

2018年12月14日

## 統合データベース講習会：AJACS町田

# NBDCの紹介とNBDCが提供するサービス ～生命科学系のデータを活用するために～

三橋 信孝、豊岡 理人

バイオサイエンスデータベースセンター(NBDC)、JST



# 目次

1. 講習を始める前に
2. 生命科学系DBの概要
3. JSTとNBDC、データ共有の取り組みについて
4. 主なサービス(カタログ、一括検索、データの一括ダウンロード)の紹介
5. ヒトに関するデータ関連のサービス
6. その他のサービス例
7. 今後に向けて

◇ハンズオンを含む実習について

- ・クリックは最小限に
- ・反応が無くともしばらく待つ

◇録画について

1

# 講習会の資料置き場

## 「NBDC AJACS町田」で検索

### 統合データベース講習会：AJACS町田

統合データベース講習会：AJACSは、生命科学系のデータベースやツールの使い方、データベースを統合する活動を紹介する初心者向けの講習会です。

今回の講習会では、生命科学系データベースのカタログ、横断検索、アーカイブ、ヒトデータに関するサービス・ツール等の紹介に加えて、文献データベース、化合物データベース、ゲノムデータベース、遺伝子発現データベースについてご紹介します。参加者全員がハンズオンでコンピュータを使いながらの講習です。

この講習会は協和发酵キリン株式会社の社内限定です。

- 対象： 生命科学分野のデータベースを利用したい、研究に役立てたい方（初心者向け）。
  - 日時： 2018年12月14日（金）10:30-17:40  
(開場および受付は、開始時間の30分前より)
  - 会場： 協和发酵キリン株式会社 東京リサーチパーク  
(東京都町田市旭町3-6-6)
- [アクセス]**
- 定員： 50名（協和发酵キリン株式会社所属の方限定）
  - 参加費： 無料
  - PC： ご自身のPCをお持ちください。  
※電源、ネットワーク（無線LAN）有。
  - 申込： 社内にて告知されたURLよりお申し込みください。（申込締切：12月10日（月）12:00まで）\*定員超過の場合は抽選とさせていただきます。

講習資料はこちら（準備中）

12月14日（金）

10:30-10:35 受入れ機関挨拶

浦川 利（協和发酵キリン株式会社 研究開発本部 创造基盤研究所）

## 講習会のプログラム&資料 (タブで開けておいて！)

<https://biosciencedbc.jp/>



<https://github.com/AJACS-training/AJACS74>

AJACS-training / AJACS74

Code Issues Pull requests Projects Wiki Insights Settings

Dec 14, 2018 Manage topics

15 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

Latest commit d514bf1d 13 days ago

| File          | Action           | Date        |
|---------------|------------------|-------------|
| 01_mitsuhashi | Create readme.md | 13 days ago |
| 02_kushida    | Create readme.md | 13 days ago |
| 03_kushida    | Create readme.md | 13 days ago |
| 04_oki        | Create readme.md | 13 days ago |
| README.md     | Update README.md | 13 days ago |

README.md

AJACS町田



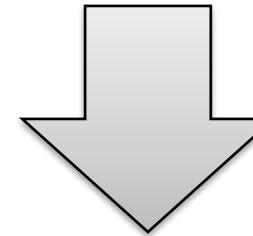
© 2018 統合データベース講習会 Licensed Under CC 表示 4.0 国際

# 目次

1. 講習を始める前に
2. 生命科学系DBの概要
3. JSTとNBDC、データ共有の取り組みについて
4. 主なサービス(カタログ、一括検索、データの一括ダウンロード)の紹介
5. ヒトに関するデータ関連のサービス・ツール
6. その他のサービス・ツール
7. 今後に向けて

# 転換期を迎える生命科学

- 生命科学の情報爆発
- 仮説検証型からデータ駆動型の科学への転換



- データ、データベースを自由に使いこなす
- 情報学、統計学、数学、計算機科学などの融合
- データの統合による知識発見

## 代表的な生命科学関係のデータベース

| データベースの内容  | データベース例                     |
|------------|-----------------------------|
| DNA塩基配列    | GenBank, ENA, <b>DDBJ</b>   |
| アミノ酸配列     | UniProt, Swiss-Prot         |
| タンパク質立体構造  | ww <b>PDB</b> , SCOP2, CATH |
| アミノ酸配列ドメイン | Pfam, InterPro              |
| アミノ酸配列モチーフ | PROSITE, Blocks             |
| パスウェイ      | <b>KEGG</b> , Reactome      |
| 遺伝病        | OMIM                        |
| 文献         | MEDLINE                     |
| 遺伝子発現      | GEO, ArrayExpress           |

