IJDC | *General Article*

Choose Your Own Research Data Management Guidance

Alex Ball
University of Bath

Kellie Snow
University of Bristol

Peter Obee
Cardiff University

Greg Simpson
University of Exeter

Abstract

The GW4 Research Data Services Group has developed a Research Data Management Triage Tool to help researchers find answers quickly to the more common research data queries, and direct them to appropriate guidance and sources of advice for more complex queries. The tool takes the form of an interactive web page that asks users questions and updates itself in response. The conversational and dynamic way the tool progresses is similar to the behaviour of text adventures, which are a genre of interactive fiction; this is one of the oldest forms of computer game and was also popular in print form in, for example, the *Choose Your Own Adventure* and *Fighting Fantasy* series of books. In fact, the tool was written using interactive fiction software. It was tested with staff and students at the four UK universities within the GW4 collaboration.

Received 23 January 2017 | *Revision received* 6 June 2017 | *Accepted* 20 June 2017

Correspondence should be addressed to Alexander Ball, The Library, University of Bath, Claverton Down, Bath BA2 7AY. Email: a.j.ball@bath.ac.uk

An earlier version of this paper was presented at the 12th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

 doi:10.2218/ijdc.v12i1.494

Introduction

One of the complexities of supporting researchers in managing their data is that there is rarely a straightforward answer to any given question. So much depends on the context: not just the researcher's institution but their funding source, research domain, the type of data with which they are working, their project role, their external collaborators (if applicable), contractual arrangements, and so on. When it comes to writing guidance for researchers, therefore, the language can quickly become a maze of caveats and conditional clauses. It is hard to express the necessary information in a clear and concise way, and even harder for researchers to navigate and understand it. A possible strategy for dealing with this is to provide minimal guidance and instead rely on the provision of an advisory service; in this way, the supporter can have a conversation with the researcher and, having understood the context of their research, provide them with advice tailored to suit. This quality of service is highly desirable, but there is a limit to how far it can scale. At times of peak demand, it is better if simpler queries can be dealt with through guidance, with the advisory service dealing with more complex cases.

This issue was discussed at a meeting of the GW4 Research Data Services Group. GW4 is a collaboration between the University of Bath, the University of Bristol, Cardiff University and the University of Exeter;[1] the Research Data Services Group is one of a number of groups that facilitate co-operation, co-ordination, and the sharing of good practice between the four institutions. The group felt that what was needed was a form of interactive guidance that could, to a limited extent, mimic the conversational approach outlined above, and either provide straightforward answers tailored to the context or, on reaching its own limitations, refer the user on to the most appropriate sources of advice or detailed guidance.

It occurred to the group that this more conversational and interactive approach to text is a defining feature of *interactive fiction*. This term refers to a form of game or story in which the player takes the role of the point-of-view character in an unfolding textual narrative, and by directing the character's actions they affect how the story develops (Montfort, 2004). Among the group there was some experience in using dedicated interactive fiction authoring tools, and so a small working group was set up to take forward the idea of using them to develop a Research Data Management Triage Tool.

Background

There is a long history of using characteristic elements of games in serious settings to encourage uptake and engagement. The most familiar examples come from the commercial sector, such as loyalty points schemes where customers accrue points that may be redeemed against goods or services, or trigger preferential treatment when they reach a certain level. There are, however, examples of these techniques being used in higher education and research.

Such examples can be put on a spectrum according to how extensively game elements have been applied. At the minimal end of the spectrum, some Citizen Science

[1] GW4: http://gw4.ac.uk/

projects provide leader boards that introduce a sense of competition among contributors; SETI@Home's Top Participants list is an example of this.[2] Moving along the spectrum, online learning modules, such as those developed as part of the MANTRA course, include puzzles and quizzes to enable participants to demonstrate their understanding.[3] At the far end of the spectrum are full games whose primary purpose is something other than entertainment, known as *serious games* (Deterding, Dixon, Khaled & Nacke, 2011). Examples include Foldit, a game in which players compete to find the optimal way to fold a protein, and thereby predict how it would fold in reality (Cooper et al., 2010). The Grenoble Ecole de Management developed 'Game of Deans' to help teams conceive and develop ideas for HE services. Since 2014 the Jussieu Inter-University Science Library of the Sorbonne Universities has been running 'Murder Party' games that provide a more imaginative form of library induction (Swiatek, 2015).

The Triage Tool idea sits at the minimal end of this spectrum, since it is using some text adventure paradigms but without any sense of winning or losing; it is *gamified* guidance rather than a serious game. There is some evidence to suggest that using gamification in teaching and learning leads to improved results, with the caveat that it should be considered as an addition rather than a replacement for traditional techniques (van Meegen & Limpens, 2010). Thus the group was keen to position the Triage Tool as an additional resource for researchers, providing an alternative route to accessing information and by no means a substitute for existing websites or advice services.

