EPQ Project Diary

A record of the processes, developments and evaluations made during the creation of my EPQ project.

|  |  |  |
| --- | --- | --- |
| **Date** | **Target** | **Solution** |
| *11 – 17 May 2020* | Choose a topic that interests me, which I can easily research, and which relates to my career plans for the future.  Come up with a question concerning this topic that I will aim to answer.  Create a detailed mind map explaining how I will answer this question, any steps I will take and sub-topics I will research. | Having considered possible topics for this project a while before starting, I chose to base my EPQ on Steganography (the act of hiding a message in plain sight so an interceptor does not suspect its existence). I had first learned of this technique during a lecture in early March, and immediately wanted to explore it further.  My initial idea for my question is ‘How can I conceal a message within a bitmap image, and then retrieve it?’ This means that I aim, through use of the Python programming language, to hide a message within a digital image using a method of steganography called Least Significant Bit (LSB) modification. This involves converting a message into its equivalent ASCII binary form, and then hiding these digits within the binary that makes up an image. I chose this question because it challenges me to put my research to use and produce a workable program as well as explain the topic, and I think that this will be enjoyable and test my skills as a programmer. As I hope to study computer science at university, I believe that this topic will give me a good insight into aspects of cyber security and software development also.  I used an online mind map creator for ease and flexibility, and divided my plan into four sections. These are colour-coded so it is easy to distinguish between them, with more generic bubbles being brighter than the ones providing extra detail. Three of these sections cover explaining the topic to the reader: a general overview of steganography and its varying forms; a summary of why the technique is used, including a comparison with cryptography (encoding a message so it becomes unreadable); and an explanation of steganography within digital images. This will explain LSB modification. The fourth section contains a plan of the steps I will take to create the Python program, in order to conceal and retrieve a message and therefore answer my question. As the mind map is based online, I can edit it in the future if my plans change. |
| *18 – 24 May 2020* | Begin the research process, using Google Scholar to find reliable articles and record some relevant facts along with their source details. | I found facts from four online articles: "An Overview of Image Steganography" by T. Morkel, J.H.P. Eloff and M.S. Olivier, from which I took 11 facts; ‘Digital image steganography: Survey and analysis of current methods’ by Abbas Cheddad, Joan Condell, Kevin Curran and Paul McKevitt, which provided two facts; ‘The history of steganography’ by David Kahn, giving me 4 facts, and ‘Steganography and Steganalysis’ by J.R. Krenn, which gave me 20 facts. Some of these were direct quotes if I thought they were explained well and could be useful, others are paraphrased to condense ideas together. |
| *1 – 7 June 2020* | Number my notes and sort them into distinct groups.  Write a paragraph summarizing some of these notes for one subheading. | I chose to both number and colour code my notes to make it easy to see which notes belong to which subheading. My groups are General Steganography (based on my first subheading, including information unspecific to the other groups), Bitmap Images (methods of steganography available in digital images and their advantages or disadvantages), Other Methods (including any method of steganography, past or present, except those in bitmap images) and Cryptography (this category is currently quite small, but will allow me to compare the methods later on.).  The paragraph I chose to begin with was in Other Methods, one of two for this subheading (the second yet to be written). I chose this one because I was able to quote from three of my sources and it wasn’t overly technical. The paragraph itself was colour coded to make it easy to distinguish between quotes, paraphrased facts and extra words used to structure the paragraph. This paragraph contains 258 words.  While working on this I noticed that a few of my sources had been referenced using the date they were first available online, rather than the date they were first published. This didn’t take long to find and change. |
| *8 – 14 June 2020* | Evaluate my sources, making notes on their reliability and context.  Write paragraphs for the rest of my subheadings, encompassing each of the facts I have found. | For each of the three sources referenced in my paragraph, I added an evaluation of their dependability, coloured purple so they can be distinguished. This included clues in provided email addresses, locations or cited sources, as well as information I could find about the authors online.  I then began to collate similar paragraphs for the rest of my subheadings. I used the same colour scheme, making it clear where my facts have originated, and wrote an evaluation for the source not used in the first paragraph (Cheddad’s article). I have finished three paragraphs. |
