

第 2 回 SPARC Japan セミナー2011

「今時の文献管理ツール」ワークショップ

Mendeley's Role in the Open Science Infrastructure

(オープンサイエンス研究基盤における Mendeley の役割)

Dr. Victor Henning

(CEO of Mendeley)

講演要旨

Mendeley は 2008 年、3 人の博士課程の学生により始まり、3 年後、利用研究者数は 100 万人を超え、文献管理数、1 億 3,000 万を超える世界最大の科学研究共同プラットフォーム、データベースとなった。API (Application Programming Interface) を利用して第 3 の開発者 (利用者) は、Mendeley を研究の評価、可視化、共有、発見のためのオープン・サイエンス・プラットフォームへと成長させている。

何がこの急成長を可能としたのか? —Mendeley ストーリーから、他の“サイエンス 2.0”プロジェクトの何を学ぶことができるか、21 世紀に構築されたこの科学研究手法は何を意味するのか?



Dr. Victor Henning

Mendeley 共同設立者兼 CEO。

Mendeley は、2009 年、ロンドン・NY を基盤に活動を開始し、2009 年の設立以来、世界に 100 万人以上のユーザーを持つ巨大共有プラットフォームに成長し、今や、1 億 3,000 万以上の文献を有するクラウドを利用した世界最大のデータベースを有する。

英国王立芸術協会 (RSA) フェロー。“消費者意思決定における感情の役割”をテーマにバウハウス大学 (ヴァイマル) 博士号を取得。過去には戦略コンサルティング会社、映画制作、音楽業界にも従事。

Mendeley が目指すもの

本日は、クラウドソーシングリサーチ、オープンサイエンス研究基盤における Mendeley の役割についてお話しいたします。

最初に、皆さまがよくご存じの World Wide Web (インターネット) の父といわれる Tim Berners-Lee が TED カンファレンスで言った言葉を引用したいと思います。「われわれ人間は常に、直面している大きな問題、たとえば、がんの治療やアルツハイマー病の脳の働きについて理解することには非常に大きな関心を払っている。しかし多くの場合、人類の知識はそれぞれ

の科学者、研究者のコンピューターの中に閉じこめられていて共有されていない。これを開放して、大きな問題の解決策を見出すことができるようにしなくてはいけない」。これこそ、Mendeley が解決しようとしている課題です。

Mendeley についてさらに知りたい方は、「WIRED」という雑誌の最新号に紹介記事が写真付きで載っていますので、ぜひ読んでいただければと思います。

Mendeley の始まり

最初に、Mendeley の発想がもともとどこから来た

のかをお話いたします。

3年余り前の2008年、Jan、Paul、私という3名のドイツのPh.D.の学生が集まりました。そのとき私たちの手元にあったのは、アイデアとSkypeのアカウントのみでした。ある時、私たちは、研究者が共通の問題を抱えているということに気が付きました。すなわち、研究者として無秩序に文献をたくさん持っている。例えば私はPh.D.の学生でしたので、コンピューター関連分野の論文を700ほどため込んでいました。このような文献をもっと簡単に整理できないかと考えたわけです。そして、PDFから情報を自動的に検索して整理するソフトが書けるはずだと思い、それを3年前に開発しました。

Mendeleyの抽出機能

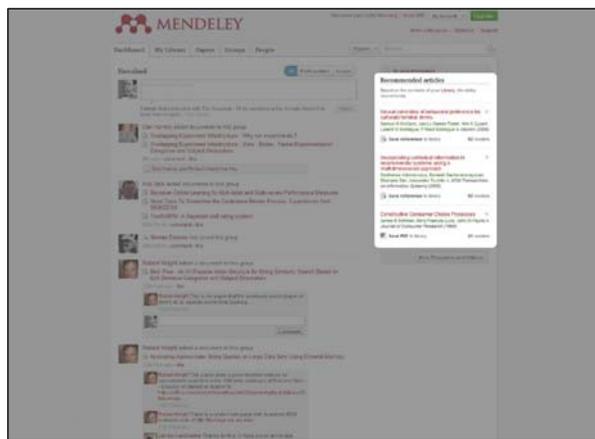
Mendeleyのコアは、Windows、Mac、Linux、iPhone、iPad上で動くフリーソフトです。コンピューターに保存されているPDFを自動的に判別し、書誌情報(著者名、タイトル、掲載誌名等)をリストアップして、PDFを構造的なデータベースに体系付けます。PDFを検索したり読んだりすることが可能ですし、さらにさまざまなメモやノートを付けたり、ハイライトしたりすることもできます。また、個々のユーザーが持つ情報を抽出し、クラウド上に研究データとして集めることができます。