Developing the Tool

Method of Interaction

Development of the tool began in earnest in late April 2016. One of the first decisions to be made was how the user should interact with the tool. In the sphere of interactive fiction, there are two main ways the player can interact with the story. In *choice* games, the user is asked to choose one of several options in order to proceed. This type of game was used in the *Choose Your Own Adventure* and *Fighting Fantasy* series of gamebooks. In *parser* games, the user interacts by typing in commands that the game engine interprets. This mechanism was used in many early computer games, such as *Adventureland* and the *Zork* series. The strengths and weaknesses of these two styles derive respectively from the fact that with choice games, all the available options are laid out explicitly on the screen, while with parser games, the options are hidden and must be guessed.

For the Triage Tool, the parser approach would allow more topics to be covered, and allow guidance to be accessed without having to navigate through menus. On the other hand, there is greater potential for frustration since the user has first to guess what topics might be covered, and second to express their query in a way the parser can understand. Parser games are also harder to write since the author must anticipate all the various commands the user might issue: not only the requested topics but all the multifarious ways in which they might be expressed.

Conversely, choice games are limited by the number of options that can reasonably

[2] SETI@Home Top Participants list: http://setiathome.berkeley.edu/top_users.php
[3] MANTRA: http://datalib.edina.ac.uk/mantra/

appear together on a screen, and the number of selections a user would be willing to make in order to get to an answer. However, there is a much shallower learning curve to using them, since the user need only point and click in order to interact. Such games are correspondingly easier to write since the author controls the available responses and can plan the effect of each one in turn.

On reflection, the group decided to use a choice-based approach. Since the tool was not intended to be a comprehensive advice service, it was felt that the ease of use and development afforded by a choice-based text would be worth the sacrifice of the potential richness of something parser based.

Development Environment

Having decided on the style of interaction, the group reviewed the various systems available for authoring such games and narrowed the field to a shortlist of two: Twine and Squiffy. Twine was first released in 2009 and has established itself as one of the most popular systems for choice-based games.[4] Squiffy was first released in 2014, and was developed to a state of relative stability over the following 17 months.[5] The choice between them was made on the basis of four criteria: collaboration, ease of installation, ease of use, and game play characteristics.

Collaboration An important consideration was that the tool would be developed jointly by the GW4 partner institutions. The group needed a system that compiled games from source code, rather than an opaque binary file, and where changes from each partner could easily be merged into the master copy. In this respect, Squiffy had the advantage, since it compiles transparently from a source file that uses user-generated internal identifiers and a Markdown-like syntax.[6]

In contrast, Twine 2 discourages direct editing of the source code; authors instead use dedicated authoring software which saves to an SGML file. While that file can be exported, shared and imported, Twine assigns sequential numeric IDs to passages; this means that if two people work on a game at once, their versions will have conflicting IDs. This makes merging the two versions non-trivial.

That being said, there is an unofficial command-line tool, Twee2, that supports a more portable version of Twine 2 code comparable to that of Squiffy.[7]

Ease of installation For the purpose of sustainability, it was also important that any of the partner institutions could compile the source code to a working Web page. On this criterion, Squiffy and Twine were equally suitable: the editing applications for both can be used online or run locally without installation. The aforementioned Twee2 variant requires a local installation of the Ruby programming language and was therefore problematic on locked-down university PCs.

Ease of use Another factor relevant for sustainability was the learning curve for using the source code language, since the responsibility for maintaining the tool would lie

[4] Twine: http://twinery.org/
[5] Squiffy: http://docs.textadventures.co.uk/squiffy/
[6] Markdown: https://daringfireball.net/projects/markdown/
[7] Twee2: http://twee2.danq.me/

with non-programmers. Here again there was little to choose between Twine and Squiffy, although Squiffy appeared to be slightly simpler at the expense of some functionality.

Game play The game play experience provided by the two systems was very similar; indeed, there were only a couple of notable differences. In Squiffy games, progress is saved automatically in a browser cookie, so if the player leaves the page and returns later, they pick up where they left off. In Twine games, any reload of the page causes the game to return to the start, though players can manually save and resume progress.

The other main difference is that Twine allows players to undo and redo their decisions, while Squiffy does not.

On balance the group decided to use Squiffy, on the basis that it could be used without having to compromise on any of the above criteria, although an undo function could have been useful.

