

メタデータ付与によるオープンサイエンスの実現

研究支援産業・ベンチャー

SATテクノロジー・ショーケース2020

■ はじめに

今般では、研究論文、そして研究データなどに必要なメタ情報を付与することで、多様な研究成果を部門、研究所の垣根を越えて相互利用していくオープンサイエンスの動きが加速しています。(2019年3月には内閣府が研究データリポジトリ整備・運用に係るガイドラインを策定)

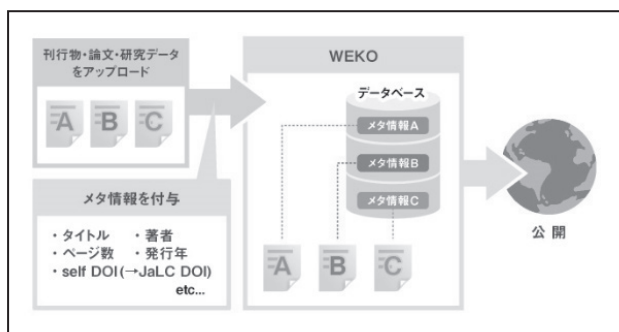
■ オープンサイエンスによる新分野創出

従来、研究データは研究者、部門、研究所の中でのみ利用されることが多くありました。また、それらを統合しようとした場合、データの統一化などに要する時間・コストが阻害要因となることがありました。しかし今般、データはそれぞれの場所に、それぞれの形式で置いたまま、メタデータを付与することで、横断的な検索を可能とする新技術が登場してきました。これにより、従来にはなかった学際的な新分野の創出が生まれてきています。以下に弊社が携わった3つの事例を紹介します。

■ 事例紹介

事例1: 刊行物、論文等へのメタデータ付与および管理

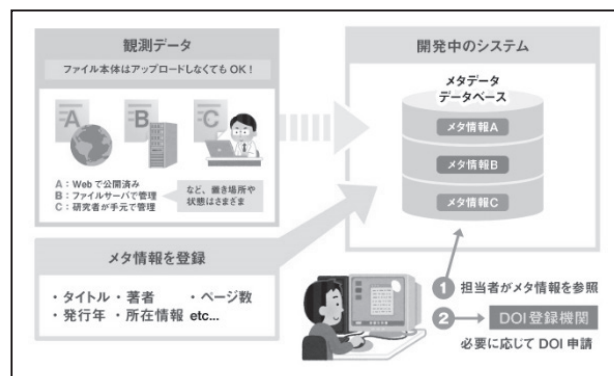
1,000件超の刊行物、論文、研究データにJaLC DOIを含むメタデータを付与した上で管理、外部への公開を行う仕組みを構築しました。管理、公開にはWEKOを利用しました。管理対象のほとんどがPDF形式のファイルでしたが、タイトル・著者などの情報に加えて、本体のPDFファイルからテキストを抽出し、その内容をメタデータの種類として登録しておくことで、刊行物や論文の中に登場するキーワードからも検索可能な状態となりました。(下図)



事例2: 観測データへのメタデータ付与および管理

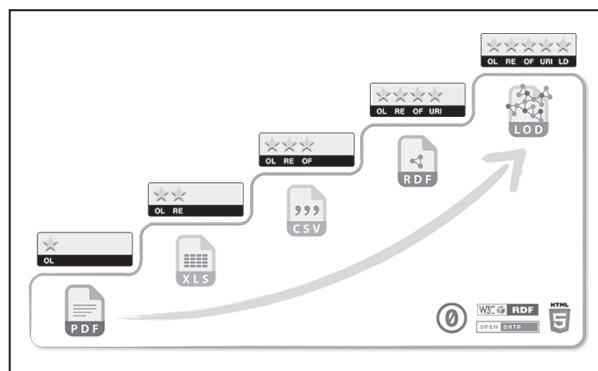
衛星・気象等のさまざまな観測データにメタデータを付

与し、管理する仕組みを開発しています。現状の業務フローを変えずに生かし、ユーザー様にとってご負担の少ない方法で実現する方向で進めています。(下図)



事例3: LOD (Linked Open Data) として公開

PDF や Excel 形式等のファイルを Web 上に配置するのは異なり、データ自体に URI を付け、SPARQL を用いて他のオープンデータと統合して利用可能とする方法です。データの利用者にとっては、Web 上のあちこちに散らばったデータを自力で集めたり、項目を関連付けたりする手間がなく、求める情報を取得したり比較することができます。データの公開者としては、自分のデータを見つけてもらいやすく、且つ活用してもらいやすくなり、データの価値が高まるというメリットがあります。(下図)



図：オープンデータの公開レベルの分類
(5つ星オープンデータ <https://5stardata.info> より)

代表発表者 川鍋 真利(かわなべ まさとし)
所 属 ペンギンシステム株式会社
問合せ先 〒305-0047 茨城県つくば市千現 2-1-6
つくば研究支援センターCB10
TEL:-029-846-6671 FAX:029-846-6672
customer@penguins.co.jp

■キーワード: (1) オープンサイエンス
(2) メタデータ
(3) LOD (Linked Open Data)