HERITAGE SCIENCE

RESEARCH ARTICLE

CrossMark

Damage function for historic paper. Part I: Fitness for use

Matija Strlič[1*], Carlota M. Grossi[1], Catherine Dillon[1], Nancy Bell[2], Kalliopi Fouseki[1], Peter Brimblecombe[3], Eva Menart[1], Kostas Ntanos[2], William Lindsay[2], David Thickett[4], Fenella France[5] and Gerrit De Bruin[6]

Abstract

Background: In heritage science literature and in preventive conservation practice, damage functions are used to model material behaviour and specifically damage (unacceptable change), as a result of the presence of a stressor over time. For such functions to be of use in the context of collection management, it is important to define a range of parameters, such as who the stakeholders are (e.g. the public, curators, researchers), the mode of use (e.g. display, storage, manual handling), the long-term planning horizon (i.e. when in the future it is deemed acceptable for an item to become damaged or unfit for use), and what the threshold of damage is, i.e. extent of physical change assessed as damage.

Results: In this paper, we explore the threshold of fitness for use for archival and library paper documents used for display or reading in the context of access in reading rooms by the general public. Change is considered in the context of discolouration and mechanical deterioration such as tears and missing pieces: forms of physical deterioration that accumulate with time in libraries and archives. We also explore whether the threshold fitness for use is defined differently for objects perceived to be of different value, and for different modes of use. The data were collected in a series of fitness-for-use workshops carried out with readers/visitors in heritage institutions using principles of Design of Experiments.

Conclusions: The results show that when no particular value is pre-assigned to an archival or library document, missing pieces influenced readers/visitors' subjective judgements of fitness-for-use to a greater extent than did discolouration and tears (which had little or no influence). This finding was most apparent in the display context in comparison to the reading room context. The finding also best applied when readers/visitors were not given a value scenario (in comparison to when they were asked to think about the document having personal or historic value). It can be estimated that, in general, items become unfit when text is evidently missing. However, if the visitor/reader is prompted to think of a document in terms of its historic value, then change in a document has little impact on fitness for use.

Keywords: Heritage management, Conservation, Heritage values, Fitness for use, Psychometrics, Statistical experimental design (DOE)

Background

In heritage management damage functions are used to model change in relation to environmental variables and other stressors, e.g. use. Damage functions are defined as 'functions of unacceptable change to heritage dependent on agents of change' [1], reflecting the fact that the term 'damage' reflects both the state of an object and the stakeholders' value of the level of change to that object that is unacceptable.

In our previous work [2], we quantitatively investigated the value users attach to library and archival heritage. The study involved 543 respondents in a variety of contexts: historic houses, reading rooms and exhibitions. Users were shown to have well-defined attitudes to objects, broadly characterised within nine categories: 'Future Value', 'Public Value and Evidence', 'Understanding the

*Correspondence: m.strlic@ucl.ac.uk
[1] Institute for Sustainable Heritage, University College London, London, UK
Full list of author information is available at the end of the article

Chemistry Central

Strlič *et al. Herit Sci* (2015) 3:33

Page 2 of 12

Present', 'Content and Learning', Personal Meaning and Identity', 'Discovery and Engagement', 'Rarity', 'Materials and Sensory Experience' and 'Connection to the Past'.

In the same work, we also investigated stakeholder attitudes to the future use of collections, and identified that ~90 % of users defined as visitors or readers, would want collection items to remain in a usable state (readable, or suitable for display) for the next 500 years. The most frequent response, around half of respondents, was 100 years. This corresponds to another study [3], where a similar question was posed to 187 experts curating or researching geological collections. This study reported ~70 % satisfaction with the same preservation horizon of 500 years. Other studies of perspectives on attitudes to the future have produced horizons of the same order of magnitude, if not necessarily the same particular horizon [4]. This is essential evidence for underpinning collection management protocols.

To effectively manage change of cultural heritage, we need to answer the question of how users interact with such change as well as whether different values affect such interaction.

This question is complex, as there are many processes of physical change, as well as values applied by users. It is also well known that such interactions might reflect the presumed 'use' of an object [5]. In this paper, we will define 'use' as the mode or context of use, for example display (objects are observed without handling) or

reading, which involves manual interaction. There may be other purposes, depending on the type of interaction in question, including, e.g. storage.

The concept of use allows us to assess whether an object is fit for a type of use. As a term, fitness for use is similar to 'condition', although it does not include the susceptibility of an object to degrade as a criterion of condition. The concept of fitness for use is more meaningful because if an object is unfit for a particular use, a change in how it is accessed is required. For example it may need to be removed from an exhibition, or use may only be allowed under supervision.

This implies that benefits can no longer be accrued from interaction with such an object, meaning that its value for the particular type of use has been consumed and that the object reached the end of its lifetime (Fig. 1), unless an investment in a conservation intervention restores its fitness.

Since accrual of benefits occurs in the interaction between objects and users, it is useful to explore how users define the metrics of fitness. A number of methodologies have been developed for this purpose, most of which are based on psychometrics [6] and standard procedures exist for estimating aspects of image quality using psychometric scaling [7–9]. Psychophysics is the field of quantitative studies of perception, examining the relations between observed stimuli and responses. It allows for the relation to be extracted, while providing a

Fig. 1 Newsprint from 1918 to 1919, labelled as 'unfit for use' due to the fragile nature of the material

Strlič *et al. Herit Sci* (2015) 3:33

Page 3 of 12

reason for it [10] and has found application in different fields, from occupational and clinical health [11, 12], to politics and law [13].

Psychophysical studies are broadly divided into two main classes. Threshold methods are based on the energy a user can just detect, such as the decibel level at which sound can be heard. These methods are used when detection of a stimulus is the factor of interest. Supra-threshold methods are used in instances where the stimulus is easy to perceive, and are used to discriminate between stimuli, such as discrimination between sounds [14].

