

DCC Methodology for Designing and Evaluating Curation and Preservation Experiments V1.1

1. Objectives

The purpose of this document is to describe a Digital Curation Centre (DCC) testbed methodology which will serve as a workflow framework for designing experiments to validate the effectiveness of curation and preservation strategies. The methodology is grounded in the following general principles: the methodology must

- conform to the fundamental standards of a scientific methodology,
- be easy to follow and implement, i.e. accommodate experimenters of all levels of technical expertise,
- be general enough to accommodate future changes and the evolution of ideas in curation and preservation theory and practice,
- be specific enough to provide concrete guidance in the immediate short term,
- be sufficiently flexible and extensible to allow for technological advances and the evolving complexity of available resources

The methodology extends that described within the Planets testbed¹ framework, to fully reflect curation activity from an end user perspective. Previous methodologies have been focused on the testing of tools which take specific action on digital objects, and gauging its performance in terms of whether or not it meets organisational objectives. The organisational objectives are, however, only partially disclosed implicitly through the importance factors which have been attributed to the properties of digital objects, management, and costs by experts in the organisation. The DCC testbed will complement previous work by evaluating usability of digital assets on the basis of end use case validation, to make objectives more transparent and accessible to end users. Use cases will be modelled by identifying what is being used (what), for what purpose (why), in what way (how), when it is being used (when), by whom (who), in what place (where). This will serve two purposes:

- it will bring home to data creators and users an awareness of exactly which of their processes will benefit from curation and preservation activities, encouraging them to get involved, and,
- it will make the aggregation of experimental results more viable across and beyond organisations.

The use case model will be made transparent and documented so that it can be used repeatedly to validate the usability of digital assets within the context of the tool or policy being tested, and so that a browser of the experiments can easily determine whether a specific experiment is relevant to use objectives.

The following section will provide a brief description of previous methodologies to add context to the evolution of preservation testbed methodology work which has formed the basis for the implementation of the PLANETS testbed. The subsequent section will describe the proposed use case model and how the previous methodology will be absorbed into the DCC methodology.

¹ <http://testbed.hatii.arts.gla.ac.uk:8080/testbed/> link to be updated in the future.

2. Previous testbed methodology

In order for the DCC testbed methodology to carry weight within the digital preservation community, it needs to subsume the advantages of previously developed and established methodology. A review of these methodologies is presented in this section.

Dutch testbed methodology²

This is the result of a three year research project established by the Dutch government, influenced by Jeff Rothenberg's proposal under the title “ Digital Preservation: Carrying Authentic, Understandable and Usable Digital Records Through Time”³. This effort was driven by an archival perspective and therefore is underwritten by the records management approach to digital preservation. The project team consisted of the Nationaal Archief and two UK firms, Tessella Support Services plc with Audata Ltd. as subcontractor. The main concerns of this work was to examine authenticity features of digital records, defining, capturing, and generating meta data for ingest, preservation and long term access, cost factors, possible technical solutions, effectiveness of the approaches available at the time. The initial study confined itself to considering digital records such as text documents, e-mail messages, simple spreadsheets and databases in the context of approaches such as migration, emulation and XML. An emphasis was put on defining a required subset of metadata that would help preservation, and experiments were designed to investigate the effects of content, context, structure, appearance and behaviour of digital records. Their methodology implements twelve stages in experiment design:

- Stage 1 Define Exploration Area
- Stage 2 Prepare for Experiment
- Stage 3 Define Requirements
- Stage 4 Develop Experiment Design
- Stage 5 Specify Resources
- Stage 6 Go/No Go Decision
- Stage 7 Develop Experiment
- Stage 8 Test Experiment
- Stage 9 Go/No Go Decision
- Stage 10 Run Experiment
- Stage 11 Evaluate Experiment
- Stage 12 Consider Results

Each stage is defined clearly by specifying what is being done, why it is being done, who is doing it, where and when, and how it will be done, as well as a description of any inputs required and outputs expected, and, what is done the resulting output. Apart from procedural specifications, the Dutch testbed recognises the essential role of documentation. It is recommended that each process, at each stage, is documented thoroughly. There is a paper⁴ describing the experiments there have been successfully carried out with the Dutch framework⁴.

