

FundFind

Exploring scholarly funding opportunities

Emanuil Evgeniev Tolev

April 1, 2013

Declaration of originality

In signing below, I confirm that:

- This submission is my own work, except where clearly indicated.
- I understand that there are severe penalties for plagiarism and other unfair practice, which can lead to loss of marks or even the withholding of a degree.
- I have read the sections on unfair practice in the Students' Examinations Handbook and the relevant sections of the current Student Handbook of the Department of Computer Science.
- I understand and agree to abide by the University's regulations governing these issues.

Signature

Date

Consent to share this work

In signing below, I hereby agree to this dissertation being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Signature

Date

Acknowledgements

I would like to thank my supervisor, Professor Reyer Zwiggelaar for his support and unique approach to the process of producing a major project.

Thanks to Dr. Richard Jensen, Dr. Joanne Walker and Professor Simon Cox for their valuable input in the formative stages of the project.

Furthermore, thanks to all Computer Science Department staff who have taught me over the last 4 years for their patience and being such an inspiration with their varied interests and opinions.

I would like to thank Cottage Labs LLP for introducing me to the Open Knowledge movement, as well as teaching me about most of the major technologies used in this project and agreeing to host a FundFind instance.

Thanks to the Open Knowledge Foundation for striking out and encouraging prototype projects, providing ample inspiration for FundFind.

I would like to thank Research Councils UK, a partnership of the seven UK Research Councils, for all the effort they have put into the Gateway to Research API, consolidating data from seven separate institutions and making it available under an Open license.

Abstract

FundFind is a web application which enables scholars, research development officers and post-graduates to share information about funding opportunities. Data about funding opportunities and funding organisations can be submitted, browsed and searched via a mobile-friendly web UI as well as a JSON API. FundFind stores data using an asynchronous approach, which enables the development of small independent modules for automatic funding data harvesting.

The project is a learning exercise, aiming to explore the field of scholarly funding, identifying potential problems with the open sharing of such data. Prototype technical solutions concretise the technical problems. Other types of problems are noted and discussed.

CONTENTS

1	Background & Objectives	1
1.1	Scholarship and funding	1
1.2	Audience and high-level requirements	2
1.2.1	Professionals looking for funding	2
1.2.2	Those looking to advertise funding opportunities.	3
1.3	Existing works	3
1.4	Scholarly funding opportunities	3
1.4.1	Funders	3
1.4.2	The value of Openness in scholarship and elsewhere	5
1.4.3	Scholars	5
1.5	The project	6
1.6	Existing works	6
2	Development Process	7
2.1	Introduction	7
2.1.1	Overview	7
2.1.2	Evolution of the Development Process	8
2.2	Modifications	9
2.3	Requirements	9
2.3.1	Requirements gathering	9
2.3.2	Final requirements	9
2.4	Design process	9
2.5	Technologies used	9
3	Design	10
3.1	Overall Architecture	10
3.2	Alternative Designs	10
3.3	Main Application	10
3.3.1	Core	10
3.3.2	DAO	10
3.3.3	Web	10
3.3.4	Importer	10
3.3.5	Config	10
3.3.6	Util	11
3.3.7	TweetBot	11
3.4	Harvesters	11
3.4.1	HarvesterBase	11
3.4.2	CliManualSubmitter	11
3.4.3	RssHarvesterBase	11
3.4.4	EmailHarvesterBase	11
3.4.5	EpsrcEmailHarvester	11
3.5	User Interface	11
3.6	API Design	11
4	Implementation	12

5	Testing	13
5.1	Overall Approach to Testing	13
5.2	Automated Testing	13
5.2.1	Unit Tests	13
5.2.2	Functional / Integration Tests	13
5.2.3	User Interface Testing	13
5.2.4	Stress Testing	13
5.3	User Testing	13
6	Evaluation	14
6.1	Approaching the field of scholarly funding	14
6.2	Whose requirements?	14
6.3	Collecting information from disparate sources - technological suitability	14
6.4	User needs	14
6.5	Exploring scholarly funding	14
6.6	In retrospect	14
	Appendices	15
A	Third-Party Code and Libraries	16
B	Code samples	17
2.1	Random Number Generator	17
	Annotated Bibliography	20

LIST OF FIGURES

LIST OF TABLES

1.1	Notes on case representation. Based on Table 1 from (?)	4
-----	---	---

Chapter 1

Background & Objectives

A note on “scholarship” - the word and its derivatives are used throughout this text to mean all types of academic work, due to “science” being sometimes perceived as exclusionary to the Arts and Humanities fields.

1.1 Scholarship and funding

Scholars need money to conduct research. To be precise, research projects need *additional* money (on top of the investigators’ salaries) in order to cover material expenses, travel costs, hiring Research Assistants or providing studentships to PhD candidates and other expenses.

In other words, established scholars need to look for funding for aims which may be hard to define, for amounts of money ranging from thousands to millions.

Postgraduate-level scholars (candidates, postgraduates, post-doctoral researchers) are also usually looking for a way to fund the next stage of their scholarly career. (Note: undergraduate-level funding was not considered in this project.)

The quintessential problem is that the information about available funding is published in many different places. Usually each funding organisation publishes its own calls for proposals and announces its own studentships on its own website(s) and mailing list(s).

