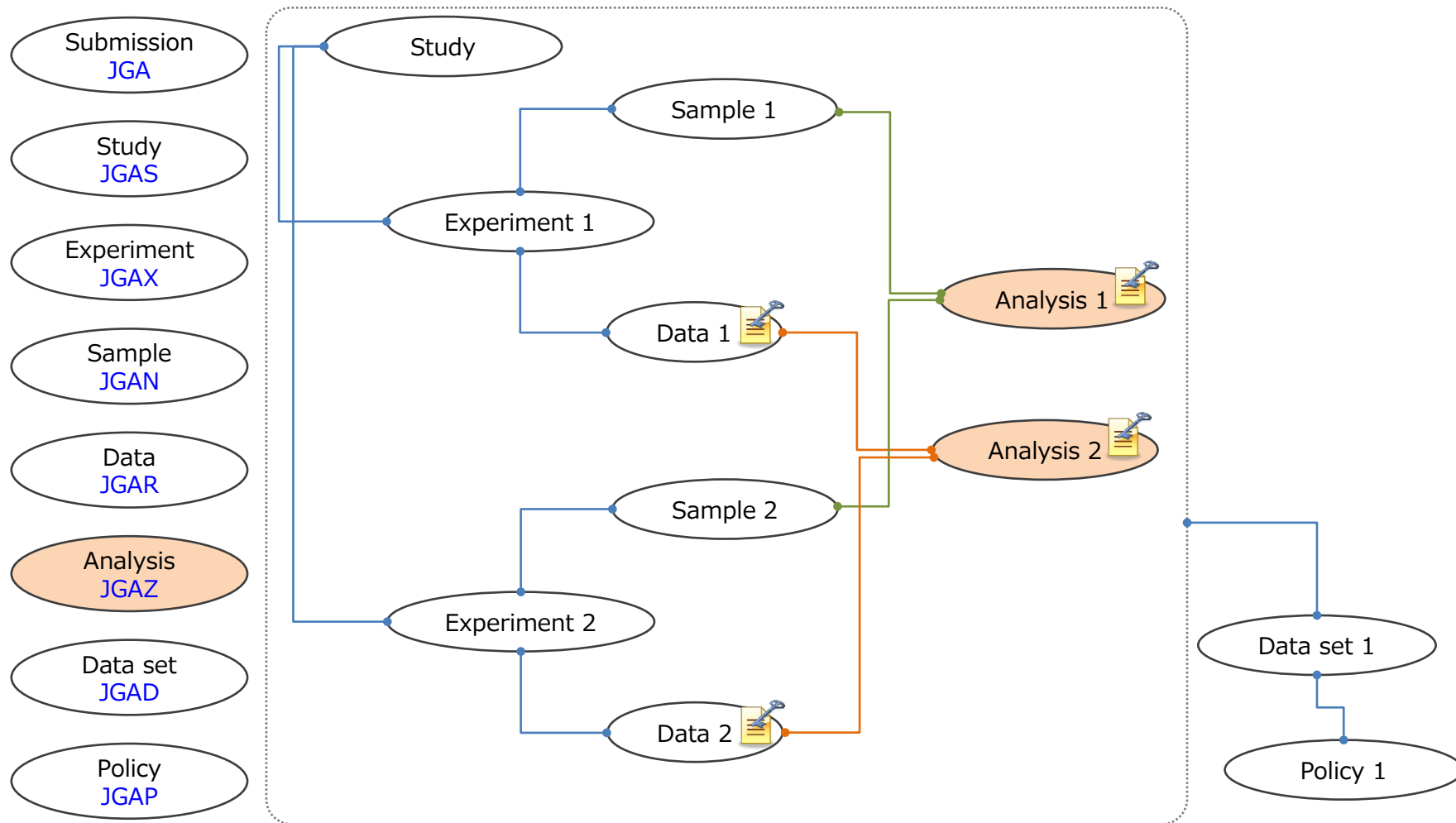


✓ Sample ≡ 個人

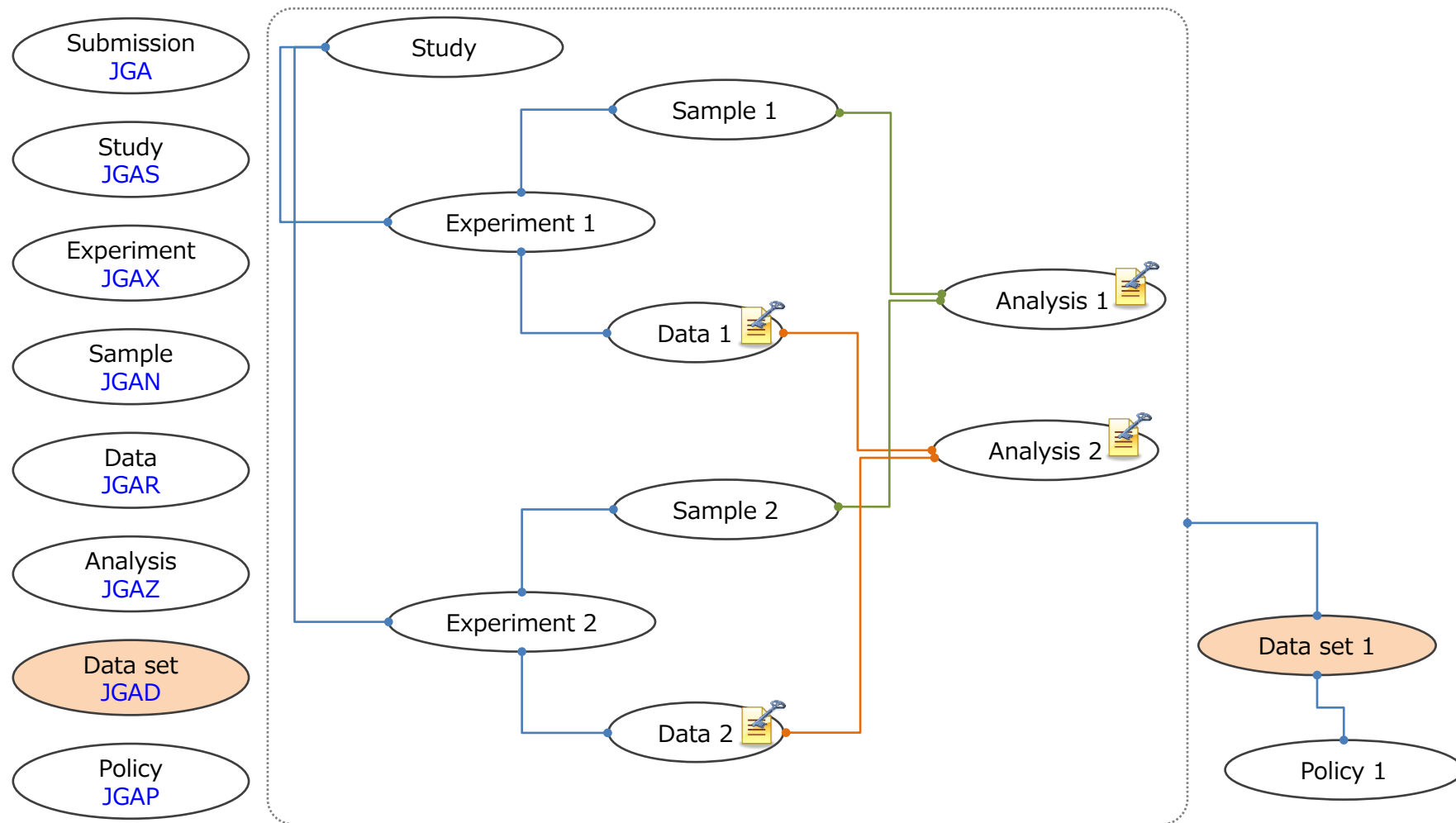
✓ 表現型情報 (gender, age