# 日本の代表的な生命科学系データベース

- ◆ DDBJ (DNA Data Bank of Japan)
  - DNA塩基配列を収集
  - 国立遺伝学研究所DDBJセンター（静岡県三島市）
- ◆ PDBj (Protein Data Bank Japan)
  - タンパク質の立体構造を収集
  - 大阪大学蛋白質研究所
- ◆ KEGG (Kyoto Encyclopedia of Genes and Genomes)
  - 生命情報をシステムとして表現
  - 京都大学化学研究所バイオインフォマティクスセンター

# 構築法から見たデータベースの分類

2

(「生命科学データベース統合に関する調査研究(H17~H19)」より)

| 型        | 情報源の種類     | 処理方法   | 処理主体  | データ形式   | 例                  |
|----------|------------|--------|-------|---------|--------------------|
| バンク型     | 測定器と登録者    |        | 不特定多數 | 構造化テキスト | DDBJ,<br>wwPDB     |
| プロジェクト型  | 測定器と実験者    |        | 特定人間  | 構造      | FANTOM,<br>モデル生物DB |
| プログラム型   | データベースレコード | 機械的処理  | マシン   | 構造      | UniGene            |
| キュレーション型 | データベースレコード | 高度情報処理 | 特定人間  | 構造      | SCOP2              |
| 知識モデル型   | 読み物        | 高度情報処理 | 特定人間  | 構造      | KEGG               |
| 総説型      | 読み物        | 高度情報処理 | 特定人間  | 構造化テキスト | OMIM               |

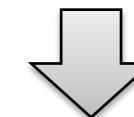
# バンク型データベースとは

(→提出されたデータを受け取り公開するDBを広くリポジトリと呼ぶ)

- ◆ DDBJ/GenBank/ENAおよびwwPDBでは、全世界で解読された塩基配列および立体構造情報を、

- 査定して受け入れ
- データベースに蓄積し
- 公開して共有する
  - accession numberを付与

→ 論文を投稿する際には、雑誌側から  
accession numberの記載を求められる



塩基配列や立体構造以外のデータについても、  
リポジトリへの登録が求められるようになっている

## 科学の世界では本来データは第3者と共有するもの

- 再現性→真理か？
- 透明性→データそのものが間違っていないか？
- 再利用→新しい観点からの解析、共同研究、教育目的、時間や費用の節約
- 新しい技術（大量データのマイニング等）の開発の促進
- 市民は2度税金を払わない  
→公的資金を投じたデータはすべての市民のもの

**国などの公的資金を投じて得られたデータは共有へ！**

# オープンサイエンスに関する国際動向と日本の対応

2013年 G8科学大臣会合 共同声明

「論文のオープンアクセス化に加え、研究データのオープン化を」  
↓

2016年 G7科学技術大臣会合  
オープンサイエンス作業部会の設置

↓  
2017年 G7科学技術大臣会合  
①オープンな研究工コシステムのためのインセンティブ  
②研究データの最適利用のためのインフラ

## ○総合科学技術・イノベーション会議

- 国際的動向を踏まえた  
　　オープンサイエンスに関する検討会  
(H26(2014).12.9～H27(2015).3.30)
  
- 国際的動向を踏まえた  
　　オープンサイエンスの推進に関する検討会  
(H29(2017).12.27～)
  - ・研究データ管理・利活用ポリシー策定ガイドライン

公的研究資金を用いた研究成果  
(論文、生成された研究データ等)について、科学界はもとより産業界及び社会一般から広く容易なアクセス・利用を可能にし、知の創出に新たな道を開くとともに、効果的に科学技術研究を推進することでイノベーションの創出につなげることを目指した新たなサイエンス

# 目次

1. 講習を始める前に
2. 生命科学系DBの概要
3. JSTとNBDC、データ共有の取り組みについて
4. 主なサービス(カタログ、一括検索、データの一括ダウンロード)の紹介
5. ヒトに関するデータ関連のサービス・ツール
6. その他のサービス・ツール
7. 今後に向けて

