C3 DMP

1. General Information	
Name applicant	**David Burn**
FWO Project Number & Title	Lost in Translation? Multilingualism in Early Song, 1350-1550
Affiliation	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other:
2. Data description	
Will you generate/collect new data and/or make use of existing data?	☒ Generate new data ☒ Reuse existing data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume *If you **reuse** existing data, specify the **source** of these data.* *Distinguish data **types** (the kind of content) from data **formats** (the technical format).*	**Goal of the DMP** The project aims to study and valorize fourteenth- to sixteenth-century songs (both monodic and polyphonic), their sources and repertoire. The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy with regard to the project (research) data. The DMP is a living document that will evolve during the project period. **Origin and types of the data** Data are collected by means of the Alamire Digital Lab: ● by state-of-the-art digitisation of manuscripts resulting in high-resolution images: Digital images of these sources will be collected by the state-of-the-art Alamire Digital Lab (ADL) using a 150-megapixel camera able to produce digital images of the highest quality, a scanback, lighting, book-stand, and all necessary ancillary equipment. ● by user-friendly and performative online databases: the images created by ADL are disseminated through the Integrated Database for Early Music (IDEM), a multi-platform online resource, and accompanied by descriptive manuscripts metadata. Over time this resource will grow to offer a rich environment for the images, including recordings, transcriptions, and further documentation. **Formats** ● High-resolution digital images of manuscripts are preserved in RAW, TIFF, JPEG2000 and PDF formats ● Transcriptions and translations of manuscript texts in RTF format **Estimated volume** ● High-resolution digital images: ADL - 1 image RAW (180MB), TIFF (300 MB), JPEG (180MB) Thumbnails (5,4MB) – total estimated volume: 5.000 pages (ca.25 manuscripts) = 3,50 TB

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. *In case your host institution does not (yet) have a privacy register, a reference is not yet required of course; please add the reference once the privacy register is in place in your host institution.*	☐ Yes ☒ No If yes: ● 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).	☐ Yes ☒ No If yes: ● 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 ☐ No If yes, please comment: IDEM: The Mirador-viewer is implemented in the IDEM database so that data is interoperable in function of efficient data-connection. The Mirador viewer is a IIIF compliant viewer application running on the LIBIS KU Leuven service infrastructure. Between the holding institutions and the Alamire Foundation (KU Leuven) an agreement shall be concluded authorising the ADL to digitise the necessary sources for further integration of these images in the Integrated Database for Early Music (IDEM) and to use them for related research and valorisation purposes. (Standard ADL agreement) Texts: a future integration of text files into the IDEM data model is planned, of which the necessary agreements (e.g. with translators and/or transcribers) will be part.
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 ☐ No If yes, please comment: ● The terms and conditions of IDEM (https://www.idemdatabase.org/terms) specify the constraints regarding use of the images: the generated images are stored and made accessible through the Mirador viewer of the Integrated Database for Early Music. There will be no limitation towards the consultation of the images without charge. However, download of the images is subject to the approval of the holding institution. Any constraints on the use of the data are defined in consultation with the holding institution of the original source. (e.g. copyright mention in case of publication) ● Texts: third-party agreements which restrict dissemination or exploitation of existing transcriptions and/or translations might exist.

4. Documentation and metadata	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	● High-resolution digital images - IDEM are accompanied by descriptive manuscripts metadata. Over time this resource will grow to offer a rich environment for the images, including recordings, transcriptions of text and music, translations, and further documentation. The database will connect the images, texts, and metadata to the sources and different data types (high-resolution digital images, transcription of texts, transcription of music, audio recordings, …).
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 ☐ No If yes, please specify: ● A customised data model was developed by LIBIS (KU Leuven), a division of Leuven Research and Development (LRD) and part of the University Library. The data model meets the research questions and the metadata needs recorded by the Alamire Foundation (KU Leuven). Special attention is given to the interrelations between the different data types and the integration of text files (transcriptions and translations) into the IDEM data model. The custom datamodel was fully documented and aligned (mapped) as far as possible with the most important standards in relation to the material, namely Dublin Core, SPECTRUM and LIDO. None of the standards could cover all the metadata needs of the research group. ● Digital Asset Management (DAM): Teneo - the system for digital preservation and Long-term digital preservation complies with the international standards for digital preservation (such as OAIS and Premis). Teneo takes care of the long-term preservation and the persistent access to complex objects by a persistent and resolvable PID. It safeguards the authenticity, unicity and integrity of data and fulfils all requirements that must be present in a digital preservation system, being: Ingest, Archival storage, Data management, Administration, Preservation planning, Access.

	● Ingesting the metadata into Rosetta: the mapping template for ingesting basic metadata related to the digital object into the long-term preservation (LTP) system uses the Dublin Core (DC) standard. Both DC elements and DC terms are available. ● The digital representations (RAW, TIFF and JPEG2000) stored in Rosetta (part of Teneo) are represented as IIIF manifest files. These manifest files describe the structure of the object and associated metadata according to the specifications of the International Image Interoperability Framework (IIIF), an open international standard for image presentation, sharing and visualisation. This allows the digital objects to be visualised easily and anywhere in the world in any IIIF compliant viewer application. Mirador was chosen as a viewer application in order to easily place images from different sources next to each other for comparative research. ● The open-source web application, Omeka, enables the public to discover the digitised manuscripts, by means of a persistent link to the Mirador viewer, and to consult all the associated data in the database. Omeka supports data export in Dubline Core via OAI-PMH and API.