デモンストレーション

実際にデモをご覧ください。これがデスクトップ版のMendeleyのインターフェースです(図1)。例えば、著者別、タイトル別、出版年別、掲載誌別に検索ができ、すべてのドキュメントの全文検索も可能です。

例として、私の研究分野でemotion(感情)という言葉で検索してみると、論文がリストアップされ、クリックするとその論文が表示されますので、PDF版を読むこともできますし、PDF内の検索も可能です。

例えば、この論文の中に仏教の話が出てきたという



(図1) Mendeleyのインターフェース

ことを思い出したとしましょう。そこでBuddhismという言葉で検索すると、ハイライトされます。ハイライトされたものをクリックして前へ前へ進むことができますし、必要に応じて自分なりのハイライトを追加することもできます。また、何かノートを追加すると、吹き出しが表示されます。このようにして、この論文に関する自分なりの考えを整理していくことができます。

学術論文を書いているときには、ビブリオグラフィやサイテーションを入れることもできます。

マイクロソフトのWordには、プラグインがあり、Insert Citationというボタンをクリックするだけで引用文献が挿入されます。この引用をライブラリ内で全文検索することもできます。

ビブリオグラフィを作らなくてはならないときには、Insert Bibliographyというボタンをクリックすればできます。引用のスタイルは変更でき、変更するとそれが自動的に反映されます。

また、グループを作成して論文やノートを共同研究者と共有することができます。ニュースフィードで、グループ内のほかの人たちが何をしているのかという活動情報を得たり、各共同研究者がグループに追加したドキュメント類を確認したりすることができます。

これはFacebookと同じように共有・アップデートができ、この部分には「今日、私は東京のNIIで話しています」と書き込んであります。

Mendeley にはプロフィールも入っていて、そこに自分の出した論文や出版物を簡単に追加することができます。My Publication というところに私のいろいろな文献が載っているのですが、それが自動的にこのウェブサイトのプロフィールに反映されます。受賞したものや獲得したグラントなどを出すこともできます。あるいは、このプロフィールをほかのウェブサイトにも組み込むことも可能です。ウィジェットがあり、クリックすることで、どの情報を入れるか選択・編集できます。つまり、Mendeley を使って、ほかのウェブサイトにも自分のプロフィールや論文などの情報を載せることができるということです。

さらに自分の分野でどんな新しい活動が起こっているかを発見することもできます。例えば、Groups をクリックして、私の専門分野である心理学を選ぶと、今そのコミュニティでどのようなタグや用語、グループが人気があるのかが分かります。コミュニティ内のメンバーも一部紹介されていて、プロフィールも確認できますので、自分の興味のある分野の研究者を探すこともできます。

グループの情報を見ると、どんな文献が追加され、読まれているのかが分かります。例えば、「Sing the mind electrical, principles of deep brain stimulation (心の電気を歌う、深層頭脳の刺激原理について)」という面白いタイトルが付いているものがあります。このページを見るとアブストラクトと共に論文の詳細が分かります。例えば、Mendeley でこの論文を 22 人が読んでいることが分かりますし、読者の専門分野や肩書き、国も分かります。サイテーションを見たい場合には、References をクリックすると、この文書の中で引用文献がどのような形で使われているかまで見られます。例えば「Remembering the past and imagining the future (過去を記憶し、未来を想像する)」という引用文献がありますが、このペーパーについて、この著書は何と言っているのかを見ることができます。もちろん、論文を検索してさらに内容を知ることができます。このようにして、自分が興味のある分野の新しい