The work on acceptable colour change focussed on threshold methods and the concept of 'just noticeable difference' (JND). JND represents "a stimulus difference that leads to a 75:25 proportion of responses in a paired comparison task" [15]. In assessment of image perception, this concept is used not only for colour changes, but all visually perceivable aspects of image quality e.g. blurring and contrast, as well as other aspects of image quality [16].

In the heritage field, this concept, including the synonymous term 'just perceptible colour difference', JND has mainly been used to understand the effect of illumination [17–20] despite the fact that the magnitude of a just noticeable difference remains ill-defined. If a more precise definition were in place [21], the timeframe for acceptable change could be defined as well. The concept of JND is well suited to measure the visual ability, rather than the viewer's perception of change in an image [22].

To determine fitness for use, concepts other than JND might be more useful, such as category scaling. The goal is to assign numbers to perceptual events, the benefit of which is that it allows for direct observer response to changes, although restricted to only a number of categories [23]. Category scaling methodologies exhibit high stability and lead to low participant stress [24]. Standard category titles have been developed which can be useful [25].

In this paper, we explore the attitudes of library and archival users to what could be defined as damage, i.e. the threshold at which change is no longer acceptable by users. We do this in the context of two types of use: reading and exhibition, and three contexts of value: pre-assigned personal or historic value, and no pre-assigned value.

Methods

Value scenarios

To examine which values might affect the way users evaluate the fitness of an object, we used the outcomes of the VALUE questionnaire study [2], which has clearly shown that different contexts of use are associated with:

- Different value profiles
- Different attitudes towards the future
- Different attitudes towards agents of change

It is evident that the value profile of a user may affect a fitness assessment and ideally, the influence of all types of values on fitness assessment might need to be evaluated. While from an academic point of view this might be justifiable, it is questionable whether the results could significantly affect collection management practice, particularly in large collections. It is therefore useful to examine only a few value types.

In this research we explored the potential to bias fitness assessment responses by describing the value of a document as a historical source and learning resource, without specifying the type of historic value (e.g. personal history, accountability, contextual, history), relating to the military, population migration, parish records, institutions including workplaces, hospitals and asylums, law and religion.

The second prompt was related to personal and community identity. To bias fitness assessment in this sense, the assessors were referred to lists of family member names, a relative's war record or information about a relative's early childhood, such as school or hospital records.

The third scenario was the 'random' scenario, where the assessors were given no specific information about the document they were assessing, thus providing un-biased data. In this scenario, the assessors considered the object as any random library or archival object, which allowed us to see how users assess fitness purely on the basis of material state of a document.

To summarize, we used the following three value scenario prompts:

1. *Random* "imagine you are assessing a random archival document"
2. *Personal* "Imagine that you are assessing documents you found during research into your family history"
3. *Historical* "Imagine that you are assessing documents by a historian interested in significant events in the history of your country.

Types of degradation

For psychometric scaling to allow for aspects of damage to be directly related to physical deterioration and to determine how this is affected by values, assessors were presented with a set of documents with deterioration features selected on the basis of the following criteria:

- Gradually accumulated during regular use (and not prior to acquisition or during catastrophic events, e.g. flood, fire, mould outbreak as such deterioration

Strlič *et al. Herit Sci* (2015) 3:33

Page 4 of 12

is the consequence of singular events rather than a continuous process)

- Easily visually detectable without prior expert material knowledge, thus allowing any archival user to participate.

In storage or display the mechanical strength or the colour of documents may change, however, it could be argued that only the latter can be easily assessed by a user without technical aids or expertise. On the other hand, during manual handling, tears and missing pieces might gradually accumulate as a consequence of reduced mechanical strength. Thus, the three aspects of degradation of interest are: tears, missing pieces, and discolouration.

Their influence on fitness for use could be assessed using actual archival documents, however, in such an experiment many variables could be difficult to control: document size, paper texture, writing style, deterioration prior to accession. Additionally, it would be difficult to prevent degradation of such archival documents during assessments.

Therefore, a suitable series of model documents was sourced from a notebook (1946) and therefore of identical appearance. These pages were further distressed to exaggerate the three aspects of deterioration (Fig. 2): discolouration, tears and missing pieces.

Three levels of degradation were considered (Table 1). Discolouration was achieved by dry heat at 190 °C for 15 or 45 min.

Experimental design

The aspects of degradation could be studied separately: if a missing piece and a tear appear on the same page, the two aspects might be assessed independently or one relative to the other. For example, a highly discoloured page with a large tear and a small missing piece could be evaluated as more unfit than another document with a large missing piece only. Therefore, combinations of degradation aspects need to be examined, to see if the aspects add up or if they have synergistic (or antagonistic) interactions.

To do so, the number of differently distressed documents used in a one-at-a-time experimental design would need to be 3^3, i.e. 27, which could make an assessment workshop long and potentially tiring for assessors.

