Sustaining Practice Assets for Research, Knowledge, Learning and Engagement (SPARKLE)

Final Report and Recommendations

Tom Jackson, Claire Knowles, Scott McLaughlin, Rachael Kotarski, Alex De Little, Matthew Warren, Andrew Horne

Funded by the Arts and Humanities Research Council

(Grant number AH/W007606/1)

Jackson, Tom, Knowles, Claire, McLaughlin, Scott, Kotarski, Rachael, De Little, Alex, Warren, Matthew, Horne, Andrew (2023) Sustaining Practice Assets for Research, Knowledge, Learning and Engagement (SPARKLE) Final Report and Recommendations, University of Leeds

1. Executive Summary

The goal of the SPARKLE project was to identify the foundations for a UK practice research infrastructure that will increase visibility, recognition, and support for world-leading practice research. Within the scoping study we build on the findings of the Bulley/Şahin reports for PRAG-UK on practice research (Research England 2021) in order to:

- Establish a greater understanding of practice research community needs for a data infrastructure.
- Build a blueprint for development of a national practice research infrastructure (SPARKLE) backed by the practice research community.
- Establish a deeper recognition of the skills gap in practice research, for both practice-researchers and repository managers, and a plan to overcome these gaps.
- Strengthen the practice research community.
- Highlight the importance of practice research beyond REF.

The development of a co-designed infrastructure will enable new levels of engagement with practice-based research.

2. Approach

SPARKLE brings together the University of Leeds, British Library, and EDINA, supported by our partners at the universities of Sheffield and York, and the Digital Preservation Coalition. We aimed to scope what it means to provide a heterogeneous and extensible infrastructure that befits the multi-faceted, multi-disciplinary, multi-format nature of practice research, reducing reliance on standardised asset-types. Our focus is on an open infrastructure that consists of three layers: the technical layer (with store, preserve, discover, and access components); the human-computer layer (with re/use and analysis capability); and the skills layer (learning resources and skills development for researchers, repository managers, library staff).

In shaping a bold vision for an innovative infrastructure for practice research, direction must be taken from the needs and current capabilities of researchers and practitioners, and the institutions that support their work. In collaboratively building infrastructure, we will nurture an engaged community that is invested in a toolset that brings greater credibility to their academic endeavour and supports their long-term skills and development. This civic approach included collaboration and sharing of insights between the PRVoices and SPARKLE projects, as we have moved towards a common goal.

Initial analysis identified 9 stakeholder categories (Appendix 1). 32 interviews were held by project post-docs and Co-Is to collect stakeholder views on the future of practice research infrastructure. As well as a broad response to the challenges, interviewees gave feedback on needs around representation of practice research, discoverability, transparency and storage (see Appendix 2 for the interview script).

Interview findings were validated in a face-to-face workshop, with stakeholders including some interviewees and a representative from PRVoices. These validated outcomes are worked into key findings and recommendations in this final report.

Finally, project partner EDINA worked with the outcomes to develop the technical plan and roadmap for a practice research infrastructure.

3. Activities

Title	Description	Date	No. of Participants
Launch Event	Hosted by the Jisc Digital Research Community and in collaboration with the PRVoices team	9.02.2022	58
Interviews	Stakeholder interviews conducted by PDRAs and Co-Is.	July–Aug 2022	32
PRVoices end-of-project community event	SPARKLE and PRVoices teams gave an update on progress. 27 attendees, with recording distributed to 40 people.	20.07.2022	27 (40)
SPARKLE workshop (University of Leeds)	Discussion with stakeholders (alongside PRVoices, and EDINA) of findings and plans.	6.09.2022	13
RMA (Royal Musical Association) conference panel	Panel + audience discussion around what a national repository might look like for music researchers.	10.09.2022	16
Embedding open research practice in the Arts and Humanities	Presentation at University of York, and discussion with library staff and humanities academics.	12.09.2022	28
Practice Research at RAM	Consultation discussion with researchers at the Royal Academy of Music.	10.10.2022	8
PGR training at RBC	Training needs session with practice research PGRs at Royal Birmingham Conservatoire	11.10.2022	6
Practice Research and Open Data (Sussex)	Presentation at Sussex Humanities Lab for Open Data Week; various UoAs.	24.10.2022	6 (20)

4. Summary of key findings

Top 3 findings

1. Use

Our research showed that practice researchers require digital features and functionality purposefully designed to accommodate the ways in which they conduct their research, and the kinds of data they produce. Practice researchers need a system that feels authentic and targeted towards helping them address the specific challenges they face on a day-to-day basis.

2. Presentation

Practice researchers require a system that allows them to present their data in a way that communicates their research journey and clearly illustrates how they have generated contributions to knowledge through their practice. They need functionality that is sympathetic to the complexity and entangled nature of their research.

3. Access and Reuse

More effective methods of searching for relevant projects and data would greatly benefit practice researchers. A system that stores data in future-proof formats and with appropriate metadata would aid access and reuse, generating greater impact.

These 3 top findings are further explicated below, drawing upon a range of specific recommendations for how they might be achieved, voiced during our interviews.

1. Use

Submissions

- The system should define a minimum level of what a submission needs to consist of, whilst acknowledging there is no 'correct' way of documenting and presenting practice research.
- Users must be able to submit developmental work as well as published outputs, as those final pieces of work are not always the closest to the knowledge.
- Users require functionality that allows them to meaningfully interlink all components of a submission, communicating their relationships.

Interoperability

- Multimodal research data must be represented by a broad range of media formats.
- Submitted files should, wherever possible, be machine readable, automatically tagged and transcribed in order to increase searchability.

Interface Characteristics

- The interface should offer quick and intuitive navigation between related research data regardless of formats or complexity.
- The ability to compare projects—through features such as split-screen—could aid analysis.
- The interface must offer high quality playback functionality for image, video and audio files, allowing them to be examined in detail.

Community

- The proposed infrastructure should be central to an online community through which users discover the work of other academics, establish connections with them etc.
- The submissions should be prepared and presented in ways that feel appropriate to all of the communities of practice that researchers commonly engage with. Practice researchers want to direct others, whether they are academics or not, to the proposed infrastructure, and not to a fragmented landscape of disparate commercial or self-hosted platforms.

Review

- Functionality for review by trusted parties could be implemented, allowing practice researchers to garner feedback before their submissions are made public.

Ethics

- Practice researchers should be offered guidance and training regarding ethical approval, including 'checklists' to ensure that appropriate clearances are in place prior to submission.

2. Presentation

REF

- The proposed infrastructure could be structured in a way that supports the composition of REF submissions, including framing statements.
- Functionality that allows users to illustrate their research journey will help REF assessors to better understand and appreciate practice research submissions.

IP and Ownership

- Practice researchers should have granular control over the extent to which their work can be accessed, downloaded and reused, through robust functionality regarding the licencing of content.
- Data with strict and limiting access controls— for instance due to privacy issues—will still need to be archived to enable access for those that do meet the conditions.
- Licencing control might include functionality for selectively releasing content on request.