## 統合データベースの沿革

2000年11月

科学技術会議 ライフサイエンス部会 ゲノム科学委員会  
「ゲノム情報科学におけるわが国の戦略について」(2000年11月17日)  
※人材養成、データベース構築、情報解析技術開発の3つの観点から推進戦略を提案

2001年4月

科学技術振興機構 (JST) にバイオインフォマティクス推進センター (BIRD) を設立

2005年8月

科学技術・学術審議会 研究計画・評価分科会  
ライフサイエンス委員会 データベース整備戦略作業部会  
「我が国におけるライフサイエンス分野のデータベース整備戦略のあり方について」(2006年5月17日)  
※戦略委員会の設置、ポータルサイトの構築、統合データベースのための技術開発、人材養成を  
緊急に取り組むべき課題として提言

2006年9月

情報・システム研究機構を中心機関とした文部科学省「統合データベースプロジェクト」が開始

2006年4月

農林水産省、経済産業省でも  
統合データベースの  
プロジェクト開始

2008年12月

総合科学技術会議 ライフサイエンスPT 統合データベースタスクフォース  
「統合データベースタスクフォース報告書」(2009年5月27日)  
※ライフサイエンス統合データベースセンター (DBCLS) と BIRD との一体的な運用の提言

2011年4月

JSTにバイオサイエンスデータベースセンター (NBDC) を設置  
ライフサイエンスデータベース統合推進事業が開始  
基盤技術開発プログラムに1件、統合化推進プログラムに10件の課題を採択

科学技術振興機構(JST) <http://www.jst.go.jp/>



## 1. 未来を共創する研究開発戦略の立案・提言

## 2. 知の創造と経済・社会的価値への転換

### 2. 4. 情報基盤の強化

(ライフサイエンスデータベース統合の推進)

我が国における**ライフサイエンス研究の成果**が、広く研究者コミュニティに共有され、活用されることにより、基礎研究や産業応用につながる研究開発を含むライフサイエンス研究全体の活性化に貢献するため、文部科学省が示す方針の下、様々な研究機関等によって作成されたライフサイエンス分野データベースの統合に向けて、**オープンサイエンス**の動向を踏まえた戦略の立案、ポータルサイトの拡充・運用及び研究開発を推進し、ライフサイエンス分野データベースの統合に資する成果を得る。

## 3. 未来共創の推進と未来を創る人材の育成

[JST第4期中長期目標 (H29年2月) より]

## バイオサイエンスデータベースセンター(NBDC)

<https://biosciencedbc.jp/>

- JSTの中の1つのセンター  
2011（平成23年）年4月に設立
- 生命科学分野のDBを使いややすくする！
  - (1) データベース研究開発戦略の立案
  - (2) データベースの充実
    - 府省間のデータベース統合
    - 基盤技術開発
    - ファンディングプログラム  
(分野ごとにデータの統合を進めるため)
- (3) データの公開  
(ポータルサイトや関連サービスも含めて)
- (4) 国際連携

## NBDCの取り組み

(3)

# 連邦型統合DB (⇒中央集中型DB)



- 基盤技術開発
  - ファンディングプログラム  
(分野ごとにデータの統合を進めるため)

