

平成 28 年（2016 年）11 月 10 日

## 「日本古典籍データセット」公開で国文研と協働 国内研究機関のオープンデータの取り組みを支援・推進

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所（NII、所長：喜連川 優、東京都千代田区）は、大学共同利用機関法人 人間文化研究機構 国文学研究資料館（国文研）<sup>(\*1)</sup> による「日本古典籍データセット」の公開に関する協働を進め、11月10日からデータセットの公開を始めました。NIIは、昨年11月に「国文研古典籍データセット」をNIIのデータセット共同利用研究開発センターの事業である情報学研究データリポジトリ（IDR）を通じて公開した際も協力<sup>(\*2)</sup>しており、今回の「日本古典籍データセット」公開は両機関の協働の第二弾となります。「日本古典籍データセット」は二次利用を歓迎するオープンデータ<sup>(\*3)</sup>として、国文研のサイトや情報・システム研究機構データサイエンス共同利用基盤施設の「人文学オープンデータ共同利用センター」準備室<sup>(\*4)</sup>のサイトから提供します。

NIIは国文研と協働し、「日本古典籍データセット」のオープンデータ化に取り組みました。古典籍画像データを一括ダウンロードするためのサイトを用意するだけでなく、相互運用性の国際基準であるIIIF（International Image Interoperability Framework）に準拠した画像データアクセス方法を採用して画像の内容をブラウザ上で簡単に確認できるようにするなど、利便性も向上しました。また、オープンサイエンスの重要な基盤であるDOI<sup>(\*5)</sup>への対応を想定した画像データの命名規則を導入するなど、オープンデータ化の次のステップである利活用を見据えた枠組みを構築しています。

今回公開するのは、昨年公開の350点を含む古典籍700点の画像データ（約16万コマ）と書誌データです。作品紹介や翻刻テキストデータ（原文をそのまま活字にしたもの）、タグ情報が付いているものもあります。詳細につきましては、別紙の国文研発行のプレスリリースをご参照下さい。「日本古典籍データセット」の公開サイトは以下の通りです。

- 国文学研究資料館  
[http://www.nijl.ac.jp/pages/cijproject/data\\_set\\_list.html](http://www.nijl.ac.jp/pages/cijproject/data_set_list.html)
- 「人文学オープンデータ共同利用センター」準備室  
<http://codh.rois.ac.jp/>

NIIは本データセットを活用した研究基盤の構築に国文研と協力して取り組んでいます。その一環として、昨年が続いて本データセットを活用した「歴史的典籍オープンデータワークショップ（アイデアソン）」を12月9日に国文研と共催（予定）します。

NIIは今後も組織の枠組みを超えて国内の研究機関のオープンデータ化を支援、推進し、日本の学術コミュニティにおけるオープンサイエンス推進の担い手として、データ公開の流れが加速するよう取り組んでいきます。

〈メディアの皆様からのお問い合わせ先〉

**大学共同利用機関法人 情報・システム研究機構 国立情報学研究所**

総務部企画課 広報チーム

TEL:03-4212-2164 E-mail : [media@nii.ac.jp](mailto:media@nii.ac.jp)

「日本古典籍データセット」について

**大学共同利用機関法人 人間文化研究機構 国文学研究資料館**

古典籍共同研究事業センター事務室 古典籍共同研究係

TEL:050-5533-2988 E-mail : [cjinfo@nijl.ac.jp](mailto:cjinfo@nijl.ac.jp)

以上

(\*1) 国文学研究資料館： 国内各地の日本文学とその関連資料を大規模に集積し、日本文学をはじめとする様々な分野の研究者の利用に供するとともに、それらに基づく先進的な共同研究を推進する日本文学の基盤的な総合研究機関。<http://www.nijl.ac.jp/>

(\*2) 「国文研古典籍データセット」を公開した際も協力： 詳細は、昨年11月4日発行のニュースリリース「国文学研究資料館の『古典籍データセット』公開に協力／NIIの『情報学研究データリポジトリ』を通じて提供」([http://www.nii.ac.jp/userimg/press\\_20151104.pdf](http://www.nii.ac.jp/userimg/press_20151104.pdf)) 参照。

(\*3) 二次利用を歓迎するオープンデータとして提供： 利用条件は、クリエイティブ・コモンズ・ライセンス表示-継承4.0 CC BY-SA (<https://creativecommons.org/licenses/by-sa/4.0/deed.ja>)。

(\*4) 「人文学オープンデータ共同利用センター」準備室： 平成28年4月1日に、情報・システム研究機構データサイエンス共同利用基盤施設にて活動をスタート。NIIコンテンツ科学研究系准教授の北本朝展が室長となり、NIIと統計数理研究所との共同研究、および国内外の人文学研究機関との連携を軸に、人文学オープンデータの共同利用という課題に取り組む。来年4月にセンター化予定。<http://codh.rois.ac.jp/>

(\*5) DOI： デジタルオブジェクト識別子 (Digital Object Identifier)。学術論文などの文献に一意的識別子を付与するとともに、URLや公開日、公開者などのメタデータを登録することで、デジタル文献への永続的なアクセスを保証するシステム。近年はDOIを研究データ等の学術資源一般にも活用する動きが盛んになっている。