To economise on the number of documents, while still being able to explore the interactions between aspects of degradation, principles of statistical Design of Experiments (DOE) were used. These allow for a significant reduction of the required experiments, as well as for variations of each aspect of degradation to be studied simultaneously, while the effect of each aspect on fitness for use can be evaluated independently. If the aspects have an additive effect, then DOE represents not only a more economic experimental design, but also allows for more precision; however, if the aspects are not additive, but interact synergistically, then DOE can allow for detection and estimation of such interactions [26]. The advantage of DOE is that it allows for the maximum amount of information to be extracted using
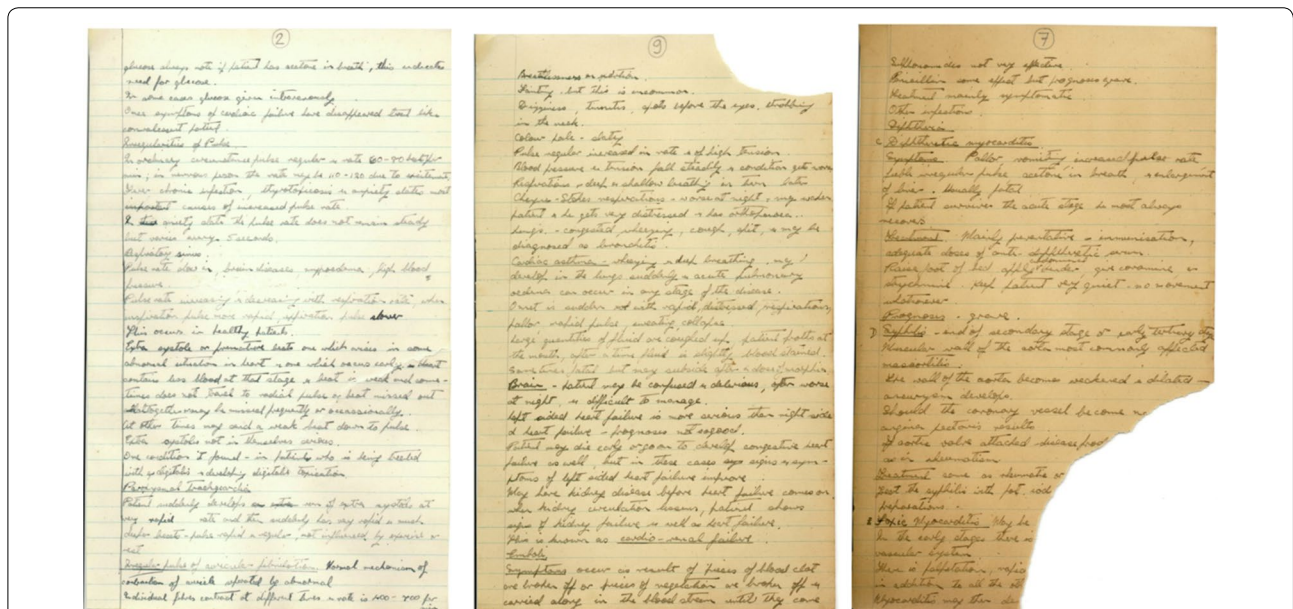

Fig. 2 Examples of differently distressed documents used in fitness-for-use workshops, progressively discoloured and with a progressively big missing piece from *left* to *right*. The document on the *left* also has a large tear, stretching across text. The documents were written in English hence users tended to read them, which they were reminded not to in order for content not to affect their responses

Strlič *et al. Herit Sci* (2015) 3:33

Page 5 of 12

Table 1 Aspects and levels of degradation of the documents used for assessment of fitness

Level	Aspect of degradation		
	Tears	**Missing piece**	**Discolouration ($\pm$SD, N = 5)**
0	None	None	As acquired: L* = 97.3 $\pm$ 0.2, a* = -5.0 ± 0.7, b* = 12.6 $\pm$ 1.5
1	In the margin, not extending across the text	In a corner, not extending across the text	ΔE_{1976} = 21 $\pm$ 6, L* = 83 $\pm$ 5, a* = 0 $\pm$ 2, b* = 27 $\pm$ 4
2	Extending across the text ~1/3 into the page	Extending across the text, ~1/6 of the page missing	ΔE_{1976} = 36 $\pm$ 7, L* = 72 $\pm$ 8, a* = 5 $\pm$ 2, b* = 35 $\pm$ 2

Total colour difference is relative to level 0

the minimum number of experiments, and for efficient handling of experimental errors [27, 28].

The simplest are factorial experiments, where all factors are varied simultaneously at a limited number of factor levels [26]. More complex DOEs involve response surface designs, such as the central composite design (CCD), which is among the most popular types of response surface designs. It consists of a factorial design with centre points, augmented with a group of axial points that allow for estimation of curvature. Such a design requires only 15 differently distressed documents, while the centre point can be repeated several times for a better estimation of the uncertainty. Three such repetitions give the total number of documents of 17. The DOE response surface designs and data analysis were carried out using Minitab 16 Statistical Software (State College, PA, USA).

The three factors of degradation, i.e. tears (T), missing pieces (MP) and discolouration (D) now need to be assigned to the three factors to produce the documents with required combinations of aspects of degradation (Table 2) at three levels: 0, 1 and 2.

The documents, single loose handwritten pages, were distressed in such a way that no aspect of degradation was removed or obscured by any other. As documents assessed for the purpose of reading were to be handled to reflect the normal process of reading, they gradually degraded during the workshops, in which case they were replaced.

Fitness-for-use workshops

In accordance with "Methods" it was examined whether thresholds determined for a particular type of use can be manipulated by the relationship that people form with the document under the influence of value factors. Hence, prior to the workshops, users were prompted to consider that the documents they were observing have a personal or historical significance, or not.

To account for the two types of use (display and handling) and three value scenarios (*Random*, *Personal* and *Historical*), six variations of workshops were developed (Table 3).

Table 2 Documents required for the fitness-for-use assessment workshops, with combinations of aspects of degradation: tears (T), missing pieces (MP) and discolouration (D), as required by the CCD response surface design, with three repetitions of the centre point

Document no.	T	MP	D
1	0	0	0
2	2	0	0
3	0	2	0
4	2	2	0
5	0	0	2
6	2	0	2
7	0	2	2
8	2	2	2
9	0	1	1
10	2	1	1
11	1	0	1
12	1	2	1
13	1	1	0
14	1	1	2
15	1	1	1
16	1	1	1
17	1	1	1

Participants were volunteers drawn from among the public at the workshop locations. All participants were given equal information about the purpose of the workshops, before any assessment was carried out. Further information was provided only after they completed the assessments. To control this, workshop assistants were trained and given the same information on how to interact with assessors.

Per scenario, 50 participants were required. The workshops were conducted at The National Archives (Kew, UK), Library of Congress (Washington DC, USA), and Wellcome Library (London, UK).

The objects were viewed individually, without taking other documents as a reference and participants did not move objects to compare one to another. The objects

Strlič *et al. Herit Sci* (2015) 3:33

Page 6 of 12

Table 3 Six fitness-for-use workshop scenarios, accounting for two different types of use and three value scenarios, with the associated codes, as used in the analysis

Code	Value scenario	Use
H–D	*Historical*	Display
H–R	*Historical*	Reading with handling
P–D	*Personal*	Display
P–R	*Personal*	Reading with handling
R–D	None (*Random*)	Display
R–R	None (*Random*)	Reading with handling

were not displayed in the numeric order (1–17, as in Table 2), but were arranged randomly.