² Testbed Digital Bewaring Research Framework 2.11, Nationaal Archief, Netherlands

³ Jeff Rothenberg and Tora Bikson (RAND Europe) , ‘Carrying Authentic, Understandable and Usable Digital records

Thorough Time’, The Hague 1999. Website: www.digitaleduurzaamheid.nl under Testbed (also in English).

⁴ Slats and Verdegem (2004), Practical experiences with the dutch digital preservation testbed. Journal of Information and Knowledge Management Systems, 34 (2):56-65.

DELOS deliverable 6.1.1: testbed methodology⁵

This research follows on from the Dutch testbed framework introduced above and was carried out as part of the DELOS Network of Excellence Work Package 6 Digital Preservation cluster initiative. The concerns expressed in this report coincide with the concerns expressed within the Dutch framework, but with the following overall objectives emphasised:

- Establish a framework of a digital preservation testbed environment.
- Produce metrics for testing and validating digital preservation strategies.
- Establish mechanisms for ensuring comparability between testbed environment including a testbed test data set (which might include programmes as well as data).

The objectives are driven by an effort to avoid repeated efforts, identify research questions, and monitor state-of-the art resources. Unlike the pilot study of the Dutch testbed concentrating on the limited record types, preservation approaches, and object attributes, the DELOS study is an effort to take this to the higher theoretical level for long term applicability and use. Their scope of study is therefore wider and more abstract.

The DELOS methodology retains the same architecture and twelve stage process as the Dutch proposal and maintains the concern for documentation via experiment and research databases, along with the same special concern for defining preservation metadata. The study also provides an overview of further context in general testbed research, including data-centric environments constructed within the information retrieval (IR) community such as TREC⁶. TREC has been a phenomenal force in taking IR technology to the next level by providing a shared data collection on which to compare and to evaluate experimental results. The shared data and evaluation methods maximise the comparability of experimental results.

In the digital library community there have been some effort to create shared data such as the D-Lib Test Suite⁷ and the Open Video Project⁸. The weight of a shared data collection is enormous in establishing the comparability of experimental results and is essential in scalability measurements. In the DELOS framework, however, the initial emphasis has been placed on the motivation to realise an infrastructure for creating a laboratory setting that enables the testing of preservation and curation approaches. They do point out the necessity of adequate object selection for the experiments from model objects to real world example. Section 2.1 of the DELOS report makes reference to a selection of papers discussing the data-centric approaches the testbed development.

The annexes of the report provide valuable templates that can be adopted for the DCC in implementing the twelve stages of the Dutch testbed experiment design framework.

DELOS Deliverable 6.4.1: Framework for Documenting the Behaviour and Functionality of Digital Objects and Preservation Strategies⁹

The DELOS Testbed framework described in the previous section defined the research questions and processes and documentation necessary for building a preservation and curation research experiment

⁵ DELOS Network of Excellence WP 6 Digital Preservation Cluster deliverable 6.1.1, <http://www.dpc.delos.info/outputs/index.php>

⁶ <http://trec.nist.gov>

⁷ <http://www.dlib.org/test-suite/research.html>

⁸ http://www.open-video.org/project_info.php

⁹ DELOS Network of Excellence, WP 6 Digital Preservation Cluster deliverable 6.4.1, <http://www.dpc.delos.info/outputs/index.php>

DCC Methodology for Designing and Evaluating Curation and Preservation Experiments

methodology. However, the framework does not specify precisely how to quantify objectives for measurement and comparison. This problem is addressed in the report produced by DELOS deliverable 6.4.1. by setting the workflow for assigning measurable quantities to the behaviour and functionality of digital assets and preservation/curation strategies. Whereas the Dutch testbed methodology uses the nine descriptive elements for each of their twelve stages for experiment design, the DELOS framework for documenting the behaviour and functionality of digital objects goes deeper into an explicit description of exploration area and requirements. This is carried out using a objectives tree which identifies the characteristics of an object and organisational objectives regarding the object (see example in Table 1, extracted from DELOS deliverable 6.4.1).