研究論文、研究活動を知ることができますし、アマゾンのように、それぞれの文書について関係している研究分野、研究課題も検索できます。

新しいフィーチャー

間もなくリリースされる新しいフィーチャーを二つご紹介します。

一つ目は、QuickSend といわれるものです。これを使うと、コンタクトリストや Eメールの情報をそのままインポートすることができます。そのコンタクトに対して文書類をドラッグ&ドロップでき、相手との文書の共有がすぐにできます。

二つ目は、Mendeley Suggest というものです。これは今あるマイライブラリに入っている情報をベースに、自動的にお勧めのものを推薦するという機能です。もちろん iPhone 向けのアプリケーションも用意しているので、移動中でもドキュメントを読むことができますし、iPad バージョンもあり、こちらの方が画面が大きくて、より見やすくなっています。アンドロイド用には、第三者が用意しているプログラムが幾つかあり、それを使うと Mendeley のライブラリをインポートすることができます。そしてドキュメントの検索、あるいは自分が所属しているグループにアクセスできます。これらはすべて無償です。

Mendeley のビジネスモデルとユーザー

皆さんは、Mendeley はどのようにして利益を上げているのかと思われるかもしれませんが、われわれはベンチャーキャピタルやビジネスエンジェル、例えば Skype の設立者などから出資をいただいています。そればかりでなく、市場が非常に大きいという点で、Mendeley はビジネスモデルとして優れていると自負しています。Mendeley のようなツールを使うユーザーとしては、学者、企業、政府系の研究者、エンジニアなど、およそ 1 億 5000 万人はいると見ています。

ビジネスモデルのベースは、個々のプレミアムアカウントです (図 2)。無料のプランは EARTH プランと

名付けています。1GB のスペースが提供され、最高 10 人までがこのグループ内で共有できます。それに対して、もう少しストレージスペースが必要な場合には、プレミアムアカウントというコースを選ぶことが可能です。SOLAR SYSTEM プランは月 4 ポンド 99 ペンス、MILKY WAY プランは月 9 ポンド 99 ペンスです。それから、BIG BANG プランは、企業等、非常に大きなアカウントが必要なお客さま向けです。それぞれストレージスペースと協働できるグループ内のユーザー数が異なります。これは個人ベースのものですが、機関・企業バージョン、あるいは研究データやアナリティクスを対象にしたものもあります。

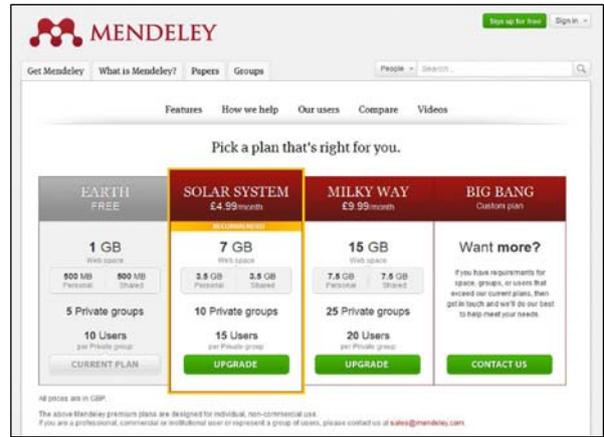
現在のところ、SOLAR SYSTEM プラン、MILKY WAY プランのユーザーはそれぞれ数千人、そして無料バージョンの EARTH プランのユーザーは 100 万人近くです。私どもはまだスタートアップ企業で、黒字化はしていません。今後 2 年のうちにぜひ黒字し、利益を上げたいと考えています。

また、2012 年 1 月には、Swets 社と共同で図書館向けの商品を発売する予定です。Swets は、図書館のジャーナル定期購読を支援する世界最大の会社の一つで、Mendeley のコラボレーションという機能を図書館に対して提供します。例えば、その図書館が定期購読している学術誌の利用率、あるいは図書館利用者の活動内容分析など、図書館にとって興味のあるデータを示すことが可能です (図 3、4)。これは来年 1 月から使えるようになる予定です。