For the two aspects of use, the participants were encouraged to interact with the documents in the following way:

1. Reading under the normal conditions as are expected in libraries and archives, where a document can be freely handled and read
2. Viewing, under the normal conditions as are expected in exhibition spaces, where a document is presented in a way that it is intelligible without handling, separated from the viewer.

The medical manuscripts usefully allowed for association with a historic event or a personal circumstance, however, since the fact whether or not a user could read the handwriting might affect their assessment, they were discouraged to read.

Each participant was asked to categorize the 17 documents for their fitness for the specified purpose. Categories from 1 to 5 were modified on the basis of those prescribed by BS ISO 20462-2:2005:

1. 'Excellent'
2. 'Good'
3. 'Quite good' (in the UK context) or 'Reasonably good' (in the US context)
4. 'Not good'
5. 'Unfit'

The data collected across the six groups of participants and scenarios were expected to indicate the threshold of damage, as represented in the various combinations of deterioration features in each of the sets of documents. Personal and historic value scenarios (H–R, H–D, P–R, P–D) were thought to provide an indication of how thresholds of damage are influenced by the two specific value contexts. In the following section we analyse the strengths and limitations of this approach.

Results and discussion

The premise of the workshops was that users have determinable acceptance levels for states of documents when considered for particular use and with specific attached values, and that these acceptance levels can be revealed through the fitness-for-use workshops. The acceptance levels represent thresholds of damage, whereby users combine their visual observation of a physical state with a judgement of what state is required for a particular type of use. Conversely, damage is defined as loss of fitness for use (unacceptable change).

Across the three participating institutions, The National Archives, Library of Congress and the Wellcome Library, 331 users responded and carried out the assessments during November and December of 2012 (Table 4). This provided a dataset slightly larger than required by DOE.

At the Library of Congress, the workshops were mostly carried out in the exhibit room, which provided an ideal context for the display scenarios, and only 15 responses were collected in the reading rooms. On the other hand, in the Wellcome Library, all of the responses were collected in the reading rooms, providing the context for scenarios H–R, P–R and R–R. At The National Archives, the workshops were carried out in a space which was separated from the reading room and from the exhibition space, so both modes of use could be explored.

Since the participants were asked to provide a written description of what they imagined the documents to represent (cf. "Value scenarios"), it was of interest to analyse whether the value prompts were effective. As a preliminary test, the responses were arranged into word clouds (Fig. 3), where the frequency of a word is represented by the size of the font used to depict the word in the cloud.

It appears that most assessors looked for a connection between the textual content of the document and a historic event or a personal circumstance. In the case of H–R and H–D, the most frequent associations were with the Civil War (in the US context) or with WWI or WWII (in the UK context). In the case of P–R and P–D, the key words are family, diary, grandfather, letters, history and war, indicating that the prompts were potentially successful. This indicates that participants responded positively to the value suggestions given to them by workshop assistants, who helped if the participants expressed difficulties with suitable associations. An analysis of the *Historical* and *Personal* value scenarios could thus be meaningful.

Frequency analysis of the expressed fitness for use (FfU) responses was performed on the individual response data, i.e. using all of the individual responses corresponding to the individual documents, per workshop scenario (Fig. 4). It is of interest to note that most responses are skewed

Strlič *et al. Herit Sci* (2015) 3:33

Page 7 of 12

Table 4 Total number of collected response sheets per scenario and institution

Workshop scenario	The National Archives	Library of Congress	Wellcome Library	Total
H–D	23	40	0	63
H–R	24	5	19	48
P–D	25	41	0	66
P–R	24	5	15	44
R–D	24	42	0	66
R–R	24	5	15	44

For workshop scenario abbreviations see Table 3

Fig. 3 Word clouds for responses to the question "What did you imagine the document to represent?" for the workshops scenarios H–R and H–D (*above*) and P–R and P–D (*below*)

towards the lower numbers, meaning that in general the users thought that the documents were on average 'Good' to 'Quite good' ('Reasonably good'). This is perhaps most evident in the case of H–D and R–D scenarios, and less so in the R–R scenario, and could indicate that when users were not given a value scenario they are the least forgiving.

A linear comparison of responses (FfU) vs. individual aspects of degradation (T—tears, MP—missing pieces, D—discolouration) was performed (Fig. 5) to investigate if there is one single aspect that has most influence on fitness for use. It is evident that it is likely that MP has the most pronounced effect on fitness as the regression between MP and fitness is strongest for all six scenarios, while the regressions between T and D, and FfU are weak.

Further quantitative analysis was performed to obtain the response surface in the form of the following general surface equation (Eq. 1):

$$FfU = c_1 T + c_2 MP + c_3 D + c_4 T^2 + c_5 MP^2 + c_6 D^2$$
$$+ c_7 T \cdot MP + c_8 T \cdot D + c_9 MP \cdot D + Const. \quad (1)$$

In this equation, terms 1–6 represent linear and quadratic terms representing the three aspects of deterioration without any interaction, while terms 7–9 represent interaction terms. If the interactions are significant, then coefficients c_7, c_8 and c_9 should be statistically significant ($p < 0.05$).