Table 1.: Exemplary implementation of an objective tree (from DELOS deliverable 6.4.1)

Top level	Level 2	Level 3	Level 4
File Char.	Appearance Structure Behaviour	Letters Paragraph Page Word functionality	Size Special characters Separation Picture inclusion Footnotes Page numbering Page margins Page break
Process Char.	Authenticity Stability Scalability Usability	Change traceability Completeness Independent from File format range Size/amount of files Availability Complexity Functionality	Authors Tables of content Rest Electricity Hardware Software Location(downloadable) Location(portable) Loading time Saving time Saving Maintenance Reopening Automatisation Search function Additional remarks
Costs	Technical Personnel	Hardware Software Personnel Initial Saving Maintenance Reopening	Required Space Electr. Energy Initial Assets Maintenance Initial Assets Maintenance Adoption Introduction Maintenance

DCC Methodology for Designing and Evaluating Curation and Preservation Experiments

Once these are identified, each item is quantified, i.e. represented as a measurable value. Alternative preservation/curation strategies are selected and tested on the objects. Then the resulting objects are examined on the basis of each item in the objectives tree for values that can be compared against the initial values. Finally, the difference of each compared value is aggregated on the basis of the weights (reflecting organisational objectives) assigned to each factor in the tree. Alternative strategies are then ranked according to the aggregated value.

The DELOS testbed methodology has been brought together with the functionality analysis in a paper presented at the International Conference on Asian Digital Libraries¹⁰. The DELOS report provides illuminating case studies on building objectives trees for Journals and Audio Files, which can be used as templates to be adopted when the DCC testbed experiment is being designed.

PLANETS Testbed methodology¹¹

The PLANETS Testbed¹² is built to meet the functional requirements specified by the experiment methodology of the Dutch testbed and it implements the functionality and behaviour objectives approach introduced by the DELOS deliverable 6.4.1 described in the previous section. A research paper on the integrated framework for preservation planning in the testbed environment was described at the 7th ACM/IEEE-CS joint conference on Digital libraries¹³. In this model, the twelve processes developed by the Dutch research has been modified to eleven stages, where the previous stages 8 (Test Experiment) and 9 (Go/No Go Decision) have been removed and stage 12 (Consider Results) has been expanded to include the three stages Transform Measured Values, Set Importance Factors, and Analyse Result.

The PLANETS testbed itself is mostly only concerned with benchmarking individual tools on selected digital objects and does not enable the implementation of aggregation of results to evaluate whether or not selected tools meet organisational objectives, which has been left to the Plato tool (<http://www.ifs.tuwien.ac.at/dp/plato/>) currently being developed to support preservation planning. The twelve stage Dutch testbed methodology has been contracted to six stages (Define Basic Properties, Design Experiment, Specify Resources, Go/No-Go decision, Run Experiment, Evaluate Experiment) in their architecture for this reason (<http://testbed.hatii.arts.gla.ac.uk:8080/testbed/>).

At the moment the experimental scenarios revolve around characterisation, migration and emulation, and each instance is described as a separate example. For example, a characterisation experiment has the following schematic process:

1. Select the digital objects and the properties of the objects to focus on.
2. Identify the initial values of the properties selected for the digital object(s).
3. Run experiment (e.g. run characterisation tool X on the digital object).
4. Identify the terminal values for the selected properties resulting from the experiment (e.g. produced by tool X).
5. Compare the initial property values to the terminal property values.
6. Mark this outcome.

¹⁰ S. Strodl, C. Rauch, A. Rauber, H. Hofman, F. Debole, G. Amato (2006). The DELOS Testbed for Choosing a Digital Preservation Strategy, Proceedings of the 9th International Conference on Asian Digital Libraries, Springer, p. 323 - 332

¹¹ PLANETS TB3-D2-MethodsForTesting_v1.1

¹² <http://testbed.hatii.arts.gla.ac.uk:8080/testbed/>

¹³ Strodl, S., Becker, C., Neumayer, R., Rauber, A. (2007) How to choose a digital preservation strategy: evaluating a preservation planning procedure. Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, ISBN:978-1-59593-644-8, 29 – 38.

DCC Methodology for Designing and Evaluating Curation and Preservation Experiments

Finding the property values or characteristics of the original object can be done in various ways:

- Using corpora with known characteristics;
- Using visual inspection of the original file;
- Using an “approved” / “trustworthy” characterisation tool to define the characteristics.

The PLANETS guideline emphasises the introduction of objective measurements in property value descriptions so that values and the judgement of the similarity between values are not organisation or purpose dependent. The organisation or purpose is recommended as an interpretive tool using Plato (<http://www.ifs.tuwien.ac.at/dp/plato/>) to rank the various experimental results registered in the testbed.