Keeping up with these sources of information becomes quite difficult - so much so, that it takes a significant portion of research development officers’ time. Worse still, this is a significant barrier for those looking into a career in academia - the fragmented information can make it difficult to build a mental picture of what the funding in one’s chosen field “looks like”. This leads to a feeling of opaqueness instead of transparency and being overwhelmed - instead of having a clear idea of what it takes to succeed in a chosen field. Sometimes it even leads to accusations of cronyism in distributing funding (private conversation as well as (? ?)).

These problems seem to stem from the fact that the distribution of scholarly funding seems to have evolved over time from a series of changes to more general money distribution in society. This is actually a well-known problem that the software development field has had to deal with - if multiple stakeholders with distinctive interests try to influence the development of a software system, changing its requirements over time, the results can be disastrous.

There is no easy way to track money - where it gets allocated to, what fields (or even topics) within scholarship get funded more than others. Up until November 2012, there was no publicly accessible *centralised* detailed account of historic funding information in the United Kingdom. The Gateway to Research project being implemented by Research Councils UK finally gives a partial (RCUK-specific) overview of historical data.

However, there is still no mechanism of getting *present, current* data about scholarly funding.

Something which allows searching the (vastly) different, *currently available* streams of funding within academia was thus identified as potentially useful by various people interested in the subject:

1.2 Audience and high-level requirements

The potential audience of this project was discovered to be quite varied. It seems multiple stakeholders stand to benefit from more transparent scholarly funding dissemination.

The requirements or potential uses that these users might want to put FundFind to are closely tied to the benefits that come with opening up funding data.

1.2.1 Professionals looking for funding

1.2.1.1 Accomplished scholars

- **Who:** Academics - lecturers, professors
- **Aims:** Wish to fund future projects
- **Benefits from opening up funding data:** Easier access and search across funding institutions through tools which aggregate funding data; Transparency and accountability of their funding bodies.
- **Possible requirements from FundFind:** Aforementioned cross-funder search; Sharing opportunities with colleagues.

1.2.1.2 Junior academics

- **Who:** Postgraduate candidates, postgraduate students, academics in post-doctoral posts (e.g. Research Associates)
- **Aims:** Wish to find suitable positions to move on with their academic career, e.g. a postgraduate with a PhD will be looking for a post-doctoral position
- **Benefits from opening up funding data:** (More) easily find and compare funding sources
- **Possible requirements from FundFind:** Find funding. To quote a postgraduate candidate: "I want to do a taught MSc in Astrophysics but I can't afford to self fund. Where can I go and who will pay me?"

Share funding information with peers (e.g. if a candidate for a Computer Vision PhD finds a good PhD studentship call for applications in Bioinformatics, they may well want to forward it to a colleague).

1.2.1.3 Research development officers

- **Who:** Professionals who find more funding opportunities for scholars
- **Aims:** Wish to record the opportunities they find somewhere and share them with scholars. Also may want to access a list of all opportunities *they* have found over time, so that they can report to their line managers more easily.

- **Benefits from opening up funding data:** Not as many as scholars, unless they are happy to share the opportunities they have found with research officers from other institutions (which they do not always want). However, a system which allows sharing of recorded opportunities with a select subset of users (such as all within the officer's institution) would be welcome as this can allow them to record a funding opportunity once, share it with "their" scholars and include it in reports.
- **Possible requirements from FundFind:** Find funding. Share funding with subset of users. See information submitted per user (themselves).

1.2.2 Those looking to advertise funding opportunities.

Funding Stream and Programme managers at various institutions such as:

1.2.2.1 Focusing on certain audience groups

Satisfying the major requirements of all these users is not a suitable scope for this project (it may be for a follow-up project).

Therefore, this project will focus on the first group of users. The main reasons are time constraints and ease of access to such users. Searching for opportunities can be significantly harder than submitting them. In a way, providing data to academic users will pave the way for providing data to a more general public (third group) for re-use and analysis.

The project will nonetheless provide basic information submission ("digestion") capabilities and basic general raw data access via an open machine-ready interface ("API") to satisfy some of the requirements of the second and third groups. The author belongs to the third group. The Open Knowledge Foundation (28) and Cottage Labs LLP (8) will be briefly consulted on the completed project. Cottage Labs has stated that the project will be valuable to Open Knowledge if executed well and is happy to provide feedback.

1.3 Existing works

FundFind is largely based on IDFind (12), another project in the Open Knowledge field, developed by Cottage Labs LLP.

1.4 Scholarly funding opportunities

This project deals with an important bit of the infrastructure of scholarship - namely, the money it relies upon. "Scholarship" in this case (and throughout this document) refers to the human endeavours in the fields of the Humanities (e.g. arts), Social Sciences (e.g. law) and Science (e.g. biology, computer science).

1.4.1 Funders

There is no easy way for people who wish to fund scholarly endeavours to advertise their funding to a lot of scholars at once, getting a higher chance of the "perfect" researcher/problem fit.

Certain ad-hoc channels have evolved around this problem - for example, scholars interested in a funder's sphere can subscribe to their e-mail list. There are problems with this - the most obvious one is scholars having to subscribe to all the funders they might be interested in. While this sort

Table 1.1: Notes on case representation. Based on Table 1 from (?)