| *15 – 21 June 2020* | Create a project diary explaining what I have done, the decisions I have made, and why I chose to do things a certain way.  Complete the unfinished paragraphs. | I have created a diary on Microsoft Word (this document), explaining the processes, developments and evaluations of my project. This includes a target for each week, an explanation of the progress made and a note of any changes not relevant to the target in a separate column called the changelog. (Changes to this document will not go in the changelog)  Then I finished the rest of the paragraphs, creating a good foundation for when I continue to research and add more notes later on. The paragraphs contain a total of around 1700 words. |
| *22 – 28 June 2020* | Write a page reflecting on the progress I have made and what this project has taught me about my learning style and preferences. | I discussed the successes and challenges encountered so far in preparing for the project from home (such as the limited resources available and being unable to meet with a supervisor); how I managed my time, by splitting the week into 1hr slots for each of my subjects; what I have learnt about my learning preferences, including the fact that I work much better with online mind maps than hand-drawn ones; the most enjoyable and most difficult parts of the project; the strengths and weaknesses of my work (a weakness being the ratio of facts coming from my reliable sources compared to the less reliable ones); and what I would change if I were to attempt this part of the project again. |
| *End of Quarantine Preparation Work* | | |
| *8 September 2020* | Complete pages 1, 5 and 6 in the AQA project log, filling out my personal details and initial project ideas. | I discussed my initial ideas for the project as a result of the preparation work I did, including my idea for the title (How can I conceal a message within a bitmap image and then retrieve it?), how I plan to research and develop my project and how it relates to my other A-levels and future ambitions. I have decided to continue with the steganography topic, as it is something that really interests me and aligns well with my future ambitions to study computer science at university. My A-Level subjects are Computer Science, Maths and Further Maths, and I believe that this project is a good example of mathematical thinking applied to the digital world. |
| *14 September 2020* | Meet Dr Dolton (my supervisor) and discuss my current ideas for the project. | My supervisor is Dr Dolton, who is also the centre co-ordinator. I will meet her every Monday to discuss the progress I have made and any questions I have. Today she gave me positive feedback on my preparation work, and clarified the timescale for the project. As a result of this I have decided to focus on the written part of my project until November, when the focus will switch to planning and then creating my artefact. |
| *15 September 2020* | Continue to research, looking specifically for facts about why steganography is useful in the world, using a new source. | I began to use a new online article (“Steganography - A Data Hiding Technique” by Kumar, Arvind and Pooja, K.M.) to look for new facts to add to my paragraphs. In the 1hr space I was unable to find more than 5 new facts – by the end of this week I aim to have scanned the whole document. |
| *22 September 2020* | Create a plan for the essay and add citations to my sources using the reference tool.  Keep researching the new source. | I created a draft document for the essay, with a plan of which pieces of research should fit under which subheadings, in order to give the project some structure. In this document I have created source references using the reference tool in Microsoft Word, allowing me to cite them later on.  Then I continued to research Kumar’s article, and noticed that some of the information was paraphrased or directly copied from Krenn’s article, without citation. Krenn’s was published 6 years before Kumar’s, so this made me sceptical of the source’s reliability. However I have continued to use it to gather facts, so that I can confirm or challenge them using other sources when I write my essay in full. I made a note of this in my earlier evaluation of Krenn’s work – his source is one with very little background information to be found also. It will also be included in this source’s evaluation in order to alert the reader of the issue. |
| *26 – 27 September 2020* | Continue to research, as well as start to insert my ideas and research into the essay draft. This should give me a better idea of how to structure the essay and which bits of research fit where. | I have looked at another two sources, one an online magazine article by L. Cameron, and the other the website of a company dedicated to countering criminal use of steganography, CUIng.org. Most notably, these contained recent examples of real-world steganography, including some using LSB modification.  I created references for these two sites in my essay draft – though I still need to review these – and copied my earlier research into the new draft. I then made notes on what needs to be modified or included in the plan. |
| *29 September 2020* | Shorten my source evaluations, and adjust the colour scheme to make my work clearer. | After speaking with Dr Dolton yesterday I identified that my source evaluations were unnecessarily lengthy, so I spent time making them more concise and relevant to my essay.  Another issue I identified while referring to my earlier paragraphs was that the colour scheme I was using (showing whether each quote was paraphrased or quoted from my facts list) was very unclear, so I adjusted it. It now shows the difference between direct quotes from my sources, paraphrased information, extra words to make the content flow and source evaluation. |
| *1 October 2020* | Evaluate the final three sources. Begin to write the introduction. | I wrote evaluations for the final three sources (Kumar, Cameron and CUIng), and colour coded the facts from these sources, so that I can easily refer to them when I add them into my essay later on.  I began to write the introduction, trying a couple of approaches in order to create suitable context to introduce my essay. |
