

This is a fictional AHRC technical plan developed and commented on by a reviewer to highlight common pitfalls

## Technical Plan

### Section 1: Summary of Digital Outputs and Digital Technologies

The digital output of the project will be a database of historical recipes from throughout Britain from the 1600s to the first world war. Recipes will be geocoded based on project research into where the recipe originated and it will be possible to search for ingredients or styles of cooking based on region / town in combination with dates to show how the use of ingredients and methods changes over time and space. A highly innovative, vibrant and interactive public website will be created through which data will be plotted on historical map layers and on interactive visualisations, and an app will also be released. User interaction will be integral to the website and users will be able to post comments about recipes including images and video clips of dishes created by following the recipes. A content management system will also be developed for the project's researchers to use to record recipe and other associated data over the course of the project and for user submitted content to be managed.

### Comments

1. Mostly this section is ok, the main problems are things that are mentioned in it that are not then further expanded upon in the subsequent sections.
2. The source data isn't mentioned:
  - a. Where are all these recipes coming from?
  - b. Is the data already available in a digital format or will researchers be transcribing data themselves?
3. The level of access isn't mentioned – is the resource going to be freely available?
4. There is unnecessary hyperbole: 'A highly innovative, vibrant and interactive public website'

### Section 2: Technical Methodology

#### 2a: Standards and Formats

The project will use a variety of open and proprietary formats that will best suit the needs of the project's outcomes. These will be migrated to suitable open standards to facilitate preservation at the end of the project. Text will be transcribed as plain text with HTML markup and will take up roughly 500Mb of space. Images will be in the JPEG format and 5Gb of server space will be set aside for them. Video files will be MOV and 20Gb of space will be available for them. Visualisations will be SVG files. The map interface will be based around Google Maps. Web pages will follow current HTML and CSS standards.

## Comments

1. Information supplied is vague. 'A variety of open and proprietary formats that will best suit the needs of the project's outcomes': What exactly are all these formats and how do they best suit the needs of the project's outcomes? This section doesn't demonstrate that the project actually knows what it's talking about.
2. 'These will be migrated to suitable open standards to facilitate preservation at the end of the project': again, this sounds impressive but is too vague. What open standards? And why couldn't these open standards just be used from the start of the project?
3. Some file formats and standards are mentioned but the reason for their choice is not made clear, for example:
  - a. Why will video files be in the MOV format when this means users will need to install the Quicktime plugin for the files to be playable in their browser?
  - b. Why was Google Maps chosen over other maps interfaces?
4. Some statistics are given relating to the size of the data but these are not especially useful. Giving file sizes can be a good thing, but it would be more useful if these were given in combination with estimates of how many images and videos the project expects to receive, and how long they expect video clips to be.
5. The handling of user submitted data is given insufficient consideration throughout the plan. Only allowing users to upload videos in the MOV format is also very limiting and will generally exclude users who are not using Apple devices. Also, the project should ideally have considered allowing users to embed media they have uploaded elsewhere, e.g. YouTube and Flickr.
6. Information about the textual data is far too vague. The recipe data will have to be structured in some way and there is no information about how the project intends to do this. Will the data be stored in a relational database? Will recipes be marked up in XML?
7. The section does mention that text will be transcribed with HTML markup but this approach is not considered best practice. HTML should be used to mark up presentation not content, and if markup is to be used to denote ingredients then XML would be better suited. Alternatively the project could record ingredients in a relational database.
8. Stating that the textual material will take up 500Mb of space doesn't tell the reviewer much about the data. More statistics need to be provided, such as how many recipes will be stored, how many towns / regions are likely to be covered, how many different ingredients and cooking styles. A lot of this information will likely need to be estimates but providing such information (e.g. it is estimated that the project will record around 10,000 recipes featuring around 5,000 unique ingredients) shows what the scope of the project is and whether the technical approach is suitable for the estimated amount of data.
9. Section 1 states that an app will be produced but there is no information about this here (or in other sections). E.g. what platforms the app will be released on, whether it will be developed as a native app or using standard web technologies and then 'wrapped'.