Crediting Collaborative Research

- Practice research is often co-created with partner organisations, community groups and specialist practitioners. Submitters must be able to appropriately credit all of these contributions.

3. Access and Reuse

Accessibility

- Whilst all web services must comply with accessibility standards, the proposed infrastructure could include additional accessibility-related functionality specifically tailored towards interactions with audio-visual media.
- The types of metadata that are allocated to each component of a submission will also be of paramount importance. This should be a combination of metadata generated by the system and allocated by the practice researcher.

Search Functionality

- Search fields for academic discipline, locality, media format, research method(s) etc. will aid users in finding practice research of relevance to their own.
- The proposed infrastructure must be optimised for search engines, so that practice research is indexed, visible, and discoverable via other routes.

Longevity

- Practice research is commonly conducted at the cutting edge of creative and technical development, so the proposed infrastructure will need to be continuously iterated and developed in collaboration with the community, to support their evolving needs.
- The proposed infrastructure could recommend file formats, compression technologies etc. with the greatest chance of maintaining compatibility and that are suitable for migration or emulation.

Metadata

- Metadata and cataloguing functionality must allow practice researchers to:
    - formulate sets of networked research data
    - represent the chronological development of their practice, and key insights.
    - evidence authorship, citations and impact

The data upon which these key findings are based can be found in Appendix 4: Thematic Report from Interview Data.

5. Top 5 recommendations

The 5 recommendations have been considered from both the perspective of practice researchers, and the underpinning infrastructure (in *italics*).

1. The proposed infrastructure must, through the implementation of evidence-based features and functionality, become an indispensable companion to practice research that users want to engage with throughout the process. Rather than adding to the already long list of digital systems that academics only use because they are obligated to do so and creating another requisite for end-of-project reporting, the opportunity exists to create a co-designed system that supports practice researchers and improves their ability to conduct and disseminate research from start to finish.

The infrastructure should be delivered in a way that integrates existing tools where possible, while reducing reliance on archiving and preservation systems that sit outside of organisational management. The delivery must also include dedicated co-design with a varied group of stakeholders.

2. The proposed infrastructure must present practice research data in a way that is meaningful, rather than simply *hosting* practice research data. There are already systems in place for the storage of large datasets (such as university data repositories) but those systems fail to effectively represent either (1) the complex, interconnected, and often messy, ways in which practice research is carried out, or (2) complex networks across communities of practice. The data structures, functionality and user experience of this new system all need to be designed from the ground up in order to meaningfully communicate how practice makes contributions to knowledge.

The use of view and access technologies such as the International Image Interoperability Framework (IIIF), which supports streaming of time-based media, granular quoting, commentary, citation and extraction, will enable this.

3. The design and development of the proposed infrastructure must embody the principles and approaches of practice research. In order to arrive at their contributions to knowledge, practice researchers typically engage in iterative and cyclical phases of experimentation and play, the outcomes of which are critically evaluated and used to inform the future development of their practice. To ensure that the proposed infrastructure is as authentic, sincere and credible as possible, it should be designed and developed using similarly agile, responsive and consultative ways of working.

Valuable content already archived must not be forgotten, an approach that pulls together existing works from across repositories and enables new relationships and commentaries to be defined will embody this and will also encompass the recommendations and roadmap from PRVoices. This will be achieved through linking and mapping between metadata sources. Just as researcher-needs evolve, so will the sources that need to be drawn together, to form a coherent structure.

4. The features and functionality of the proposed infrastructure must facilitate and engender best practice approaches but without being overly prescriptive. Using the ways of working outlined above, we need to design and develop a system that aids practice researchers in implementing the most effective approaches to archiving their data, layering it with meaning and making it accessible for reuse. However, we cannot even suggest that there is one 'correct' way of conducting practice research. Doing so would be completely oppositional to the ethos of practice research and likely alienate our intended users.

This further underlines the need for flexibility in our technical approach over time, a modular component-based infrastructure, and agile approach to development.

5. The success of the proposed infrastructure is completely dependent upon appropriate provision for advocacy, training and support. Our research has revealed that key stakeholders have significant skill and knowledge gaps that need to be addressed in order to make this initiative a success. We cannot assume that our intended users will adopt the infrastructure, understand its significance or utilise its functionality appropriately without meeting their specific needs regarding training and skills development.

This needs to be incorporated throughout the project, through partnership with the community, and developing a train-the-trainer approach.

6. Roadmap

SPARKLE will take an open, user-centric and iterative approach towards an interoperable, trusted repository infrastructure for practice research. The key challenge will be developing and maintaining engagement with a diverse, heterogeneous group of researchers using a variety of data formats and extant embedded tools and paradigms.

SPARKLE will develop a powerful, feature-rich set of appropriate tools for researchers to both search existing archives (acknowledging initial community inertia in migration to adopting the new SPARKLE platform) and to upload, edit, annotate and submit their own archives for others to view and (re)use. The nature of the community means that the development of the tools will have to be a cooperative endeavour, building engagement via a continuous iterative feedback loop between the project team and the researchers as the tools are developed. Metadata enrichment, interoperability and discoverability will be the underpinning foci informing development practices.

An Agile co-design approach will therefore be effective in this context as there is a need for more fully specified requirements. SPARKLE's Agile approach will be an iterative process that puts conceptual interface designs and prototype functionality in front of users in order to refine the user experience.

The Agile core team will comprise a mixed set of skills comprising software development, user interface design, audio-visual expertise, and user engagement (full details in the appendix). The team mixture will change over time but is expected to comprise 3-6 individuals over the course of three years.

The focus of the team will be to

- Provide an intuitive and powerful interface to the repository
- Provide a stable, highly accessible and searchable repository on non-commercial hardware
- Develop and build a community of researchers using the repository
- Work to migrate large archives of practice research material onto the repository
- Develop innovative techniques to annotate and index the material in the repository

To support the team academic research staff in the fields of Informatics and Artificial Intelligence (AI) will be employed to use developing technologies to innovate in the field of recognizing and annotating video and audio practice research materials.

This roadmap is expanded upon in Appendix 3: SPARKLE Technical Delivery.