etc) や匿名化された donor ID



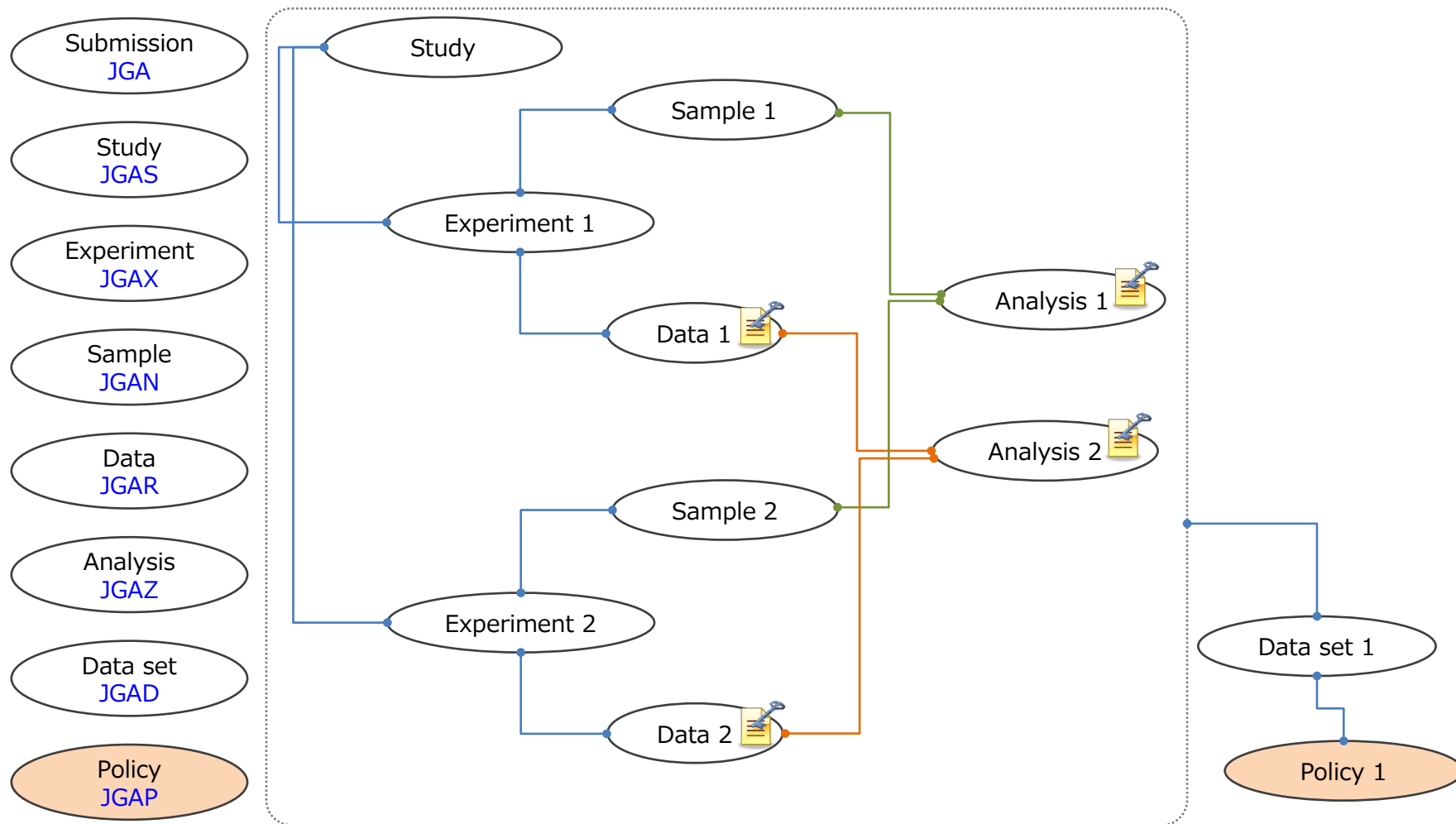
✓ 個人に対応する (生) データファイル (fastq, bam, アレイデータ) を格納