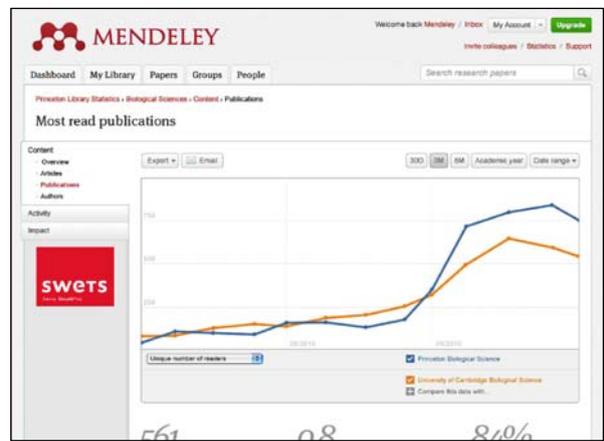
そして他のビジネスモデルは、研究論文の iTunes 版です。現在、私どもは多くの学術出版社から引き合いがあり、Mendeley ウェブサイトへのメタデータやコンテンツの掲載を進めています。それらの出版社の多くが、直接 Mendeley 経由でコンテンツを販売したいと考えています。

Mendeley Advisors に支えられた急成長

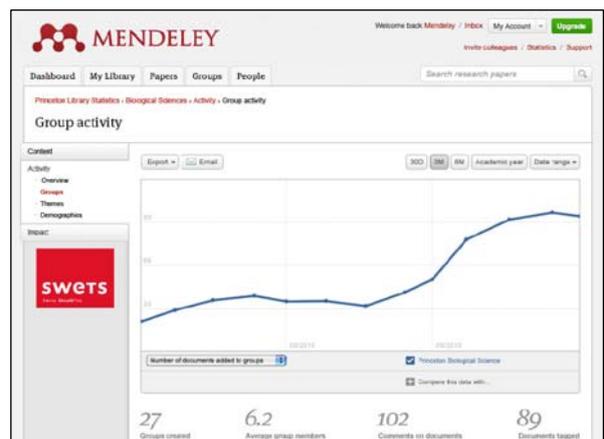
過去 3 年で Mendeley は大変な勢いで成長してきました (図 5)。現在、ユーザー数が 130 万という世界最



(図 2) 個々のプレミアムアカウント



(図 3) Most read publications



(図 4) Group activity

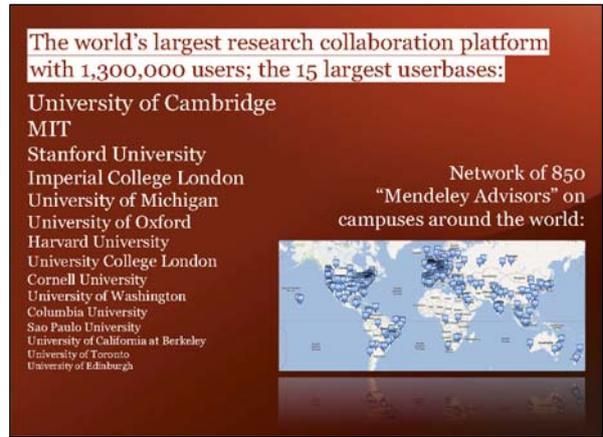
大規模の研究コラボレーションプラットフォームとなっています。Mendeley ユーザー（研究者）の数が多く機関を 15 ほどここにリストアップしました。ご覧のように世界でトップクラスの機関が、真っ先に Mendeley を利用してくれています。ここに載っているすべての機関と直接契約しているわけではありませんが、ユーザー数がこれだけあるということを理解してください。

このような急成長が可能だったのは、850 名の Mendeley Advisors のネットワークがあったからです。このアドバイザーは世界各地にいて、Mendeley のユーザーで、かつ Mendeley のプロダクトを非常に気に入ってくれている人たちです。利用者が多ければ多いほど、より多くの情報を共有できるというメリットから、Mendeley の普及に協力したいという申し出をしてくれました。