10. Section 1 states that 'historical map layers' will be used but only Google Maps is discussed in this section. Where are these historical maps coming from? How will they be incorporated into the resource?
11. Stating that web pages will follow 'current HTML and CSS standards' may sound good, but it's not clear to the reviewer whether the person who wrote the plan knows what the current standards are. It is better to be more explicit, e.g. 'web pages will adhere to the HTML5 and CSS3 standards'.
12. Insufficient information is provided about the visualisations. What sort of visualisations will be created? How will they be created? Section 1 states that the visualisations will be 'interactive' so they presumably won't just be static SVG files but will change to reflect choices made by the user. Presumably some sort of existing visualisation library or package will be used by the developer to generate the visualisations and information about this should appear in this section (or the following section as the library could be considered software rather than a format or standard).

## 2b: Hardware and Software

The website and content management system will be hosted on LAMP servers based at project partner the University of Edinburgh who will supply the project with two virtual servers: a 'development' and a 'live' server. The resource will be developed using the Joomla content management framework. Images will be edited with Adobe Photoshop and videos with Final Cut Pro.

### Comments

1. The information provided about the server setup is absolutely fine.
2. The use of the Joomla framework is also fine, although some indication as to the version of the framework that will be used would have been good, as would some indication as to why this framework was chosen over other alternatives.
3. It would also have been useful to mention any of the Joomla extensions that the project will be using to create the resource, and whether these are free or cost money.
4. If the data is to be stored in a database then this should be mentioned here. The 'M' in 'LAMP' stands for 'MySQL' but it is unclear whether this database is actually going to be used by the project.
5. It is unclear why images will be edited as the only images mentioned are ones uploaded by users as part of comments (no mention is made of digitised pages of recipes being made available in addition to the text versions, for example). There doesn't really seem to be much justification for using Photoshop here as it's a very expensive piece of software.
6. The same is true about videos. Final Cut Pro costs a lot of money and there is no evidence that the project will be editing videos that are uploaded by users. There is no mention of the project creating their own videos either.

## **2c: Data Acquisition, Processing, Analysis and Use**

An initial project website will be set up by the developer during the first month of the project. This version of the website, along with project presences on social networking sites such as Facebook and Twitter, will be managed by the Co-I. During months 1-3 the developer, in collaboration with the rest of the project team, will create a scoping study for the content management system and the public website. S/he will work on a first version of the CMS during months 4-6, launching it during month 6 with further iterations which will introduce further functionality being made every few months over the three years of the project.

The 10 project RAs, who will have begun collating data on their laptops from fieldwork to libraries and archives around the UK in month 2, will receive training in the use of the CMS by the developer in month 6. Data uploaded to the CMS by the RAs will be analysed and processed by the developer to convert it into formats suitable for display on the project website, which the developer will be working on during years 2 and 3 of the project.

A 'beta' version of the online resource will be made available to selected users midway through year 2 of the project. This version will feature full access to the recipe records but limited search and browse functionality. The launch of the 'beta' version will coincide with the project symposium.

The developer will continue to refine the online resource, adding functionality such as the map interface and the visualisations throughout year 3 of the project. An official launch of the final resource that will be available to all users will take place in the final month of the project to coincide with the project conference, at which point users will be able to access the maps and visualisations and post comments on the recipes.

The online resource, the CMS, one IIIF server, and the Solr indexing system will be located on Linux web servers managed by University of Glasgow IT Services. The web servers will be backed up nightly to an Ultrium LTO2 unit located remotely from the server. The project archive will be stored on an Active Directory network, supported by 4 domain controllers, located in two separate 'server' rooms at each end of campus. Each server has a RAID disk subsystem and is backed up nightly to devolved backup systems. Both server rooms are protected by both UPS and generators. The backup system creates and maintains two copies of each system state backup which are held on near-line disk, on-site tape and off-site tape. 7 versions of each AD state are retained for 90 days. The data schema, system specification and procedures for data creation and management will be described in a detailed set of documents.