- ✓ 複数の Data もしくは Sample を解析したデータを格納
例: 変異データ (vcf) や表現型情報をまとめた表



- ✓ Policy が適用される Data と Analysis をまとめる (データ提供の単位)
- ✓ データ提供開始後、含まれているデータの種別を示すために一般に公開される

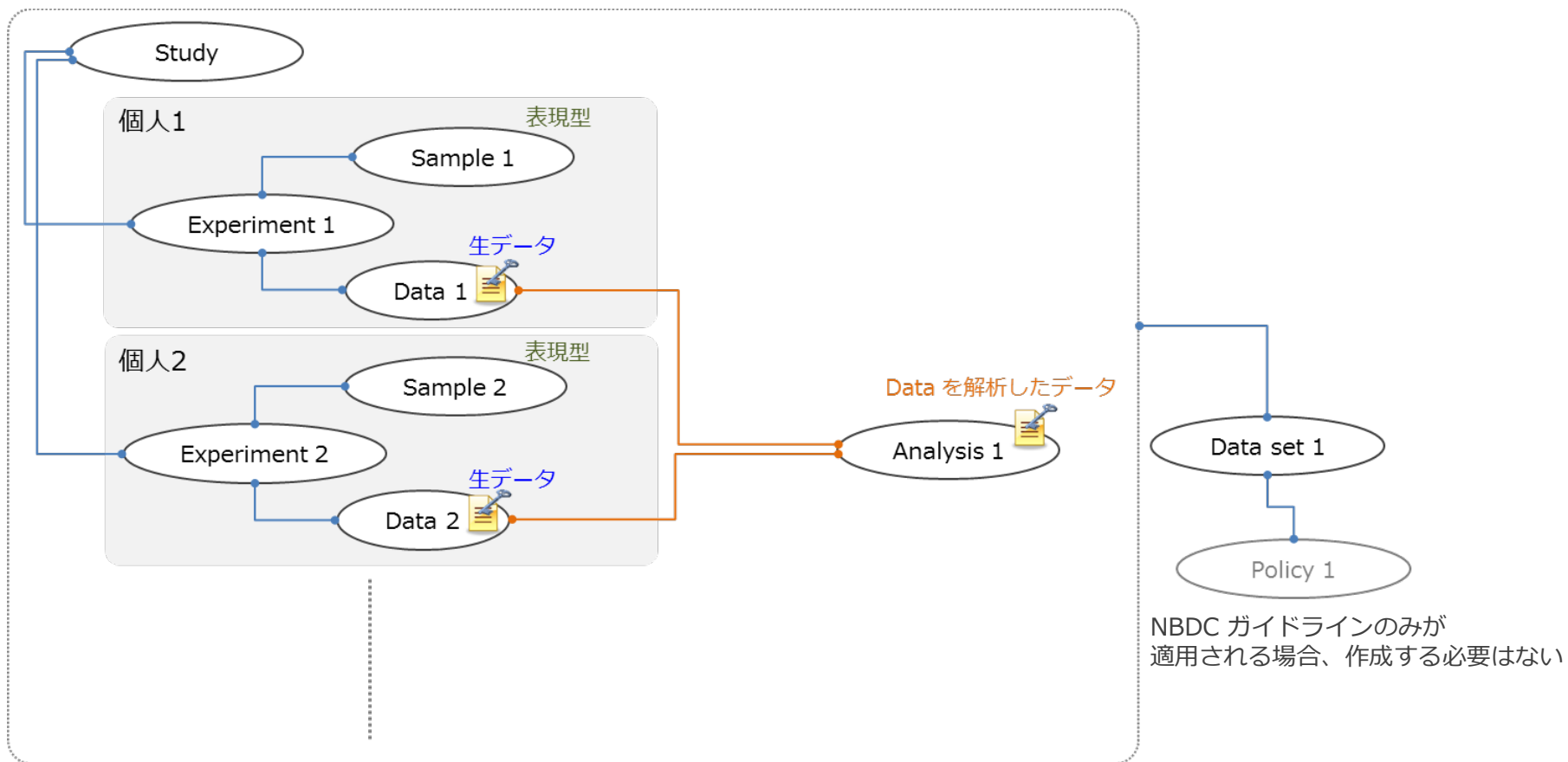


✓ Dataset に適用される利用制限事項

独自の利用制限事項 (例: 特定の研究のみ) を NBDC ガイドラインに追加したい場合に作成

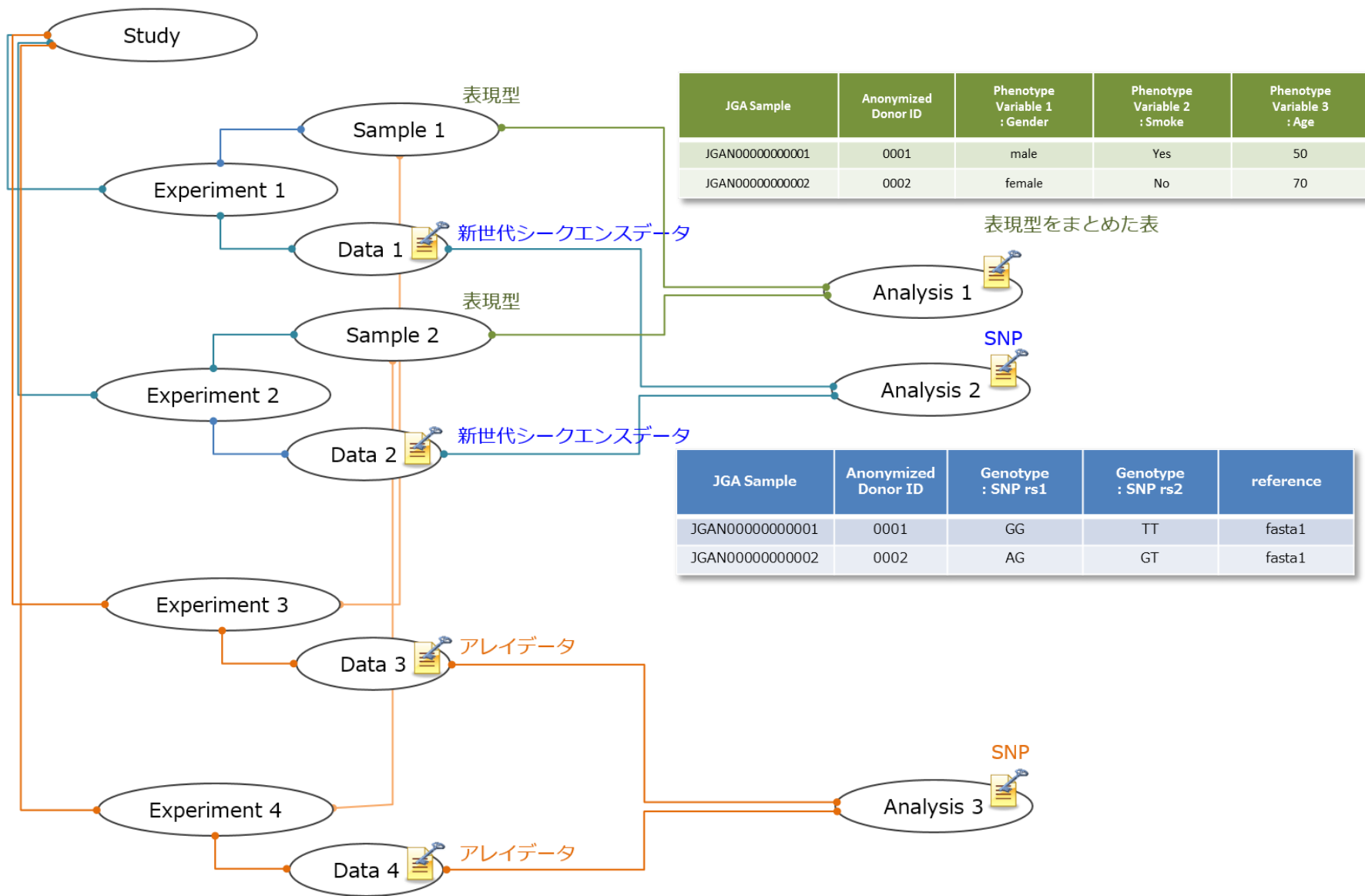
JGA データ構成の例

JGA データ 例1: シンプルなケース

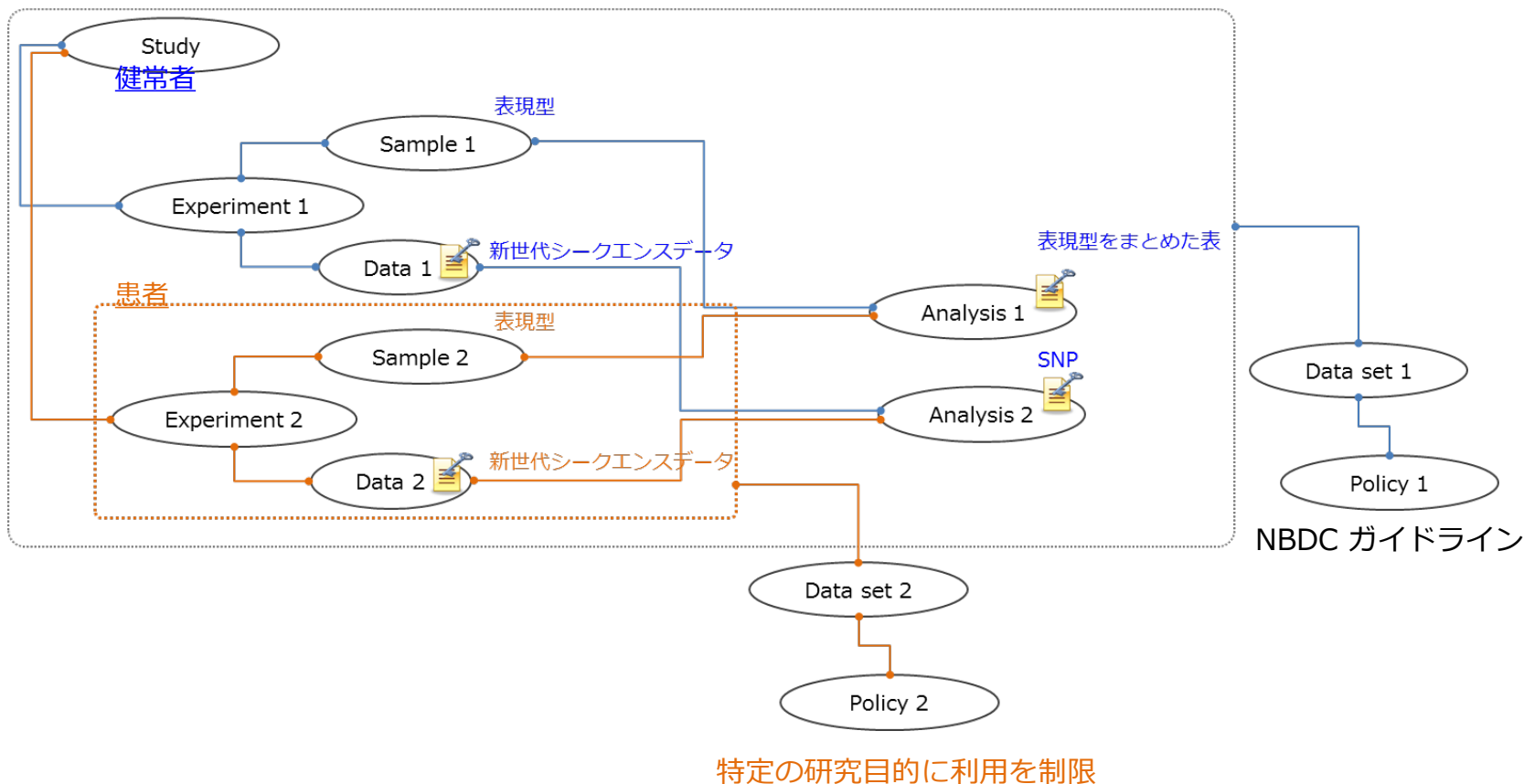


- ✓ Experiment – Sample – Data セットが「個人」に対応
- ✓ Data: 生データ、Analysis: 解析したデータ

JGA データ 例2: NGS とアレイデータ



JGA データ 例3: 異なる利用制限事項

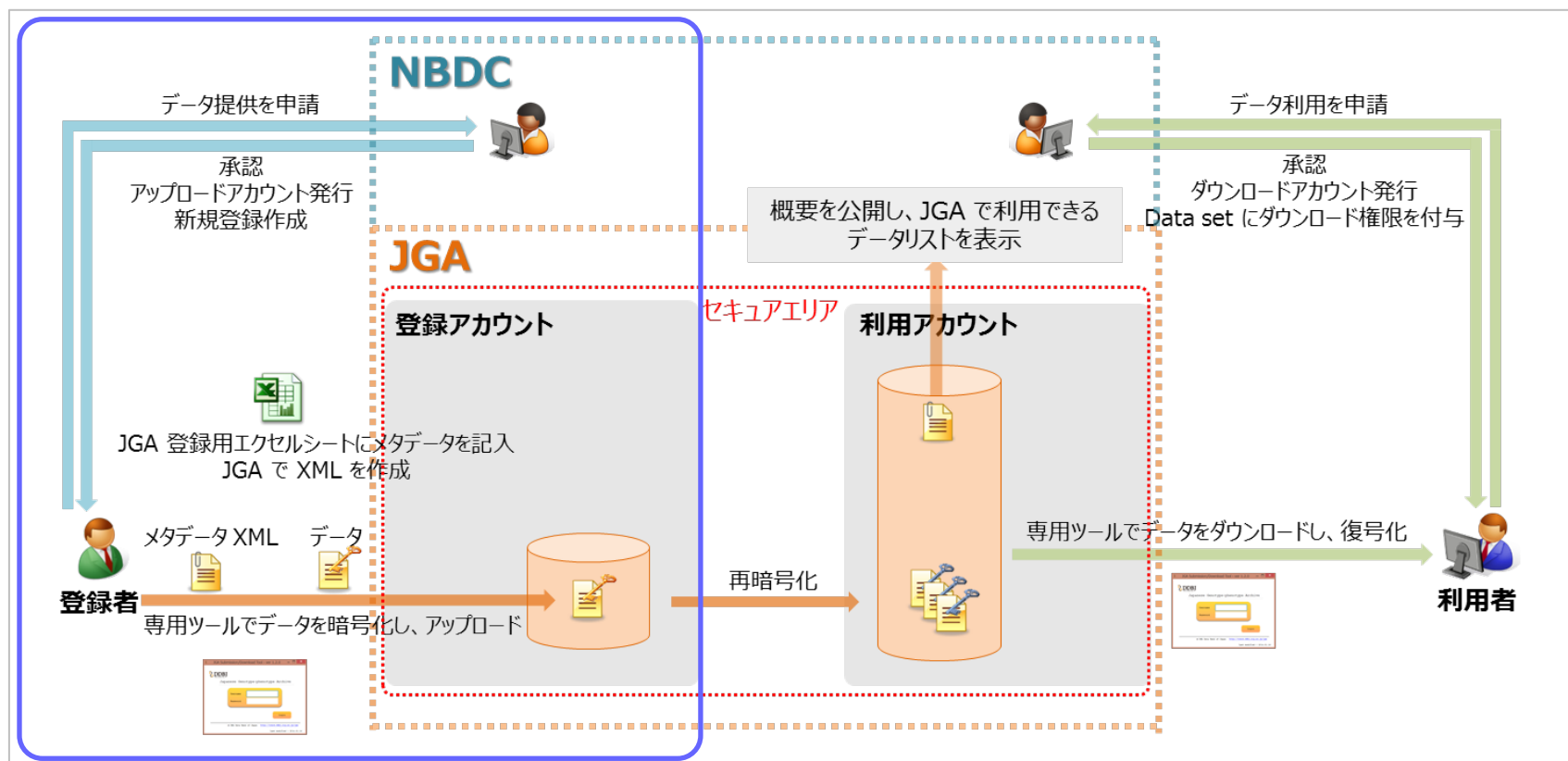


✓ 患者由来の個人データに対して、特定の研究目的に利用を制限する場合

NBDC ガイドライン + 独自の利用制限事項

JGA へのデータ登録

JGA へのデータ登録



<http://trace.ddbj.nig.ac.jp/jga/submission.html>

- ✓ NBDC にデータ提供を申請
- ✓ 承認されるとアップロードアカウントが発行される
- ✓ メタデータは登録用エクセルに記入、JGA で XML を作成 (メールでのやり取り)

JGA へのデータアップロード

