

積極的なデジタル化でコンテンツの提供力を強化、 今後はAIや機械学習を活用したユースケースの拡大をめざす

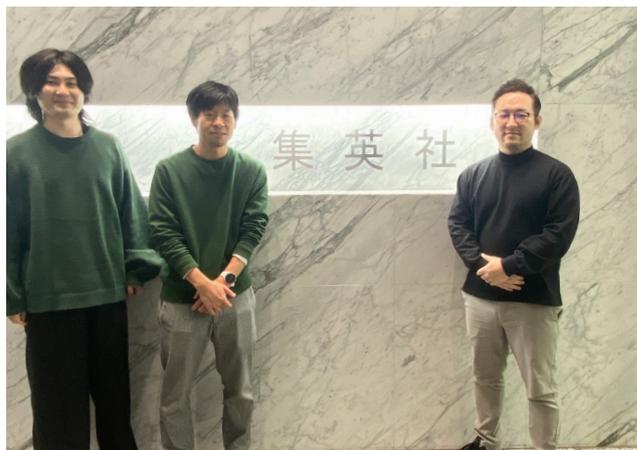
集英社 SHUEISHA

豊富なコンテンツを強みに 多角化とデジタル化を推進、 新たなプロジェクトも積極的に展開

集英社は1926年に設立され、間もなく100周年を迎えます。現在は、祖業である紙の出版物にとどまらず、デジタル配信や映像化、リアルでのイベントや、物販・ECなど、コンテンツを通してさまざまな展開を推進しています。このような取り組みの幅の広がりにあわせて、部署を横断する新たな社内プロジェクトの立ち上がりが増えていきます。また、集英社といえば「少年ジャンプ」「りぼん」などのマンガ誌のイメージが強い一方で、元々は小学館の娯楽雑誌部門が分離して設立された会社であり、女性誌・取材誌も多くのタイトルを展開しています。また現在は、女性誌も紙の雑誌にとどまらず、「SPUR.JP」「MAQUIA ONLINE」など14の媒体/ブランドをWeb上でサービス展開しています。「HAPPY PLUS（通称ハピブラ）」は、これら女性誌のコンテンツを包括するポータルサイトです。紙の雑誌の宣伝サイト「s-woman.net」を前身とする「HAPPY PLUS」は、「s-woman.net」をリニューアルする形で2012年11月にオープンし、会員数300万人（2022年7月末時点）を抱えるサイトに成長しています。「s-woman.net」は、女性誌の告知・販促を主な目的とした媒体でした。一方、「HAPPY PLUS」は、それらのコンテンツをWebで配信し、収益化することを目的としたものです。

データブリックスを採用して 次世代のデータ基盤を構築、 圧倒的なコストパフォーマンスが 決め手

同社の主要事業の1つである女性誌は、各媒体のデジタル配信にも力を入れています。デジタル配信では、PV（ページ・ビュー）やUU（ユニーク・ユーザー）を主要KPIとしてトラッキングし、ビクターの行動データの分析を通じて、これらのKPIの改善に取り組んでいます。Web関連で生成されるデータ量は膨大で、1日あたり数億レコードにも及びます。2016年までは、各媒体のWebサービスごとに個別にデータ収集/分析を行っており、2017年に、統合基盤



左)株式会社集英社 デジタルソリューション部 プラットフォーム室
中島 洋介 氏(株式会社JOINT CREW所属)

中)株式会社集英社 デジタルソリューション部 プラットフォーム室室長
兼 マーケティング室室長 横田 孝俊 氏

右)株式会社集英社 情報システム部 情報マネジメント室係長
須藤 明洋 氏

として BigQuery ベースでデータウェアハウスを構築しました。しかし、全体的なアーキテクチャに課題が生じた点に加えて、エラー発生時のオペレーションなどにも課題があったため、次世代のデータ基盤を検討することになりました。

販売部や情報システム部門で既に利用していた Azure Synapse とデータブリックスの比較検討を行った結果、データブリックスの採用を決めました。決め手の1つはコストパフォーマンスです。先述のように、Webのシステムは購買データなどと比較して膨大であるため、Azure Synapse を採用した場合のコストの増大とパフォーマンスの劣化が大きな懸念となりました。また、マイクロソフトとデータブリックスが共同で実施した OpenHack に合計 2 回、計 6 名が参加し、直接データブリックスに触れることによって、使いやすいと高いパフォーマンスに感動を覚えました。結果として、BigQuery からデータブリックスへの移行を決定しました。

導入後、大幅なコスト削減に加えて 生産性の改善と 強力なデータガバナンスが実現

導入後の効果として、移行中の現在でも、期待どおりのコストパフォーマンスが実現しており、今後も大幅なコスト削減が見込まれます。全てノートブック上で編集できるため、開発者の全体的な開発体験が大幅に向上しています。また、Databricks アシスタントを

駆使することにより、3倍以上の生産性改善を実感しています。加えて、データブリックスの Unity Catalog の活用により、移行前の課題の1つであったデータ欠損が解決し、強力なデータガバナンスが実現しています。

今後は、AIや機械学習を活用したユースケースの拡大を予定しています。具体的には、Eコマースにおける購買予測やユーザーの状況

や環境にあわせた動的なコンテンツの提供です。また、基盤としては、既存の CDP システムとの部分統合や、Google Analytics から直接データブリックスへデータを繋ぐ計画も検討しています。また、昨今注目を集めている LLM（大規模言語モデル）の活用においては、各記事の特性／特徴に応じてエディターが手動でタグ付けを行っているプロセスの完全な自動化も検討しています。