After the response surface model is developed on the basis of actual participant responses (FfU), the actual
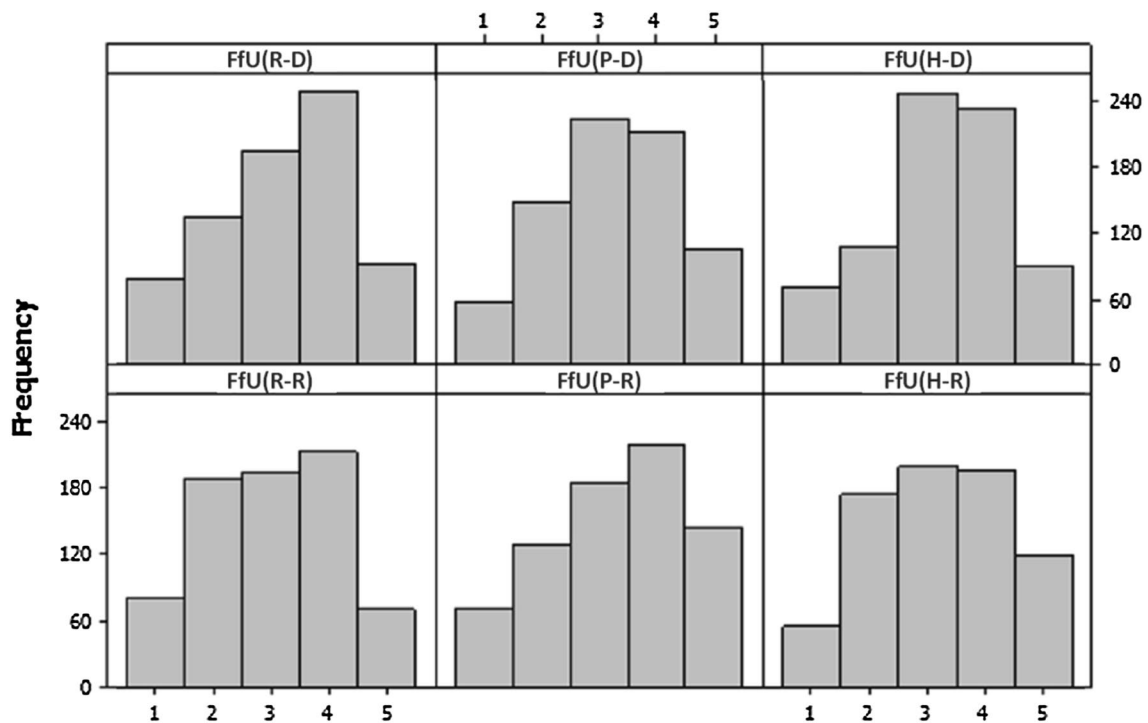
Strlič *et al. Herit Sci* (2015) 3:33

Page 8 of 12

Fig. 4 Frequency analysis of responses (*FfU* fitness for use), where each document represents one response (i.e. 17 per participant), per scenario. The ratings were *1* 'Excellent', *2* 'Good', *3* 'Quite good'/'Reasonably good', *4* 'Not good', *5* 'Unfit'
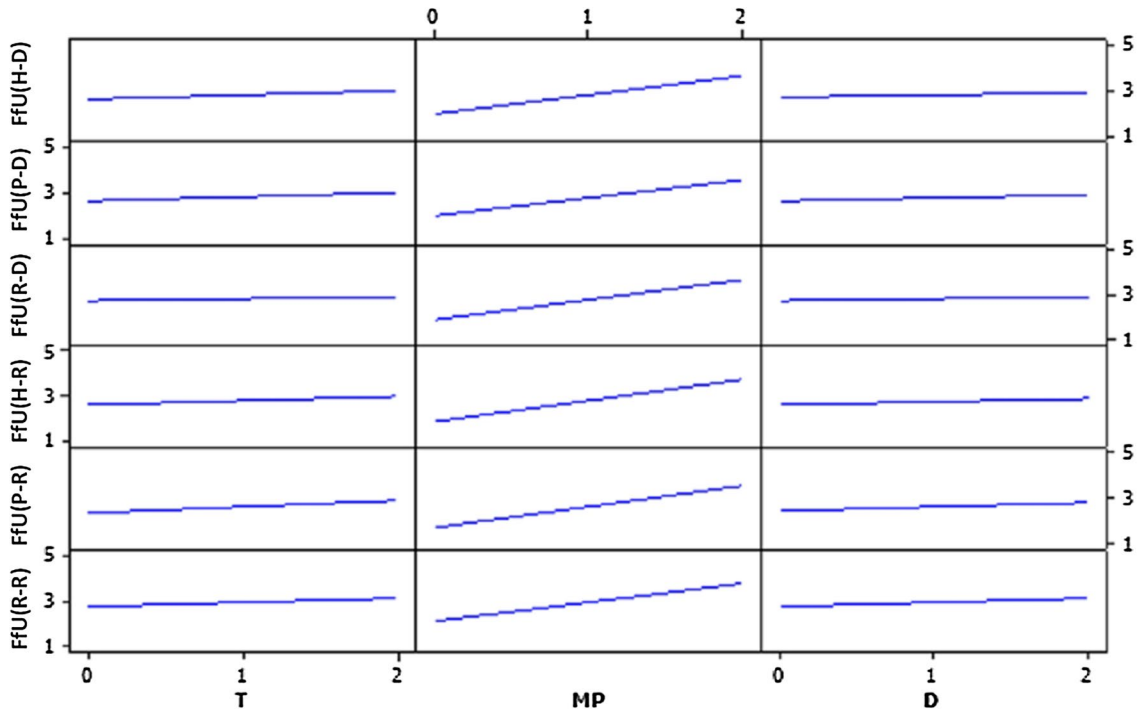

Fig. 5 Regression lines between FfU values (1–5) and levels of degradation (1–3) for tears (T, *left column*), missing pieces (MP, *middle column*) and discolouration (D, *right column*), for the six scenarios: H–D, P–D, R–D, H–R, P–R, R–R, from *top* to *bottom*

Strlič *et al. Herit Sci* (2015) 3:33

Page 9 of 12

responses can be compared with the calculated (modelled responses), and the quality of the correlation between actual and modelled FfU values can be explored using linear regression, and expressed as R^2, i.e. the squared regression coefficient—the closer it is to 1, the better the model.

The developed response surfaces have R^2 values (Table 5) that might be considered low for physical sciences, but indicate a moderate-large effect size according to Cohen's conventions, often cited in the psychology literature [29]. It is possible to appreciate that certain models better describe the responses, as the R^2 values are higher.