The screenshot shows the 'JGA Submission Tool - ver 1.2.0' window. On the left, a file explorer displays a directory 'C:\jgadata\jgatest-0001' containing several files: data1.bam, data1.fastq, data2.bam, data2.fastq, and five XML files (Analysis, Data, Dataset, Experiment, Sample, Study, Submission). A blue arrow points from the XML files to the right panel. The right panel shows a 'Submission ID' dropdown set to 'jgatest-0001' and an 'Encrypt & Upload' button. A table below the button is empty, with the text '表には中身がありません' (Table is empty). A legend on the left identifies the files as 'データファイル' (Data files) and 'メタデータ XML' (Metadata XML). A blue arrow also points from the 'Encrypt & Upload' button to the text 'Encrypt & Upload で暗号化 & アップロード' (Encrypt & Upload for encryption & upload).

メタデータ XML とデータファイルを選択

JGA 登録を選択

データファイル

メタデータ XML

表には中身がありません

Encrypt & Upload

Encrypt & Upload で暗号化 & アップロード

1. Select files to be uploaded in left panel.
2. Select a Submission ID in the Submission ID menu in right panel.
3. Click the "Encrypt & Upload" button.
4. Uploading and encrypting start.
5. "upload succeeded." is shown when uploading finishes.

```
[16:02:28][INFO] set Submission ID list.  
[16:02:36][INFO] set uploaded file-list.(empty)
```

アップロードアカウント
にログイン

- ✓ JGA 専用ツールでログイン
- ✓ メタデータ XML とデータファイルを暗号化し SSH でサーバにアップロード
- ✓ メタデータの整合性、データファイルの存在と md5 値がチェックされる

プレフィックス	メタデータオブジェクト	数字の桁数	例
JGA	Submission	11	JGA00000000001
JGAS	Study	11	JGAS00000000001
JGAN	Sample	11	JGAN00000000001
JGAX	Experiment	11	JGAX00000000001
JGAR	Data	11	JGAR00000000001
JGAZ	Analysis	11	JGAZ00000000001
JGAD	Data set	11	JGAD00000000001
JGAP	Policy	11	JGAP00000000001

✓ JGA メタデータオブジェクトに対してアクセション番号を発行

JGA データの利用

利用可能な JGA データ一覧