|  |  |
|--|--|
|  <p><b>菅野 伸夫</b> Sumio Sugano<br/>東京医科歯科大学総合遺伝研究所・世界遺傳組</p> |  <p><b>豊田 伸郎</b> Toshiro Toyoda<br/>東京医科歯科大学総合遺伝研究所・世界遺傳組</p> |
| <p>2017- 血疾患ヒトゲノム異常の生物学的機能剖析を目的とした多施設オーラリストアの貢献</p>  | <p>2017- 血疾患ヒトゲノム異常の生物学的機能剖析を目的とした多施設オーラリストアの貢献</p>  |
| <p>2014-16 血疾患ヒトゲノム異常の生物学的機能剖析を目的とした多施設オーラリストアの貢献</p>  | <p>2014-16 血疾患ヒトゲノム異常の生物学的機能剖析を目的とした多施設オーラリストアの貢献</p>  |
| <p><b>研究実績詳細</b><br/><a href="http://keru.hgc.jp/">DIBERO</a> <a href="http://keru.hgc.jp/">http://keru.hgc.jp/</a></p>                      | <p><b>研究実績詳細</b><br/><a href="http://pgdb.jp/">PGDB</a> <a href="http://pgdb.jp/">http://pgdb.jp/</a></p>                                      |
| <p><b>田端 香美</b> Sayuri Tabata<br/>リガクDNA解析室 斎美</p>   | <p><b>田端 香美</b> Sayuri Tabata<br/>リガクDNA解析室 斎美</p>   |
| <p>2017- 個別ゲノム検討代行に向けた<br/>植物ゲノム多様解釈基盤の構築</p>  | <p>2017- 個別ゲノム検討代行に向けた<br/>植物ゲノム多様解釈基盤の構築</p>  |
| <p>2014-16 植物ゲノム情報連携のための<br/>組合せ収集基盤の構築</p>  | <p>2014-16 植物ゲノム情報連携のための<br/>組合せ収集基盤の構築</p>  |
| <p>2011-13 ゲノム情報に基づく植物データベースの開発</p>  | <p>2011-13 ゲノム情報に基づく植物データベースの開発</p>  |
| <p><b>研究実績詳細</b><br/><a href="http://pgdb.jp/">Plain Genome DataBase Japan (PGDB)</a> <a href="http://pgdb.jp/">http://pgdb.jp/</a></p>      | <p><b>研究実績詳細</b><br/><a href="http://phenome.info/">phenome</a> <a href="http://phenome.info/">http://phenome.info/</a></p>                    |
| <p><b>糸屋 駿志</b> Hiromi Matsuya<br/>電化研究所バイオナビゲーションセンター ユニットリーダー</p>  | <p><b>糸屋 駿志</b> Hiromi Matsuya<br/>電化研究所バイオナビゲーションセンター ユニットリーダー</p>  |
| <p>2014-16 生命と環境のフェノーム統合データベース</p>   | <p>2014-16 生命と環境のフェノーム統合データベース</p>   |
| <p><b>豊田 伸郎</b> Toshiro Toyoda<br/>電化研究所バイオナビゲーションセンター データマネージャー</p>   | <p><b>豊田 伸郎</b> Toshiro Toyoda<br/>電化研究所バイオナビゲーションセンター データマネージャー</p>   |
| <p>2011-13 生命と環境のフェノーム統合データベース</p>   | <p>2011-13 生命と環境のフェノーム統合データベース</p>   |
| <p><b>研究実績詳細</b><br/><a href="http://phenome.info/">phenome</a> <a href="http://phenome.info/">http://phenome.info/</a></p>                  | <p><b>研究実績詳細</b><br/><a href="http://phenome.info/">phenome</a> <a href="http://phenome.info/">http://phenome.info/</a></p>                    |

|  |
|--|
| <b>大庭 修一</b> Shuichi Onami<br>東北大学理学系研究科システム科学センター チームリーダー   |
| <b>研究室概要</b>   |
| 2016 - 17 生命活性物質と細胞・生産過程開拓組合データベース   |
| 2012 - 14 生命活性システム科学のデータベースの統合化  |
| <b>研究室概要</b>   |
| SBD: System Science of Biological Dynamics <a href="http://ssbd.ebic.riken.jp/">http://ssbd.ebic.riken.jp/</a>   |
| <b>SSB Database</b><br><small>RIKEN Institute of Biological Sciences</small>   |
| <b>桃永 謙士</b> Kuniyuki Tokunaga<br>東京大学大学院理学系研究科 教授<br>准教授  |
| <b>研究室概要</b>   |
| 2014 - 16 個別化免疫系におけるヒトゲノムリエーションデータベース  |
| 2015 - 13 ヒトゲノムリエクションデータベースの開発   |
| <b>研究室概要</b>   |
| ヒトゲノムリエクションデータベース<br><a href="https://www.biosciencedbc.jp/index_japanese.html">https://www.biosciencedbc.jp/index_japanese.html</a>   |
| <b>迄井 威</b> Takehi Iwatsubo<br>東京大学大学院理学系研究科 教授<br>准教授   |
| <b>研究室概要</b>   |
| 2011 - 13 ヒト免疫画像データベース   |
| <b>研究室概要</b>   |
| ヒト免疫画像データベース<br><a href="https://human.ebi.biosciencedbc.jp/hum0043-v1">https://human.ebi.biosciencedbc.jp/hum0043-v1</a><br><a href="https://human.ebi.biosciencedbc.jp/hum0031-v1">https://human.ebi.biosciencedbc.jp/hum0031-v1</a> |
| <b>松田 文也</b> Fumiaki Matsuda<br>新潟大学大学院理学系研究科遺伝学・分子生物学センター センター長・教授  |
| <b>研究室概要</b>   |
| 2015 - 13 大規模ゲノム疫病研究の統合情報技術の構築   |
| <b>研究室概要</b>   |
| Human Genetic Variation Browser<br><a href="http://www.genome.med.yado-u.ac.jp/SnpDB/index.html">http://www.genome.med.yado-u.ac.jp/SnpDB/index.html</a>   |