7. Appendices

Appendix 1: Stakeholder mapping

Practice research academics

Research support staff

Practice research PGRs

REF admin staff

Non-academic practice researchers

Research data repository experts

Practice research organisations

Developers/IT

Digital humanities expert

Appendix 2: Interview script

Warm-up questions

- Are you comfortable with the term 'practice research' and what it means? Would an explanation, or clarification, be helpful?
    - If yes: practice research - an umbrella term that describes all manners of research where practice is the significant method of research conveyed in a research output. This includes numerous discipline-specific formulations of practice research, which have distinct and unique balances of practice, research narrative and complementary methods within their projects. [The term is non-capitalised in general usage, in common with other research fields.] - Bulley/Şahin report p. 1
- Are you comfortable with the terms 'data' and 'repository' and what they mean? Would an explanation, or clarification, be helpful?
    - If yes: data is a broad term that essentially means any information that can be stored, so it can be text, images, video, audio, and more, and could be data about practice, data that represents practice outputs (documentation etc.), data that explicates (research narratives, annotations, contexts, 300wd statements), and metadata.
        - A data repository is also known as a data library or data archive. This is a general term to refer to a data set isolated to be mined for data reporting and analysis. The data repository is a large database infrastructure — several databases — that collect, manage, and store data sets for data analysis, sharing and reporting. - https://digitalguardian.com/blog/what-data-repository
        - Metadata is (usually) text that describes and gives information about other data, often for search and storage processes. E.g., author name, file type, upload date etc.
- Tell me a little about your work, and how it relates to practice research.
    - This question situates the interviewee and influences later questions.
- What data repositories do you currently engage with in your work?

- Researchers: How do other researchers access your work? Not just in professional artistic setting (gallery, concert, CD, etc.) but as a way of reaching other researchers. Is there a difference between how you engage with institutional repositories (e.g., for REF) and less formally?
- Library/REF-admin: how are outputs stored and disseminated? Are there multiple different systems that may be formal (e.g. institutional repository) and informal (e.g., personal/community website)?

Open questions

- In relation to practice research, what issues, if any, do you currently face with the data repositories you are required to engage with?
  - Do you have any suggestions for how those issues might be overcome? How might the repositories be improved in order to meet your needs of practice researchers?
- What characteristics, features or functionality would the ideal data repository for practice research possess?
- If you could design the dream data repository for practice research, what would it look like?

Representation

For the researcher, this may include the methods by which their practice research is represented for dissemination and storage; i.e., representation, documentation, etc. For those in the library this may include constraints of the systems they work with (e.g., storage/dissemination, submitting to REF, etc.) in terms of data representation, metadata expectations, file types, or on a technical level even searchability (e.g., can the system search within this file type? Yes if it's a PDF, probably not if it's audio/video? and is that an issue?)

- How successfully do you think current data repositories represent practice research, and the contributions to knowledge that are made through practice?
- How might data repositories be designed to allow practice researchers to communicate their research in more meaningful ways?
- What do other academics need to be able to access and experience the practice research when it is disseminated?
- What might be the most effective ways of representing any non-textual outputs from research?
- How flexible, or rigidly structured, do you think the ideal data repository for practice research would be?
- Would you like to describe, or even draw, the ideal interface for representing how contributions to knowledge are made through practice?

Discoverability

This is the ways that research can be found by other researchers or the public. This may include the ways that a researcher "packages-up" their research to make it easier to find and/or navigate; e.g., documentation, annotation etc. This may also include systems/repository design that both make discoverability easier, and also best capture the intent of the researcher (especially when the research needs to be aimed at a narrow set of stakeholders.

- How do people discover or find practice research on the data repositories you engage with?

- What features or functionality might allow practice research to be found more successfully? What would the ideal search tools look like?
    - Prompt: this might include search tools, but also back-end analysis tools that allow access via a wider variety of search tools.

Transparency

- What problems does it create for practice researchers if data repositories fail to represent their practice effectively and transparently, as well as their writing? What impact might it have upon the ways in which their work is interpreted by other academics?
- Should a data repository for practice research only disseminate the finished practice, or do other materials related to the research process and development need to be included as well?
- Should all the materials related to practice research be open and accessible through data repositories, or do some elements need to remain private for any reason?
    - If yes: Should the data repository still accommodate those private materials? And if so, how?

Storage

- How do you think practice research should be stored and catalogued?
- What structures should be in place to effectively organise and collate practice research?
- Do you have any recommendations regarding file formats, metadata, data structures etc.?

Appendix 3: SPARKLE Technical Delivery

Andrew Horne

The Agile Team

To deliver the SPARKLE project a team will be required that is self-governing and fully capable to deliver functionality without constant external input, though those external dependencies might exist. Not all roles are expected to be full-time. The key roles will be as follows:

Product Owner

The Product Owner will be responsible for representing the practice researchers' interests, and report to the Sparkle stakeholders. They will identify the key user stories that the team will deliver, and prioritize the functionality that will most benefit the user community. The Product Owner would be employed by the SPARKLE team directly.

Project Manager

The Project Manager will run the software team and negotiate any challenges and obstacles associated with the hardware and software. They will represent to the Product Owner the art of-the-possible in terms of technical solutions. The project manager and the software engineers below would be employed by Edina.

User Experience (UX) Designer

The UX designer will deliver mock-ups, screen designs and sample user interactions that will both define how a user story will be achieved through interaction with the application and give users examples of interactions that will allow them to more clearly define their wants.

User Interface Developer

This developer will work closely with the UX designer to deliver on the UX design, and to develop key libraries of functionality that will support the many and varied scenarios that SPARKLE presents.

Software Infrastructure Developer

This developer will define metadata and structure for the underlying archive, ensuring it is searchable efficiently and applying technical solutions to generate metadata where it is not available (e.g. speech to text and other solutions)

Hardware Infrastructure Developer

This developer will ensure the archive is available on hardware that is appropriate with the necessary functionality for the Software Infrastructure Developer to sort and structure it. The hardware engineer would be employed by EPCC and provide access and resources to the EIDF (see appendix).

Audio-Visual Engineer

This engineer will provide technical expertise to ensure that the formats of the objects stored in the repository are compatible formats with a minimum loss of quality.

Hardware Storage

Edinburgh International Data Facility

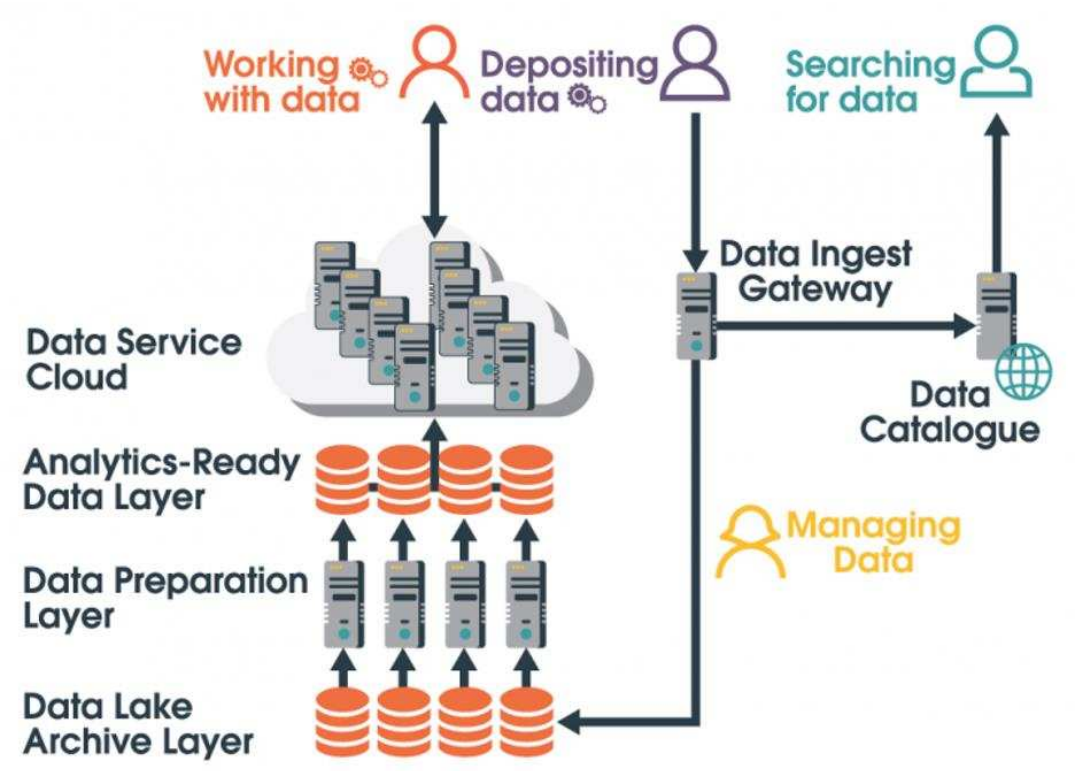
It is proposed that SPARKLE use the Edinburgh International Data Facility (EIDF) as their hosting solution for the archive of practice research.