NBDC National Bioscience Database Center NBDCヒトデータベース

ホーム データの利用 データの提供 ガイドライン NBDCヒトデータベース委員会

利用可能な研究データ一覧

データ利用方法は[こちら](#)をご覧ください。

Research ID	研究題目	公開日	データの種類	研究方法	手法	参加者 (対象集団)	提供者	アクセス制限
hum0001.v1	SCA31罹患者のゲノム解析データ	v1:2013/12/01	NGS (Whole)	配列決定	Illumina (HiSeq2000)	1検体 (日本人)	森下真一	制限 (Type 1)
hum0003.v1	関節リウマチ患者及び健康人における HLA領域の塩基配列比較解析	v1:2013/07/01	NGS (Target Capture)	HLA領域 配列決定	Illumina (MSeq)	33検体 (セルライン)	井ノ上逸朗	オープン
hum0004.v1	上皮成長因子受容体遺伝子異常を有する 肺癌の体細胞性遺伝子変異プロファイル 解明のための分子疫学的研究	v1:2014/07/11	NGS (Exome)	配列決定	Illumina (Genome Analyzer Iix)	97症例 (日本人)	後藤功一	制限 (Type 1)
hum0005.v1	難聴の遺伝性解析と臨床応用に関する研究	v1:2013/12/27	NGS (Target Capture)	9遺伝子領域 配列決定	Life technologies (Ion PGM)	17検体	宇佐美真一	オープン
hum0006.v1 (JGAS000000000004)	脳腫瘍のゲノム・遺伝子解析と その臨床病理学的意義の解明	v1:2014/01/31	NGS (Exome)	配列決定	Illumina (HiSeq 2000)	6症例23検体 (日本人)	斎藤延人	制限 (Type 1)