### **Comments**

1. There are some problems with the technical milestones in this section, including:
  - a. The project website will be set up by the developer in month 1, but the following section states that a developer may not be recruited until the end of month 1
  - b. 10 project RAs will be collating data in month 2 of the project but the content management system (CMS) won't be released until month 6. No indication is

given as to how the RAs will manage their data before the CMS is available, or if this data might be batch uploaded into the system once it is available. If this isn't possible then uploading 4-5 months of data from 10 people is going to be a time consuming process.

- c. User interaction (comments, images, videos) is mentioned as a fairly big thing for the project, but the systems for allowing this will only be available in the final month of the project. It is unclear who will manage user submissions, whether these will be need to be moderated, how this will work or what happens once the project ends
2. The creation of the app version isn't mentioned at all in this section, which is a major oversight
3. The workflow for getting data published seems rather muddled and inefficient. This section seems to suggest that data uploaded to the CMS will then need to be manually analysed and processed by the developer. If this really is the case then it seems like a very inefficient approach. If the data from 10 RAs needs to be manually converted by the developer there is a serious risk of a bottleneck developing. The developer shouldn't have to be involved in the process at all – it should be possible for the developer to build the CMS in such a way as to automate these tasks.
4. There is no information about how the geocoding will be handled and where this fits in with the workflow
5. Issues of monitoring and quality control are not addressed. The data appears to be passing through the developer but this does not appear to be for quality control purposes, and the developer is not the person who should be checking the quality of the data anyway. Ideally the PI or some sort other team member with detailed knowledge of the content should be performing some sort of quality checks on the data produced by the RA.
6. No information is given on the sorts of documentation that will accompany the resource.
7. Timescales for development appear to be feasible but you'd need to know exactly how much effort has been assigned to the developer to form a proper opinion on this (this information would be found elsewhere in the bid documentation). Plus, there is no indication as to the contents or complexities of the app, or when the developer would create this or how much effort has been assigned for its creation.
8. Evaluation sessions are not mentioned in this section. The 'beta' version will be launched in Year 2 but there's no indication that feedback from the users will be solicited or if any suggestions will be acted upon. Similarly, the final resource will launch in the last month of the project and no evaluation or testing sessions are mentioned here. It would appear that the website will be launched and the developer will then have no more involvement with the project. This is a very risky strategy as major problems with the resource may only be uncovered once users start to use it.
9. It would have been useful if the plan was a little more specific about how the developer will employ the standards and formats and the hardware and software in order to produce the CMS and the website. For example, more information about how the visualisations and the historical maps will be generated from the data should have been included here. What types of visualisations do the project intend to utilise?

10. It would have been useful if a possible workflow for managing the data could have been mapped out in this section (either as a diagram or as text). E.g. “The 10 RAs will be working at archives and libraries across the UK, discovering recipes in original sources (books, manuscripts, newspapers, letters) and creating records for these recipes through the CMS via their laptops. The CMS will incorporate a Google Maps interface to enable latitude and longitude values associated with the recipe to be found and stored as part of the record. The recipe record will be created in XML and accompanying metadata will be stored in a relational database. The CMS will provide a simple text editor through which the RA may enter the necessary XML information, including tags for ingredients and cooking methods, and a web form where metadata can be added. The RA may save and edit a record as they see fit until all of the necessary data has been compiled, at which point they may choose to send the record to the editor for approval and inclusion in the resource.”
11. A big paragraph is included about backups, but it looks like this has been copied and pasted from a different project. All the information relates to servers at Glasgow but Section 2b states that the technical infrastructure will be set up at Edinburgh. Also, the backup section introduces technologies not mentioned elsewhere in the document – IIF servers and Solr indexing. A reviewer would not be able to rely on the information found in this paragraph.

### **Section 3: Technical Support and Relevant Experience**

The PI and Co-I both have considerable previous technical experience. The project aims to recruit the developer before the start of the project but at the very latest s/he will start working by the end of month 1. The developer will be based at the University of Edinburgh and will work closely with other project partners at Edinburgh and the University of Glasgow.

