Towards the DATA-KBR-BE platform:
an iterative approach to publishing 'Collections as Data' at KBR, the Royal Library of Belgium

Sally Chambers[1]&[2], Thuy-An Pham[1], Julie Birkholz [1]&[2] and Frédéric Lemmers[1]
[1]KBR, Royal Library of Belgium
[2]Ghent Centre for Digital Humanities (GhentCDH), Ghent University

Keywords: Collections as Data; Digital Cultural Heritage; Digital Humanities Datasets; Data-level access; FAIR (Findable, Accessible, Interoperable, Reusable) data; Open Science Trusted Digital Repository

Abstract

Inspired by the 'Collections as Data' movement[1] (Padilla et al, 2019) as an approach for cultural heritage institutions to prepare their digital collections for analysis using digital methods, in 2020, KBR, the Royal Library of Belgium, embarked on a 48 month project[2] called DATA-KBR-BE (2020-2024). The aim of the DATA-KBR-BE project is to optimise KBR's ICT infrastructure to stimulate sustainable data-level access to KBR's digitised and born-digital collections for digital humanities research. Even though digital cultural heritage collections in libraries, archives, and museums are increasingly being used for digital humanities research, traditional ways of providing access to such collections, for example through digital library interfaces, are less than ideal for researchers who are looking to build datasets around specific research questions. The key outputs of the DATA-KBR-BE project will include: 1) the design of a sustainable data extraction workflow; 2) the design and implementation of an Open Data Platform (data.kbr.be); 3) an inventory of KBR's digital collections; 4) the publication of the open humanities datasets stemming from the interdisciplinary research scenarios and beyond, and following their publication, 5) a hackathon using the datasets.

In 2021, our presentation at DH Benelux (Chambers, et al, 2021) focused on the interdisciplinary experiments in the area of social history, literary studies and media history to be carried out in the context of DATA-KBR-BE. In 2022, our artificial intelligence and data science team, Ghent University's IDLab, presented their NewspAIper demonstrator for undertaking AI-Based Metadata Enrichment of KBR's Historical Newspaper Collection (Ali, et al, 2022). This year's contribution to DH Benelux focuses on a researcher-centred and iterative approach to gather requirements for the design and development of the DATA-KBR-BE Platform.

Towards the DATA-KBR-BE platform

---

[1] See, for example: 'Always Already Computational: Collections as Data' and 'Collections as Data: Part to Whole'.

[2] DATA-KBR-BE is financed by the Belgian Science Policy Office (Belspo) as part of the Belgian Research Action through Interdisciplinary Networks, BRAIN 2.0 programme. Originally foreseen as a 24-month project, in February 2022, the project was extended until 15th March 2024.

In addition to the [Collections as Data initiative](#), DATA-KBR-BE is inspired by data platforms at other cultural heritage institutions, such as the national libraries of [Luxembourg](#), [the Netherlands](#), [Scotland](#) and [The British Library](#). By the end of the project in Spring 2024, the DATA-KBR-BE team will have designed, implemented and tested the usability of the new data.kbr.be data platform. A number of thematic datasets will be curated and published on the data.kbr.be platform by the end of the project to support the interdisciplinary research scenarios that have been selected for the project: a) **Collective Action Belgium,** led by [GhentCDH](#), which focuses on social history in the Interbellum and World War Two period and aims to trace the dynamics of contention, strikes, demonstrations and other forms of collective action in Belgium as reported in Belgian newspapers; b) **the feuilleton in Belgium**, led by [ACDC](#), which focuses on literary studies in the period 1830–1930 and aims to map the publication of literature in Belgian newspapers across the first century of the Belgian nation state; and c) **the History of Belgian Journalism**, led by [ULB](#)-[KBR](#), which focuses on media history from 1886 until now and aims to trace the history of Belgian journalism through the lens of critical discourses about journalism in Belgian newspapers. The project is compliant with the **FAIR (Findable, Accessible, Interoperable and Reusable)** principles of research data management and will ensure that the resulting datasets are deposited in a Trusted Digital Repository.

Methodological approach

Work on conceptualising the future DATA-KBR-BE platform began at an early stage in the project, during an initial *[Brainstorming Workshop](#)*, held in November 2020. A first important step in this process is to review some of the existing (National) Library data platforms. Questions such as: *What data is offered? How? What format? What did people like, dislike about the platforms* were explored. The outcomes of this workshop were used as the basis for iteratively developing a checklist of needs.

To facilitate this requirements building process, the DATA-KBR-BE project team made use of the "Collections as Data Checklist", which is currently being developed by the International GLAM Labs Community. This emerging checklist, which was introduced during the [Towards implementing Collections as Data](#) in GLAM institutions webinar in October 2022, is intended to provide "an easy to apply method to encourage, especially small and medium-size cultural heritage institutions, to publish their digital collections as 'Collections as Data'". The DATA-KBR-BE project team was able to use the checklist as a framework to help structure the development of the functional and technical requirements for the DATA-KBR-BE platform. This work will be continued as part of the "[Paths to Participation - Collections as Data: State of the Field and Future Directions](#)" workshop, which will take place in Vancouver, Canada in April 2023.

This checklist of needs will be used as a basis for undertaking a series of semi-structured conversations with key colleagues within KBR, e.g. ICT department, digitisation department, collection managers and KBR's digital data strategy team, to further iterate the document. Once mature enough, it will be also shared with the researchers beyond the DATA-KBR-BE project, including the project's Scientific Advisory Board as well as the DH Benelux community for

further validation and feedback. After which, a series of wireframes to visualise the emerging DATA-KBR-BE platform will be designed for additional feedback, before a first version of the platform is developed. It is anticipated that the platform will both adopt KBR's existing visual identity and become part of the family of KBR services such as BelgicaPress and BelgicaPeriodicals.

Next Steps, Future Research

Once the technical and functional requirements have been signed off by KBR's Management Team and the DATA-KBR-BE Scientific Board, the work to implement the platform can begin. This will be undertaken in close liaison with the KBR's ICT Department. It is anticipated that by DH Benelux 2024, it will be possible to demonstrate a first version of the platform and potentially organise a hackathon at the conference using the published datasets.

Bibliography

Ali, D., et al. (2022) "NewspAIper : AI-Based Metadata Enrichment of Historical Newspaper Collections." *DH Benelux 2022 : RE-MIX : Creation and Alteration in DH, Abstracts*, 2022. http://hdl.handle.net/1854/LU-8755465

Chambers, S., et al. (2021) "Collections as Data : Interdisciplinary Experiments with KBR's Digitised Historical Newspapers : A Belgian Case Study." *DH Benelux 2021, Abstracts*, 2021. http://hdl.handle.net/1854/LU-8712404

KBR, Royal Library of Belgium (2020). DATA-KBR-BE: facilitating data-level access to KBR's digitised and born-digital collections for digital humanities research. https://www.kbr.be/en/projects/data-kbr-be/

Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E. & Varner, S. (2019). *Final Report : Always Already Computational: Collections as Data*. http://doi.org/10.5281/zenodo.3152935 & https://osf.io/mx6uk/wiki/home/