- DDBJ Sequence Read Archive (SRA)サイトにおいて、the European Sequence Read Archive (ERA)およびthe Sequence Read Archive (SRA)に登録済みデータを検索できます。
- Database Center for Life Science (DBCLS) SRAサイトからの検索も可能です。

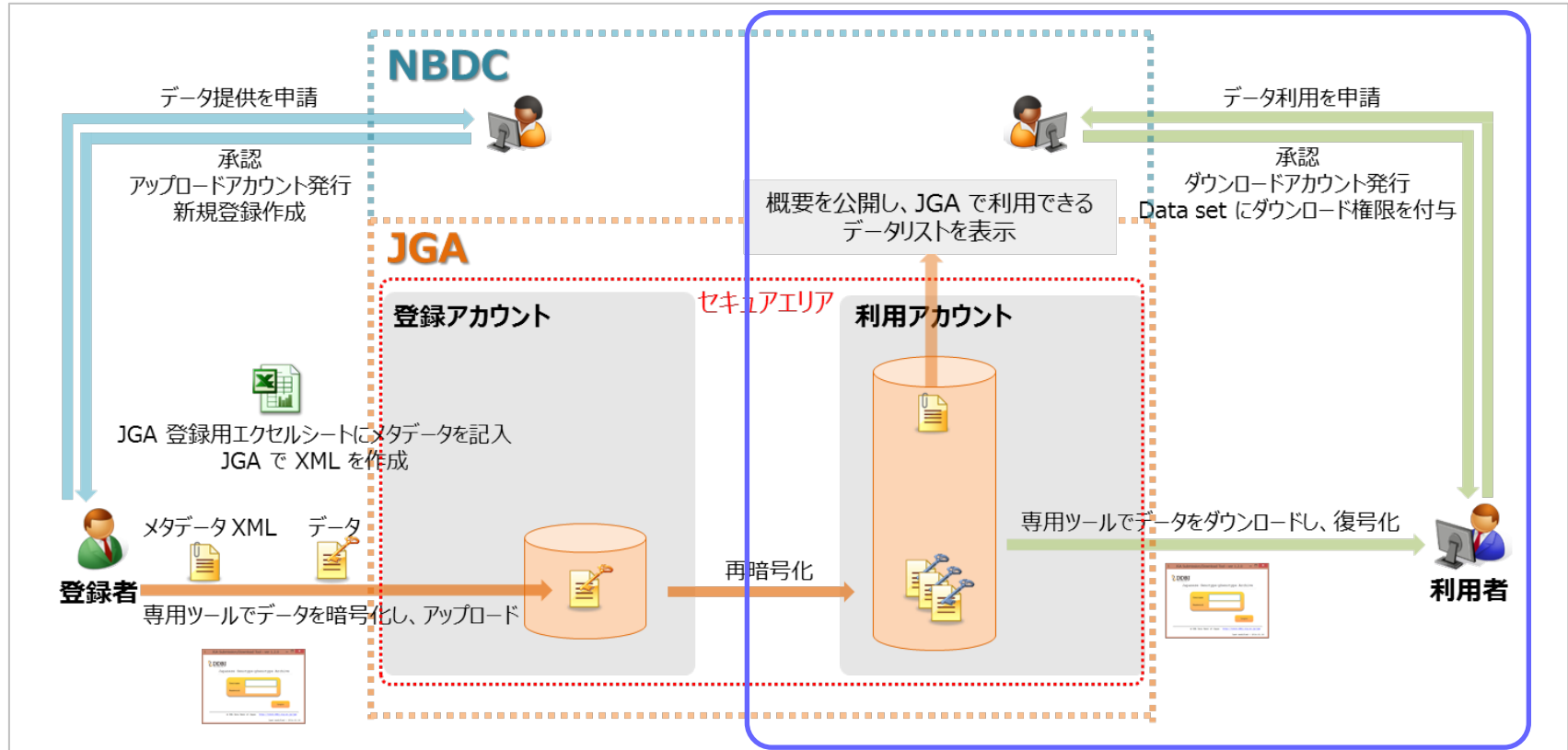
<http://humandbs.biosciencedbc.jp/data-use/all-researches-jp>

JGA Meta Viewer

Study List

Menu	Studies	
Studies Datasets JGA Home	Accession	Title
	JGAS000000000001	Identification and Characterization of Cancer Mutations in Japanese LungAdenocarcinoma without Sequencing of Normal Tissue Counterparts