Potential users	Their aims	Benefits of opening up funding data
Research Councils in the UK	Wish to attract the best scholars in their own field, but also look to establish new cross-disciplinary projects	Already committed to opening up historical funding data, would like to advertise current funding data more widely.
JISC (21) (digital infrastructure in education/research)	Wish to attract talent to work on scholarly infrastructure in the UK, i.e. give their aims and projects more exposure. Scholars rarely choose scholarship infrastructure as their main field.	Have always wanted wider reach, as well as vigorously supported Open Data. Have a lot of funding opportunities from minor to large, would like to advertise them better.
Various charitable Foundations, Institutes, Endowments and so on which fund scholarly activities (e.g. the Wellcome Trust (58), Shuttleworth Foundation (39), Knight Foundation (22))	Want the best scholars competing for their money.	Need to reach more scholars to build up an ever-improving pool of applicants (in terms of quality). Also, as private institutions, would like to demonstrate greater transparency with their funding dissemination and allocation.
Commercial organisations which perform R&D activities	Simply need quality applicants when using talent outside the organisation	Might make it significantly easier to reach a lot of good scholars. Such organisations may only need external help with R&D intermittently and thus may not have an audience of scholars following their every move.
There is a group which has markedly different needs from the first two groups - people interested in analysing how scholarly money is allocated and/or spent		
Software developers, journalists and others interested in Open Knowledge and visualising data for the purposes of transparency or advancing the digital economy (28) The author is included in this group.	Want transparency and accountability. May want to visualise how scholarship money is being spent or otherwise help society engage with scholarly spending.	Open Data essential for re-use for their aims. Usually do not have first hand information about what happens behind closed doors when funding is distributed and used by universities. Think that this is strange, since often both funder and fundee are public

of works for national funding bodies like the UK Research Councils, there are many more private funders - various non-profit and/or Non-Governmental Organisations and even private companies. A cross-disciplinary funding opportunity would have to be advertised across multiple channels, which would take time/effort on the part of funders therefore making such opportunities costlier to set up.

It also fails fantastically when it comes to the globalisation of scholarship. There are only two ways to get to know about funds coming from the other side of the world:

1. directly: be subscribed to that particular funder's news (the number of funders has not been estimated globally to the best of the author's knowledge, not to mention the huge variety of forms which news feeds take - from RSS through e-mail to just publishing on a website somewhere).
2. indirectly: hear about the opportunity from others. "Others" could mean colleagues - research development officers, other scholars - but these people have to be directly or indirectly related to the funding source. The funder will only reach a very small subset of the world's scholars directly, however, and this is crucial for a network model.

The other currently possible meaning of "Others" is commercial companies who collect scholarly funding information, package it into databases with a front-end and sell it to research institutions. While the author is not against making a profit out of information, there has been overwhelming evidence in the past two decades that commercially exploiting information *by virtue of hoarding it and then restricting access to it* is not a good strategy for anybody but the hoarder.

1.4.2 The value of Openness in scholarship and elsewhere

The Free Software, Open Source and more recently, Open Knowledge (which Open Access is a part of) movements have all demonstrated that the value of information can be greatly multiplied simply by having more people access it, reuse it and even be creative with it (e.g. visualise or summarise). The Open Knowledge Foundation argues this (28). Seminal works like "The Cathedral and the Bazaar" (14) have also argued this.

Governments seem to have grasped the benefits as well. The UK Government has mandated that all research funded with public money (which is a lot in the UK - almost everything through the seven Research Councils) should be published as Open Access by 2014 (20). It has also funded initiatives such as Open UK governmental data in the form of data.gov.uk (45) and the recent Finch report (2) (47).

Private organisations are not far behind and may in fact be leading the way - one of the largest funders of science in the UK and around the world, the Wellcome Trust, has had a progressive and strict Open Access policy since 2006 (10).

1.4.3 Scholars

Openness is recognised as beneficial - and many scholars are supposed to be able to benefit from having access to all this scholarship for free instead of having the libraries pay extortionate amounts to academic publishers. The very essence of scholarship is building on what has been discovered before.

However, while previous knowledge is a core requirement of science, sustenance is a very core human requirement, and in the modern world, this often translates to salaries *and* (sometimes "or") grant money, at least for scholars. Why should *this* information be commercially exploited for the

benefit of a few corporations when it could be used to better (globally) connect the minds to the money, to put it bluntly?

There is another point - information about funding opportunities may be available but may be too generic or hidden beyond layers and layers of website navigation, reflecting the mental model of the funding organisation instead of the mental model of the researcher (RCUK are a case in point (34)). Thus, the goal of this project from the perspective of scholars can be summarised as “bringing it together”.

1.5 The project

This project is about making an open-source web application named “FundFind” which lets stakeholders (just the scholars at first) share information about funding opportunities under open terms (the Open Definition defines “open” well (25)).

Follow-up work may increase this scope to include funding organisations or governments - they will already be able to submit funding information directly to a centrally hosted instance of the software produced by this project, but they will probably have special requirements related to this.

The project does include functionality to allow scholars to access the information, and allow developers / journalists / analysts and other stakeholders to analyse and mash up the information.