The EIDF is a collection of computational, data management and safe-haven services supported by the Data Driven Innovation Programme of the Edinburgh and South-East Scotland City Region Deal.

The EIDF offers the scale and level of compute needed by the SPARKLE programme, with the benefit of being developed by EPCC, part of the University of Edinburgh, which is the UK's leading centre of Supercomputing and Data Science expertise.

What is the Edinburgh International Data Facility?

Computing services



Most users of the EIDF work in the Data Science Cloud, which offers a rich set of data science and analytics tools: from browser-based notebooks to full desktop environments. The Data Science Cloud is also a gateway to more powerful EIDF services like the Ultra2 large-memory system and the Cerebras CS1, and to EPCC's more powerful high-performance computing (HPC) systems like ARCHER2, the UK National HPC Service.

Data management services

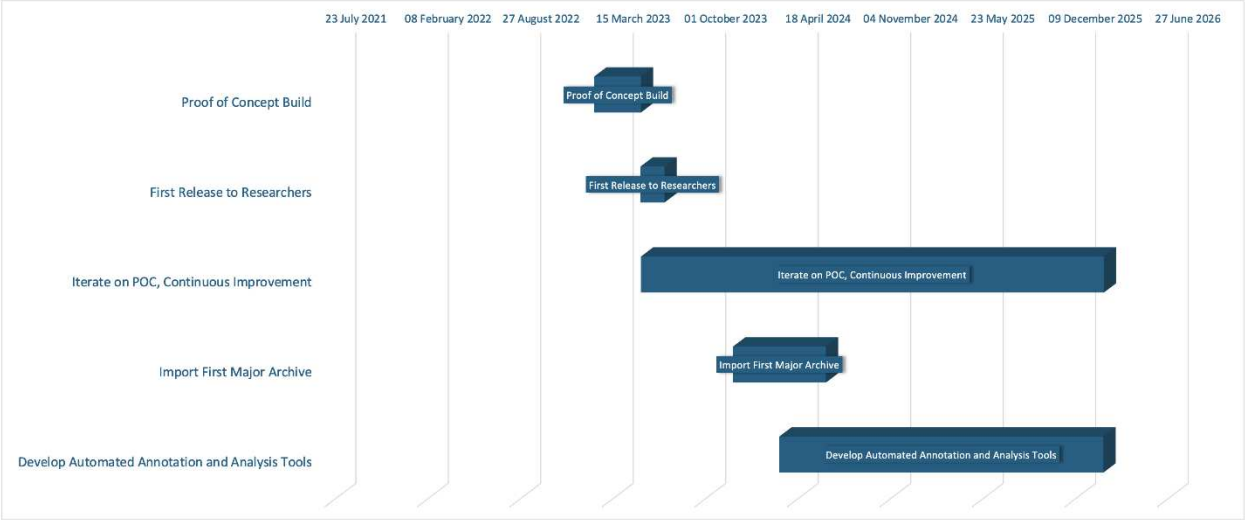
The Data Science Cloud sits on top of an Analytics-Ready Data Layer (ARD Layer), where EIDF data can be shared and re-used for science and innovation. This ARD Layer will grow over time as more and more data are collected in the EIDF. Innovators and researchers

looking for data can search and browse through the Data Catalogue to discover just what analytics-ready data EIDF has, and how they can get access.

EIDF data managers work with data depositors at the Data Ingest Gateway, ensuring that incoming data are safely stored in the Data Lake Archive Layer, and well-described in the Data Catalogue. Data in the Data Lake are stored for the long-term using best practices in digital preservation.

EIDF data wranglers work in the Data Preparation Layer, often in collaboration with data depositors and others, to turn archived data from the Data Lake into analytics-ready data products in the ARD Layer. They are then ready for data innovators to create new, exciting datasets that can be stored and shared all over again.

Indicative Roadmap of activities, if starting Spring 2023



Appendix 4: Thematic Report from Interview Data

Sustaining Practice Assets for Research, Knowledge, Learning and Engagement (SPARKLE)

University of Leeds

Summary of the findings of the Thematic Report on Stakeholder Interviews

Alex De Little & Matthew Warren

July 2022

Notes on the Thematic Report

Report Structure

This report is in two halves. The first focusses on Interviews with Practice Researchers and practitioners (end users). The second draws on interviews with those working primarily with repositories and research management (though some also have a specific interest in practice research).

Anonymisation

All interviewees (except by explicit request) have been anonymised and in the Thematic Report are replaced by a code that gives a little context for the interview. The naming convention is:

Interview number (UID) _ Interviewee Category (see below) _ Interviewer (see below)
e.g. 1_RDR_MW

Interviewee Categories

Interviewees were identified as belonging to one or more of several stakeholder categories:

- Research Data Repository
- Non-university Data Repository
- Digital Humanities
- Development/IT
- REF Administration
- Practice Research Academia
- Practitioner
- Postgraduate Practice Research

Interviewers

Interviews were carried out by three interviewers:

ADL = Alex De Little

MW = Matthew Warren

SMcL = Scott McLaughlin

Access requests

Access to the Thematic Report and anonymised versions of the underpinning interview, recording and document materials, needs to be requested and arranged with the Co-Investigator, Scott McLaughlin, at the University of Leeds.

Abbreviations

PR = Practice Research(er)

UI – User Interface

REF = Research Excellence Framework 2021

Part 1: Practitioners and Practice Researchers

Alex De Little

Features and functionality of a repository for practice research

Open vs prescriptive

- This is a complex space of competing and sometimes incompatible imperatives, but the repository needs to have functionality that carefully balances these needs.