For the collaborative version control environment, the group looked for an external service rather than an institutional one to ensure equitable access to the code by all partners. GitLab was selected since it allowed repositories to be private initially and opened up at a later point.[8]

Planning and Writing the Content

Having decided on the software to use and set up a collaboration environment, the group sketched out a structure for the Triage Tool. It had been decided at the outset that the tool would be directed at postgraduate researchers. Generally speaking the level of research data management information required by this group is at the introductory level, and therefore requires a less discipline- or institution-specific focus. This would aid writing the content of the tool across multiple institutions. A need had also been identified by all four partners for more guidance specifically tailored to this group, and it was anticipated the text adventure format would work well for a student audience wishing to 'explore' the topic.

The idea was to provide broad topic areas on the first screen; on selecting an area, the user would then be shown a list of questions that the tool could answer on that topic. Some questions would lead to answers or referrals to other sites, others to further questions. The group identified frequently asked questions concerning research data management and grouped them into five topical areas: Data Management Plans, storing data, organising data, documenting data, and sharing data.

Writing the tool was completed in two phases, with a review after the first phase to steer activity in the second. Two areas were selected for development in the first phase: organising data and documenting data. These were chosen as having least variation in guidance across the four institutions. Bristol developed the former and Bath developed the latter.

The initial review of the tool was conducted within the Research Data Services Group, but by those outside the working group, in July 2016. The key items of feedback were as follows:

[8] GitLab: https://about.gitlab.com/

 doi:10.2218/ijdc.v12i1.494

Figure 1. Initial screen of version 0.2 of the Research Data Management Triage Tool

- The usual behaviour for Squiffy was to add new text to the end of the page, resulting in a long transcript. This was felt to be messy and confusing, so it was decided to clear the screen periodically instead.
- It was felt that the level of detailed information provided by the tool should be reduced to lessen the maintenance burden.
- The level of interactivity should be increased to further differentiate the tool from existing Web guidance.
- It was felt that people should be asked for their institution and funder only at the point where the guidance diverged, rather than at the start.

Having taken this feedback on board, the existing content was revised, and the remaining sections allocated to working group members. As each section was completed from the perspective of the first member's institution, the remaining members reviewed the content and contributed their own institution-specific guidance.

A full prototype of the tool was completed in early January 2017, at which point GW4 branding was applied (see Figure 1).

The way in which the prototype tool behaves is as outlined above: the tool asks the user questions and lists possible responses, each encoded as a link. Some links lead to further screens, others replace the response with relevant information. Links are also embedded within some of the answer text; on selection, they insert more detailed information on the topic adjacent to the link, rather than at the bottom of the page.

When a user comes to guidance that varies according to their funder or institution, they are presented with a list from which to select the relevant value. The tool remembers these selections using internal variables, so that if the users navigate to a different question they do not have to choose again.

Each screen has a ‘restart’ link at the top. This returns the user to the first screen and clears any internal variables set. In addition, any screen that does not simply link to further screens has one or two links at the end prefaced with ‘Do you have any other questions about. . . ?’ These allow the user to explore the other questions answered within the current topical area, or select a new topical area, by returning to previous screens. In contrast to the ‘restart’ link, no internal variables are cleared.

User Testing

Some preliminary user testing was held in late January 2017 with staff and postgraduate research students at the University of Bath. Participants were asked to use the tool to find the answers to research data management questions; they were invited to choose their own questions but sample ones were provided as a fallback. The tester observed their progress and noted down any points at which the tool surprised, confused or frustrated the participant.

After 10 to 15 minutes using the tool, participants were asked four questions:

1. Which aspects of the tool did you like or dislike?
2. Was the tool self-explanatory? Was there anything you wish you had known at the start?
3. Is there anything it doesn’t do that you would like it to do?
4. Would you use it again, or recommend it to a peer?

The results from this preliminary round of testing gave some consistent messages. On the positive side, all participants said they liked either the look and feel of the tool, or the way it gave clear and concise answers to questions. Most approved of the conversational way it led them to those answers. None found it confusing or hard to use.

On the negative side, almost all participants expressed a concern about the navigation. A few missed the links to previous screens at the bottom, and others did not realise how they differed from the ‘restart’ link. Many said they would prefer to see a breadcrumb trail or a ‘back’ or ‘undo’ button.

On a related point, users were sometimes surprised by the effect of some of the links. Within the same list, some links might be replaced with simple answers while others might lead off to a separate screen to give room for more complex answers. This confounded the expectations of users tackling their query in a non-linear way, that is, trying several avenues simultaneously. Several participants suggested that links to external resources should be opened in a new window, or that external links should be explicitly marked; they did approve, however, of the way the tool allowed them to resume their session when they returned to the page.

Two other common points were that the tool needed clearer links back to the institution’s research data support Web guidance or email address, and that a few questions did not sit intuitively within the topical areas on the initial screen.

