

COMP 2420 – Intro to Data Mngt, Anal & Security

1. IT and Data Ethics

Roger Clarke

Xamax Consultancy, Canberra
Visiting Professor, A.N.U. and U.N.S.W.

<http://www.rogerclarke.com/DV/Comp2420.html#L1>
<http://www.rogerclarke.com/DV/Comp2420-1> {ppt, pdf}

ANU RSCS – 24 April 2020

IT & Data Ethics

- 1. Introduction**
IT's power, impact, implications
Ethics
Codes of Ethics of the ACS and others
- 2. Ethics in Data Science / Data Analysis**
Ethical Issues, and the Harm Arising
Case Studies: Robo-Debt, Data Scrubbing,
Transparency, Digital Surveillance Economy
- 3. Applied Ethics**
Codes
Guidelines
A Business Process with Embedded QA

IT's Power has Impacts and Implications

- **Computing**
Power, Miniaturisation, Cost, but Inherent Insecurity
- **Communications**
Capacity, Availability / Ubiquity, Cost
- **Data**
Capture, Digitisation, Storage, Access, Openness
(Id)entification, Authentication, Biometrics, Surveillance
- **Robotics / Actuator Technologies**
Diversity, Effectiveness *in situ*, and in controlled
environments, Conflicts in uncontrolled contexts
- **Cyborgisation / Orthotics**
Wheelchairs, blade-runners, but rights?

Ethics

A body of principles governing right and wrong

cf. **Morality**

Each individual's own sense of right and wrong

Ethics

A body of principles governing right and wrong

May be applied retrospectively, to enable abstract, *ex post facto* judgements about good and evil

and/or

May be applied prospectively with volitional or motivational power that influences actors' behaviour

Some Ethical Issues

Ethical Issues arise from conflicts among stakeholders' interests

Economic

- Income Distribution
- Casualisation of Labour
- Work-Dependence of Income (cf. 'a living wage')

Environmental

- Habitat Destruction
- Climate Change

Political

- Location and Tracking

Technological

- Nuclear Power
- Robotic Warfare

Social

- Capital Punishment
- Unfair Discrimination
- Gender Equality
- Continuous Disruption (Workplace, Occupations)

Some Ethical Issues

Ethical Issues arise from conflicts among stakeholders' interests

Economic

- Income Distribution
- Casualisation of Labour
- Work-Dependence of Income (cf. 'a living wage')

Environmental

- Habitat Destruction
- Climate Change

Political

- Location and Tracking

Technological

- Nuclear Power
- Robotic Warfare

Social

- Capital Punishment
- Unfair Discrimination
- Gender Equality
- Continuous Disruption (Workplace, Occupations)

Categories of 'Persons-at-Risk'

Ethical Issue: Data Exposure may be Life-Threatening

Social Contexts

- Celebrities and notorieties at risk of extortion, kidnap, burglary
- Short-term celebrities such as lottery-winners, victims of crime
- Victims of domestic violence
- Victims of harassment, stalking
- Individuals subject to significant discriminatory behaviour
- People seeking to leave a former association, e.g. ex-gang-members

Political Contexts

- Whistleblowers
- Dissidents

Organisational Contexts

- Corporate executives
- Government executives
- Undercover operatives
- Law enforcement and prison staff
- Mental health care prof'ls, counsellors

Legal Contexts

- Judges, lawyers and jurors, particularly in highly-charged cases
- Witnesses, especially people in protected witness programs
- Ex-prisoners re-integrating with society

The Codes of Ethics of the IT Profession(s)

- Australian Computer Society (ACS)
http://www.acs.org.au/content/dam/acs/rules-and-regulations/Code-of-Professional-Conduct_v2.1.pdf
- Association for Computing Machinery (ACM)
<https://www.acm.org/code-of-ethics>
- IEEE
<https://www.ieee.org/about/corporate/governance/p7-8.html>
- Engineers Australia
<https://www.engineersaustralia.org.au/sites/default/files/resource-files/2020-02/828145%20Code%20of%20Ethics%202020%20D.pdf>

The ACS Code

As an ACS member you must uphold and advance the honour, dignity and effectiveness of being a professional.

This entails, in addition to being a good citizen and acting within the law, your conformance to the following ACS values.