| *5 – 11 October 2020* | Adjust the layout of the project diary, and rewrite certain entries in more detail. | I concluded that my project diary was unclear and repeating itself in places. As a result, I decided to remove the changelog, opting for just three columns instead, while explaining in detail everything that was a part of this in the solution column. I also rewrote a large portion of what was already in the solution column to make it more specific. |
| *12 – 16*  *October 2020* | Complete the introduction within the essay. | I have finished a complete paragraph to introduce my topic, considering the developments of technology over the past few decades and how cases of data theft are common within the media. This paragraph is entirely my own work, which I think will help guide the reader into the more complex ideas later on in my essay. |
| *19 – 23 October 2020* | Start to write the essay, incorporating my older work from over the summer, the new facts I have gathered and following the structure I planned in my mind map. | So far, I have written roughly 1100 words, and am about halfway through my first draft of the Secondary Research section. This has included a lot of the preparation work, most of which having to be rewritten in order to flow properly and allow for suitable explanations. The sources I have used so far have been evaluated the first time I made reference to them, and then simply cited each other time. Throughout the essay I have made notes to myself as to what to improve: where the content doesn’t flow, where I need to do more research and so on. For example, I have identified that I need to find sources to explain how steganography within text files works and what redundancy means, as so far I have been filling in these details myself without knowing where the information is coming from. I will address these issues once I have completed the draft, so that I can see if any issues recur.  I have considered that I may need to adjust my plan for the methodology, as it will follow a different structure to the rest of the essay. I will discuss this with my supervisor as soon as possible (I cannot currently meet her due to Covid-19). |
| *2 – 8 November 2020* | Rewrite the introduction, making it clearer what my aims are in the project and what I will be discussing.  Continue to write the draft. | I have identified that my introduction was lacking in clarity, both in the explanation of steganography and my discussion of what I hope to achieve. I decided to add more detail to both parts; however this has left me unsure as to whether I need to find a source to back up the new information.  Then I continued to write my draft, this time focusing on the modern applications of steganography. I am currently at roughly 1700 words. However, I soon realised that I may need to re-structure part of it so that it makes sense to the reader. Specifically, I had hoped to discuss the reasons that steganography is used *before* going into detail about steganography within bitmap images. This is because the research relating to images will help the reader to understand my methodology, which would come next. However, doing so has made some of my discussion of steganography examples more confusing, especially to a reader who has not come across certain terminology before. I am unsure whether to move the whole digital images section, summarise it briefly early on and leave the technical aspects until later, or reduce the complexity of the reasons for steganography part, at the expense of some of the examples. I will discuss this with my supervisor, in order to get a second opinion on what does and doesn’t currently make sense, before I decide which approach to take. |
| *9 – 10 November 2020* | Talk to my supervisor about whether she understood the Methods of Steganography section, and whether I should re-arrange my essay.  Write page 9 of the AQA project log. | I asked my supervisor to read my draft, and comment on whether it made sense to her as it is. As a non-expert, she replied that the current structure did make sense, and that it would be better for me to stick with my original plan. Including a brief explanation at the point where I discuss my source will help the reader to understand, so I can leave the digital images section until just before my methodology. This will allow me to keep referring back to it and hopefully make the methodology make more sense to the reader.  I wrote the Planning review section of my project log, explaining my current motives, deadlines, comments from my supervisor and modifications I have made as a result. |
| *17 November 2020* | Finish the Modern Applications section of my draft. | After considering my supervisor’s feedback, I continued to write, using my research to finish this section and a paragraph on cryptography. However, it soon became clear to me that this is severely lacking in evidence, and that which I do have needs backing up or challenging using other sources. This has caused the paragraph to feel unplanned and messy. To counter this, I have begun to make a list of possible new sources to use throughout the essay once I have finished the first draft (there is only one section left to write.) This week I have struggled to find time to work on the EPQ and so will need to find a better balance. |