1.6 Existing works

There is no single piece or combination of pieces of software which enables involved stakeholders to do what this project would allow them to do upon completion, to the best of the author’s knowledge.

However, this project fits well within a current framework of Open Knowledge-related projects, a nice selection of which are hosted and developed by the Open Knowledge Foundation (27) (26).

This project will also require data to be useful. While one of its core aims is to enable crowd-sourcing of scholarly funding information, it should also try to make use and “digest” existing information. This could mean looking at current funding opportunities (33) (1) (5) (13) or at loading information about past opportunities, such as what the Australian National Health and Medical Research Council has funded over the past 23 years (24). Of course, part of the point of this project is that it is not easy to get such information in a nice universally readable (e.g. machine-readable) format. Research about useful data would be an optional objective subject to good feedback on the core software output and enough budget (of time).

Chapter 2

Development Process

2.1 Introduction

2.1.1 Overview

An Agile, iterative approach to the development of the software was initially chosen.

The main reasons for this choice were

1. Agile's perceived advantage in pinning down changing requirements.
2. Less formal communication with the users, who are all busy professionals. It was assumed they would not have sufficient time for writing functional specifications, even in collaboration with the author (as can sometimes be assumed in industry).
3. Linked to the user communication reason was a concept which many Agile methodologies used - user stories. These are short sentences (or rarely, paragraphs) which describe a single piece of functionality - a single "thing" that the user wants to do with the system.

However, it turned out that

1. The requirements did not change. Issues with funding data have been present for a while and a system which tries to prototype fixes through Open Data for some of them needs (a) developer(s) who know these issues inside out. Of course, the requirements gathering can still be done in an Agile way, but needs to be mostly completed before development starts. Defining the problem, in this case, had to come before solving it.

It could be argued that this was only a problem since the developer was also the "initiator" of the project. If a particular funder, Higher Education Institution or an organisation like the Open Knowledge Foundation wanted a project like this done, they would already know of certain problem characteristics - such as drawbacks of having so much funding data "closed". Then, more targeted exploration could be done and technical development would also have more precise targets to meet.

This means that the development process looked a lot more like **evolutionary prototyping** with a lot of requirements gathering between the prototype versions rather than an Agile iterative approach.

2. The assumption that users won't have time for anything but informal problem specification turned out to be correct. Thus, the "stories" concept proved to play an important role when gathering requirements from the target audience in short meetings, on the go or at hackdays.

3. The project's stories have been publicly accessible at the PivotalTracker management tool continuously throughout the exploration phase of the project which was used to record the aforementioned informal requirements. The "Icebox" still contains the list of features as prioritised by the various users who have been consulted during the exploration phase.

2.1.2 Evolution of the Development Process

Initially, Agile practices such as Pair Programming (38), full Scrum (37) and Feature-Driven Development were deemed unsuitable since they are aimed at groups of developers. Behaviour-Driven Development (with Cucumber (3)) was considered, but requires more commitment from the users - they need to learn an English-like domain-specific language which describes how the system is going to behave which didn't sit well with assumption #2 above and turned out to be the right decision as that assumption proved correct.

Thus, initially, with no methodology fitting the precise needs of the project and the Agile Manifesto (6) in mind, a methodology of an Agile *character* was planned, using the Extreme Programming lifecycle as a base (11).

The basic steps which were defined were:

“

1. Meet with a variety of potential users from the chosen user group (§1.2.2.1), trying to pick them so that each one has a different professional perspective on scholarship.
2. Define *and prioritise* the functionality of the project with each user in the form of the "stories" mentioned above. Record the results using the chosen project management tools (§??).
3. Release planning - prepare an initial timeline of which project features are to be released when and denote "release points". This work can be repeated as required - the order and the features themselves might change on the basis of user feedback over time (this is how Agile projects try to deliver more value by responding to changing requirements). The total size of the work and the size of each feature will most probably remain the same. This piece of work produces the "Project Timeline" and the current version is included in §??.
4. Iterate - every week is a self-contained unit of work consisting of coding, documentation and write-up work. Before starting to code each week, break up that week's story into tasks so goals can be tracked more effectively and there is something to be accomplished each day.
5. Repeat items 1 and 2 - meaning regular meetings with previously chosen users - one per user per month, demonstrate the latest features and record the feedback.

Emanuil Tolev in FundFind Progress Report (?)

”

Within the coding part of each iteration, Test-Driven Development is employed using both unit and integration tests (the latter via browser automation with Selenium (55)). This will be the evidence (on a basic technical level) for the "correctness" of the software. A basic form of Continuous Integration (23) will also be used - the Jenkins (52) software is currently being evaluated for the job. It basically runs all automated tests continuously, helping to detect regressions in code quality.

This methodology makes for a considerably simpler lifecycle than the Extreme Programming one (a visualisation of which can be accessed at (11)). This is intentional as there are only about 11 (potentially 13) weeks which allow for programming work as can be seen in §??.

The length of one iteration will be one week. Considering the tight deliverable deadlines (see §??), this will keep work on the project focused and will force the definition of user stories of a manageable size.