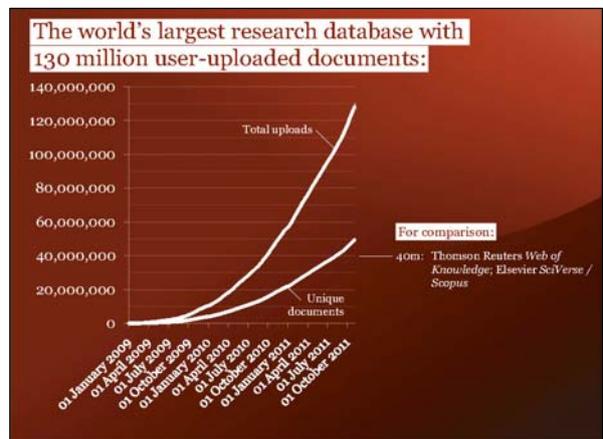
Mendeley のアドバイザーの人たちは、新しい機能等が出ると、優先的にアクセス可能ですし、プレミアムアカウントも無料で使えるほか、Mendeley の T シャツなどももらえます。そして、各大学のキャンパス等でローカルユーザーグループを作成し、Mendeley の活用方法をプレゼンしたりすることが許されています。

膨大なアップロード文献

私どもが最も驚いたのは、アップロードされるドキュメントの多さです（図 6）。Mendeley が立ち上がったのが 2009 年 1 月で、ドキュメント数は 0 から始まりましたが、現在までに 1 億 3000 万本がアップロードされています。1 日当たり 50 万本という勘定になります。また、このうち 5000 万が重複掲載されていないドキュメントです。トムソン・ロイターの Web of Knowledge、あるいはエルゼビアの SciVerse などは 4000 万という規模なので、私どもはそれらを 25% 上回る規模になっています。Mendeley は必ずしもすべて全文が掲載されているわけではありませんが、これは恐らく SciVerse も同様だと思います。



(図 5) Mendeley のユーザー（研究者）



(図 6) アップロードされるドキュメントの数

Mendeley のデータベースの特徴は、単にメタデータ、アブストラクトだけが載っているのではなく、デモグラフィックスというコラムがあり、利用者の情報が分かるようになっていることです。また、ドキュメントをアップロードしたユーザーが使ったキーワード、タグも整理しています。その文書に関連した内容の文書がどこにあるかや、読まれている状況に関するリアルタイムの統計数も見ることができます。

このようなデータベース、情報のすべては Mendeley Open API で利用可能となっています。dev.mendeley.com でデベロッパーのアップロードキーを登録し、クリエイティブコモンズというライセンスで情報を扱うことができますので、アップロードが自由にできます。

大学図書館との連携

大学や図書館とも統合を進めています。その一つが JISC DURA Project です。これは、イギリスの政府組織である JISC (Joint Information Systems Committee) という委員会の資金を得て、ケンブリッジ大学と共同で行っているプロジェクトです。

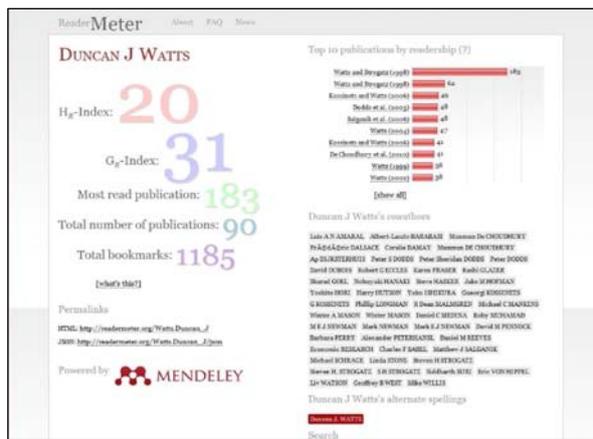
Mendeley のユーザーが例えばケンブリッジ大学の教授である場合、Mendeley のプロフィールに自分の出版した論文を掲載することができます。そうするとケンブリッジ大学のリポジトリに自動的に登録されるようになっています。ケンブリッジ大学のリポジトリも私どもの方でインデックス化しているので、自動的に Mendeley から見られるようになります。このようなプロジェクトをほかの大学とも進めて、ほかの大学のリポジトリとの統合化も行いたいと思っています。

「リポジトリの中で出版されたものと、そうでないものが明確に区別されているのか？」というご質問ですが、このプロジェクトにはマクミランという出版社が第3のパートナーとしてかかわっていて、マクミランが提供しているリポジトリマネジメントツールにより、どの論文がリポジトリに入るかを管理しています。これに関するご質問はブログスポットがありますので、こちらにお寄せいただければ回答いたします。(http://jisc-dura.blogspot.com)

