

# Beyond the Archival Turn: Bridging the Gap between Digital Archives and Critical Editions.

## Full DMP

### Admin details

**Project Name** Beyond the Archival Turn: Bridging the Gap between Digital Archives and Critical Editions. - Full DMP

**Principal Investigator / Researcher** Wout Dillen   
<https://orcid.org/0000-0002-7862-2403>

**Funder** Fonds voor Wetenschappelijk Onderzoek

**Institution** University of Antwerp

## 1. General Information

### Name applicant

Dirk Van Hulle

### FWO Project Number & Title

Beyond the Archival Turn: Bridging the Gap between Digital Archives and Critical Editions.

### Affiliation

- Universiteit Antwerpen

## 2. Data description

### Will you generate/collect new data and/or make use of existing data?

- Generate new data

The data described below is generated by project members, affiliated researchers, and affiliated holding libraries and archives.

### Describe the origin, type and format of the data (per dataset) and its (estimated) volume.

The data will consist of:

1. Scanned images of archival materials concerning the writing process of Samuel Beckett's Murphy (notes, manuscripts, typescripts, published versions, etc). in .tiff and .jpg formats. Expected total volume: ... (Vincent).
2. Manual transcriptions of the images in [1], in .xml format. Expected total volume: ... (Vincent).
3. Data files to transform the documents in [2], and present them in a web environment alongside the images in [1], in .xsl and various web technology data formats (.html, .css, .js, .json., etc.). Expected total volume: ... (Vincent)

### 3. Ethical and legal issues

**Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register.**

- No

Privacy Registry Reference:

Short description of the kind of personal data that will be used:

**Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s).**

- No

Reference to ethical committee approval:

**Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data, and which restrictions will be asserted?**

- Yes

Because the copyright to the materials on which the generated data is based belongs to the Beckett Estate (see below), we are legally unable to provide unrestricted access to them. In practice, (1) we are required to restrict access to the data by putting up a paywall that pays for the license; and (2) we are unable to grant our users the right to reuse our data outside of the negotiated publication environment (accessible through <https://www.beckettarchive.org>).

**Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?**

- Yes

The scanned images and transcriptions mentioned in the Data Description section are based on extant source materials produced by the Irish writer Samuel Beckett. Since Beckett only died in 1989, his works will remain in copyright, under the care of the Beckett Estate, for the foreseeable future.

Thanks to carefully negotiated third party agreements with the Beckett Estate and the holding libraries and archives that currently hold these documents, we are able to publish our data on the Internet (accessible through <https://www.beckettarchive.org>) – albeit behind a paywall to pay for the license. Participating institutions (and their members and visitors), however, are exempt from paying a license fee. Since the copyright to Beckett's works remains with the Beckett estate, we are also unable to grant our users the right to reuse our data outside of this publication environment.

#### **4. Documentation and metadata**

##### **What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?**

The website that houses our data (<https://www.beckettarchive.org>; hereafter designated as "the edition") provides extensive documentation on the methodology used to create the data (<https://www.beckettarchive.org/editorial.jsp>); on how to use the edition (<https://www.beckettarchive.org/manual.jsp> and <https://www.beckettarchive.org/librarymanual.jsp>); and on how to cite the edition (<https://www.beckettarchive.org/cite.jsp>). The edition also provides more technical documentation concerning the edition's metadata (<https://www.beckettarchive.org/technical.jsp>). Finally, the edition also points to a website that explains in more detail which subset of the TEI-XML standard was used to create the edition (<http://uahost.uantwerpen.be/bdmp/>), and that provides documentation about its schema specifications (<http://uahost.uantwerpen.be/bdmp/index.php/the-tags/4-validation/>).

##### **Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.**

- Yes

The project's metadata is encoded in XML, in accordance with the guidelines provided by the Text Encoding Initiative (TEI; <https://tei-c.org>). As considered best practice in the community, the project uses a subset of the TEI's complete tagset, and published the schema of this subset to further the data's computer readability (see <http://uahost.uantwerpen.be/bdmp/index.php/the-tags/4-validation/>).

#### **5. Data storage & backup during the FWO project**

##### **Where will the data be stored?**

'FLW Server:' a dedicated server owned by the UAntwerp's Faculty of Arts, located in the university's datacenter.

##### **How will the data be backed up?**

The data is backed up automatically on the servers of the university's datacenter.

**Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.**

- Yes

The FLW server currently has a capacity of 200GB, and can be expanded further in communication with the UAntwerp's datacenter.

**What are the expected costs for data storage and backup during the project? How will these costs be covered?**

The FLW server was bought by the Faculty of Arts and is available for use by the faculty's researchers. There are currently no further costs associated to its storage and backup.

**Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**

General data security is provided by the university's datacenter. A software developer hired by the faculty of arts acts as a gateway between the researchers and the FLW server's services (including direct access to the data where necessary).

**6. Data preservation after the end of the FWO project**  
**Which data will be retained for the expected 5-year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues...).**

All the data described in the Data Description section is part of the project's core data, and will remain available for at least 5 years after the end of the project.

**Where will these data be archived (= stored for the long term)?**

The data are archived on the FLW server, and there are ongoing discussions with the University Library team to ensure that the edition is catalogued and preserved for the long term.

**What are the expected costs for data preservation during these 5 years? How will the costs be covered?**

For the first five years after the project has ended, we currently expect no additional preservation costs.

**7. Data sharing and reuse**

**Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?**

- Yes. Specify:

As described in the section on Ethical and Legal Issues, copyright

restrictions prevent us making the data publically available, and for opening up the data for reuse beyond the scope of the edition itself.

**Which data will be made available after the end of the project?**

All data described in the Data Description section will be made available during (and after the end of) the project within the terms specified in the Ethical and Legal Issues section.

**Where/how will the data be made available for reuse?**

N/A

**When will the data be made available?**

- Upon publication of the research results

**Who will be able to access the data and under what conditions?**

The data is made available upon payment of a subscription fee to pay for the edition's license. Both individual and institutional licenses are available. Participating institutions also obtain a free institutional license to the data, ensuring their members and visitors to access the data through their institutional networks.

**What are the expected costs for data sharing? How will these costs be covered?**

N/A

## **8. Responsibilities**

**Who will be responsible for the data documentation & metadata?**

The PI bears the overall responsibility for data documentation and metadata.

**Who will be responsible for data storage & backup during the project?**

The PI bears the overall responsibility for data storage and backup during the project.

**Who will be responsible for ensuring data preservation and sharing?**

The PI bears the overall responsibility for ensuring data preservation and sharing.

**Who bears the end responsibility for updating & implementing this DMP?**

The PI bears the overall responsibility for updating & implementing this DMP.