5. Data storage & backup during the C3 project	
Where will the data be stored?	● High-resolution digital images - IDEM: The generated data are stored and made accessible through the Integrated Database for Early Music, a joint effort of the Consortium and LIBIS (KU Leuven Research and Development). IDEM includes a long-term preservation repository Teneo which makes use of the data centres of KU Leuven ICTS for storage and mirroring. It archives and provides persistent access to born-digital or digitised data objects and datasets that need to be kept for the long term, and includes metadata, usage policies and access tools. Using this service guarantees the safe storage of the data with back-ups in a redundant environment. IDEM has been established with Hercules grants (2009 & 2013) and is ensured by an FWO Large Scale Research Infrastructure grant. Other funding from the KU Leuven and/or Alamire Foundation (structural and project resources) provides the necessary guarantees for long term storage. Stability, functionality and development of IDEM is a top priority for KU Leuven / Alamire Foundation. The continuity of IDEM is embedded in their structural operation as well as in their projects, in order to guarantee sustainability beyond the current project period. ● Manual transcriptions and translations of early music lyrics: because of their small volume, RTF-files can be replicated easily. A master copy of the transcriptions and translations will be stored on IDEM servers, from which point additional copies will be made on computationally powerful machines. As the dataset is so small in footprint, no additional storage facilities are required for this data.

How will the data be backed up?	● IDEM: By using Teneo the storage and the back-up of the data is guaranteed. Moreover, the evolution of digital formats is monitored by a community of experts in order to take proper preservation actions when a format threatens to become obsolete. Its scalable and extensible architecture makes it a sustainable preservation infrastructure that can deal with the growing amount of digital data and wide range of formats. ● The security of the data is guaranteed by LIBIS using services provided by ICTS, KU Leuven's Department of ICT. ICTS maintains a continuous mirror back-up policy in 2 physically separated data centres. The transfer and accessibility of the data is ensured by recently installed glass-fibre connections. ● The data on the server in the Library of Voices is backed up at KU Leuven ICTS.
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 ☐ No If no, please specify:
What are the expected costs for data storage and backup during the project? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	● IDEM has been established with Hercules grants (2009 & 2013) and is ensured thanks to a FWO Large Scale Research Infrastructure grant. IDEM makes use of Teneo which guarantees long term preservation and access to data in a fully OAIS compatible ecosystem. The costs of access and long-term data preservation of IDEM data is 1700 euro per TB per year. For this project the generated data volume is estimated at 3,500 GB or 3,5 TB. A small additional cost is foreseen for the creation and programming of a new ingest model for a new data type. This brings the total cost for long-term data preservation and persistent data access to about 20.000 euros. This cost includes the investment cost for 3,5 TB of storage volume, the ingest cost and cost of metadata as well as the recurrent storage cost spread over three years. Costs will be covered by funding from funding from the KU Leuven and/or Alamire Foundation (structural and project resources).

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	● The security of the data is guaranteed by LIBIS and by using services provided by ICTS, KU Leuven's Department of ICT.

6. Data preservation after the end of the FWO project FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.	
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, ...).	● IDEM, including a long-term preservation repository Teneo, will be continued with funding from the KU Leuven and/or Alamire Foundation (structural and project resources).
Where will these data be archived (= stored for the long term)?	● LIBIS: repository Teneo ● Texts (transcriptions and translations): local server at Library of voices and ICTS mirror. In a second phase, for selected audio files, the IDEM repository TENEO will be used.
What are the expected costs for data preservation during these 5 years? How will the costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	● Cost/TB LIBIS (ca. € 1.700 euro/TB/year or ca. € 29.750 /5 years/3,5 TB), which will be covered with funding from the KU Leuven and/or Alamire Foundation (structural and project resources).

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 ☐ No If yes, please specify: ● IDEM: free consultation, download restrictions depending on the conditions and requirements of the holding institutions (cf. 3) With an explicit agreement of the holding institution, it is possible to facilitate the download of the images from the IDEM database thus guaranteeing an open access treatment for the cleared images. ● Transcriptions and translations: restricting factors regarding the sharing of this data might be defined in an agreement with a third party.
Which data will be made available after the end of the project?	● IDEM via website: www.idemdatabase.org
Where/how will the data be made available for reuse?	☒ In an Open Access repository ☐ In a restricted access repository ☐ Upon request by mail ☐ Other (specify):
When will the data be made available?	IDEM data will gradually be made available in the course of the project on www.idemdatabase.org.
Who will be able to access the data and under what conditions?	● Partners of the consortium will have access to those data elements necessary for them at the various stages of the research according to their specific assignment in the project. Transfer of this data (methods, formats, platforms, …) and conditions of use are to be discussed within each phase of the project. ● IDEM: open for public consultation via website: https://www.idemdatabase.org.

What are the expected costs for data sharing? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	● IDEM: there are no costs for accessing IDEM. Costs may occur for obtaining images from the holding institutions and costs for copyrights depending on the use.

8. Responsibilities	
Who will be responsible for the data documentation & metadata?	C3 consortium: KU Leuven (Research Unit Musicology, Alamire Foundation)
Who will be responsible for data storage & back up during the project?	LIBIS and the Alamire foundation, making use of the KU Leuven ICTS storage services
Who will be responsible for ensuring data preservation and sharing?	Alamire Foundation in collaboration with LIBIS and KU Leuven ICTS
Who bears the end responsibility for updating & implementing this DMP? *Default response: The PI bears the overall responsibility for updating & implementing this DMP*	KU Leuven - Musicology: David Burn