CASPAR Validation methodology¹⁴

The above list of methodologies are not meant to be an exhaustive list of methodology. For example, closely related frameworks include the CASPAR validation methodology. The Caspar framework suggests testing preservation strategies using three metrics:

- sound theoretical underpinning;
- changes in
 - ✓ software and hardware,
 - ✓ environment (e.g. organisational and legal requirements),
 - ✓ the knowledge base of the designated community;
- and a development of standards for the audit and certification of a repository's trustworthiness.

The theoretical underpinning mentioned includes models for mapping digital objects to information objects, identifying gaps in Representation Network and other external factors such as measuring the rate at which hardware degrades. The CASPAR strategy is presently focused on datasets from three communities: cultural heritage data, contemporary art, and science data.

3. DCC testbed methodology - use case driven workflow

The previously proposed methodologies are already anchored to a scientific framework. As such the framework can be subsumed by the DCC testbed methodology as a core activity. However, it is recommended that the DCC methodology further encourage taking a flexible and holistic approach, and, developing a fuller evaluation methodology. The use case driven methodology we will present here is intended to support both the CASPAR and Dutch/DELOS methodology by modelling the various changes across knowledge bases listed in CASPAR, and the organisational objectives listed in the DELOS framework, in terms of use cases that represent the community knowledge base and organisational objectives.

The PLANETS approach is constructed to confine itself to scoring the performance of tools or services. This is carried out by using a similarity function to measure initially observed values of pre-defined properties of selected objects against the values of the same properties observed after the experiment transforms or analyses the same objects. It is recommended that no contextual information is integrated into the evaluation process to maximise objectivity. The context is introduced at the ranking stage introduced within PLATO. The context is represented through the organisational perspective, objectives and goals which define importance of different object properties, as well as policy, costs, and resources. The tools are then ranked according to how well they meet these requirements. The PLATO guideline is driven by experts who take charge of the experiment from selection and characterisation of data to the evaluation of the experimental results. At the heart of the PLANET/PLATO framework is the assumption that each preservation or curation activity can be evaluated independent from context by measurable values. That is, it is assumed that, given a preservation strategy which has as a component the conversion of PDF to text, the state of all other variables can be fixed while we choose the best tool to perform the conversion. This however is not the real world

¹⁴ <http://www.casparpreserves.eu/search?SearchableText=validation&x=17&y=9>

DCC Methodology for Designing and Evaluating Curation and Preservation Experiments

situation: the processes which produced the PDF file can affect the outcome of the conversion. A tool which does a good job on one set of data can be completely unreliable on another set of data. To test a tool in a more robust way each preservation/curation process has to be considered within context. Ideally, two tools can only be compared if they are tested on the same data embedded within the same organisational policy given the same resources. Likewise, two policies can only be compared if they are tested using the same tools on the same data given the same resources.

The suggestion of the current methodology is to move from the notion that tools and services can be tested away from context and that such contexts can only be implemented by organisationally driven experts. We would like to introduce a use case driven approach to capture contexts which can be useful across a broad range of organisations and user communities.

The fundamental objectives which underlie preservation are:

- Primarily to preserve the digital object's value as an information object, e.g. this could include
 - its role as a scientific database,
 - its value as a piece of evidence
 - information source
 - etc.
- Consequently to preserve the ability to reliably manipulate the object as part of an information processing activity, e.g.
 - manipulating a database for experiments or further analysis
 - re-publishing a previously published article
 - re-purposing previously composed text
 - producing evidence in court
 - etc.
- Predictably, to accommodate use cases which have not yet been identified but might be performed in the future, e.g.
 - data produced in one community being used in another community for a completely different purpose from the original reasons for its collection.

The new testbed methodology should accommodate not only testing the effects of limited preservation activities (such as the comparison of different migration tools on specified data objects) but also testing how different processes, policies, and resources affect the viability of using the objects in selected scenarios. The only way to truly validate a preservation strategy is by showing that the strategy results in usable assets (Figure 3). Figure 3 is not meant to imply that there is no interaction between creators, curators and end users. In fact, it is most likely that there will be some interaction, and, ideally, the communication between these actors and agents should be encouraged and developed actively (Figure 4). Figure 3 is intended to convey the fact that the preservation strategy should not *depend* on interaction, and that the experiment methods should be devised to test the role and effectiveness of each agent/actor separately.

