

# ナショナルバイオリソースプロジェクト

## 第1回情報ワーキンググループ

### 議事概要

#### 1. 日時・会場

平成20年7月15日(火) 14:00~16:00

東京八重洲ホール8階811会議室

#### 2. 出席者

##### 委員

|            |   |
|------------|---|
| (主査) 城石 俊彦 | 情報・システム研究機構国立遺伝学研究所<br>系統生物研究センター 教授      |
| 芹川 忠夫      | 京都大学大学院医学研究科附属動物実験施設 施設長                  |
| 田畑 哲之      | かずさディー・エヌ・エー研究所 副所長                       |
| 中村 太郎      | 大阪市立大学大学院理学研究科 准教授                        |
| 深海 薫       | 理化学研究所バイオリソースセンター情報解析技術室 室長               |
| 山崎 由紀子     | 情報・システム研究機構国立遺伝学研究所<br>生物遺伝資源情報総合センター 准教授 |

##### 文部科学省

河野 広幸 研究振興局ライフサイエンス課専門官

##### 事務局

文部科学省研究振興局ライフサイエンス課  
ナショナルバイオリソースプロジェクト事務局

#### 3. 議事

1. 開会
2. 情報ワーキンググループについて
3. NBRPにおけるデータベース整備の現状について
4. データベース整備に係る中核機関と情報センターの業務分担について

5. ゲノム情報等整備プログラム、基盤技術整備プログラムの成果データの公開プロセスの見直しについて

6. その他

7. 閉会

#### 4. 配付資料

資料 1 : ナショナルバイオリソースプロジェクト情報ワーキンググループ委員名簿

資料 2 : ナショナルバイオリソースプロジェクト情報ワーキンググループ設置要綱

資料 3-1 : データベース整備に関するアンケート結果（中核的拠点整備プログラム）

資料 3-2 : データベース整備に関するアンケート結果（ゲノム情報等整備プログラム・基盤技術整備プログラム）

資料 3-3 : データベース整備に関するアンケート結果（情報センター整備プログラム）

資料 3-4 : データベース整備に関する主な問題点（アンケート結果抜粋）

資料 4 : ナショナルバイオリソースプロジェクト募集要領（平成19年5月）抜粋

資料 5 : 今後のスケジュール（案）

## 議事要旨

### 1. 開会

- ・開会の挨拶が城石主査からあった。
- ・委員の紹介および配付資料の確認が行われた。

### 2. 情報ワーキンググループについて

- ・資料2に基づき、情報ワーキンググループの設立趣旨・目的について文部科学省より説明があった。
- ・委員からの質問はなかった。

### 3. NBRPにおけるデータベース整備の現状について

- ・資料3-1、2、4に基づき、データベースに関するアンケート結果について文部科学省より説明があり、続いて資料3-3に基づき、情報センター整備プログラムの現状について山崎委員より説明があり、その後で質疑応答が行われた。内容は以下のとおりである。
  - このワーキンググループでは具体的に何を議論するのか。（田畑委員）
  - 各機関で抱えている問題をもう一度掘り起こすこと、知財の問題とデータベースの問題をもう一度洗い直すこと、データベースの公開のスピードアップ方策を検討することを目的としている。（城石主査）
  - 個々にはよく対応しているが、このままでは雑然とした状態が続くだけなので、いくつかの категорияに整理し、各中核機関に依頼すべき事項を整理していくのがよい。（田畑委員）

### 4. データベース整備に係る中核機関と情報センターの業務分担について

- ・議事3の説明等を踏まえ、業務分担について質疑応答および意見交換が行われた。内容は以下のとおりである。

#### <データベースに搭載するデータの知財について>

- 我々が扱っているのは2次著作物に当たるので、1次著作物の権利を引きずるのだが、そもそも1次著作物がライセンスという考え方なしに来ているために、何をよりどころすればいいのか分からない。（深海委員）

