



HAL
open science

Check Your Tech - Considering the Provenance of Data Used to Build Digital Products and Services

Dympna O'sullivan, Damian Gordon

► **To cite this version:**

Dympna O'sullivan, Damian Gordon. Check Your Tech - Considering the Provenance of Data Used to Build Digital Products and Services. 16th International Conference on Social Implications of Computers in Developing Countries (ICT4D), Jun 2020, Manchester, United Kingdom. pp.195-204, 10.1007/978-3-030-65828-1_16 . hal-03272526

HAL Id: hal-03272526

<https://inria.hal.science/hal-03272526v1>

Submitted on 28 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Check your Tech - Considering the Provenance of Data Used to Build Digital Products and Services

Dympna O'Sullivan and Damian Gordon

School of Computer Science, Technological University Dublin, Kevin Street,
Dublin 8, Ireland

dympna.osullivan@tudublin.ie
damian.x.gordon@tudublin.ie

Abstract. Digital products and services are producing unprecedented amounts of data worldwide. These products and services have broad reach and include many users and consumers in the developing world. Once data is collected it is often used to create large and valuable datasets. A lack of data protection regulation in the developing world has led to concerns about digital colonization and a lack of control of their data on the part of citizens in the developing world. The authors of this paper are developing a new digital ethics curriculum for the instruction of computer science students. In this paper we present two case studies we have developed with a focus on data ethics in a developing world context. Each case study is accompanied by a list of specific questions to be used by the instructor to allow students to evaluate the implications of introducing new digital products and services in a developing world context as well as a generic case studies checklist that allow deeper reflection on the intended and unintended consequences of introducing new technologies.

Keywords: Digital Ethics, Digital Colonization, Computer Science Education.

1 Introduction

The World Economic Forum predicts that the digital universe will reach 44 zettabytes by 2020 [1]. The current number of smartphone users in the world is 3.5 billion, on average, Google processes more than 40,000 searches every second, there are 2 billion active Facebook users and 156 million emails are sent each day. In 2020 there are projected to be 200 billion Internet of Things devices. A significant proportion of this data is generated by users in the developing world. The ITU report from 2019 reports that 47% of individuals in the developing world have internet access, and in the least developed countries 19 percent of individuals were online in 2019 [2].

Introducing new technologies into developing countries is generally a positive development, leading to new infrastructure and new business models that stimulate economic development. One consequence of the significant pace of technological development is the creation of large digital datasets. Large datasets are of huge economic value and form the backbone of many computing applications, particularly those that use machine learning, predictive analytics and business intelligence. Such techniques allow applications to synthesize billions of data points and make inferences

about users, how they behave now and how they will behave in the future. As technology companies have come to utilize data in products and services, extensive data protection laws and privacy regulations emerged to ensure ethical use of that data. These same protections, however, do not exist uniformly in the developing world. These conditions have led to accusations of digital colonialism, which has been defined by Kwet as “the decentralized extraction and control of data from citizens with or without their explicit consent through communication networks developed and owned by Western technology companies”[3].

Digital colonialism is one facet of a larger field of digital ethics that seeks to understand the impact of technology on our societies and the environment at large [4, 5]. It encompasses a range of issues and concerns from privacy and agency around personal information, digital literacy, big data including governance and accountability, the dominance of a small number of large network platforms, pervasive technology, the Internet of Things, surveillance applications, Artificial Intelligence and algorithmic decision making. It concerns how people, organizations, society and technology interact. With digital ethics comes the added variable of assessing the ethical implications of things, which may not yet exist, or things, which may have impacts we cannot predict.

The authors of this paper are leading an Erasmus+ project, Ethics4EU [6] that seeks to explore these issues via the teaching of ethics in computer science curricula. Research in the last decade shows that there is a paucity of ethics teaching in computer science, unlike other science disciplines [7]. One of the core project objectives is the development of a wide variety of educational content for the instruction of digital ethics to university students. Given the ethical issues surrounding digital data, a number of resources are being developed in this area. In this paper we present sample educational content that was developed as part of the project, specifically synthesized or fictionalized case studies, that focus on data collection from users in developing countries. The case studies are designed to serve as a way to improve computer science students’ ability at consequence scanning – a way to consider the potential consequences - intended and unintended -of a new technological product or service on people, communities and the planet [8].

2 Methods

A case study is a suitable vehicle for examining this topic as case studies explore specific real-world phenomena that focus on interpreting events, and exploring the societal context in which the case occurs [9]. The qualitative nature of these cases can be seen as novel when introduced in computer science courses which are typically more quantitative in nature. They can be used to both explore and evaluate specific problems and challenges of introducing new technologies into developing countries, as well as exploring digital ethics in a more general context. The materials allow for detailed examination of technological, organizational and social implications.