DCC Methodology for Designing and Evaluating Curation and Preservation Experiments

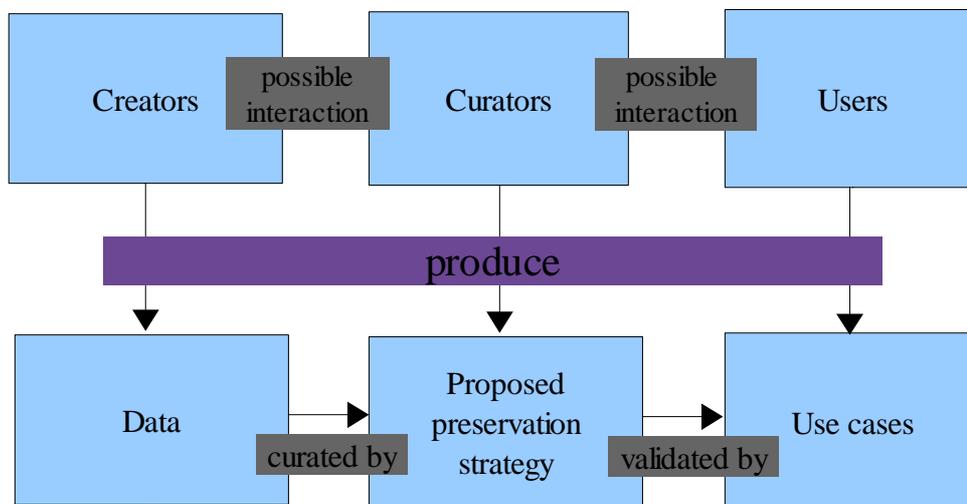


Figure 3. Experimental methodology

Already in previous methodology, the comparability of the values, resulting from a selected set of experiments, to organisational benchmark goal values, once the organisational objectives are identified, is soundly represented. The construction of the objectives tree which identifies the organisational objectives (introduced in DELOS deliverable 6.4.1- also see Table 1), however, is the most difficult and time consuming step in preservation planning. The DCC methodology aims to alleviate this difficulty by presenting experimental validation on the basis of use cases. Organisational objectives are often designed to make resources usable within selected use case scenarios (under the constraints of resources and policies). These use cases are understood and are applicable across a broad range of organisations. Experimental results validated by use case models would support users in efficiently and effectively deciding which experiments are relevant to the community. Even the same designated community use the same data in different ways and it is essential to be able to gauge the weaknesses and advantages of the strategy with respect to these different uses within and beyond organisational objectives.

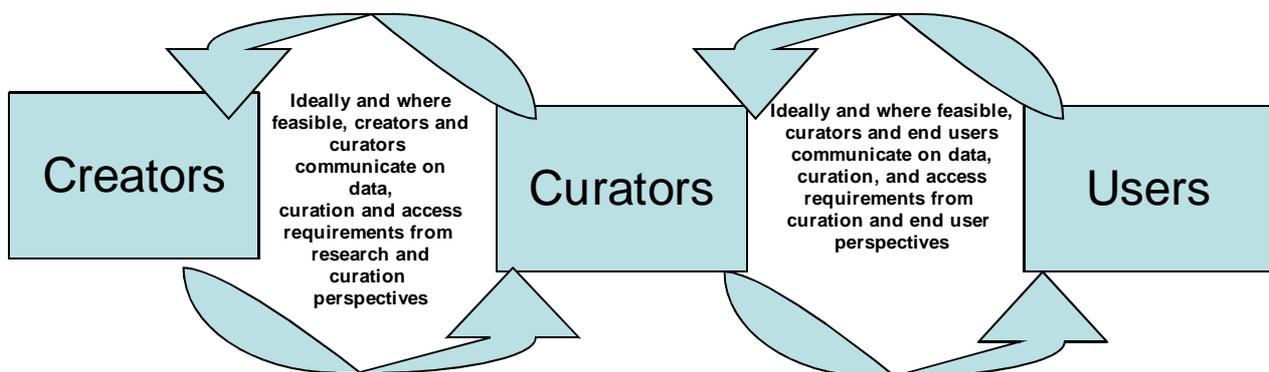


Figure 4. Interaction between actors and agents.