| *22 November 2020* | Re-formulate the Modern Applications section of my essay, bringing the cryptography part to the beginning in order to allow for better flow. | I have identified that the cryptography part felt like a tangent at the end of the Modern Applications section, and that it would be better suited at the beginning of this section when I explain the rapid growth of technology over the past few decades. This made more sense chronologically, as this is when encryption became much more necessary, and also because it allowed me to reference it later in the essay without having to partly explain beforehand. Next week I aim to begin the digital images section of my draft. |
| *23 – 29 November 2020* | Begin the Digital Images section, creating a framework of the key ideas that need to be understood before I can explain LSB modification in detail. | It became clear to me quite quickly that the sources I have looked at so far are focused on LSB modification (the technique I will use) and assume that the reader has an understanding of the concepts of binary, pixels and how we represent colours on a computer before viewing their source. So I decided to explain these ideas myself first of all, which really helped me to identify which parts needed to be explained first and which built on earlier ideas. Once my draft is complete I can then look for new sources which explain these concepts in detail, and modify my explanation around those.  As this is quite a technical area of the subject, I have made use of diagrams in order to allow the reader to visualise the ideas. One of these I created myself, which will allow me to show the reader how I have applied steganography to it. However, this has brought up the issue of how I should refer to these images – whether or not I need to clarify that I created them. I will ask my supervisor about this tomorrow. The next part, where I apply LSB modification to the diagram, does have some research prepared, so hopefully that part won’t be as rough as this section has been. |
| *30 November – 6 December 2020* | Finish the Digital Images section of my draft, completing the Secondary Research part. | I completed my explanation of LSB modification including some of its advantages, disadvantages and ways it can be foiled. The draft currently has roughly 3500 words, but is still severely lacking in sources to back up statistics and compare with one another. This last paragraph was more successful than the first from this section in this regard, and I was able to introduce the final of my current sources. Next I need to re-read the draft, checking for grammatical or logical errors, and identifying clearly where I need more evidence so I can begin even more research. I think writing the first draft using the research I already had has really helped me to find these gaps. |
| *7 – 13 December 2020* | Continue to research in order to fill the gaps in my Secondary Research half of the draft.  Create a plan for the Python project and include this in the Methodology part of the draft. | I spent some time researching context for the History section, including notes on Sir Francis Walsingham, Elizabeth I’s principal secretary who developed a notable intelligence system during the Renaissance period. Then I created a plan for my artefact, using the same mind map creator as before to outline what each section the code will do, without going into detail about the specific programming constructs I need to use. However, I am unsure that this mind map is easy to understand and may change it to a list or other format so it is more chronological. While creating the mind map I had to research to find a suitable library to import, so that I could access the properties of the image. Python Imaging Library was the solution I decided upon, because it is simple and only requires one download, compared to other examples I saw which contained matrices and other complicated constructs as well as multiple downloads. While downloading PIL was simple, I will have to make sure to do it on every device that runs the code, and provide instructions in the program to explain to anyone else how to do this. |
| *14 – 18 December 2020* | Add to my Secondary Research, to finish the History section and add two new diagrams illustrating the process of statistical analysis. | The last part of the History section was the Microdot. I added a diagram and another source in order to explain the technology in more detail.  During this week, I was introduced to a program in Excel that allowed me to quickly and easily create a coloured image by typing binary codes into each cell. This was something I had hoped to include in my project before, but I could not find an efficient method to do so. The program allowed me to create a simple image of a house, and then re-create it by only using the LSBs of each binary value. This clearly and effectively illustrates how statistical analysis (a method to counter LSB modification) works and the fact that LSBs are not totally useless. |
| *21 December 2020 –*  *4 January 2021* | Research some of the new sources I have found and add to my essay.  Begin to code, taking screenshots of the process and explaining them in my Methodology. | I researched some of the sources I had made a note of earlier and implemented their facts into my essay. Most of these sources will be used in a range of places throughout the essay, and then whatever gaps are left can be covered directly by more specific sources later on.  However, the majority of time spent over the Christmas period was used to begin my program. It allows the user to select whether they want to embed or extract a message, providing their own message, image and key. At this stage, I am almost halfway, having completed the embedding section, except for the fact that it doesn’t actually work yet. I have completed all the code I believe I need for this part, but I just need to identify my mistake so that the program can work as intended. Then I can move on to extracting the message. Finally, I took some screenshots of the parts that I believe aren’t going to change, and pasted them into the methodology, along with explanations of what the code is doing at each stage. |