2.2 Modifications

2.3 Requirements

2.3.1 Requirements gathering

2.3.2 Final requirements

2.4 Design process

2.5 Technologies used

Chapter 3

Design

3.1 Overall Architecture

3.2 Alternative Designs

3.3 Main Application

3.3.1 Core

This is a core class of any Flask Python web framework project. It monitors all application code for changes and reloads components on-the-fly if changes are detected.

3.3.2 DAO

Data Access Object module – a database abstraction layer. Contains all the model classes of the application as well as functions which deal with the elasticsearch index (data store).

3.3.3 Web

This module is the “web server” of the application – it defines the routes and the corresponding functions which handle them (in a RESTful way), runs constantly and outputs exceptions to standard output. If more logic is needed to respond to a certain request than will fit comfortably in one short function, then the function can load another module (e.g. Importer).

3.3.4 Importer

A module which contains the Importer class. This handles data collection – submissions to the IDFind project. The Importer class gets created and used by the Web module.

3.3.5 Config

This class is a wrapper around a simple JSON dictionary of key-value pairs which contains hard-coded configuration information.

3.3.5.1 Default Settings

3.3.6 Util

3.3.7 TweetBot

3.4 Harvesters

3.4.1 HarvesterBase

3.4.2 CliManualSubmitter

3.4.3 RssHarvesterBase

3.4.3.1 BbsrcRssHarvester

3.4.3.2 EsrcRssHarvester

3.4.4 EmailHarvesterBase

3.4.5 EpsrcEmailHarvester

3.5 User Interface

3.6 API Design

Chapter 4

Implementation

Chapter 5

Testing

5.1 Overall Approach to Testing

5.2 Automated Testing

5.2.1 Unit Tests

5.2.2 Functional / Integration Tests

5.2.3 User Interface Testing

5.2.4 Stress Testing

5.3 User Testing

Chapter 6

Evaluation

- 6.1 Approaching the field of scholarly funding
- 6.2 Whose requirements?
- 6.3 Collecting information from disparate sources - technological suitability
- 6.4 User needs
- 6.5 Exploring scholarly funding
- 6.6 In retrospect

Appendices

Appendix A

Third-Party Code and Libraries

If you have made use of any third party code or software libraries, i.e. any code that you have not designed and written yourself, then you must include this appendix.

As has been said in lectures, it is acceptable and likely that you will make use of third-party code and software libraries. The key requirement is that we understand what is your original work and what work is based on that of other people.

Therefore, you need to clearly state what you have used and where the original material can be found. Also, if you have made any changes to the original versions, you must explain what you have changed.

Appendix B

Code samples

2.1 Random Number Generator

The Bays Durham Shuffle ensures that the psuedo random numbers used in the simulation are further shuffled, ensuring minimal correlation between subsequent random outputs (?).

```
#define IM1 2147483563
#define IM2 2147483399
#define AM (1.0/IM1)
#define IMM1 (IM1-1)
#define IA1 40014
#define IA2 40692
#define IQ1 53668
#define IQ2 52774
#define IR1 12211
#define IR2 3791
#define NTAB 32
#define NDIV (1+IMM1/NTAB)
#define EPS 1.2e-7
#define RNMX (1.0 - EPS)

double ran2(long *idum)
{
    /*-----*/
    /* Minimum Standard Random Number Generator */
    /* Taken from Numerical recipies in C */
    /* Based on Park and Miller with Bays Durham Shuffle */
    /* Coupled Schrage methods for extra periodicity */
    /* Always call with negative number to initialise */
    /*-----*/

    int j;
    long k;
    static long idum2=123456789;
    static long iy=0;
    static long iv[NTAB];
```

```
double temp;

if (*idum <=0)
{
    if (-(*idum) < 1)
    {
        *idum = 1;
    }else
    {
        *idum = -(*idum);
    }
    idum2=(*idum);
    for (j=NTAB+7; j>=0; j--)
    {
        k = (*idum)/IQ1;
        *idum = IA1 *(*idum-k*IQ1) - IR1*k;
        if (*idum < 0)
        {
            *idum += IM1;
        }
        if (j < NTAB)
        {
            iv[j] = *idum;
        }
    }
    iy = iv[0];
}
k = (*idum)/IQ1;
*idum = IA1*(*idum-k*IQ1) - IR1*k;
if (*idum < 0)
{
    *idum += IM1;
}
k = (idum2)/IQ2;
idum2 = IA2*(idum2-k*IQ2) - IR2*k;
if (idum2 < 0)
{
    idum2 += IM2;
}
j = iy/NDIV;
iy=iv[j] - idum2;
iv[j] = *idum;
if (iy < 1)
{
    iy += IMM1;
}
if ((temp=AM*iy) > RNMX)
{
```

```
        return RNMX;
    }else
    {
        return temp;
    }
}
```

Annotated Bibliography

- [1] AHRC, “AHRC Funding Opportunities,” <http://www.ahrc.ac.uk/Funding-Opportunities/Pages/Funding-Opportunities.aspx>, accessed November 2012.

The root funding page of AHRC, the Arts and Humanities Research Council.

- [2] Alok Jha, “Open access is the future of academic publishing, says Finch report,” <http://www.guardian.co.uk/science/2012/jun/19/open-access-academic-publishing-finch-report>, June 2012, accessed November 2012.