Submissions

- There should be an agreed minimum level of what a submission should consist of, and it should allow for multimodal formats: though caveats may be need to made to limit file types to the most accessible or universal forms (see 'File Formats' below). The repository needs to support both private and public access to data as appropriate to the ethics of the original capture of the data.

File Formats

- The repository must support a broad range of media, should favour the most accessible and universal forms, and should be machine-readable. Where a submission can only be made in a format that is proprietary or otherwise difficult to access, or has poor reverse- compatibility, efforts should be made

to provide a copy in a more accessible format, with whatever compromises this entails,

Accessibility

- The repository needs to take accessibility issues into account from the start. There will need to be consulation about tensions between accessibility and the need to include a broad range of formats (with varying levels of accessibility).

Search function

- Search is complex in practice so nuance and context matters possibly more than usual when dealing with multimodal submissions. Getting the search function right is a key priority to align with goals of flexibility and sharing.

Categorisation, Cataloguing and Metadata

- Categorisation is complex and contested across different disciplines and stakeholder groups (e.g. professional/commercial/academic/etc). Even seemingly straightforward things like ‘media’ reveal problematic complexities when the categories are too simplistic. The metadata has to be really clear and in-depth: this is probably an area that will need most training for users.

Networking and relationality

- Relationality is key to practice, and the repository needs to have the tools to capture that; within a researchers’ practice, across multiple researchers, across multiple fields of unrelated practice etc. It is vital for each submission to capture, and make searchable, the range of potential relations and connections.

Communication of work and audience; community

- How can the repository represent and communicate to different stakeholders both within and outwith academia.

Discoverability

- Discoverability is key, and public searchability is important to reach the many non-academic stakeholders.

IP and ownership of work

- While some submissions will have standard publisher-controlled IP, many artist-researchers are self-published or exist in a model where there is no equivalent of ‘publisher’, so IP is often in the hands of the author, which may be a different dynamic to typical research landscape. Also,

embedding/hotlinking should be possible and public (within ethical constraints, see above).

Peer Review

- Peer-review of practice research is an open question at the moment. While it should not be central to the repository function at this moment, it is important to ensure that the repository could (at least) support peer review access in future.

Crediting collaborative research

- Collaborators should be included as part of the basic functions and needs a flexible system to allow different forms of collaboration. Ethics, Diversity and Inclusion
- The repository must avoid repeating historical problems, and centre chosen identities and ability to search by identities/affiliations/community. The repository should include ethical clearance information.

Issues with existing repositories

- A litany of issues with existing repositories' approach to practice.

Defining what the repository should store and who should be included

- This is a 'research repository' not a 'practice repository'. Practice is the mode of research**.**

The politics of a repository for practice research

- The repository needs to avoid repeating historical mistakes around colonisation of knowledge. Institutional relationships: practice researchers are often part-time and may work for several different institutions concurrently, so the repository should centre the researcher over the institution and ensure portability of outputs. The repository and the existing Practice Research community
- The repository works for the community of knowledge.

Open access

- The repository needs to be open access.

Public vs Private research artefacts

- Researchers decide what they should include in the repository. Ethics must be central.

Ensuring take-up of the repository / Usership among practice researchers

- The repository needs to be easy to use and be attractive to researchers. Incentivise researchers, communities of researchers, and institutions, and ultimately the public.

Tensions between the practice and its documentation and representation

- Similar to several other points below, the solution for this is more in the hands of researchers than the repository, but we welcome any ways the repository can lean into accommodating this need. Development of the repository should be done hand-in-hand with the developing training resources for end-users.

Relationship between physical/embodied and digital

- So much practice research is wedded to physical objects and experiences in real-space, and while researchers need to find ways to capture the research aspect of that, the repository needs to accommodate the many ways that researchers may capture this.

Challenges with preserving digital practice research

- Repository needs to be able to capture context in a meaningful way. While it's up to the researchers to do all they can to make context clear, the tools need to be there to make the connections to other research within and outwith the repository.

Longevity, financial support, and maintenance

- A key driver of this for researchers is that their work will have longevity, especially the captures of practice (audio/video/image/text).

Obsolescence

- It is difficult to keep pace with technological developments (especially competing with corporations), and because artists tend to work on cutting edge of technology. Ethics and accessibility should drive the decision-making about what individual submissions need to include to minimise obselescence,

The challenge of reproducing the experience or affect of the practice research

- For many practice research disciplines, the experience/presence of the work is vital, and while it's impossible to replace this, the repository must be flexible enough to include any/all representations and captures of this experience. While the repository should support (and make searchable) a broad range of formats, it will be up to the researcher to capture the shareable knowledge in the most fitting ways.

Tensions of a centralised platform

- We need flexibility to ensure that researchers can maintain the uniqueness of their work. Centralisation can be a form of neutrality that erodes distinctiveness.

Part 2: Repositories and Research Managers, Developers, and Administrators

Matthew Warren

Repository Content

The dynamic nature of the outputs and the careers of researchers needs to be represented in the metadata, and context is key.

Summary of findings

- **Metadata**—There is a need to be able to represent practice research in a flexible way in metadata and cataloguing. The main point to be raised were:
    - **Collections/projects** — There is a desire to be able to form groups of entries into collections that represent projects or sets of networked practice research.
    - **Dating** — Related to the idea of collections is the idea of dating. The ability to reflect the continually developing lives of practice research projects is needed. This also needs to include the ability to see not just the most recent version but also to trace its development. There was the suggestion that accessible archiving of previous versions should be a priority. The 'Last Modified' date was also suggested as an important piece of metadata
    - **Citations and impact -** There is a need to establish a way of recording and quantifying impact, citations are not an appropriate measure (re-performances, adaptations, etc.)
    - **Authorship** — There is a need for more flexible categories of authorship, acknowledging the diverse and highly collaborative roles at play in much practice research. *N.B. it may be necessary to develop a taxonomy of credit.*
    - **Institutional affiliation** — There are views both that clear presentation of institutional affiliation is important to institutional and individual participation, and that a perception of a neutral, non-institutionally affiliated space may also be important for other individuals' participation. In terms of metadata, the key things to bear in mind might be to do with how researchers moving from one institution to another might be recorded in ongoing collections/projects.
- **Standards —** There was a view that the development/adoption of clear standards against which the quality of data submissions (including and especially the recording of metadata) could be clearly and transparently judged is important. The need here is to improve the quality of data held in the repository. N.B Quality of data is not for the repository to judge, we can *only* have some comment on useful metadata.
- **Multimedia formats —** There is a need for multimodal materials to be stored, presented, and contextualised in a clear and easily navigable way. Open questions remain over physical objects. The importance of text was still frequently emphasised as helpful for discoverability, use, and highlighting the new knowledge embedded in the practice.

- **Documenting Context —** It has been highlighted that the context of practice research is important but challenging to capture and present, yet efforts need to be made to capture and present context given how it is also central to understanding practice research.