Mendeley/PLoS Binary Battle

もう一つ私どもが行っているのは、アプリケーションの開発を進めるためのコンペです。PLoS (Public Library of Science) と組んで、Mendeley/PLoS Binary Battle を開催し、応募された中の最優秀アプリケーションには1万1ドルの賞金を提供しています。Binary Battle という名前のコンペですので、賞金は1等賞が1万1ドルになっています。2進法でいくと、17ドルになるのでしょうか。とにかく賞金は1等賞が1万1ドルです。

審査員は、Web 2.0 という言葉を造語した Tim O'Reilly さん(オライリーパブリッシング)、クラウド



(図7) Reader Meter

コンピューティング発明者の1人といわれる Werner Vogels さん(アマゾン CTO)、James Powell さん(トムソン・ロイターCTO)など、非常に面白い方々に務めていただいています。

このコンテストで開発されたアプリケーションの一部をご紹介します。

一つ目は、Reader Meter というものです(図7)。著者の名前を入れると、Mendeley のデータベースからその人の書いた出版物を引っ張ってきます。Mendeley 上でその人の各著作物を読んでいる人がどのぐらいいるかという数字も出してくれます。例えば著者 Duncan Watts について、共同著者すべての名前や、H-Index、G-Index といった統計の数字、また最も読まれている著作物を表示してくれます。Publication をクリックすると、著者の書いたものを読んでいる読者がどこにいるか、まだどんな分野の人が分かれます。

二つ目は、kleenk というものです(図8)。Kleenk では、Mendeley のライブラリを引っ張ってきて、自分のライブラリのドキュメントとのリンクを張ることができます。このリンクは、それぞれのドキュメントがどのように関連しているのかをセマンティックに記述し、その人のライブラリにあるドキュメントがどのような関連性を持っているのかというマップを作成してくれます。

三つ目は、Collabgraph というものです(図9)。

Mendeley のライブラリをインポートすると、ライブラリに入っているドキュメントすべてについて、それぞれどんな関連性があるかを示すグラフを作成します。

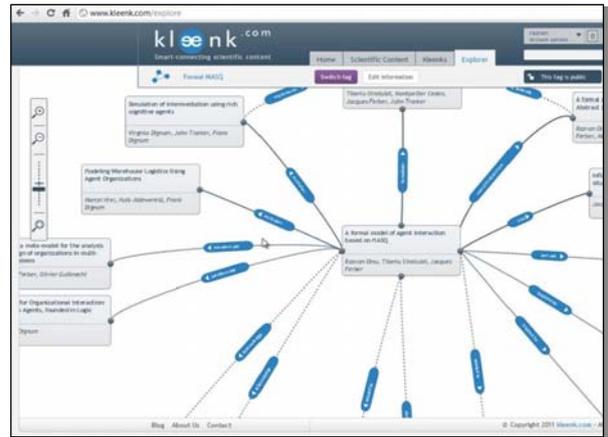
四つ目は、PaperCritic というものです (図 10)。これは Mendeley のデータベースに載っている論文に関してレビューを書いて評価するものです。例えば独創性という点ではどうかといった評価をすることが可能ですし、また、ツイッターでその論文に言及しているものを表示することができます。つまり発表された研究論文に関して、クラウドソーシングの情報を提供します。

五つ目は、openSNP といわれるものです。実はこのアプリケーションが Binary Battle の最優秀賞を取り、賞金 1 万 1 ドルを獲得しました。23andME というサービスをご存じかもしれませんが、これはゲノム情報を解読してくれるサービスです。自分のゲノムのデータを openSNP にアップロードすると、同様にアップロードした人の中で、似たようなゲノムデータを持っている人を発見することが出来ます。その情報が自動的に Mendeley および PLoS 上の最新のリサーチペーパーに関連付けられ、遺伝学分野の研究者が、家系や健康に関する情報等を知らせてくれます。また、このサイトから遺伝子情報をダウンロードすることもできます。