The coefficients modelled on the basis of Eq. 1 are compared for the different workshop scenarios in Fig. 6. It is evident that there are a few consistent features across all the scenarios. There is no statistically significant interaction between discolouration (D) and missing pieces (MP), except in one scenario, R–R. Discolouration seems to be evaluated mostly independently of the other aspects of physical degradation, and significantly contributes to only one scenario, i.e. R–R, reading of documents of no pre-assigned value.

The other interactions, T × D and T × MP are mostly statistically significant, although they contribute little to the overall score, as the coefficients are small. It seems therefore that tears and missing pieces seem to be also mainly observed as independent aspects of degradation.

Another interesting observation is that tears mostly contribute insignificantly to the overall score, although they seem to be more important in the context of manual handling (scenarios H–R, P–R, R–R). Readers seem to be concerned with tears if they need to handle an object: tears have an impact on how a reader holds a document—e.g. how easy it is to turn a page or pick a document up without damaging it further.

Overall, it is evident that the quadratic term MP × MP contributes most to the overall score, indicating that users

are mostly concerned with the textual content of a document. This is in agreement with the data presented in Fig. 5.

We therefore modelled the responses by using only the terms MP and MP × MP. The quality of this model is less good than the one based on Eq. 1, though only marginally so in most scenarios, as evident from the regression coefficients in Table 5. In Fig. 7, we can examine the differences between the coefficient values for the six scenarios. All are statistically significant.

The *Historic* value scenario significantly differs from other scenarios in that both the Constant and the MP term coefficient are significantly different from the other four scenarios, while the quadratic term coefficient is similar for all of them. With the linear term being negative for the H–D and the H–R scenarios, this means that the FfU responses for these value scenarios will be overall more similar, leading to the conclusion that with a historic document in mind, users mind least about how distressed the documents are, regardless of the access context, i.e. display or reading.

There seems to be a further minor, although statistically significant difference between H–D and H–R, i.e. for display purposes, the aspects of degradation are least important and most documents were considered 'Good' or 'Quite good' ('Reasonably good'), as is also evident from the corresponding frequency plot in Fig. 4.

The other scenarios, P–D, P–R, R–D and R–R, are similar. This leads to the conclusion that users were unable to identify with the personal value prompt, and this is no longer taken into account in the analysis to follow.

The conclusion is that although tears contribute to fitness to a small extent, the model based on MP as the most important aspect of physical degradation to general library and archival users, as presented in Fig. 7, explains most variance in the data. This allows us to easily estimate the level of degradation at which users assess a document as not being fit for use (Fig. 8).
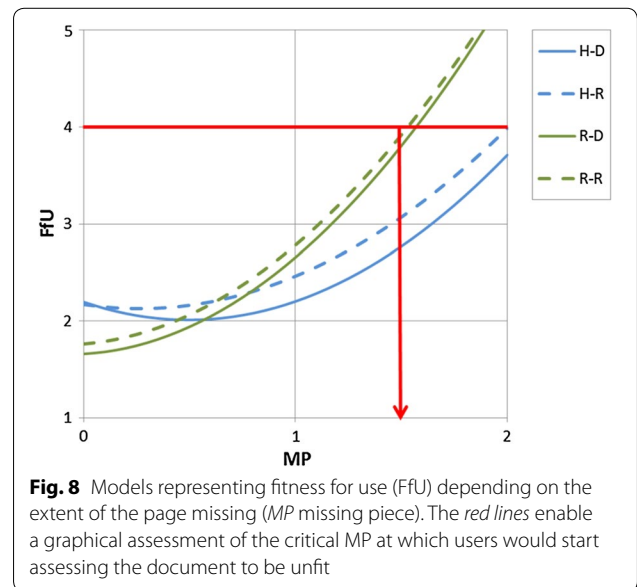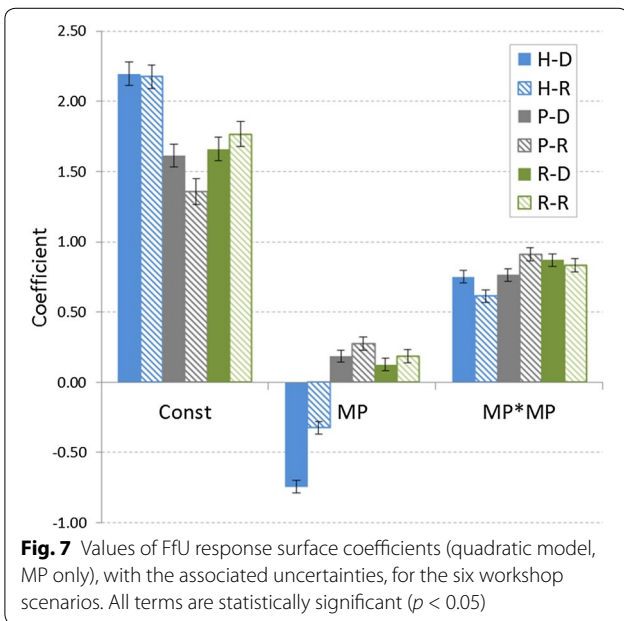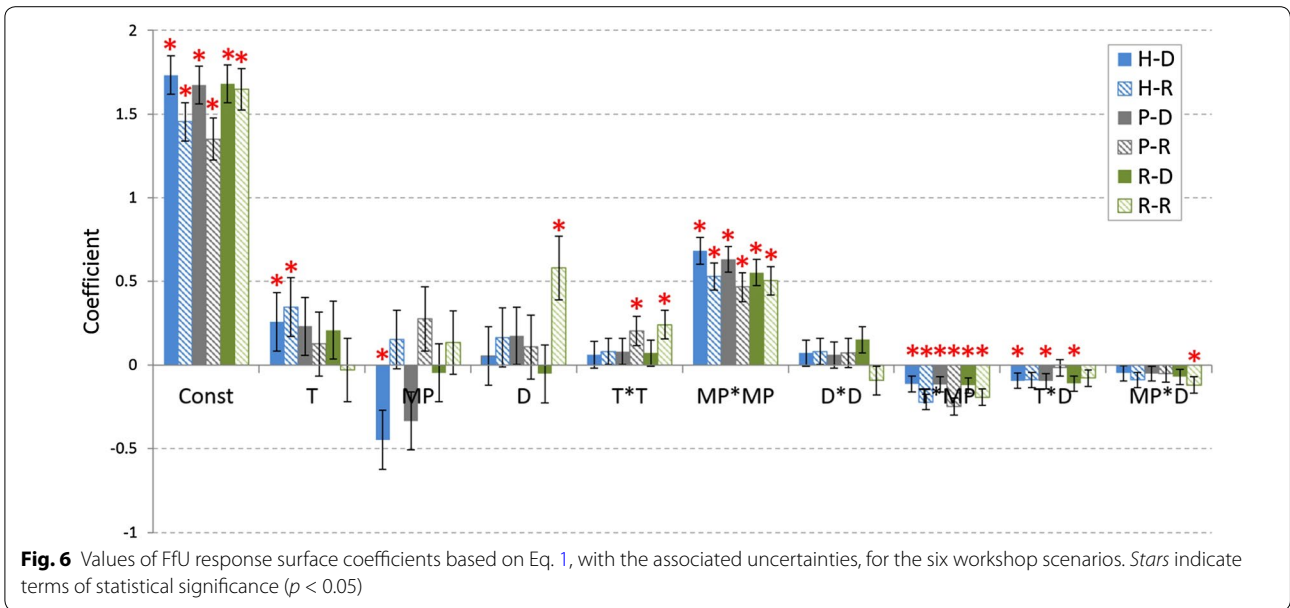
However, scenarios H–D and H–R show a different picture to the rest. Regardless of the size of the missing piece (the largest missing piece was ~1/6 of a page), assessors never consider a historic document to be unfit, regardless of the context of use, i.e. display or reading. In the context of collection management, the physical state of a document of historic significance is not considered to be particularly important to its fitness.