情報技術 基盤技術開発プログラム DB統合化の実現に向けて基盤となる技術の開発を実施

小原 雄治 Yuki Kohara  
情報・システム研究機構ライフサイエンス  
統合データベースセンター センター長

統合のためにDBCLSが提供するサービス  
統合(トータル)→

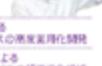
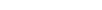
2011-12 データベース統合に開拓する基盤技術開発

新規型の大規模集中型のデータベース統合でなく、全く新しい「フェデレーション(連邦)型」のデータベース統合を行なうため、RDBを中心としたもので、DBIや JDBCなど複数の異なるデータベース、統合化を進めるプロセスの分野別データベースを分離して、互いにつながるのシステムを開拓。統合化に必要な機能開発も同時に実行しています。また、すでに構築のほまっている連携API(データエクスポート)をはじめとして大規模データの利用に関する技術開発も行なっています。

連携 API 開発実績: <http://hogehogehdbi.jp/> ライフサイエンス統合DBAPI: <http://lifebio.medicoredb.jp/>  
GigaDB: <http://giga.ncbi.nlm.nih.gov/> Netflix: <http://netflix.ncbi.nlm.nih.gov/> Alter: <http://altered.ncbi.nlm.nih.gov/> YeastGenome: <http://yeastgenome.org/>

生命知識の創成と発見

2011年4月からeGRCOとの連携開始して、  
「基盤技術開発」を実施しています。

|   |  |
|---|--|
| <p><b>生命科学系コンテンツ</b> 統合構造プログラム<br/>生物種、様々な目的やプロジェクトを超えて、幅広く<br/>生物学分野のDNA結合を実現</p>  |  <p>生物種データベースプログラム 研究会<br/>長崎 敏和 Toshiaki Nagasawa<br/>先エーザイ株式会社 プラットフォーマー</p>  |
| <p><b>有田 正樹</b> Masanori Arita<br/>情報システム研究会統合構造生物学研究会 副会長<br/><br/>2018 - 物質循環を考慮したメタプロトクスデータベース<br/>2014 - 16 生物種メタボロームモデル・データベースの構築</p> <p><b>金谷 康彦</b> Shigeo Kaneyoshi<br/>東海大学理学部生物化学科 教授<br/><br/>2011 - 13 メタプロトクス・データベースの開発</p>          | <p><b>木下 聰子</b> Kyoko Fujii-Kinoshita<br/>東海大学理工学部 教授<br/><br/>2017 - 種群科学ポータルの構築</p> <p><b>成松 久</b> Hiroyuki Narimatsu<br/>東海大学理学部生物化学科研究会部門 研究会研究会幹事<br/><br/>2014 - 16 種群統合データベースおよび 生物種構造通りの開発<br/>2011 - 13 種群統合データベースと研究支援ツールの開発<br/></p> |
| <p><b>石浜 勝</b> Yasushi Ishihama<br/>京都大学生物機能高分子研究所 教授<br/><br/>2018 - プロトオームデータベースの構築深化と<br/>進化基質活性化</p> <p><b>日本標準規格</b><br/>jPOST (Japan Proteome Standard<br/>Repository/Database) <a href="https://jpstdb.org/">https://jpstdb.org/</a> </p> | <p><b>栗原 順嗣</b> Genji Kurusu<br/>京都大学蛋白質研究所 教授<br/><br/>2017 - 蛋白質構造データリンクのデータ標準化と統合化</p> <p><b>中村 春香</b> Harumi Nakamura<br/>京都大学蛋白質研究所 教授<br/><br/>2014 - 16 蛋白質構造データバンクの標準化と独自の運用<br/>2011 - 13 蛋白質構造データバンクの国際的な構築と統合化<br/></p>            |
| <p><b>津 真吾</b> Shinya Chi<br/>九州大学大学院生物資源科学 研究員<br/><br/>2017 - ユビキノン/クオツ統合データベースの開発と<br/>構築手法</p> <p><b>日本標準規格</b><br/>CNP-Atlas <a href="http://chip-atlas.org/">http://chip-atlas.org/</a></p>  | <p><b>黒川 哲</b> Ken Kurokawa<br/>情報システム研究会統合構造生物学研究会 副会長<br/><br/>2017 - データサイエンスを加通させ<br/>生物種統合データベースの標準化開発<br/>2014 - 16 ゲノム・ゲノム連携蛋白質・疾患・医薬品の<br/>統合データベース<br/>2011 - 13 ゲノム情報に基づく疾患・医薬品・<br/>生物学データの統合<br/></p>  |
| <p><b>金久 寿</b> Minoru Kaneko<br/>京都大学化学生命系 研究教授<br/><br/>2017 - ゲノム・疾患・医薬品のネットワークデータベース<br/>2014 - 16 ゲノム・セフノ・タイプ・疾患・医薬品の<br/>統合データベース<br/>2011 - 13 ゲノム情報に基づく疾患・医薬品・<br/>生物学データの統合<br/></p>  | <p><b>KEGG MEDICUS</b> <a href="http://www.genome.jp/kegg/medicus/">http://www.genome.jp/kegg/medicus/</a> </p>   |