- 1次著作物とは具体的にどのようなものなのか。著作物と言っても、寄託者が掲載依頼してくるものと、中核機関が積極的に集めているものがあり、データの由来が違うものを同じレベルで話していいものか。（芹川委員）
- 中核機関が独自に集めたものに関しては、中核機関が判断できると思う。しかし、寄託者がリソースの寄託とともに付加情報と一緒に預けた場合、その情報がNBRPのデータベースに出ているのはいいが、関係のないサイトに、しかも形が変わって出ているときに、中核機関が説明できるのか問題である。BRCではゲノムデータに関しては今後の問題なので対処のしようがあるが、既に預かっている系統情報に関しては、遡って対処できるか問題である。情報についてもMTAの取り交わしを検討しているが、整理に難航している。（深海委員）
- 学術用のデータベースでは、著作権はあまり主張しないものが多い。リソースを使ってもらうための情報である。データベースに限らず、改ざんは違法行為であるため、それを避けるために一定のプロテクトをする必要はあるが、中核機関のデータだということが分かる形で他のサイトで使われ、中核機関とリンクするのはいいことなので、できるだけオープンにする方向で進めたい。情報の世界ではMTAは存在しないが、気になる部分は押さえておいた方がいいかもしれない。（山崎委員）
- コンテンツに関しては、最低限出典がきちんと記載されていることが絶対条件だろう。外からのオファーの場合、そこは明言しているのか。（城石主査）
- そういう方向で提供しようということで始まったところだ。（深海委員）
- NBRPのデータの一部が他のデータベースに無断で転用されている事例はあるのか。（城石主査）
- 今のところは聞いたことがない。（山崎委員）
- オファーが来たときは、出典を記載することを条件にして拒否しないという方向性がいいのではないか。（城石主査）
- データベースは公開した時点でオープンになるのだから、あとは利用者のモラルの問題ではないか。NBRPではできるだけ早く公開することを原則にしているので、他で利用されるのはやむを得ない。情報を公開しておいて、それを勝手に持っていくなというのは通じない。出典を書くようお願いすることはできるが、MTA等を取り交わしても法的効果はないと思う。仮にやるとすれば覚書レベルになるのではないか。（田畑委員）
- NBRPではできるだけ公開するのが基本的な方針だが、寄託者と齟齬があってはいけないので、寄託者に一度確認しておくように中核機関に対して発信しておいた方が

いいのかもしれない。また、コミュニティに広く使われることが一番重要であるが、出典はきちんと書いてもらうように要求していくことも必要である。ただ、著作権に関するところは専門家の意見を聞いた方がいいので、実費徴収・知財ワーキンググループともクロストークをしておいた方がいいのではないか。（城石主査）

- 情報ワーキンググループからの宿題ということで、情報の知財の対応については実費徴収・知財ワーキンググループにおいて検討する。（河野専門官）

#### <内部データベースと公開データベースの関係>

- 公開のデータベースをきちんとしようとする、内部のデータベースもきちんとしなくてはいけない。系統維持を日々行っているデータベースについて、バックアップ等の要望はあるのか。（城石主査）
- ショウジョウバエでは、全面的に情報センターが支援して内部の管理運用データベースを作った。大掛かりな仕組みではなく自分たちがハンドルできるものでいいなど、ケースバイケースだが、メダカも要望があればできると思う。（山崎委員）
- 酵母では、情報センターで作ってもらったフォーマットに従って入力し、データのアップは自分たちでやっている。（中村委員）
- ラットでは、内部データは日々変わっているので、毎月ミーティングをして、最新の生データの確認を行うとともに、内部で共有フォルダを作り、業務に従事している各人が見られるような環境を作ってマネージしている。（芹川委員）
- 一番ネックになるのはデータ入力のインターフェース。作業が面倒になれば、全て意味がなくなる。（深海委員）
- 一番難しいのは、複数の方が作業を分担しているとき。ショウジョウバエでは、実際に運用すると不具合が結構出てきて、使えるようになるまで1～2年はかかった。（山崎委員）
- 遺伝研のマウスでは、凍結胚等はバーコード管理が全部進んでうまくいっている。民間でも結構安いプログラムができており、うまく使えないか。（城石主査）
- 使いやすい例があれば、そういうものを型にして使っていただく。ただ、公開用データベース以上に機関毎に違いが出てくると思う。内部のデータベースは一緒だが、インターフェースが非常に複雑になり、運用は情報センターもそれほど関与できない。現場サイドで一番使いやすい形に作っていただくしかない。（山崎委員）
- コマースの汎用プログラムはないのか。（田畑委員）