A Guardian article describing the Finch report which looked into Open Access and its benefits for academic publishing and society at large.

- [3] Aslak Helleoy, “Cucumber - Making BDD fun,” <https://cukes.info/>, accessed April 2013.

Cucumber is a system for Behaviour-Driven Development, which was considered as one of the options for development methodologies for this project.

- [4] O. Balci, W. S. Gilley, R. J. Adams, E. Tunar, and N. D. Barnette, “The Spiral Model,” <http://courses.cs.vt.edu/csonline/SE/Lessons/Spiral/index.html>, accessed November 2012.

A textbook section-length explanation of the Spiral software development model with a graphic visualisation and a table enumerating the various artefacts of the model. Although this an online educational resource from an academic institution, there is no publication date available for the project itself or the section on the Spiral model.

- [5] BBSRC, “BBSRC Research Funding,” <http://www.bbsrc.ac.uk/funding/funding-index.aspx>, accessed November 2012.

The root funding page of BBSRC, the Biotechnology and Biological Sciences Research Council.

- [6] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber, J. Sutherland, and D. Thomas, “Manifesto for Agile Software Development,” <http://agilemanifesto.org/>, accessed November 2012.

A succinct set of statements enumerating the core values behind the Agile software development collection of approaches to software development.

- [7] Browshot, “Browshot - Powerful web screenshots,” <https://browshot.com/>, accessed November 2012.

Homepage of the Browshot service. This allows a developer to make screenshots of web pages in any screen size, as any device including Personal Computers and mobile devices, in a variety of resolutions.

- [8] Cottage Labs LLP, “cottage labs - we don’t sell products, we trade our skills,” <http://cottagelabs.com/>, accessed November 2012.

The homepage of a partnership of freelance software developers working together to produce open source software for the benefit of Higher Education and commercial companies around the world. Dedicated to Open Knowledge and Open Scholarship.

- [9] Cottage Labs LLP and Open Knowledge Foundation, “FacetView - Pure Javascript Frontend for SOLR and ElasticSearch,” <http://okfnlabs.org/facetview/>, accessed November 2012.

Homepage of the Facetview project, a faceted Javascript browser (front-end) to indexing servers such as elasticsearch.

- [10] Craig Brierley, “Wellcome Trust strengthens its open access policy,” <http://www.wellcome.ac.uk/News/Media-office/Press-releases/2012/WTVM055745.htm>, June 2012, accessed November 2012.

A press release by the Wellcome Trust stating it is strengthening its Open Access policy and related punishment for breaking the terms of the policy, as well as linking to a full version of the policy.

- [11] Don Wells, “Extreme Programming Project,” <http://www.extremeprogramming.org/map/project.html>, accessed November 2012.

A diagram showcasing the Extreme Programming (Agile software development practice) project lifecycle. Different sections of the diagram provide links to more detailed explanations of the various parts of the project lifecycle.

- [12] Emanuil Tolev and Cottage Labs LLP, “IDFind - Got an ID? Not sure what it is? What you can do with it? Well, use this!” <https://github.com/CottageLabs/idfind/>, accessed November 2012.

The current homepage and version control repository page of the IDFind project. It tries to guess what kind of identifier an unknown string is.

- [13] EPSRC, “EPSRC Research Funding,” <http://www.epsrc.ac.uk/funding/Pages/default.aspx>, accessed November 2012.

The root funding page of EPSRC, the Engineering and Physical Sciences Research Council.

- [14] Eric S. Raymond, “The Cathedral and the Bazaar,” <http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/>, Sept. 2000, accessed November 2012.

A discussion on why / how JSON maps better to intuitive human understanding of the world than XML does and why this is important for a data representation format.

- [15] FAST, “Organisation Details - IBM UK Research Laboratories,” <http://www.fastuk.org/atcommunity/orgview.php?id=3073>, accessed April 2013.

"The Foundation for Assistive Technology was founded in 1998 to tackle the inadequate design of assistive technology products and services." This listing of IBM as a "source of funding" in their R&D section seemed as an appropriate reference to IBM Research UK as a commercial R&D vendor in the context of the project.

- [16] Gerrard Consulting, "The W-Model," <http://www.gerrardconsulting.com/?q=node/531>, accessed November 2012.

Explains the W software development lifecycle model and how it follows on from (or has evolved from) the V-model. This model could also be referred to as the double-V model.

- [17] GitHub Inc, "GitHub - Homepage," <https://github.com/>, accessed November 2012.

Currently popular decentralised version control system, usually used for software source code (but suitable for any artefacts although less suitable for binary formats or large files).

- [18] Google Inc., "Google Maps," <https://maps.google.co.uk/>, accessed November 2012.

A geographical mapping service by a commercial company.

- [19] Google Inc. and various authors - an open-source project, "Google Refine, a power tool for working with messy data (formerly Freebase Gridworks)," <http://code.google.com/p/google-refine/>, accessed November 2012.

A piece of open-source software for handling large semistructured datasets, making en-masse changes to them and cleaning them up. Google has now relinquished control and ownership of the project to the community, it is now the OpenRefine project.

- [20] Ian Sample, "Free access to British scientific research within two years," <http://www.guardian.co.uk/science/2012/jul/15/free-access-british-scientific-research>, July 2012, accessed November 2012.