The PI and Co-I have consulted widely with several other projects of similar aims and complexity in order to gain a better understanding of how technology can be best harnessed in order to make a truly groundbreaking digital resource. Advice on data management has been sought from the AHDS.

### **Comments**

1. The PI and Co-I may both have lots of experience but there is no evidence of this. Note however that evidence may be available the full bid documentation.
2. The project will need to recruit a developer at the start of the project and the plan infers that it might not be possible to find someone to start until the end of the first project month. This is not an ideal situation and there is a real risk that the project might not be able to find a suitable developer. This risk has not been considered in this section and it should be. What would happen if a developer cannot start until month 6 of the project? Or later? What will the 10 RAs do? How will this affect workflows and outcomes?
3. There is no mention of any technical support from the institutions involved in the project. Who will manage backups? Who is responsible for the project servers?



4. Most of the second paragraph is too vague to be meaningful. Who exactly have the PI and Co-I consulted with? What did they actually learn? It looks like this has been thrown into the plan just to sound important.
5. This section states that advice has been sought from the AHDS, but it is likely that this sentence has been pasted in from an earlier project as the AHDS has not existed since 2008.
6. No indication is given as to whether any other technical advice has been sought. Who has provided advice on using historical map layers? What about visualisations or app development? A lot of the information in the plan is rather vague and this would suggest that whoever wrote the plan doesn't necessarily have a deep understanding of the technologies involved.

## **Section 4: Preservation, Sustainability and Use**

### **4a: Preserving Your Data**

Upon completion of the project the digital outputs of the project will be migrated to open standards for preservation as discussed in Section 2a. Outputs will be tagged with appropriate metadata to facilitate their discoverability. Long-term preservation of digital data is a considerable challenge; however, how best to preserve digital data is not the focus of this project and other projects within our partner institutions are already making significant progress in how this issue can be addressed.

#### **Comments**

1. No indication is given as to how long the data will be preserved for (3 years following the end of the project is the minimum)
2. The 'open standards' supposedly discussed in Section 2a aren't actually documented so it's impossible to say whether they are suitable.
3. It's not clear exactly what digital outputs will be preserved and which won't
4. It's not clear where the data will be preserved or who will be responsible for it
5. The information about 'appropriate metadata' is too vague to be of any use. What does the project consider to be 'appropriate metadata'? A reviewer will have no idea.
6. The last sentence is completely irrelevant to the project
7. It would have been better if the project had considered depositing their data with an archive. The recipe data would probably have been accepted by the Oxford Text Archive or possibly the UK Data Archive.

### **4b: Ensuring Continued Access and Use of Your Digital Outputs**

The website will continue to be hosted by the University of Edinburgh beyond the end of the project and user comments and contributions will continue to be enabled. In order to minimise

the cost of sustaining the resource in the longer term the CMS will not be retained following the end of the project and the focus instead will be on the public facing website.

### Comments

1. Stating that the website will continue to be hosted is a good thing, but this section needs to state how long such an agreement is in place for (as with preservation the minimum is 3 years following the end of the project).
2. Does this agreement also extend to updating the website infrastructure as newer versions of Joomla become available and older ones are no longer supported? If no updates are made to Joomla then security risks may emerge. But if Joomla is updated then some parts of the code written for the project may break. This is where good technical documentation as should be discussed in Section 2c becomes invaluable.
3. User comments and uploads will still be enabled but this raises serious questions for sustainability. Presumably such uploads will need to be moderated, especially as the commenting feature will only become available in the last month of the project. There is no indication as to who will manage user comments and uploads once the project ends.
4. The last sentence states that the CMS will not be retained, with the implication being that this is a good thing for sustainability as there is less technology to be managed. In reality getting rid of the CMS would be a very bad move as it would make it difficult for the data to be managed. What happens if a recipe has to be taken down or an error needs to be corrected? Removing the CMS would make such updates much harder.
5. There is no indication as to who would make such updates if they are needed once the project comes to an end
6. Any rights issues are not mentioned in this section. Will users be able to reuse the project data? If so is it being published under a license of any sort, for example a Creative Commons license?