- 汎用性のあるもののカスタマイズは、決められたプロトコルに従ってしかできない。それを崩していくとなると、最初から作った方が早いということもあり、一概には言えない。マウスや血清では結構よく使われているシステムがあるが、それ以外ではなかなかない。ショウジョウバエで作ったものを他の生物種用にカスタマイズすることは可能だが、そこまで複雑になるのなら、かえて自分たちがやりたいところから入ってもらった方がすっきりするのではないか。ただ、こういうものは必要なので、全面的に支援はする。(山崎委員)
- モディフィケーションが出たときに、誰が対応するのか。ずっと情報センターに対応してもらえる仕組みがあればいいが、リソースが続く限り情報のデータベースも続くわけで、結構悩み深い。(芹川委員)
- がっちりしたシステムを作ってしまうとお金が掛かるので、むしろ自分たちがハンドリングできる程度に留めておきたいという要望もある。(山崎委員)

#### <国際的連携をする上での問題点～命名法について>

- 生物種によっては国際的な Nomenclature (命名規約) に沿った形でやっていないところがあるのだろうか。(城石主査)
- あまりに複雑になりすぎているので、専門家がやっても、間違いだとか、こちらの方がベターというのがある。(山崎委員)
- Nomenclature の対応を情報センターにお願いすることができるのか。生物種ごとにいろいろな条件があるので、中核機関でやらざるを得ないのではないか。もしそれを放棄するのなら、中核機関として成り立たない。(芹川委員)
- 植物ではそんなにうるさい命名法はない。遺伝子は少しうるさいが、系統に関しては付け放題ではないか。(田畑委員)
- 酵母では、分裂酵母の系統は1種しかなく、あとはミュータントだが、きちんと3文字プラス数字で決まっている。遺伝子の方はゲノムプロジェクトから付けられた番号とリンクさせているので、そういう問題はほとんどない。系統の方は研究者が各自勝手に付けているが、今のところ混乱はない。ただし、今後は問題が起こってくる可能性があり、少なくとも我々が作成するDNAクローンに関してはきちんとしたルールの下にやっていかなければいけない。(中村委員)

#### <データベースに関する技術講習について>

- データベースに関する技術講習について、情報センターに具体的な依頼は今まであったのか。(城石主査)
- トマトは、コントリビュートを使ったページの自主管理を始めたばかりで、データベースに関してはこれからである。技術講習に関しては、何をやるかはリソースごとにばらばらなので、今年は個別にやる。ある程度行き渡ってから全体の講習会が展開できるのではないかと思うが、何かメニューを作ったとしても、全くつまらないという人とぜひやってほしいという人がいて、かなり違う。今のところは個別の方がいいのではないか。(山崎委員)
- 情報センターがどこまでケアするかで講習の中身も相当変わってくるので、カテゴリー分けをしておいた方がいいのかもしれない。(城石主査)
- 今年の3月に行ったデータベース研究会では、初めてリソースの方が一堂に会して意見交換をした。それまでリソース機関対情報センターの関わりしかなかったが、他のリソース機関の情報整備について発見があり、リソース機関から評判が良かった。今後も続けていきたい。【山崎委員】

## 5. ゲノム情報等整備プログラム、基盤技術整備プログラムの成果データの公開プロセスの見直しについて

- ・議事4に引き続き、業務分担について質疑応答および意見交換が行われた。内容は以下のとおりである。

### <マウス(奈良先端大)からの要望について>

- BRCに持ち帰り、吉木先生(マウス代表者)に確認する。シロイヌナズナのcDNAは既にblast検索ができるようにシステムづくりをしているので、技術的に問題ない。(深海委員)

### <技術研修の広報について>

- 基盤技術整備プログラムで開発した技術の研修を行う場合にNB R Pのホームページから広報するのは、喜んでやるべきことだと思う。成果の公開方法等については今後検討していきたい。(城石主査)

### <ゲノムデータの公開について>

- 論文発表待ち、コンソーシアムとの取り決めによりゲノムデータの公開時期が遅れがちになっている。しかし、NB R Pで解析したゲノムデータは可能な限り早めに公開

するのが原則であり、公募の条件にもなっている。従って、強権発動して、とにかく公開するように強く言うことは可能だが、間違っただデータを公開されても困る。猶予期間も含めて、議論したい。(城石主査)