| *5 – 10 January 2021* | Continue to research.  Fix the code so that it successfully adjusts the colour values of each pixel. | I have added more sources into my essay, having now used all of the ‘general’ ones I had listed.  I managed to fix the code so that now it seems to be working as intended. The problem with steganography is that testing whether it works is not easy, as it is designed to be an imperceptible change. The best I can do without having an extraction tool yet is make the code change the randomly selected pixels to a bright pink, so I could see that it was actually affecting them. Once that was working, I then reverted the code back to using the bits from the message, and that ran without error. However, the only way I can truly test whether it is successfully concealing the correct bits is to complete the extraction half and see if that works.  This week I filled out my mid-project review and sent it to my supervisor. |
| *11 – 17 January 2021* | Continue to research.  Continue to code the extraction half of the artefact. | This week, I researched three different sources related to colours; in particular how many colours humans can see and how they are displayed on a computer.  I then worked a lot on the program, having almost completed the extraction half. Currently, the program is producing a bunch of random characters instead of the message I put in – though each one correlates to the same character. Furthermore, I discovered that by removing the first bit, and adding a 0 or 1 to the end of each character, I am left with the correct character. However, when I tested it again using a shorter text document it worked as intended. Next week I will need to figure out the cause of this so that I can fix – and hopefully finish – the program. |
| *18 – 24 January*  *2021* | Continue to add sources into the essay.  Identify the issue within my program so it is ready to be tested. | This week, I introduced two new sources into my essay, both of which focusing on cryptography. This can be a very technical subject, so I had to rely on sources which explained in a simple manner, and found that educational videos were best. However, these took a long time to integrate into the essay, so I will need to pick up the pace in getting my secondary research finished.  I managed to find the issue in my program – when converting the length from decimal to binary, the program was adding too many filler zeroes at the start if it was more than 8 bits. This explains why the smaller file worked, but the larger one was affected. Having fixed this, the program seems to be working, except for one small error. Every now and then, a character would be extracted incorrectly, leaving an ASCII character that had a code with one bit different from the correct code. Such a small change can only be down to errors when copying or saving the file. I researched both methods and then used an online tool to test them myself, concluding that the method used to save the image was causing it to be saved slightly differently each time. I have been unable to find a way to prevent this, but did notice that the affected characters were different using different images, with larger images producing far less errors. By replacing my original cards image with one of a larger size, I was able to reduce the errors to just four within the extract (that is out of 2367 characters – almost negligible). It is a shame that the program isn’t perfect, but I need to move on to testing and documenting it. |
| *25 - 31 January 2021* | Continue to research.  Write an abstract for the essay.  Revise the screenshots in my methodology, and add more to finish explaining the program, as well as some to show off the final result.  Add a revised plan for the Python project. | I have now addressed all the gaps I had previously highlighted in the Modern Applications section of the essay. Now I only have the Digital Images section left to fill.  I wrote a 100-word abstract for the essay, which summarises the key concepts of steganography and LSB modification, and my aims in creating the Python project.  Next I continued with the methodology, and, having made substantial changes to the program since I last updated it, I ended up retaking almost all of the screenshots, adding suitable explanations of each stage. Once this was done, I added evidence of the test data used, as well as the stegoimage produced and the final text file (which contains evidence of my issue from last week). I had more free time than usual this week to work on the EPQ, and managed to write 3000 words in this section! The next stage will be to include details of the actual development process instead of just the final product, including when I modified the pixels to be pink, and testing the error handling with invalid inputs.  Finally, I went back to my original plan and used it to create a more accurate plan that better represented the final program. I included both in the essay, as the new plan may help the reader to navigate my code, but it wasn’t what I referred to when I was developing it. In total, the entire essay now has almost 8000 words. |