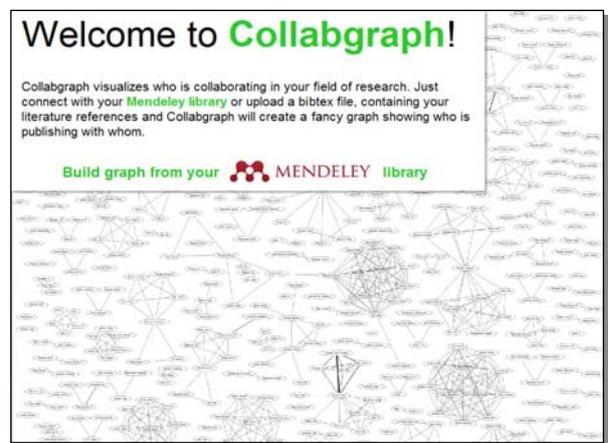
この openSNP は、より多くの人がゲノムの解析をしたり、また、データをアップロードできるように、賞金 1 万 1 ドルを使うと約束しています。このようなツールがオープンサイエンスの発展にどのように貢献しているかということを示す、大変良い事例でしょう。

今後の展望

既に Mendeley を、文書内のアノテーション、ハイライト等の作成に活用してもらっていますが、その中で、自分のライブラリにあるドキュメント間のセマンティックなリンクをしたいというリクエストがありました。例えば、今ある人が読んでいるドキュメントとほかのドキュメントの間にリンクを張って、こちらの



(図 8) kleenk



(図 9) Collabgraph



(図 10) PaperCritic

ドキュメントはこのドキュメントの言っている内容を支持している、あるいはその内容は違うと主張しているというような関連付けをするというものです。

考えてみると、Mendeley のユーザー同士や分野の垣根を越えたこのようなリンクは、セマンティックなアノテーションをクラウドソーシングして、グローバルな科学のプレーンのようなものを、我々が創り出しているのではないのでしょうか。第三者のデベロッパーの開発した API を使えば、これらの情報すべてが利用可能になります。だからこそ、アマゾンの CTO の Werner Vogels が、「Mendeley は科学の在り方を変えていくことができると強く信じている」と言っているのだと思います。

◆
●Q 1 日本語の論文や、韓国語、中国語などの論文も使えますか。

●Henning 一般論として Mendeley はどんな言語でもご利用いただくことができます。データベース、カタログに既に入っています。ただ、利用者の大半が今はまだアメリカ、イギリスの人たちで、掲載されている文献や論文等は、ほとんどがアメリカやイギリスからのものになっています。

現在、私どもの会社に 1 人、中国人のデベロッパーがいますが、残念ながらそれ以外は韓国語や日本語を話せる人間がいません。ただ、既にニューヨークにオフィスがありますので、将来、東京にもオフィスを構えることになるかもしれません。

●Q 2 アドバイザーになって、日本語のものを向上させるために協力してもいいですか。

●Henning もちろんです。既に日本にも何人かアドバイザーがいます。ウェブサイトの地図を見ると分かるかと思いますが。

●Q 3 私は EARTH ユーザーです。ユニークというものが出たのですが、私の論文が二重に登録されていて、いつまでたってもマージされません。今後、マージのリクエストを受け付ける機能を搭載する予定はありますか。

●Henning 残念ながら、まだ私どものアルゴリズムが間違っていることがあります。しかし最終的にはこれもクラウドソーシングで解決したいと考えています。ユーザーにこれは同じものだと言っていただいて、解消することを狙っています。

●Q 3 アカウントを共有するときに、一部分ずつ別々の人と共有できますか。また、自分の論文が引用されているサイテーションのリストを作ることはできますか。

●Henning 一部を共有することはできます。まず、異なるグループを作成します。同じドキュメントをいろいろなグループに入れると、それぞれ別のコピーになりますので、一つのコピーに付けたアノテーションはほかのグループでは見られません。

二つ目の点については、まだできません。今、Mendeley のデータベースに載っているサイテーションを見るということをしています。一部のドキュメントに関しては、リファレンスという項目がありますので、そこから引用を見ることができます。リファレンスのところで、「これは Mendeley ラボの実験的な機能」であることをご説明しています。