## バイオサイエンスデータベースセンター(NBDC)

<https://biosciencedbc.jp/>

- JSTの中の1つのセンター  
2011（平成23年）年4月に設立
- 生命科学分野のDBを使いややすくする！
  - (1) データベース研究開発戦略の立案
  - (2) データベースの充実
    - 府省間のデータベース統合
    - 基盤技術開発
    - ファンディングプログラム  
(分野ごとにデータの統合を進めるため)
  - (3) データの公開  
(ポータルサイトや関連サービスも含めて)
  - (4) 国際連携

## 生命科学分野のDBに対する要望

3

1. どんなDBがどこ(URLは?)にあるのか?

欲しいデータは既に誰かが作成、  
公開しているのでは?



NBDCのサービス

DBのカタログ



2. 複数のDB内を横断的に検索できないか?

DBごとに訪れて検索するのは手間  
文献や特許など異なるDBを一括検索できれば



DBの一括検索



3. データセットを取得できないか?

インターフェースに囚われずに自由に使いたい  
手持ちのデータと組み合わせてみたい

データの一括  
ダウンロード

# データを利用する際の問題点



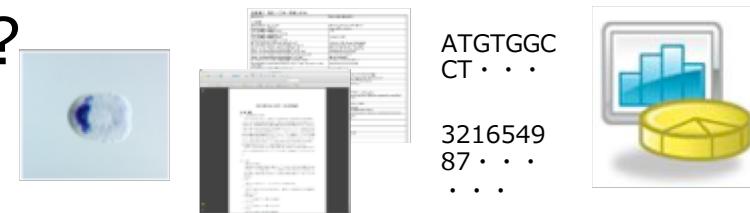
- 求められるデータが入っているか？

使いたいデータが見つからない  
公的プロジェクトの成果が非公開

データを集めるための仕組み  
・**公的な研究**には寄託を義務付け

- どのようなデータが入っているのか？

多種多様なデータ  
データ項目の説明がばらばらor無  
分野ごとにはばらばらな用語



- どのように使って良いのか？

利用に際しての条件がばらばらだし、  
そもそもどこに書いてあるのか？

商用可能か？  
契約が必要？  
使用料は必要？  
論文に引用してもいいのか？

データ**フォーマット**の標準化  
データの説明(**メタデータ**)の標準化

**利用条件の標準化**  
**and明示**

- データ作成者の著作権表示
- 商用↔非商用
- 公開条件
- 引用方法



# クリエイティブ・コモンズ (CC)

## 再利用と共有を促進する仕組み

### CCライセンスの種類

作品の利用（再配布やリミックス作品の公開、実演等）のための条件は4種類あります。



#### Attribution(BY)

表示

作品のクレジットを表示すること



#### NonCommercial

非営利

営利目的での利用をしないこと



#### NoDerivs

改変禁止

元の作品を改変しないこと



#### ShareAlike

継承

元の作品と同じ組み合わせのCCライセンスで公開すること

4種類のマークを組み合わせて 6種類の利用条件を選択することができる



CC-BY ジャパンクリエイティブ・コモンズ・ジャパン (<http://creativecommons.jp/licenses/>)

高

利用の自由度 中