- 個々に考え方が違う。かずさでは個々の研究者が自分で解析から論文発表時期まで決めているが、アメリカでは公的資金でサポートされているときには、ショットガンのローデータから次の日にはアSEMBル関係なしに全部出すことになっているし、ヨーロッパではもう少しプロテクトしている。一番分かりやすいのは、出たデータは瞬時に公開しなさいとすることではないか。このプロジェクトの趣旨から言うと、コンソーシアムとの取り決めというはまずい。一部でもNBRPのお金が入っている場合には、日本人の常識的な線ですぐ公開ということが守られるべきだろう。そこは周りの人がウォッチしてプレッシャーをかけていくぐらいしかないのではないか。(田畑委員)
- 解析が終了してから1年ぐらい経てば、基本的な誤りを十分チェックできるのではないか。(城石主査)
- NBRPのシーケンシングは理研か遺伝研でやる場合がほとんどなので、少なくともローデータの収集がいつ終わったかは分かっているわけで、その後は大体想像がつく。今回の場合、信頼して任せていたのだろうが、そういう例が出てきてしまうと、これからは無条件に信頼するというわけにはいなくなるだろう。(田畑委員)
- そのため、暗黙のプレッシャーだけでは済まなくて、ある程度の猶予期間は与えて、遅くともこのぐらいまでにはアップしてくださいというリコメンデーションを出す必要があるのではないか。(城石主査)
- 推進委員会でウォッチし、あまり遅いようだと状況を尋ねる等、何らかの対応を取ればよいのではないか。(田畑委員)
- 公開というのは、自分のところで配列を見せればそれでいいのか。かずさのように、シアノベースをきちんと作ってから公開するとなると非常に敷居が高いと思うが、ローデータをそのまま出すのも公開だし、きちんと整備してデータベースにして出すのも公開だし、いろいろな段階がある。(深海委員)
- 少なくとも何らかの形で生のデータが見られるようにする必要はあり、ずっと閉じられた状態ではよくない。また、NBRPでやっているということから言えば、特定の限られた人だけが見られるのではよくない。少なくともNBRPのホームページから直接入れるところで情報が見られるようになっているべき。(城石主査)



- 例えば理研や遺伝研にシーケンスをお願いしたときに、どんな状態で返ってくるのか。  
(田畑委員)
- ケースによって違うが、理研も遺伝研もある程度まで加工するノウハウは持っている。  
しかし、NBRPではこれまで加工等のサポートはなかなかできなかった。そこは基本的に理研や遺伝研が中心になって、ある程度仕上げて、中核機関にデータが戻されているという状態だと思う。(城石主査)
- 戻されても、中核機関でそのデータをさらにフィニッシングするのは難しいだろう。  
実際はblastがかかったデータが来て、それを見てアノテーションを付けていく作業を中核機関でする。それはそう時間はかからないので、戻された時点から何カ月後、半年後などに確認するような仕組みを作っておけば大抵は大丈夫ではないか。(田畑委員)
- NBRPのゲノムの公開は、DDBJに登録しなくてもいいのか。DDBJへの登録は大変なので、そこをサボってしまう人もいると思う。(深海委員)
- 情報センターに来たものに関しては、そういうつもりがなくても全部DDBJに登録する。(山崎委員)
- ゲノム解析情報の公開については具体的なルールを作る必要がある。(城石主査)
- DDBJへの登録で公開が終わるとなると、それなりに手間と労力がかかる。そこまで求めるからには、それに対する措置も必要になる。(深海委員)
- NBRPの予算が幾らでもあればいいが、結局、そこまでケアできる体制で申請をしていただくしかない。【城石主査】
- もし1年以内にとすると、NBRPでは4年目までしかゲノム解析はできず、その1年後(5年目)に出てきたとしても、遺伝研で自腹を切ってやるしかないといった状況にもなる。(河野専門官)
- 開始時期を7月1日から変えられないのか。【芹川委員】
- 委託契約の制度の中でやっているのだから、早くやっても7月1日になる。【河野専門官】

## 7. その他

- ・資料5に基づき、今後のスケジュールについて文部科学省より説明があった。
- ・委員からの質問はなかった。

## 8. 閉会