1. The Primacy of the Public Interest

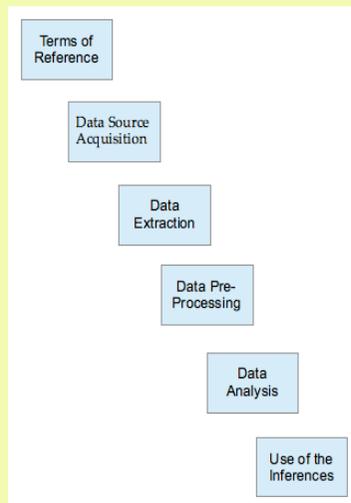
You will place the interests of the public above those of personal, business or sectional interests.

2. The Enhancement of Quality of Life

You will strive to enhance the quality of life of those affected by your work.

...

2. Data Science / Data Analytics



Ethical Issues in Data Science

- Data Expropriation for Unintended Purposes
- Data Quality Assurance
- Data Security
- Data Analysis Quality Assurance
- Unfair Discrimination
Redlining, Weblining, 'Algorithmic Discrimination'
- Decision-Making delegated to Artefacts
- Transparency of Decision-Rationale
- Due Process / Procedural Fairness
- The Digital Surveillance Economy and 'Surveillance Capitalism'

Categories of Harm

- **Data** Loss, Alteration, Access, Replication
- **Property** Damage
- **Personal** Injury
- **Asset Value** Loss
- Financial Loss
- Reputation or Confidence Loss
- Opportunity Cost

Values Associated with Data that may be harmed by Data Analytics

- **Inaccessibility** (**C**onfidentiality)
 - Data Access
 - Data Disclosure
 - Data Interception
- **Quality** (**I**ntegrity)
 - Data when Collected
 - Data when Used
 - Modification
 - Corruption
 - Staleness
- **Accessibility** (**A**vailability)
 - Data Existence
 - Data Loss
 - In Volatile Memory
 - In Non-Volatile Memory
 - Theft, Destruction, Malfunction
 - Data Inaccessibility

Case Studies of Ethical Issues



#1: Robo-Debt



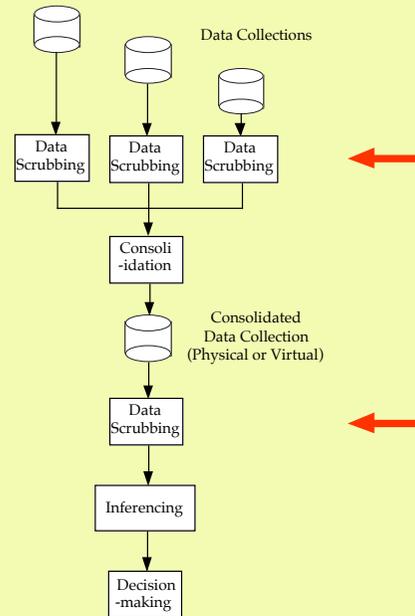
- **ATO** collects data relating to the financial year
- **Centrelink** relies on more finely-grained data: the fortnightly income of each welfare client
- **Centrelink** divided **ATO's** annual figure by 26, and assumed it applied to each fortnight
- **Centrelink** inferred that many clients had **mis-reported their income and been overpaid**
- **Centrelink** declared those people owed money
- x30 Leap in case-load, so complaints were ignored
- Centrelink hired heavy-handed debt-collectors
- People suffered badly for 3 years as a result

Case Study #2: Data Scrubbing / Cleaning / Cleansing

- What it is
- What good it can do
- What bad it can do

Big Data & Big Data Analytics

Process View



Data Scrubbing / Cleaning / Cleansing

- **Problems It Tries to Address**
 - Missing Data-Items
 - Low and/or Degraded Data Quality
 - Failed and Spurious Record-Matches
 - Differing Data-Item Definitions, Domains, Applicable Dates
- **How It Works**
 - Internal Checks
 - Inter-Collection Checks
 - Algorithmic / Rule-Based Checks
 - **Checks against Reference Data – ??**
- **Its Implications**
 - Better Quality and More Reliable Inferences
 - **Worse Quality and Less Reliable Inferences**