In this section we introduce two case studies we have developed as a part of a wider curriculum on digital ethics for computer science students. The case studies concern the impact of the introduction of new technology on persons and society in the developing world with a specific focus on how data is collected and used by third

parties as part of the introduction of new digital products and services. Each case study comprises a detailed narrative and set of questions (or “Talking Points”) to be used by an instructor in delivering the content. We also introduce a general checklist that can be used by students to evaluate scenarios involving the development of new digital products and services.

Please note that these case studies have been developed specifically as teaching tools; each is based on a synthesis of several real cases, and are designed to generate detailed and diverse discussions by student groups about the ethics of these scenarios. The use of synthesized case studies has a long history in the teaching of Law courses [10], sometimes to circumvent issues like confidentiality and legal privilege, which are clearly very important considerations when discussing Data Ethics scenarios. To highlight the fictitious nature of the case studies, fictional placenames and company names are very often used to underscore the fact that these case studies are fictional.

Synthesized case studies closely resemble a teaching approach that is already used in Computer Science, the “toy problem”, which is an approach used in the teaching of computer programming, where a scenario is created as an expository device to help students explore challenges around a specific topic [11]. These problems often distill some key features or challenges into simplified scenarios, and sometimes they combine several distilled features into one problem that would be unlikely to occur in a real world setting, but are very useful in teaching students about the challenges in a specific domain. Thus, these case studies are designed to highlight specific features or challenges that serve as the basis for the talking points to discuss with the students.

The first case study focuses on how new technology can result in divisions between a nation and external agencies, whereas the second case study explores how new technology can create divisions within a nation, and particularly how historical and cultural factors can play an important role in people’s perspectives on technology, and its impact.

2.1 Case Study 1: The Online Avalanche in Ishmaelia

Description:

Ishmaelia¹ is a developing country in East Africa, and in the past ten years its people have fallen in love with the online world; whether it is social media, search engines, or subscription television services; they cannot get enough of it. These online services are being provided mainly by a consortium of international companies, predominantly founded and based in the United States of America. The ease of using these services and technologies has had a very detrimental impact on the country’s local media, which is slowly fading away. Concurrently, much of the content that has already

¹ “Ishmaelia” is a fictional African country created by Evelyn Waugh for his 1938 novel “Scoop”

been created by the local media outlets has been bought up by these large organizations, and can only be accessed by subscribing to their services.

The Consortium argue that they are “*spending billions of euro in creating vital technological infrastructure that is bringing Ishmaelia into the 21st Century*”. It is true that they are collectively, and individually, providing a great deal of Internet infrastructure, as well as cloud-based data centres, and specialised software. They have also been digitising archives of newspapers and indigenous film, television and radio; and are also working closely with a number of national museums and universities to digitise (and preserve) important historical and archaeological artefacts that they are making freely available to scholars.

They also sponsor a number of national sports teams and individual athletes. One of the key goals of the Consortium is to harvest the vast quantity of data that is being generated by the people of Ishmaelia when they interact with these services. This data is a very valuable commodity, and can be used by the Consortium to expand their global reach; the data can also be processed to generate further consumer services, as well as being sold onto third-party organisations for marketing purposes. Ultimately, this results in the Consortium generating a large profit, the majority of which is transferred to the United States, and does not stay in Ishmaelia.

Over the past 18 months, the government of Ishmaelia have been trying to encourage the development of indigenous technology companies by providing taxation breaks and small state grants. They have had some very minor successes with this initiative, but in general these indigenous companies are not succeeding since the Consortium won't allow them access to their Internet infrastructure, and the government of Ishmaelia cannot afford to roll out its own publicly-funded Internet infrastructure. Those few companies that have been successful have had to emulate the marketing tactics, and data collection strategies of the Consortium.

The *Ishmaelian Digital Rights Activists Group* (IDRAG) is a growing political body who are deeply concerned with the digital trends that are occurring in their country, where they feel their people's personal information is being exploited by outside organizations. They also feel that their people's privacy is at risk from services that collect a range of information each time a service is accessed, including: geographical information, specific device information, a record of all of the other applications the users have launched (whether these applications are Consortium-owned or not), and a record of all key-clicks and mouse-clicks done while the application is running. According to IDRAG, this is essentially “*constant, 24 hour surveillance of everything you do, think, read and buy*”. They also point out that the Consortium have an unfair advantage, since not only do they have they technological infrastructure already in place, but they are also collecting more and more data that allow them to create services that are increasingly attractive to consumers. IDRAG has been working with the few successful indigenous technology companies to develop alternative services using Open Source Software as foundation for development.