Additional Resources

Report on IIIF - https://www.nationalcollection.org.uk/sites/default/files//2021-02/Practical%20Applications%20of%20IIIF.pdf

Technical and Organisational Infrastructure

UI, interoperability and preservation are core needs but can be infrastructure agnostic. BUT it's not clear what the specific needs are for preservation (e.g. does it go as far as emulation etc? It may do, given earlier requirements on context)

Summary

- **Environmental Costs** — Concern was raised in one interview over the environmental impact of massive data storage and the importance of balancing this against potential benefits was brought up.
- **Hosting —** The need for the hosting of the data is for powerful, reliable, inexpensive, and (as far as possible) perpetual hosting. Microsoft, Amazon, and Google were all recommended, with a cautionary note there must be an exit strategy in case this hosting becomes unavailable, and all data needs to be moved.
- **Ethics and legality —** it was pointed out that a clear legal basis for making data available is necessary. This will likely be a key component of both the development and the informed use of the repository.
- **Interoperability —** The need was raised for flexible interoperability
    - for practice researchers to build their own platforms
    - for different institutional repositories to interact with the data held with SPARKLE
- **Presentation —** There is a need for SPARKLE to present work in a way that encourages practice researchers to deposit in it and others to access it. It was suggested that it would need to look like somewhere that hosts art. It was clear from many interviews that the UI - the front end - of the repository was hugely important to its usefulness in terms of practice researchers depositing work and of users accessing it. (See also 'Participation')
- **Staffing and Curation** — A key decision will need to be made on the degree of staffing required to maintain SPARKLE. Key questions are whether to peer review submissions and whether the repository should be aggregated from existing repositories or be the primary place to store practice research data, which then can feed institutional repositories. The general feeling across the spectrum was that some element of curation and/or peer review would be needed to establish the repository as a useful tool as well as to help advocate for recognition of the research element of practice by practitioner-researchers and the wider world [the project investigators note that this suggestion might be conflating the repository system with dissemination systems (e.g. journals) that sit on top of the repository and which would be better situated to carry out curational activities]

    - It was also noted that the repository being managed at a national level would facilitate support that couldn't be offered at an institutional level.
- **Predecessors —** There was the repeated suggestion that SPARKLE should learn from similar existing and expired projects, for instance the University of the Arts London repository and the Performing Arts Data Service.
- **Preservation** — The importance of considering preservation from the outset of both the repository and practice research projects was consistently underlined. Key questions for researchers (see also 'Training Needs') included how and when to collect material for preservation. The key repository questions around this are: what formats to accept (e.g. non-proprietary); how a sustainable, diversified funding model might be achieved; and what the data exit strategy might be (both from any external data storage that will at some point become unavailable; and from the organisation itself, should SPARKLE have to shut down. An example of this often cited was the closing of the Arts and Humanities Data Service (including the Performing Arts Data Service and the Visual Arts Data Service (now managed by the University for the Creative Arts).
    - A key consideration here was being able to make the use case for the value of preserving practice research.

Additional Resources

Visual Arts Data Service - https://vads.ac.uk/

Kultur Project (archived) - https://web-archive.southampton.ac.uk/kultur.eprints.org/

Towards a National Collection - https://www.nationalcollection.org.uk/

Beagrie and Houghton report - https://repository.jisc.ac.uk/5382/1/BADCReport_Final.pdf

Participation

UI, embedding in workflow and integration across institutions are key to fostering use/adoption.

Summary

- **Researcher Participation —** SPARKLE will need researchers to use it. The main issue here is the time taken to engage with the repository: practice researchers will be reluctant to engage if they have to find the time to manually deposit their work in an additional place or navigate complex user interfaces. Another concern was whether it was beneficial or not to have the repository space affiliated to the researcher's institution. There were views on both sides of this, recognising both the benefits of a neutral space and of potential bad feeling around practice research having to go outside the institution to deposit material.
- **Institutional Participation —** SPARKLE will need institutions to use it. A key issue here is inter-institutional competition, where universities may be less inclined to buy into a repository that also supports their REF competitors.
    - It shouldn't be assumed that either practice researchers or institutions have the time and motivation to build their own front end onto the repository.
- **User Participation** — The ambition was raised to embed the repository in the practice research community of researchers and users as the place to come together to share, find, and discuss practice research. This should include supporting

independent practice researchers as well as academics, again raising challenges to the idea of institutionally based repositories.

Training needs

Training needs to emphasise the ongoing management needs of creative process, in addition to the capacity of the tools to do this. Details on the requirements for preservation are lacking.

Summary

There are significant training needs, some of which may be integrated into SPARKLE itself. It was noted that the form of the training should be exemplars and case studies rather than potentially restrictive guidelines.

- **Data Management —** In particular, a need for data management to be seen as integral throughout a research project. A particular concern here was with an understanding of data preservation and digital curation.
- **Practice Research —** The need for researchers to have a thorough understanding of what constitutes PR and PR data.
    - Other issues brought up included open access, intellectual property, ethics, and the importance and facilitation of access to work.
- **Preservation —** There was a clearly highlighted need for data to be well preserved, but what does this mean to different stakeholders?
    - **Organisational Longevity —** There is a need to secure consistent funding, requiring the case to be made for the value of data. Diverse funding sources were also recommended to increase the stability of the support for the preservation of data. It was also recommended to have good contingency and succession planning.
    - **Points of Intervention —** The need to preserve not just the end product. To have intervention points throughout a project, it was suggested that data management plans should include an informed consideration of why and how interventions would be made to preserve the development of the research.
    - **Formats for Preservation —** The need to use formats that are likely to have longevity and be suitable for migration or emulation.
    - **Migration and Emulation —** Migration risks data loss; emulation is more technically complex. Linked to Q above: If you want more context, you need emulation, not migration.
- **Limitations of Training** - It was pointed out that in areas such as digital curation, documentation, and preservation, training has limits and not everyone can become sufficiently skilled in all aspects of research documentation.

Additional Resources

JISC Digital

https://www.jisc.ac.uk/guides/using-digital-media-in-new-learning-models/digital-images

https://www.jisc.ac.uk/guides

https://digitisation.jiscinvolve.org/wp/spotlight-on-the-digital/

https://www.slideshare.net/JISCDigi

Appendix 5: Use Cases and User Stories

Use cases

As a practice researcher	
I need to	share my ideas so that people using similar ideas can find my work
I need to	share my art (score/pdf, audio, video, other) so that my work is discoverable by other researchers
I need to	connect single outputs so that others can see related threads of my overall research narrative
I need to	present a clear timeline of my research
I need to	find others doing similar things to me, both their ideas and their resources (violin, orchestra, etc)
I would like to	upload examples and work in progress so that I can share important research milestones separate from the final output (because the final artwork may obscure some aspects of research that are more visible in process.)
I would like to	share my work with the public
I need to	stream audio and video of work I find by other researchers
I might want to	be notified when another researcher uploads new work.
I might want to	save links or citations to others' work, alongside comments (visible to me only)
I might want to	limit access permissions for certain data (for ethical or security reasons) to specific users or user groups
As a student	
I need to	search for research relating to my own practice or study topic so that I can better understand the field of current research to which I am contributing.
I need to	stream audio and video of other work I find
I need to	save citations and links to outputs
I might want to	contact the author of resources I find to ask for more information
As a REF manager / UoA lead	
I need to	see my researchers' current outputs to assess what might be submitted for REF
I need to	see the relationship between a researcher's different work to understand how a narrative may be written about them, or to tie several researchers into a single narrative
I need to	explore how my researchers' work may relate to trends or outputs from researchers in other institutions.
I need to	submit outputs and research narratives (e.g. 300wd statements) to REF in a locked format so they cannot be altered post-submission
I need to	have access to metrics in easily usable form (e.g. visualisation and .csv export) across both individual researchers and groups that I define.
As a REF/AHRC panellist	