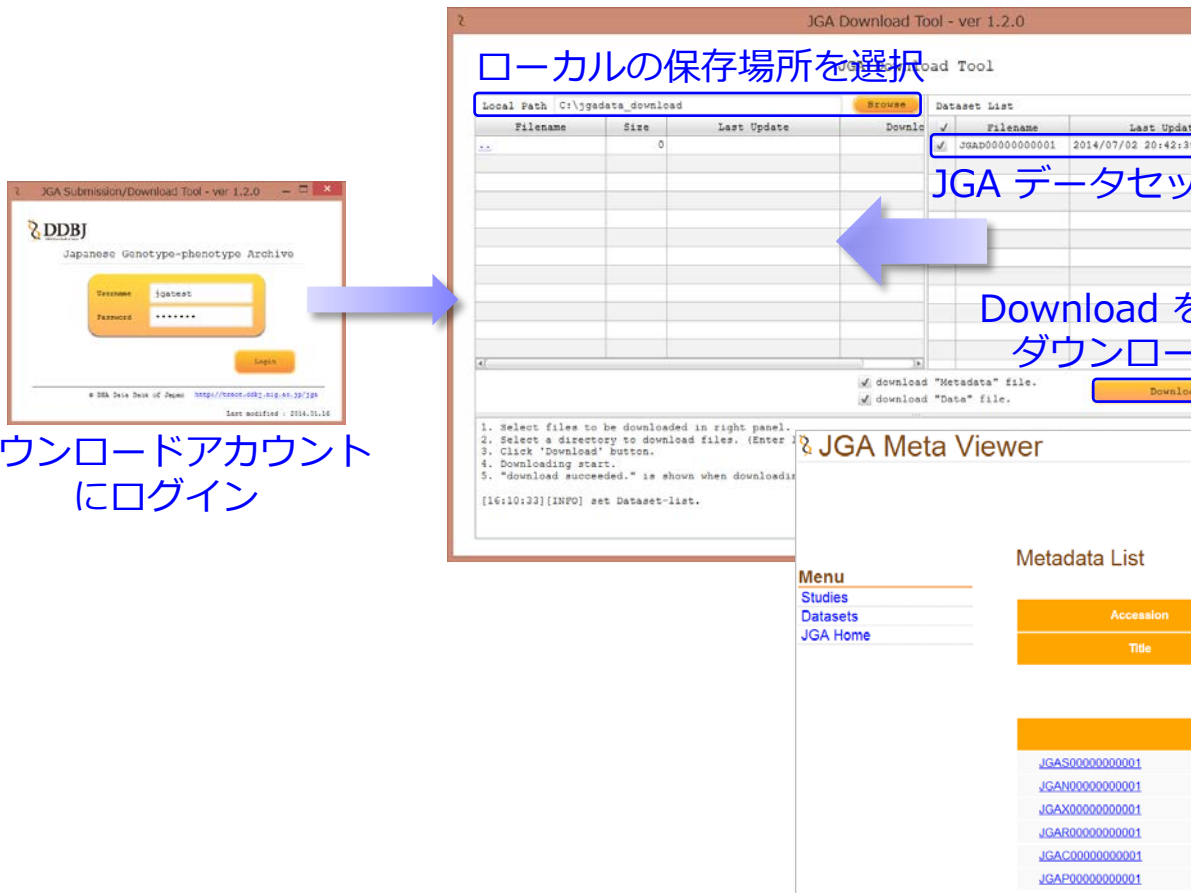
<https://ddbj.nig.ac.jp/jga/viewer/view/studies>



<http://trace.ddbj.nig.ac.jp/jga/submission.html>

- ✓ NBDC にデータ利用を申請
- ✓ 承認されると有効期限付きのダウンロードアカウントが発行される

JGA データのダウンロード



ローカルの保存場所を選択

JGA データセットを選択

Download をクリックし、ダウンロード & 復号化

ダウンロードアカウントにログイン

JGA Meta Viewer

Login User: download0001 Logout

Menu
Studies
Datasets
JGA Home

Metadata List

Accession	Dataset
JGAD00000000001	JGAD00000000001
Title	user0001-0000 datasets

Related Metadata

Accession
JGAS00000000001
JGAN00000000001
JGAX00000000001
JGAR00000000001
JGAC00000000001
JGAP00000000001

<https://ddbj.nig.ac.jp/jga/viewer/view/studies>

- ✓ JGA 専用ツールでログイン
- ✓ ダウンロード権限が付与されている Data set をリストから選択
- ✓ ダウンロードされ、復号化される
- ✓ メタデータはウェブサイトにログインして閲覧



The screenshot shows the 'Submission' page of the Japanese Genotype-phenotype Archive (JGA). The page header includes the DDBJ logo, navigation links for 'Home', 'Studies', and 'Submission', and a search bar. The main content area is titled 'JGA への登録' (Registration to JGA) and contains a list of links for various data types and submission details. A sidebar on the right is titled 'データ登録/利用を申請' (Apply for data registration/usage) and links to the National Bioscience Database Center. Below the main content, there is a section titled '登録概要' (Registration Overview) with a brief description of the registration process and a link to the submission page.

<http://trace.ddbj.nig.ac.jp/jga/submission.html>

✓ マニュアルを整備中



The screenshot shows the 'お問い合わせ' (Contact) page. It features a form for sending a message. The form includes fields for '氏名*' (Name), 'メールアドレス*' (Email address), '件名*' (Subject), and 'Submission ID'. There is a radio button menu for '問い合わせ先*' (Contact destination) with options: Sequence Read Archive, Trace Archive, Omics Archive, BioProject, BioSample, Japanese Genotype-phenotype Archive (selected), and その他 (Others). A large text area is provided for the message, with a placeholder text 'メッセージを記入してください'. A 'メッセージを送信' (Send message) button is located at the bottom left of the form area.

<http://trace.ddbj.nig.ac.jp/contact.html?db=jga>

✓ お問い合わせページ