DCC Methodology for Designing and Evaluating Curation and Preservation Experiments

The use case validation ideally should involve real time users, but can also be a set of automated simulations in an expedited time frame¹⁵ or controlled user evaluation using case models assigned to users as tasks. A use case model will minimally consist of identifying

1. what digital resource is being used (what)
2. for what purpose is it being used (why)
3. in what way is it being used (how)
4. by whom (e.g. researcher, student) is it being used (who)
5. within which designated community is it being used (where)
6. when (e.g. is it regularly being used or is it being used only one-time) is it being used (when)

The identification of these six elements will facilitate the aggregation process necessary for selecting a preservation strategy in context. The use case scenario is not meant to influence the definition of property values being tested, but give more context to the selection of object properties and later, in the evaluation stage, add the possibility of another level of analysis involving the tool's effectiveness in the usability of the object, as evaluated by formal use case application. There are four different perspectives on evaluation in the current context:

- how much the property values change after an experiment,
- how good the tool is when compared to a state-of-the-art tools constructed to do similar tasks,
- whether it meets predefined thresholds for organisational objectives,
- whether it meets selected thresholds required for specified uses.

In the PLANETS framework evaluation is carried out with respect to the first criterion only. This will be expanded to consider the second and third criteria when PLATO is released. In the use case scenario development model of the DCC methodology we aim to enrich the evaluation by also considering the fourth criterion and providing the resources for establishing a correlation between all four aspects of evaluation.

The DCC methodology will absorb the methodology developed by the PLANETS testbed as a bedrock at the core of its scientific experimental design, but encapsulate this with use case scenarios and user validation as a defining feature to make the experiments more transparent and transferable to the greater digital curation community. By retaining the object property value measurement and weighting suggested in the DELOS framework but augmenting it with use case validation, we hope to produce a methodology that can work across several types of objects including those that depend more heavily on low level features of the document and human cognitive behaviour (e.g. images and text) to those that depend less on these types elements but on the usability of the information (e.g. scientific datasets). To summarise the work flow:

1. Develop use case scenarios and model use cases according dimensions including what, why, how, who, where, when described above;
2. Define basic properties of the experiment, such as name, description, purpose and focus;
3. Design the experiment;
4. Specify the required outcomes i.e. quality criteria or characteristics that need to be maintained;
5. Go/ no go decision based on whether experiment is feasible;
6. Run the experiment in the PLANETS testbed environment;
7. Validate the results of the experiments by implementing use case process on results,
8. Record the results of 4 and 5 in a DCC report, and the database of the testbeds.

¹⁵ For example, akin to those employed in the simulation of emergency responses (<http://www.mel.nist.gov/div826/msid/sima/simconf/mns4er.htm>)

DCC Methodology for Designing and Evaluating Curation and Preservation Experiments

The DCC methodology will allow us to

- understand experiments on the basis of tasks that we can easily identify as relevant or not relevant to organisational purposes,
- establish a reliable source of information for establishing correlations between data use case and data properties, so that property value thresholds can be set to accommodate different use,
- demonstrate benefits to data creators and users to get them involved in curation and preservation activities,
- accumulate and build use cases to extend curation experiments for cost modelling and studying the interaction between use cases and organisational policies.

Apart from use case validation, the testing of curation and preservation strategies in the DCC methodology described above is intended to emphasise the importance of documentation and cross referencing material related to the experiment. Thorough cross referenced documentation of all processes, decisions and research results and products involved in an experiment is central to the comparability of the experiment to other experiments and the repeatability of the experiment, i.e. the possibility of validating the methods and results.

As mentioned in DELOS deliverable 6.1.1, it is also essential to have shared test data for experimentation, as well as new test data for verification and further analysis of results. It is suggested an initiative be implemented to build and make shared data available for use in various experiment design.

The next steps for the development of the proposed framework is to scope the range of use cases that might be available for modelling, to carry out some experiments to be validated by a selection of these use cases, and to evaluate the viability of the methodology. One way we might build uses is by examining preservation scenarios that have already been developed across the testbed projects that have been described in this paper (e.g. CASPAR User Requirements and Scenarios¹⁶), as well as others that may be in development.

Finally, it should be mentioned that the methodology presented here should be interpreted in the most general terms. It is intended as a high-level guideline for designing further concrete experiments in detail. As mentioned at the beginning of this document, it should be subject to continued development and refinement to accommodate future changes and goals. A static methodology with no scope for improvement would not benefit the DCC in the long term.

¹⁶ <http://www.casparpreserves.eu/Members/metaware/Deliverables/user-requirements-and-scenario-specifications/view?searchterm=scenario>