I need to	access shared outputs for assessment
I might want to	share access with other panellists, panel chairs, or admin
As a Librarian - administrator (there may be a difference between institutional admin and repository admin, if it's independent)	
I need to	Take down resources that are not compliant with our policies
I need to	Create collections of resources
I need to	Showcase individual resources or collections of resources
I need to	Have super-access to other resources or links shared by others in order to add or remove content or metadata
I need to	Create / organise pages with training / information for users with embeddable videos
I need to	have access to metrics on site and resource usage
I need to	Suspend temporarily sharing rights of one or more accounts
I need to	provide users with clear understanding of licencing terms
I need to	manage the licence options available to sharers
I need to	have all the same privileges as academic staff for uploading / sharing / playlisting on behalf of users, including services.
I need to	manage which users can upload content that is publicly accessible
I need to	approve/edit/refuse resources from users who cannot make resources publicly available
I would like to	extract file types and file information on file upload into metadata
I would like to	see an audit trail of metadata and file activity against users and dates
I would like to	include content from other repositories e.g. theses and data repositories
I would like to	limit the amount of metadata sharers need to complete and enable autofilling of metadata where possible
I would like to	provide users with clear understanding of policy terms of use that can be read and confirmed in an accessible way without deviating the user from the process of sharing
I would like to	create on request "collective authors"
As a collaborator / As a partner organisation	
I need to	Have the same user rights as staff assigned to me temporarily for the sharing of files or links related to a project connected to a collaboration with the University and my status as an external partner visibly recognisable

User stories

Practice Researcher 1

Kris is a practice researcher who works with augmented reality. Their practice involves the development of experimental AR experiences that users access via mobile devices and dedicated AR headsets. In order to develop this work, Kris produces digital data, such as 3D models, software files, sound recordings and computer code, but also developmental data such as planning materials, scenographic layouts, sketches etc. Although Kris' practice is the AR experience, their contributions to knowledge are actually embodied within the ways in which users respond to the experiences. They therefore produce a lot of data that is documentation. They screen-record the AR as it runs and video record the user encounters

with it. This documentation is then used as the basis of follow-up interviews with the users, which are also video recorded and then transcribed.

Like a great number of practice researchers, Kris develops their practice using a cyclical and iterative approach. They have devised three main 'phases' of their practice, each of which is a different AR experience that investigates different kinds of user interaction. Each 'phase' is also developed through iterative cycles of experimentation and play. The outcomes of each cycle are thoughtfully analysed and the findings are used to inform the subsequent cycles. Kris embraces the unpredictable nature of their practice and considers it an important aspect of their research design. However, Kris is also attempting to construct a cogent narrative that illustrates how these cycles of experimentation and play, and their responsiveness to the unexpected outcomes they bring about, have generated new knowledge regarding the use of AR in performative contexts.

The main challenges Kris faces are:

- Identifying what 'type' of data each file is: practice, planning materials, documentation etc. For the purposes of examination, Kris wants to distinguish between data that should be assessed and data that plays more of a supporting/evidentiary role.
- 'Flagging' the moments in the development of the practice that generated findings. Illustrating how these findings informed future cycles of practice and, ultimately, resulted in the contribution to knowledge.
- Presenting the messy development of their practice in a way that makes sense to others: examiners, reviewers, the REF etc.
- Representing the 'interconnectedness' of their data, both in terms of how the various files relate to each other, and the chronology in which they were developed. Kris is finding this particularly challenging when using digital systems that only support linear file structures.
- Sharing their data in a way that is transparent, encourages discoverability, reuse etc. but without overwhelming anyone accessing it. Kris doesn't want to lose any of their data, but is also concerned that people accessing it will find it too challenging to navigate and therefore won't engage with it.
- Annotating their data in a way that adds to its meaning and significance. Finding a way to embed their analysis and reflections, but without adding even more complexity to an already challenging set of data.

Practice Researcher 2

Charlie is a music composer and they have begun working with a viola player to develop a technique central to a new composition. The playing technique, and the compositional techniques/understanding that align with it and emerge from it, are the key epistemic/knowledge object that Charlie wants to be able to preserve and communicate to other researchers; both academic and professional.

The first thing Charlie wants to do is create provenance and relate this to a field. To show that they are the author of these ideas. They want to be able to put up something to capture the date/s that they did this, and the people they worked with: in this case, the viola player's professional website and ORCID (they are also an academic researcher). Charlie wants to input some kind of metadata that allows them to position this relative to a field. This might be a 300-word abstract and some keywords, but also links to other projects/repositories/YouTube/Facebook/etc. The links and keywords will point both outward

to other researchers and projects, but also inward to Charlie's own research history and ongoing projects.

Charlie had a two-day workshop with the performer where they captured 20GB of video and audio (with some text and image files also, and some specialist files relating to the audio recording/mixing software). This is the raw data - Phase 1 of the project. At this point it's not clear to Charlie how much of this data they will want to make available, whether publicly or to other researchers. Charlie will want to add to this repository some smaller annotated versions of the audio/video files, and probably also some text files and image/pdf files or of score drafts or sketches, all of which would be public-facing. In terms of additional analysis tools, Charlie is not aware of any sort of tools that analyse the video or audio because anything in there that they think is important they are going to annotate and upload in a more specific context (e.g. shorter and more direct video examples). That said, for accessibility reasons it may be important to include transcription in some way, and it may be that as the researcher, the onus is on Charlie to make transcriptions of these things.