On the other hand, if the document was considered to be a random archival document, the point at which users assess it as unfit for use can be defined, and is estimated graphically in Fig. 8. This point is the same, regardless of the context of use, i.e. display or reading.

With a missing corner of a page not containing any text, most users were of the opinion that the document was 'Quite good'. However, if the text was evidently affected at

Table 5 R^2 values for the different surface response models, for the 6 workshop scenarios

Scenario	FfU model	
	Equation 1	MP only
H–D	0.3817	0.3562
H–R	0.4533	0.4029
P–D	0.3907	0.3693
P–R	0.4536	0.3836
R–D	0.4191	0.4017
R–R	0.4148	0.3569

Strlič *et al. Herit Sci* (2015) 3:33

Page 10 of 12

Fig. 6 Values of FfU response surface coefficients based on Eq. 1, with the associated uncertainties, for the six workshop scenarios. *Stars* indicate terms of statistical significance ($p < 0.05$)

Fig. 7 Values of FfU response surface coefficients (quadratic model, MP only), with the associated uncertainties, for the six workshop scenarios. All terms are statistically significant ($p < 0.05$)

Fig. 8 Models representing fitness for use (FfU) depending on the extent of the page missing (*MP* missing piece). The *red lines* enable a graphical assessment of the critical MP at which users would start assessing the document to be unfit

which the state of a document became unacceptable (i.e. damage [1]). At this point, the value of MP is ~1.5 (Fig. 8). While on the basis of this experiment it is not possible to estimate how much missing text this level of deterioration represents, the estimation that any missing text will significantly affect the level of satisfaction is meaningful.

Conclusions

We explored attitudes of users to visually observable material change in paper documents, with the aim of understanding what extent of degradation is characteristic of objects that are no longer fit for use in the specific contexts of display or reading with handling, and when different types of values are elicited. In collection management practice, such extent of degradation could be seen as the end of lifetime for the particular mode of general access, and environmental and access practices could be adapted to optimise the lifetime as required. By taking the views of users into account we also make collection management more publically accountable.

Strlič *et al. Herit Sci* (2015) 3:33

Page 11 of 12

Fitness for use was assessed in user workshops, where 331 participants were confronted with a number of differently distressed objects. The aspects of degradation explored were those of interest to practical collection management: gradually accumulating and possibly preventable or, at least, slowed down. This ensures that the assessed fitness thresholds are of practical significance and applicability. Fitness is affected by a number of aspects, such as:

- The values, reflected in the attitudes of assessors to the objects.
  If an archival object is perceived to be of historical value, even large missing pieces do not make it unfit for some uses, as opposed to those documents for which no value is elicited. It is questionable, however, whether the value modalities can be taken into account in the management of large collections of objects of similar significance due to resource issues.
- The aspects of degradation studied.
  For archival documents, changes in colour and tears contribute to the overall assessment to a minor extent, whereas missing pieces contribute most.
- The purpose, or context of object use.
  There are statistically significant differences between how fitness is assessed for archival documents that are intended to be displayed and for those intended to be read. In the latter case, aspects of mechanical degradation (tears, missing pieces) are more important.
- The document's information content.
  In the case of documents for which no particular value is elicited, obviously missing text in a document leads to that document being assessed as unfit for use.

Having defined the threshold fitness for use for paper-based documents, and the long-term planning horizon in which this might be acceptable, a dose–response function is required that enables us to calculate the accumulation of tears and missing pieces as a consequence of frequency of use, as well as loss of strength as a consequence of natural ageing. We will explore this in Part II of this series of papers.

Abbreviations
CCD: central composite design; DOE: design of experiments; FfU: fitness for use

Types of physical deterioration
T: tears; MP: missing pieces; D: discoloration

Workshop types
H–R: workshop scenario assuming the object being of Historical (H) value, assuming the object being read while manually handled (R—Reading);
H–D: workshop scenario assuming the object being of Historical (H) value, assuming the object being read from a distance (D—Display);
P–R: workshop scenario assuming the object being of Personal (P) value, assuming the object being read while manually handled (R—Reading);
P–D: workshop scenario assuming the object being of Personal (P) value, assuming the object being read from a distance (D—Display);
R–R: workshop scenario assuming the object being of no pre-assigned, Random (R) value, assuming the object being read while manually handled (R—Reading);
R–D: workshop scenario assuming the object being of no pre-assigned, Random (R) value, assuming the object being read from a distance (D—Display).