Case Study #3: Transparency

- **Accountability** depends on clarity about the Decision Process and the Decision Criteria
- In practice, Transparency is highly variable:
 - Manual decisions – Often poorly-documented
 - Algorithmic languages
Process & criteria explicit (or at least extractable)
 - Rule-based 'Expert Systems' software
Process implicit; Criteria implicit
 - Empirical software / AI/ML / Neural Networks
Process implicit; Criteria not discernible



Case Study #4: The Digital Surveillance Economy

That combination of institutions, institutional relationships and processes, which enables **corporations** to **exploit data** that arises from the monitoring of **people's electronic behaviour** and on which consumer marketing corporations have become dependent

The Textbook Message

"[marketers'] strategy should be to bribe users to give [marketers] the appropriate demographics, which in turn can be passed onto advertisers ..."

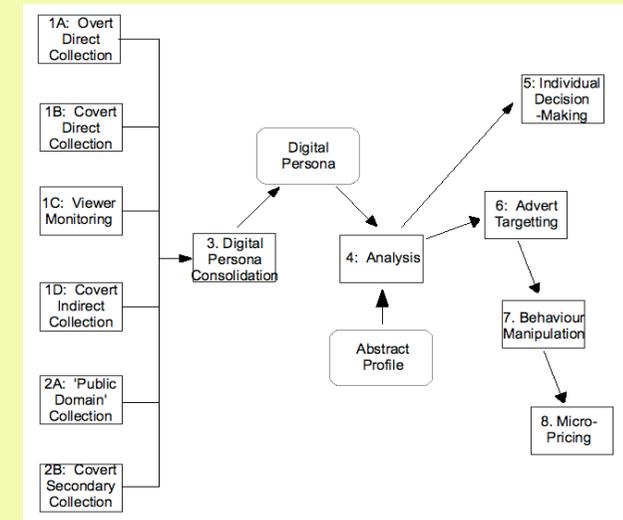
"[The game is about] inducing consumers to give [marketers] the information they want."

"we expect that many consumers will be happy to sell information about themselves for a nominal amount ..." (pp. 35-36)

Shapiro C. & Varian H.R. (1999)

'Information Rules: A Strategic Guide to the Network Economy'
Harvard Business School Press, 1999

The Digital Surveillance Economy



(4) Analysis

- Inferences are drawn about an individual, e.g.
- Compare each digital persona with one or more 'abstract consumer profiles', which may be:
 - *ad hoc* / rule-of-thumb / heuristic
 - based on studies and experiments re personality-types, attitudes and/or long-term and short-term interests
 - the 'Big Five' psychographic classification: extraversion, neuroticism, agreeableness, conscientiousness, and openness

(6) Ad Targeting

- Web-sites that attract consumers' attention rent out space in the user's browser-window
- **Based on each consumer's digital persona, ads are selected and/or customised**
- The process reflects the persona's recorded demographics, preferences, attitudes and interests
- Such 'narrowcast' ads are much more effective than old-style, mass-market 'broadcast' ads
- A highly-developed, real-time process auctions ad space in consumers' browser-windows

(8) Micro-Pricing

- Each consumer's digital persona is rich, and enables marketers to gauge the point at which buyer-resistance is likely to arise
- **So the offer can be pitched just below the individual's resistance-point, thereby extracting the maximum revenue from each person**
- This disadvantages most consumers compared with longstanding pre-set pricing – which is based on what the-market-as-a-whole will bear
- This is little-understood by consumers, who naively accept the marketer's pitch that the consumer is being given a 'special offer'

Summary of Ethical Issues in Data Science Data Analytics as Fuzzy Inferencing

- Data Selectiveness
Only some data is captured
- Data Specificity
Collected for a purpose
- Data Quality
Quality costs, so compromises occur
- Data Suitability
Collected for one reason, used for another
- Process Complexity
Few understand it
- Result Obscurity
Few can explain it

*'If you torture data long enough
it will confess to anything'*



attr. Ronald Coase (1981)

"How should economists choose?" Warren Nutter Lecture
orig. Darrell Huff (1954) 'How to Lie With Statistics'