The Consortium is aware of the bad publicity that they are engendering (in spite of the many good works they have contributed to this country), they have therefore decided to create a suite of educational services that they will make freely available to primary and secondary schools. This will include the development of a sophisticated

virtual learning environment, as well as creating a rich collection of educational content (including free textbooks) that is localised to national needs and this will also be supplied freely. The Consortium is about to start to roll out this initiative, and is also supplying free hardware, free installation, and free support service in exchange for having full access to all the data generated by these educational systems.

Suggested Talking Points:

1. Given that the Consortium have created the majority of Internet infrastructure in Ishmaelia, don't they have the right to refuse indigenous technology companies access to it? Why?
2. If the people of Ishmaelia are no longer supporting their own local media, should the government intervene to stop their existing media being bought up by the Consortium? Why?
3. When accessing the Consortium's services, the *Terms & Conditions* clearly state that all of the data generated, as well as other contextual data, is being copied to Consortium servers, what they are doing is clearly legal, but is it ethical? Why?
4. IDRAG has been working with the few successful indigenous companies to develop alternative services, and they call on the government to provide them with substantial funding to complete this project, if you were the official who had the decision-making power in this scenario, would you fund this initiative? Why?
5. IDRAG are arguing that the new educational services are simply a way of Ethics Washing, as well as a way of collecting more data about people (starting at a younger age), and this new service is also getting children accustomed to the Consortium's specific software so that they will be more likely to use this type of software in the future. IDRAG call on the government to stop this rollout. If you are the official in charge of making this decision, will you stop the rollout or not? Why?

2.2 Case Study 2: Mobile Mania in Qumran

Description:

Qumran², a developing country in the Middle East, that was historically part of the British Empire, and was a British protectorate until the late 1970s. As a consequence of these historical ties, most of the population have some proficiency in the English language. Qumran has seen an explosion in the use of smartphones in the past seven years. The country hasn't got a significant landline infrastructure, but with the advent of affordable Wifi and smartphones, internet usage is growing rapidly. A national survey was recently undertaken, and it found that Internet usage is currently

² "Qumran" is a fiction Middle Eastern country created by Antony Jay and Jonathan Lynn for their 1982 episode of "Yes Minister" entitled "The Moral Dimension"

at 57% of the population. There is a sharp divide in terms of age profile of users, with almost 90% of educated, younger adults (18- to 34-year olds) using these services; in contrast, amongst older adults (older than 34 years old) having a rate of 29% usage.

The first section of the survey asked respondents what services they access using the internet, and there was almost unanimous agreement that it serves as a very easy way to keep in touch with family members and friends who are geographically dispersed. A typical survey response stated that *"I can keep up with friends and family who are in other parts of the county, and I can see what they are up to. I also get to hear all the gossip from everyone as well"*.

The younger adults who participated in the survey also highlighted the significant educational benefits of Internet usage, commenting on the access to an enormous repository of resources, including online courses, MOOCs, textbooks, newspapers (from around the world), and magazines. They also noted that since the majority of Internet content is in English, they are ideally situated to get the most out of this content (as many of this age group are highly proficient English speakers). This demographic also use these services for online shopping, to access government services, and social media.

The older adults are cognizant of the benefits of Internet usage, but amongst this demographic there is a growing concern evidenced in the second section of the survey (as well as in the national media, and by some politicians) that there may be notable drawbacks to Internet usage. Of particular concern is their perception that the Internet is responsible for the erosion of the Qumrani culture, where younger people are eschewing traditional dress, customs and values; and seem more concerned about the latest gossip concerning American celebrities than they are about local and national issues. According to the survey, the most egregious aspect of this concern is the upsurge in the use of the English language which had been in the decline after the Qumrani Secessionist Revolt of 1978, and subsequent withdrawal from British protectorate status, but is now popular again because of prevalence of English language on the Internet.

Also in the second section of the survey 64% of the older adults expressed significant concern about the nature of the personal data they are required to share to access these services, including their date of birth, sometimes photographs, and sometimes credit card details for services that claim to be free. They also identified the issue of how their data is being used by these technology companies, how it is being collated with other people's data, and how it is being shared with third-party organizations. In contrast only 25% of younger adults expressed similar concerns, and in fact many younger adults expressed the opposite opinion, many commented that their *"personal data is a fair exchange for access to these important services"*.