An article in the electronic edition in the Guardian newspaper detailing the UK Government's policy on Open Access publications in 2014.

- [21] JISC, "JISC - We are the UK's expert on information and digital technologies for education and research," <http://www.jisc.ac.uk/>, accessed November 2012.

A UK governmental institution focused on enhancing the usage of digital technology in education and research in the UK.

- [22] Knight Foundation, "Knight Foundation - Informed & Engaged Communities," <http://www.knightfoundation.org/>, accessed November 2012.

The homepage of a global charitable foundation which supports (financially and otherwise) ideas that aim to change the world. They are particularly interested in quality journalism, media innovation, engaging the various communities around the world / in society and in fostering the arts.

- [23] Martin Fowler, "Continuous Integration," <http://martinfowler.com/articles/continuousIntegration.html>, accessed November 2012.

An article describing the principle of Continuous Integration in software development and illustrating why a developer or a development team might consider adopting this practice.

- [24] NHMRC / Australian Government, “NHMRC research funding datasets 1990 - 2012,” <http://www.nhmrc.gov.au/grants/research-funding-statistics-and-data/funding-datasets>, accessed November 2012.

The NHMRC research funding data for the period 1990 to 2012 covering all new and continuing grants funded and active in that 23 year period.

- [25] Open Knowledge Foundation, “Open Definition,” <http://opendefinition.org/>, accessed November 2012.

The Open Definition - defines what “open” is in relation to content and data.

- [26] —, “Open Knowledge Foundation Github Organisation Page,” <https://github.com/okfn>, accessed November 2012.

The homepage of the Open Knowledge Foundation organisation on Github, the currently popular version control system.

- [27] —, “Open Knowledge Labs creates services and tools that unlock the digital commons for everyone.” <http://okfnlabs.org/>, accessed November 2012.

The homepage of a collection of software projects which have sprung up to support or resulted from the Open Knowledge Foundation’s pursuit of spreading the Openness cause in society.

- [28] —, “Our Vision / We Believe in the Power of Openness,” <http://okfn.org/about/vision/>, accessed November 2012.

Notes on the Open Knowledge Foundation’s vision of how openness can be beneficial for society (meaning all of current human civilisation).

- [29] —, “Radial Bubble Tree Visualization,” <https://github.com/okfn/bubbletree>, accessed November 2012.

Homepage of BubbleTree, a higher-level Javascript graphical visualisation library.

- [30] —, “Recline.js - relax with your data,” <http://okfnlabs.org/recline/>, accessed November 2012.

A library for building applications which handle potentially large amounts of the data and intend to allow and focus on more or less advanced data manipulation.

- [31] —, “Where Does My Money Go? - Showing where your taxes get spent - Country & Regional analysis,” <http://wheredoesmymoneygo.org/bubbletree-map.html>, accessed November 2012.

A BubbleTree visualisation of UK public spending.

- [32] S. Palmer, “An Introduction to Feature-Driven Development,” <http://agile.dzone.com/articles/introduction-feature-driven>, Nov. 2009, accessed November 2012.

A good introductory overview of what the Feature-Driven software development methodology is about, why it is considered an Agile methodology and when it is applicable (it evolved as an Agile approach for larger projects).

- [33] RCUK, “RCUK India - Opportunities provided by UK Research Councils,” <http://www.rcuk.ac.uk/international/Offices/OfficeinIndia/Funding/Pages/Home.aspx>, accessed November 2012.
The root funding page of RCUK / International Office / India.
- [34] Research Councils UK, “Welcome to Research Councils UK - Excellence with Impact,” <http://www.rcuk.ac.uk/Pages/Home.aspx>, accessed November 2012.
The home page of Research Councils UK.
- [35] Ron Knott, “Fibonacci Numbers and the Golden Section,” <http://www.maths.surrey.ac.uk/hosted-sites/R.Knott/Fibonacci/fib.html>, June 2012, accessed November 2012.
An explanation of the Fibonacci number sequence and related mathematical concepts. The Fibonacci sequence is being used as the complexity estimate for each user story in this Agile-managed project.
- [36] Scott Ambler + Associates, “Introduction to User Stories,” <http://www.agilemodeling.com/artifacts/userStory.htm>, accessed November 2012.
A explanation of what “stories” or “user stories” are when the term is used in the context of Agile approaches to software development.
- [37] Scrum Alliance Inc., “Scrum Is an Innovative Approach to Getting Work Done,” http://scrumalliance.org/pages/what_is_scrum, accessed November 2012.
A succinct description of the Scrum framework of Agile software development practices.
- [38] J. Shore, “The Art of Agile Development,” http://www.jamesshore.com/Agile-Book/pair_programming.html, Feb. 2010, accessed November 2012.
The online edition of a famous work (book) on Agile Development. This section focuses on and provides a good description of Pair Programming, an Agile software development practice.
- [39] Shuttleworth Foundation, “Shuttleworth Foundation - Supporting exceptional people to change the world,” <http://www.shuttleworthfoundation.org/>, accessed November 2012.
The homepage of a global charitable foundation which funds individuals with its scholarships and then multiplies the money they invest in projects which aim to bring about social change for the better.
- [40] SoftDevTeam, “V-shaped lifecycle model,” <http://www.softdevteam.com/V-shaped-lifecycle.asp>, accessed November 2012.
Explains the V-shaped software development lifecycle model. Also has a nice succinct summary of advantages and disadvantages compared to other models (mostly Waterfall) which seems to fit previous knowledge I have of this model from less reliable sources such as Wikipedia and more reliable ones such as the teaching accompanying my Computer Science degree at Aberystwyth University.
- [41] The Document Foundation, “LibreOffice Calc - a Spreadsheet That Meets Any Need,” <http://www.libreoffice.org/features/calc/>, accessed November 2012.
A piece of open-source spreadsheet viewing and editing software, part of the LibreOffice office productivity suite.