3. Applied Ethics

- Quality Assurance
 - Data
 - Analytical Processes
- Laws
- Codes
- Guidelines

Specific Industry and Professional Codes

UNSD (1985) 'Declaration of Professional Ethics'
United Nations Statistical Division, August 1985, at <http://unstats.un.org/unsd/dnss/docViewer.aspx?docID=93#start>

ASA (2016) 'Ethical Guidelines for Statistical Practice'
American Statistical Association, April 2016, at <http://ww2.amstat.org/about/pdfs/EthicalGuidelines.pdf>

DSA (2016) 'Data Science Code Of Professional Conduct'
Data Science Association, undated but apparently of 2016, at <http://www.datascienceassn.org/sites/default/files/datasciencecodeofprofessionalconduct.pdf>

UKCO (2016) 'Data Science Ethical Framework'
UK Cabinet Office, v.1, 19 May 2016, at <https://www.gov.uk/government/publications/data-science-ethical-framework>
<http://www.rogerclarke.com/DV/DSEFR.html>

Guidelines for Responsible Data Analytics

Safeguards re Data Acquisition

- | | |
|--|--|
| <ol style="list-style-type: none"> 1. The Problem Domain
Understand the relevant real-world system 2. The Data Sources
Understand each source of data 3. Data Merger
Investigate whether it's tenable 4. Data Scrubbing
Investigate whether it helps | <ol style="list-style-type: none"> 5. Identity Protection
Nymise sensitive associations of data with entities 6. Data Security
Investigate, minimise, manage and mitigate risks 7. Identifier Compatibility
Investigate the risks of erroneous data merger 8. Content Compatibility
Investigate the compatibilities among data sources |
|--|--|

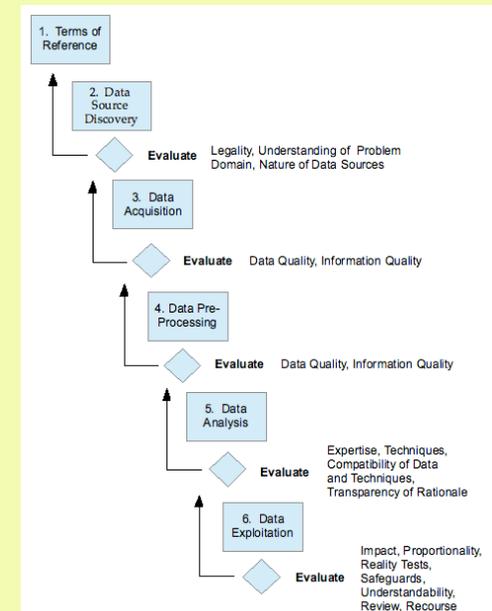
Guidelines for Responsible Data Analytics

Safeguards re Data Analysis

- | | |
|---|--|
| <ol style="list-style-type: none"> 1. Expertise
Ensure qualifications, training, real-world understanding 2. The Nature of the Tools
Understand the data analytics techniques and tools 3. The Nature of the Data Processed by the Tools
Understand the assumptions the tools make re missing values, allowed values, scales, precision | <ol style="list-style-type: none"> 4. The Suitability of the Tool and the Data
Ensure the data fits the tool 5. Inappropriate Data
Don't apply data analytics tools if the data isn't up to it 6. Humanly-Understandable Rationale
Don't use a tool unless you understand the answers it gives |
|---|--|

A Data Analytics Business Process with Embedded QA

<http://rogerclarke.com/EC/BDBP.html>



IT & Data Ethics

1. Introduction

IT's power, impact, implications

Ethics

Codes of Ethics of the ACS and others

2. Ethics in Data Science / Data Analysis

Ethical Issues, and the Harm Arising

Case Studies: Robo-Debt, Data Scrubbing,
Transparency, Digital Surveillance Economy

3. Applied Ethics

Codes

Guidelines

A Business Process with Embedded QA

COMP 2420 – Intro to Data Mngt, Anal & Security

1. IT and Data Ethics

Roger Clarke

Xamax Consultancy, Canberra

Visiting Professor, A.N.U. and U.N.S.W.

<http://www.rogerclarke.com/DV/Comp2420.html#L1>
[{.ppt, .pdf}](http://www.rogerclarke.com/DV/Comp2420-1)

ANU RSCS – 24 April 2020