

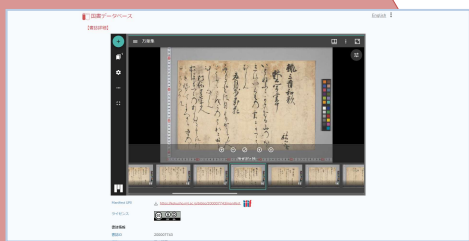


古典籍テキストデータを活用した データ駆動型人文学のための研究資源構築プロジェクト

国文学研究資料館 菊池信彦



国文学研究資料館が10年かけて作成した
国内外の古典籍（江戸以前の和古書）の
画像データベース「国書データベース」
（約30万タイトル）



OCR（光学式文字認識）技術等を
活用して、テキストデータを作成

NII RDCの支援領域

TEIに則った
構造化データを作成

古典籍テキストデータおよび
そのTEI/XMLデータを利用する
人文学（「典籍防災学」等の総合知）を支援



構造化データの
作成、維持・管理、共有に
NII RDCを活用

データ駆動型人文学