In spite of these disagreements, there was general agreement on the negative impact of the Internet on morality in the third section of the survey. Even some of the most ardent advocates in the younger adult demographic rate the influence of the Internet on morality as being *"Neutral or Slightly Negative"*, whereas 73% of older adults rate its influence as *"Somewhat or Significantly Negative"*. Older adults highlighted specific concerns about the impact of the Internet on children, including the amount of time they spend using it, the lack of censorship of adult materials that they may inadvertently access, and the perceived increase in narcissism and perceived decline in

empathy in children. They also expressed the view that since the Internet can be accessed anonymously, it can lead to immoral behaviour such as cyberbullying, digital piracy, and identity theft.

Another area of significant agreement was in the final section of the survey on the importance of Internet access for commercial activities, and the majority believe that the Internet will be crucial for the future growth of the Qumrani economy. Some of the key benefits that the survey participants mentioned include: the possibility of remote working, improved productivity, online accounting systems, eMarketing, and new business opportunities. As such, the government have decided to develop a comprehensive Computer Growth Policy, which includes the development of online government services, as well as giving instructions to all of the colleges and universities to allocate more places for students doing computer science courses, and finally they have decided to reduce their Corporation tax to attract large multinational technology companies to house their offices and Data Centres in Qumran.

Suggested Talking Points:

1. Do the older adults in the survey have a reason to be concerned about the use of Internet access by younger adults who seem to be rejecting traditional Qumrani culture (particularly given the historical context)? Why?
2. Do the older adults in the survey have a reason to be concerned about the use of Internet access by children who have access to such a wide range of uncensored content? Why?
3. Do you think the statement "*personal data is a fair exchange for access to these important services*" is reasonable? Why?
4. Do you think that children who undertake illicit activities online such as digital piracy end up undertaking illicit or immoral activities in the real world? Why?
5. Do you think the Qumrani government is making the right choice by attempting to attract multinational technology companies? Why?

As well as the set of case study specific questions (or "Talking Points") presented at the end of each case study, we have also developed a generic case studies checklist that allows a student to examine any scenario using a range of criteria that describe digital products or services - the data collected and used, the features of the product or service, the organization, stakeholders (developers, consumers, society at large) and the consequences of the technology - intended and unintended. The checklist is intended for deeper reflection on specific aspects of the case studies and is to be used in conjunction with the "Talking Points" outlined above. The Case Studies checklist is based on work by Yin [12] and is shown in Figure 1.

CASE STUDIES CHECKSHEET

A task sheet for students to work through several times and internalise.

Name of Case Study: _____

Evaluation criteria	Notes
What is the case study about?	Introduction:
What is the organisation?	Introduction:
What are the technology issues?	Introduction:
Who are the principal actors?	Introduction:
What types of data were collected?	Data Collection:
From which sources did they come?	Data Collection:
How was the data recorded?	Data Collection:
What was the situation previously?	Main Features:
What interventions have been introduced?	Main Features:
What were the general outcomes of this intervention?	Main Features:
Are there any legal, social or ethical issues associated with this intervention?	Main Features:
Is there a chronological or other logic sequence for analysis?	Main Features:
What is the nature of the organisation?	Organisation:
What is its history?	Organisation:
How is it structured?	Organisation:
How has it changed as a result of intervention?	Organisation:
Who are the principal actors in detail?	People (Ecology):
What are their positions within the organisation?	People (Ecology):
What are their technical skills?	People (Ecology):
Does the target population for this intervention include more people?	People (Ecology):
What technology was present? What software? What hardware?	Technology:
What technical level of expertise exists within the organization?	Technology:
What new technology has been introduced for this intervention?	Technology:
How has the new technology effected the organisation?	Technology:
What are the possible consequences of this technology - intended and unintended?	Technology:
How successful has the intervention been?	Evaluation:
What new outcomes have been identified?	Evaluation:
What went well in this intervention?	Evaluation:
What did not go well in the intervention?	Evaluation:
What alternative approaches could have been taken?	Evaluation:

Fig.1 Case Studies Checksheet

3 Discussion and Conclusion

New digital products and services are being introduced into the developing world at a rapid pace. Many of these technologies are developed by external technology companies and collect large amounts of data about users and consumers who use the products and services. In this paper we have presented educational content that we have developed in order to examine some of the implications of the introduction of such new technologies and specifically on the issues around data generated and collected by new digital products and services. We have presented two case studies, a set of case study specific questions (“Talking Points”) and a generic case study checklist to be used in the instruction of computer science students to allow them to reflect on the consequences - intended and unintended - of new technologies for use in developing world contexts. Although the content is developed for computer science students it could be adapted for other educational disciplines.