As this process and this project continues, Charlie wants to be able to refer back to aspects of this, and this may be the most challenging thing. If they upload 20GB of video and audio now, in six-months they may have either a finished product or a more-finished-but-not-completely-finished piece uploaded (Phase-2 repository) where they want to be able to refer back to specific moments or time indices within this data. In the most liberal version of this they might also want to be able to just say “Oh, and this is like that thing that happened in Phase-1of the project” and have the system work out what they mean from a description, but that’s not technologically viable yet. Assume that in Phase-2 they do this process again; uploading more materials that are now more advanced in some way: bearing in mind that there may also be a fork, a Phase-2a or Phase-3/4/5/*n* that do not linearly follow on from Phase-2. In any future phase Charlie may want to refer back to things that happened in Phase-1. Charlie thinks it should be up to the researcher to be clear about what the reference is; as an example, they’d want to refer back to video-N from 01:00–11:24 and have that playable from the Phase-2 repository. A good reference for this is how YouTube allows time indices to be entered in the description of a YouTube video, which automatically converts into a link to start the video at that time. To be able to do that for video and audio files would be really useful. Equally, to be able to do that for (e.g.) image/PDF/text will also be very useful. Charlie imagines doing this for text might be fairly straightforward (i.e. creating a link in a Phase-2 document to a specific line/section of text from a Phase-1 document), buto be able to do this with images/pdfs would be amazing: ie. within a Phase-2 document, to be able to refer to a point (a box? An element?) in an image in Phase-1; to take an image from Phase-1 and annotate it within a Phase-2 document?

To further describe what Charlie is uploading in Phase-1. You can see in the link [redacted] that it's a set of videos and audio in a folder structure. This is the raw data. There will later also be another folder of annotated videos and audio (so Phase-1 will need to be editable by them, but also probably lockable at some point by REF admin or similar). There may also be a PowerPoint slideshow that collects specific moments and annotations from this process. Charlie imagines that for other researchers, this is the place where there will be most variety. The basic files might all be fairly straightforward in terms of being some audio or video or some image, or some text (where there may also be raw data files in much more unusual formats, but those wouldn't be expected to be readable by everybody, only those with that specialist software), but the different ways that people annotate and mind-map and refer to that raw data, to package it up in a way that communicates knowledge outcomes forward and prioritises/categorizes all of those things; that will be very different for very for different people. For Charlie, it will be annotation of the video and audio.

One last suggestion. Something Charlie would really like to be able to do is have a director's commentary on the video/audio. To be able to have a video that they upload which has its existing audio track, but also that they can have an additional audio-commentary added that can be turned on and off, i.e. a video that has two separate audio tracks that can be toggled or even played together (this would require separate volume controls probably).

What Charlie will currently do with this (i.e. non-SPARKLE version):

The 20GB of data will sit on Charlie's google drive, or maybe get moved to their institutional OneDrive space; they don't want to approach their institutional library research-data team with it until they have a better idea of what will go into a repository and how it will be structured, but ultimately that is where it will go. Charlie will wait until the whole piece is complete and has been performed so I can make a single repository of the whole process, with the correct structuring etc, but this may be in several years, and the research will not be discoverable or disseminable until then.

Charlie can't indicate provenance through their artist website yet because that's only for finished artworks (and not really research-facing), and there isn't really a way to upload anything to their institutional CRIS because they don't have a 'thing' yet: they can probably add an entry for this as a 'Professional Activity' entry, but the system is designed for publications so that would be a 'cludge' at best. The octopus.ac approach to structuring research stages is a useful model here, if too methodologically strict for most practice research.

Charlie usually makes some short annotated videos that they disseminate across social media for other researchers and professionals, but in this case they don't want to be 'scooped' until the piece is complete, so they'll keep those videos on their OneDrive for now and take notes about the current state of the knowledge (key ideas and their documentation/examples etc).

Ultimately they will most-likely disseminate this knowledge through a conference paper, article, or other text form, which will be supported by a multimedia repository page at their institution.

REF Outputs Coordinator – At a university

A member of a team responsible for coordinating the collection of output copies from Units of Assessment (Units) in advance of final submission to REF 2021. Offering support to confirm that:

- Output copies provided by Units met REF 2021 output submission guidance (eg. on permitted submission formats, inclusion of mandatory descriptive information, etc.) as appropriate to the output type;
- All referenced and/or associated additional output components/contextual information/evidence were included;
- The descriptive metadata that would accompany the output via the REF submission system was accurate, including date of publication/first public dissemination;
- Evidence of the date of first public dissemination was included in the submission and/or available as appropriate to the output type.

Working with contacts to clarify any issues and obtain necessary output copies, information, and/or clarification in advance of submission.

REF 2021 guidance stipulated that submitted output copies:

- Could be submitted either in 1) physical format (eg. actual output copy, USB, DVD, etc.) or 2) electronic format through the REF submission system (eg. as a URL or PDF), depending on the output type - but not across both formats;
- Where an output was being provided to REF via a website/URL:
    - The website/URL was required to resolve to the version of the output that first became publicly available within the REF 2021 publication period (and institutions needed to be able to evidence this);
    - This 'point-in-time' copy could not require the input of a login and/or password or specialist software and needed to preserve the anonymity of REF reviewers/panel members.

Practice-based research outputs can be quite complex, and may include:

- Multiple *components*, which could encompass:
    - different *output* types; and/or
    - distinct dissemination dates for each component (though an overall single date was still required for REF submission purposes);
- Additional contextual material, in which case it was important that it was evident to REF assessors which material was 'output' vs contextual;
- Evidence of publication/first dissemination (eg. published performance programmes, etc.);
- Many different file types (images, video files, audio files, screenshots, websites, pdfs. etc.);
- Links to external websites;
- Material that could not be posted publicly online (e.g. due to copyright, etc.);
- Physical format material (e.g. booklets, programmes, promotional materials, etc.).

Providing outputs electronically via a URL directing to existing online research repositories was considered for REF 2021 output submissions. However, the necessity of including and maintaining evidence and navigability specifically tailored to the REF 2021 assessment exercise and assessors meant that – particularly for multi-component research outputs – it often made sense to instead recommend that Units submit an output's digital material physically e.g. a USB storage device.

In such instances, where links/URLs to external materials or content were included on the USB, additional copies of that content (e.g. archived screenshots, video files, audio files, etc., ideally date-certified) were also saved on the storage device, with the aim of solving the issue of REF assessors potentially encountering 'dead' links or the risk of web-content changing after the submission period.

Overall, key considerations/themes that became apparent during the REF 2021 exercise include:

- The importance of design and the overall aesthetic of the output's presentation;
- The importance of templates, forms, and language that accurately reflect practice-based research, roles, and activities;
- The importance of being able to illustrate links between different practice research elements/activities/collaborators – both internal and external;
- The uniquely longitudinal nature of some practice-based research;
- The diversity of relevant file types that need to be considered.

While recognising that any future practice research repository could not (and should not) be designed to privilege presenting outputs for REF purposes, functionality that may be useful for future REF exercises includes:

- Date of dissemination/publication metadata for individual items/components (month and year, at minimum);
- Re-use/copyright information;
- Information on the output type;
- Information on the file version, format, upload date etc.;
- Information on the history of the record (what's been added, changed, removed, locked, etc. by whom and when – e.g. for audit purposes);
- Ability to have administrative viewing, editing/uploading, locking privileges for the record(s) for certain users/outputs;
- Ability to share a link to either the overall output 'portfolio', or a single item/component;
- Ability to run reports on outputs/collection for reporting purposes;
- Ability to identify (and extract info on) collaborations – both internal and external to the University.