- [42] The Smart Method Ltd, “The Waterfall Development Methodology,” http://www.learnaccessvba.com/application_development/waterfall_method.htm, accessed November 2012.

A brief explanation of the Waterfall software development methodology. Also (briefly) compares Waterfall to Agile approaches.

- [43] E. Tolev, “PivotalTracker - FundFind,” <https://www.pivotaltracker.com/projects/688967>, Nov. 2012, accessed November 2012.

Publicly visible project management for the software component of this project. By default, it will show what the author is currently working on and what is in the backlog. Click on “Icebox” (in the top left-hand corner of the web page, next to the “CURRENT” and “BACKLOG” links) to see all recorded user stories and chores for getting the project done. Stories in the Icebox are ordered in descending order in terms of priority, i.e. the more important ones according to aggregated user opinion are at the top.

- [44] Twitter Inc., “Bootstrap - Sleek, intuitive, and powerful front-end framework for faster and easier web development.” <http://twitter.github.com/bootstrap/>, accessed November 2012.

The homepage of the Twitter Bootstrap User Interface library.

- [45] UK Government - Unknown individual name, “DATA.GOV.UK - OPENING UP GOVERNMENT,” <http://data.gov.uk/>, July 2012, accessed November 2012.

The UK Government Open Data portal containing various datasets released under Open licences.

- [46] Various, “What are the pros and cons of XML and JSON?” <http://stackoverflow.com/questions/3536893/what-are-the-pros-and-cons-of-xml-and-json>, accessed November 2012.

A discussion on why / how JSON maps better to intuitive human understanding of the world than XML does and why this is important for a data representation format.

- [47] Various (a UK Government Working Group), chaired by Dame Janet Finch, “Accessibility, sustainability, excellence: how to expand access to research publications,” <http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf>, June 2012, accessed November 2012.

The Finch report - looked into Open Access and its benefits for academic publishing and society at large, commissioned by the UK Government.

- [48] Various authors, “Behaviour-Driven Development,” <http://behaviour-driven.org/>, Jan. 2009, accessed November 2012.

A description of Behaviour-Driven Development. This is a Wiki-style web resource, so the year and month of publication are taken from the last edited date and may change by the time the reader accesses this resource.

- [49] Various authors - an open-source project, “D3 - Data-Driven Documents,” <http://d3js.org/>, accessed November 2012.

Homepage of D3.js, a lower-level Javascript graphical visualisation library.

- [50] —, “elasticsearch - You know, for Search,” <http://www.elasticsearch.org/>, accessed November 2012.

The homepage of the elasticsearch indexing server.

- [51] —, “Flask - Web development, one drop at a time,” <http://flask.pocoo.org/>, accessed November 2012.

Homepage of the Flask Python web framework project, helps build RESTful web applications in Python.

- [52] —, “Jenkins - An extendable open source continuous integration server,” <http://jenkins-ci.org/>, accessed November 2012.

Homepage of the Jenkins Continuous Integration server, a tool which helps a developer or a team of developers practice the Continuous Integration software development principle.

- [53] —, “jQuery is a new kind of JavaScript Library.” <http://jquery.com/>, accessed November 2012.

The homepage of the jQuery Javascript and User Interface library.

- [54] —, “OpenRefine,” <https://github.com/OpenRefine/OpenRefine>, accessed November 2012.

A piece of open-source software for handling large semistructured datasets, making en-masse changes to them and cleaning them up. Used to be the Google Refine project.

- [55] —, “SeleniumHQ - Web application testing system,” <http://seleniumhq.org/>, accessed November 2012.

The homepage of an open-source software project which enables the automation of web browsers and other testing techniques for the purposes of testing web applications.

- [56] Various authors - an open-source project hosting open data, “OpenStreetMap - The Free Wiki World Map,” <http://www.openstreetmap.org/>, accessed November 2012.

An open-source, open data geographical mapping service.

- [57] Various authors / Wikipedia, “Faceted search,” http://en.wikipedia.org/wiki/Faceted_search, accessed November 2012.

The Wikipedia page on faceted search. Would not make much sense to a reader unfamiliar with the topic, but should do with the example in the report body.

- [58] Wellcome Trust, “Wellcome Trust - We are a global charitable foundation dedicated to achieving extraordinary improvements in health by supporting the brightest minds.” <http://www.wellcome.ac.uk/>, accessed November 2012.

The homepage of a global charitable foundation which funds lots of scholarly activities in a variety of areas.