All of the synthesized case studies that are being designed as part of the Ethics4EU project are created in pairs. The first usually more straightforward, focusing on tensions between external and internal factions (in this case, The Consortium versus the Ishmaelian culture), whereas the second one is always more nuanced and often looking at internal tensions between two factions (in this case, the older Qumrani people versus the younger ones). In this way, they work well as individual case studies, but also when taken as a pair they provide an interesting contrast.

After piloting these case studies with a small classgroup, some benefits of the synthesized case studies became clear; students commented that because they knew the scenarios were fictitious, they felt more comfortable elaborating new details about the cases, and they also felt more comfortable hypothesizing motivations of particular actors in the scenarios. They also commented that the case studies opened their eyes to some of the problems associated with technology that they had not thought of before. A few commented that the use of pre-existing fictional placenames made them curious to follow-up on those references, and to explore some literature.

In future work we intend to develop a larger range of educational content for the instruction of digital ethics. Content will focus on pertinent issues such as privacy, computer security, surveillance and facial recognition, the Internet of Things, AI and algorithmic decision making including biases such as racial and gender biases often present in large datasets and the environmental implications (specifically the carbon footprint) of storing excessive quantities of data in data centres.

We intend to evaluate the educational materials with students in the classroom, gathering feedback from students on the educational instruments and evaluating their before and after understanding of the ethical issues raised in the case studies. We also intend to develop an additional instrument for assessing the ethical implications of introducing new technologies for the developing world based on Hofstede's Cultural Dimensions Theory [13].

The theory is a cultural model that attempts to represent the social values of different countries using six dimensions (the dimensions are: *Power Distance*, *Individualism*, *Masculinity*, *Uncertainty Avoidance*, *Time Orientation*, and *Indulgence*). This model is commonly used to assist computer science students in the design of online content [14], and would therefore be something that the students would be already familiar with, and they would readily be able to apply this model in the context of these case studies.

ACKNOWLEDGEMENTS

The authors of this paper and the participants of the Ethics4EU project gratefully acknowledge the support of the Erasmus+ programme of the European Union. The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

- [1] World Economic Forum, How much data is generated each day, <https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/> , last accessed 2020/03/06.
- [2] Measuring digital development Facts and figures 2019, International Telecommunication Union Publications, <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf>, last accessed 2020/13/06.
- [3] Kwet, M, Digital Colonialism: US Empire and the New Imperialism in the Global South. *Race & Class* Volume 60 (4) (2019)
- [4] Russo, F. Digital Technologies, Ethical Questions, and the Need of an Informational Framework. *Philos. Technol.* 31, 655–667 (2018)
- [5] Floridi, L. Foundations of Information Ethics" In: *The Handbook of Information and Computer Ethics*. John Wiley & Sons, Inc.. 1-23 (2009)
- [6] Ethics4EU, Erasmus+, www.Ethics4EU.eu, last accessed 2020/03/06.
- [7] Spradling, C., Soh, L.K., Ansoorge, C. Ethics Training and Decision-Making: Do Computer Science Programs Need Help?. In *Proceedings of the 39th SIGCSE technical symposium on Computer science education*, 153-157, 2008.
- [8] Doteveryone, Consequence Scanning. An Agile Practice for Responsible Innovators <https://www.doteveryone.org.uk/project/consequence-scanning/>, last accessed 2020/03/06.
- [9] Martin DA., Conlon E., Bowe B. A Constructivist Approach to the use of Case Studies in teaching Engineering Ethics. In: Auer M., Guralnick D., Simonics I. (eds) *Teaching and Learning in a Digital World. ICL 2018. Advances in Intelligent Systems and Computing*, vol 715. Springer
- [10] Dyer, B., Hughson, M.A., Duns, J., Ricketson, S. "Teaching Note: Creating a Corporations Law Case Study", *Legal Education Review*, 8 (1997)
- [11] Pearl, J., "Heuristics: Intelligent Search Strategies for Computer Problem Solving, Addison Wesley. (1984)
- [12] Yin, R.K., *Case Study Research and Applications: Design and Methods*, Sage Publications, 2017.
- [13] Hofstede, G., "Culture's Consequences: International Differences in Work-Related Values", SAGE Publications (1984)
- [14] Ahmed T, Mouratidis H, Preston D. Website design guidelines: High power distance and high context culture. *International Journal of Cyber Society and Education*. 2009 Jun 30;2(1):47-60.