Further user testing is planned to confirm these messages. There will then be a further round of revision to address the issues before the tool is launched.

Discussion

One of the issues that arose during the development of the tool was maintaining differentiation between it and the guidance pages already available on the respective institutions' websites. Since the tool is providing information on a web page, rather than acting as a serious game, there is a significant overlap of mission with the guidance pages; but there is clearly no benefit in having text from the website reproduced verbatim within the tool.

The fundamental difference in approach is that the tool provides interactive filtering of the information. The user selects various options, and is presented with a clear statement of the guidance that applies to them; they never see the irrelevant options or caveats. This helps to remove confusion and doubt, though it is of course incumbent on the tool authors to ensure that users are not presented with an over-simplification. A good example of this is in the tool's answer to 'What should my Data Access Statement look like?': after selecting a sequence of options, the user is presented with a single form of words they can copy out and complete with relevant details.

From this springs more nuanced aspects of the user experience. Instead of getting to the right topic through a menu structure, the user navigates by answering the tool's questions; this gives a more conversational feel to the process, which some users may prefer. If an issue has several facets under which it might be organised – for example, disposing of sensitive non-digital data – it is possible to lead the user to it by several routes quite naturally, without having to duplicate it at several points in a static hierarchy or favour a particular decomposition of the facets.

It is also possible to provide guidance at several levels of detail: the user reads a high level summary at first, and then digs into detailed points as they need to. At a coarse level, this can resemble an accordion menu, where clicking on a heading reveals the text beneath, but one can use this feature more subtly. For example, the tool mentions encryption as a way of protecting sensitive data; someone unfamiliar with encryption can select that word to insert additional sentences explaining it, while others can read on without hindrance from unwanted exposition.

This interactive filtering allows users to be presented with highly detailed information: since they do not see the detail that does not apply to them, they cannot get lost in or distracted by it. But just because they *can* be presented with such detail does not necessarily mean they *should*. Research data management is a fast-moving area and increasing the level of detail in the tool increases the burden of keeping the information up to date. Since any efforts in this direction are committed first and foremost to institutional Web guidance, the tool tends towards providing less detail and linking back to the existing guidance where possible.

Quite apart from the character of the Triage Tool itself, the group found benefit in the process of developing it collaboratively. When providing guidance at an institutional level it is all too easy to lose sight of what is general good practice and what is driven by local policy and infrastructure provision. Developing the tool encouraged members of the group to look again at that boundary. It also provided a useful starting point for sharing expertise and analysing possible gaps in guidance at each institution.

Conclusions and Next Steps

As mentioned above, the immediate next steps for the Triage Tool are to complete more extensive user testing across all four partners and adjust the tool to address the issues raised. Once all partners are satisfied, the tool will be published online and the respective institutions' research data management Web pages will link to it. At that point, the source code for the tool will be made available from the GW4 Research Data Services Group area of GitLab.[9]

For the purposes of sustainability, at least one member of staff at each institution has administrator rights over the source code repository. That member manages write access to the repository at their institution, and is able to help the other institutions restore their access should it become necessary. Each member is responsible for updating the guidance specific to their own institution as well as the generic guidance. One detail still to be determined is how the tool will be hosted, but once this is agreed, a release procedure will be put in place for compiling and publishing updates to the tool.

The Triage Tool provides a different way of accessing information, and it may not be to everyone's taste. Some people will prefer to navigate through a traditional hierarchy of pages and see the full, unfiltered information laid out for them, and find reassurance that they are not missing out on anything. However, the testing performed so far suggests that many find a clear and simple message more reassuring, and this a strength of the Triage Tool approach. The authors believe it serves a need, particularly for those looking for a quick answer to a quick question.

References

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., ... Foldit players. (2010). Predicting protein structures with a multiplayer online game. *Nature*, *466*, 756–760. doi:10.1038/nature09304

Deterding, S., Dixon, D., Khaled, R. & Nacke, L. (2011). From game design elements to gamefulness: Defining 'gamification'. In *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments (MindTrek '11)* (pp. 9–15). ACM. doi:10.1145/2181037.2181040

Montfort, N. (2004). *Twisty little passages: An approach to interactive fiction*. Cambridge, MA: MIT Press.

Swiatek, C. (2015). *Playing games at the library: Is this serious?* Paper presented at the 2015 LIBER Annual Conference. Retrieved January 23, 2017, from https://storify.com/cswiatek/playing-games-at-the-library

van Meegen, A. & Limpens, I. (2010). How serious do we need to be? Improving information literacy skills through gaming and interactive elements. *Liber Quarterly*, *20*(2), 270–288. doi:10.18352/lq.7993

[9] GW4 Research Data Services Group on GitLab: https://gitlab.com/gw4-rdm