| *1 – 7 February 2021* | Integrate evidence of the problems encountered when coding into the methodology. | Unfortunately, I didn’t have quite so much time to work on the project this week, however I think I managed to finish a significant section. I didn’t manage to research any new sources (though I did look again at an old one and add a few new statements), because I was focused on adding notes on my challenges while coding the artefact, and on my testing process. By adding detail in the methodology, I was able to simplify this diary a little to ensure it wasn’t too technical. I believe that the methodology is now finished, though I will have to review it. I also spent a significant amount of time cleaning up grammar errors and keeping a consistent past tense in the methodology.  My next steps will be to write a conclusion for the essay, insert a bibliography, continue with the research and try to reduce the word count as much as possible. After next week it will be half term, so I aim to have the secondary research finished by then. |
| *8 – 14 February 2021* | Write a conclusion summarising the essay and the extent to which I achieved my aims.  Insert a bibliography, table of figures and appendix including screenshots of the code as a whole.  Continue to research and fill gaps in the document. | This week I wrote a conclusion for the essay, including a summary of my secondary research, methodology and final outcome. I added a table of figures and a bibliography using the tools in Microsoft Word, and an appendix into which I pasted screenshots of my artefact, including all the details and error handling omitted from previous screenshots.  Then I continued with the research. I had overestimated the amount I had left to do last week, and over the course of this week I was able to finish addressing all the gaps I had identified previously. My target now is to go through the entire essay and reduce the word count as much as I can, which will be a challenge. |
| *15 – 21 February 2021* | Reread the essay and reduce the word count.  Adjust the format of the essay, removing the colour code and adding final touches. | I spent this week rereading my essay, editing the grammar in places to make sure it made sense, and reducing the word count as much as possible. Now, excluding the abstract and bibliography, the essay has about 7500 words, which I think is a good total. I then finalized the format of the essay, removing the colour code, checking the bibliography and table of figures were accurate, and that the images were all in the right place. |
| *22 – 28 February 2021* | Fill in the project product review part of the log.  Create a brainstorm for the presentation, including a rough idea of what each slide will contain. | This week I wrote the project product review part of the project log, evaluating my progress and how closely the project has followed my original plan. I learned this week that school will be reopening on the 9th of March, and so my presentation will take place in person on the 16th. My supervisor sent my class a sheet giving general guidance on what to include in the PowerPoint, and I used it to create a simple handwritten brainstorm adapting it to my project, discussing what I would include on each slide and what I would reserve for the script. |
| *1 – 7 March 2021* | Begin to create a script and PowerPoint for the presentation. | This week I started the script for the presentation, using my brainstorm for guidance as to what to include. I have prepared three of the five slides, talking about my aims, research and some of my more interesting findings. Next week I will finish it and then practise to check it is the correct length, and I will be presenting it the week after. |
| *8 – 14 March 2021* | Finish preparation for the presentation and fill out the presentation record part of the log. | I have now finished the script for my presentation, having edited it multiple times to check that it isn’t too long and that it is clear and understandable. After rehearsing with a friend I identified that I need to ensure I put enthusiasm into the speech and don’t talk monotonously, as well as slow down to make sure that the audience has time to understand. I feel like I will be able to improve this in the presentation. The PowerPoint is also complete, and I filled out the presentation record part of the project log, detailing this process and the feedback I received. I am beginning to learn parts of the script so that I am not reliant on my notes when I present. |
| *15 – 21 March 2021* | Present my topic to my supervisor and members of my class.  Fill in the Summary and Reflection part of the log. | I believe that the presentation was a success. I gave it on Tuesday morning in front of my supervisor and a few members of my class. After the speech I was asked three questions, which I had no trouble answering. While I haven’t yet had any feedback from my supervisor, a friend mentioned that she had been able to understand my explanation of the topic, which had been a concern of mine.  After the presentation I filled out the Summary and Reflection page of the project log, discussing all the skills I have gained from this project as well as what I would change if I were to attempt it again. The project is now very close to being complete. |
| *22 – 28 March 2021* | Update the Summary and Reflection section of the log in line with guidance from my supervisor.  Revisit the diary once more to check everything is ready for submission. | My supervisor responded that the reflection I had done was written clearly and effectively, but needed some more evidence of parts of the project that didn’t go so well. In the lesson she provided a PowerPoint with some ideas for what to write about, and I used my first evaluation alongside this to write a new reflection incorporating this feedback.  Then I looked over this document one last time, checking that each section was accurate, before submitting it and my log to my supervisor. |