Author details
[1] Institute for Sustainable Heritage, University College London, London, UK. [2] The National Archives, Kew, Richmond, Surrey, UK. [3] University of East Anglia, Norwich, UK. [4] English Heritage, London, UK. [5] Library of Congress, Washington DC, USA. [6] Nationaal Archief, The Hague, The Netherlands.

References
1. Strlič M, Thickett D, Taylor J, Cassar M. Damage functions in heritage science. Stud Conserv. 2013;58:80–7.
2. Dillon C, Lindsay W, Taylor J, Fouseki K, Bell N, Strlič M. Collections demography: stakeholders' views on the lifetime of collections. In: Ashley-Smith J, Burmester A, Eibl M, editors. Climate for collections conference, Munich, Doerner Institut, 7–9 November 2012, Postprints. London: Archetype; 2013. p. 45–58.
3. Robb JE. Quantitative assessment of stakeholder attitudes to geological collections for improved collections management, MRes Thesis, UCL Centre for Sustainable Heritage. 2012.

Strlič *et al. Herit Sci* (2015) 3:33

Page 12 of 12

4. Lindsay W. Time perspectives: what 'the future' means to museum professionals in collections-care. The Conservator. 2005;29:51–61.

5. Jones S, Holden J. It is a material world: caring for the public realm. London: DEMOS; 2008.

6. Clark K. From significance to sustainability. In: Clark K, editor. Capturing the public value of heritage—the proceedings of the London conference, 25–26 January 2006. Swindon: English Heritage; 2006. p. 59–60.

7. Mason R. Assessing values in conservation planning: methodological issues and choices. In: De la Torre M, editor. Assessing the values of cultural heritage. Los Angeles: J. Paul Getty Trust; 2002. p. 5–30.

8. Cossons N. Capturing the values of places—opening remarks. In: Clark K, editor. Capturing the public value of heritage—the proceedings of the london conference, 25–26 January 2006. Swindon: English Heritage; 2006. p. 70.

9. Throsby D. The value of cultural heritage: what can economics tell us? In: Clark K, editor. Capturing the public value of heritage—the proceedings of the London conference, 25–26 January 2006. Swindon: English Heritage; 2006. p. 40–3.

10. Jowell T. From consultation to conversation: the challenge of better places to live. In: Clark K, editor. Capturing the public value of heritage—the proceedings of the London conference, 25–26 January 2006. Swindon: English Heritage; 2006. p. 7–13.

11. Langford M. Colour and conventions. In: European colour photography. London: The Photographer's Gallery Ltd. 1978.

12. Engeldrum PG. A theory of image quality: the image quality circle. J Imaging Sci Technol. 2003;48:447–57.

13. Baird JC, Noma E. Fundamentals of scaling and psychophysics. Wiley Series in Behavior. New York: Wiley; 1978.

14. Ehrenstein WH, Ehrenstein A. Pyschophysical methods. In: Windhorst U, Johansson H, editors. Modern techniques in neuroscience research. Berlin: Springer; 1999.

15. British Standards Institution. BS ISO 20462-1:2005(E) Photography—psychophysical experimental methods for estimating image quality. 2006.

16. Bradley JS, Reich R, Norcross SG. A just noticeable difference in c-50 for speech. Appl Acoustics. 1999;58:99–108.

17. Ashley-Smith J, Derbyshire A, Pretzel B. The continuing development of a practical lighting policy for works of art on paper and other object types at the Victoria and Albert museum. In: Bridgland J, editor. ICOM-CC 13th Triennial meeting, Rio de Janeiro. London: James & James Ltd.; 2002. p. 3–8.

18. Crawford BH. Just perceptible colour difference in relation to level of illumination. Stud Conserv. 1973;18:159–66.

19. Michalski S. The lighting decision. In: Fabric of an exhibition, preprints of textile symposium. Ottawa: Canadian Conservation Institute. 1997. pp. 97–104.

20. Brokerhof AW, Reuss M, MacKinnon F, Ligterink F, Neevel H, Fekrsanati F, Scott G. Optimum access at minimum risk: the dilemma of displaying Japanese woodblock prints. In: Conservation and Access: 2008 IIC Congress, London. 2008.

21. Derbyshire A, Ashley-Smith J. A proposed practical lighting policy for works of art on paper at the V&A. In: Bridgland J, editor. ICOM Committee for Conservation 12th Triennial Meeting, volume 1, Lyon. London: James & James Ltd.; 1999. p. 38–41.

22. Richardson C, Saunders D. Acceptable light damage: a preliminary investigation. Stud Conserv. 2007;52:177–87.

23. Ehrenstein WH, Ehrenstein A. Pyschophysical methods. In: Windhorst U, Johansson H, editors. Modern techniques in neuroscience research. Berlin: Springer; 1999.

24. British Standards Institution: BS ISO 20462-1:2005(E) Photography—psychophysical experimental methods for estimating image quality. 2006.

25. British Standards Institution: BS ISO 20462-2:2005(E) Photography—psychophysical experimental methods for estimating image quality. 2006.

26. Box GEP, Hunter JS, Hunter WG. Statistics for experimenters : design, innovation, and discovery. 2nd ed. Hoboken NJ: Wiley; 2005.

27. Yamini Y, Saleh A, Khajeh M. Orthogonal array design for the optimization of supercritical carbon dioxide extraction of platinum(IV) and rhenium(VII) from a solid matrix using cyanex 301. Separ Purif Technol. 2008;61:109–14.

28. Vanaja K, Shobha Rani RH. Design of experiments: concept and applications of Plackett Burman design. Clin Res Regul Aff. 2007;24:1–23.

29. Cohen J. A power primer. Psychol Bull. 1992;112:155–9.