Legal and Ethical Issues

Fundamentals of Data Management

Includes material from
Legal Issues in Research Data Collection & Sharing

Paweł Kamocki, EUDAT. Retrieved May, 2014.
http://www.eudat.eu/training-materials-downloads

Dr Rob Baxter
Software Development Group Manager, EPCC
r.baxter@epcc.ed.ac.uk
+44 131 651 3579 | +44 7971 437749

Course outline

- What do we mean by “legal issues”? “Ethical issues?”
- What is copyright? What’s a “database right”?
- I’m a computational scientist – why the heck should I care?

- After completing this lesson, you should:
  - Understand the basics of copyright with respect to data
  - Understand the licensing of data and databases
  - Understand the issues around using personal data in research

Caveat lector!

- I am not a lawyer!
- This does not constitute legal advice!
- This material comes with no warranty , express or implied, as to fitness for any purpose whatsoever!
- Your mileage may vary!

Science is about…

- **Community**
  - the accumulation of scientific understanding emerges through the interactions among individual contributors
- **Universality**
  - entry into scientific work and discourse open for all persons of competence
- **Disinterestedness**
  - the nature of what is being sought should not be of personal interest to the researchers in order not to skew their methods
- **Originality**
  - the touchstone of acceptance of priority claims, and the source of collegiate reputations upon which rewards are based
- **Scepticism**
  - appropriate attitude towards all priority claims

Reproducible science

- **Reproducibility**
  - the ability of an entire experiment or study to be reproduced, or by someone else working independently

"[a]n article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."

- Jon Claerbout, Professor Emeritus, Stanford University

- 47 out of 53 medical research papers on cancer were irreproducible – *Nature*, 2012

The Law is about…

- **Exclusive rights for authors, producers, data subjects**

The owner of the copyright in a work of any description has the exclusive right to…

ch. 48, I (1) CDPA

The data subject's right of access and to rectification, erasure or blocking may not be excluded or restricted by a legal transaction.

s. 6 (1) BDSG

The producer of the database has the exclusive right to reproduce and distribute the database as a whole or a qualitatively or quantitatively substantial part of the database and to make this available to the public.

art. 87b (1) UrhG

Science vs Law

Science

Publications, Presentatio

- **C**ommunalism
- **U**niversalism
- **D**isinterestedne
- **O**riginality
- **S**cepticism

Reproducibility
Ethical rules
Gift economy

Law

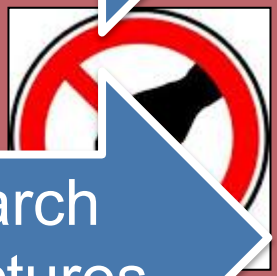
ves, Regulations, Conventions…

- Copyright
- tents
- bases
- onal data protection

You can't touch this!

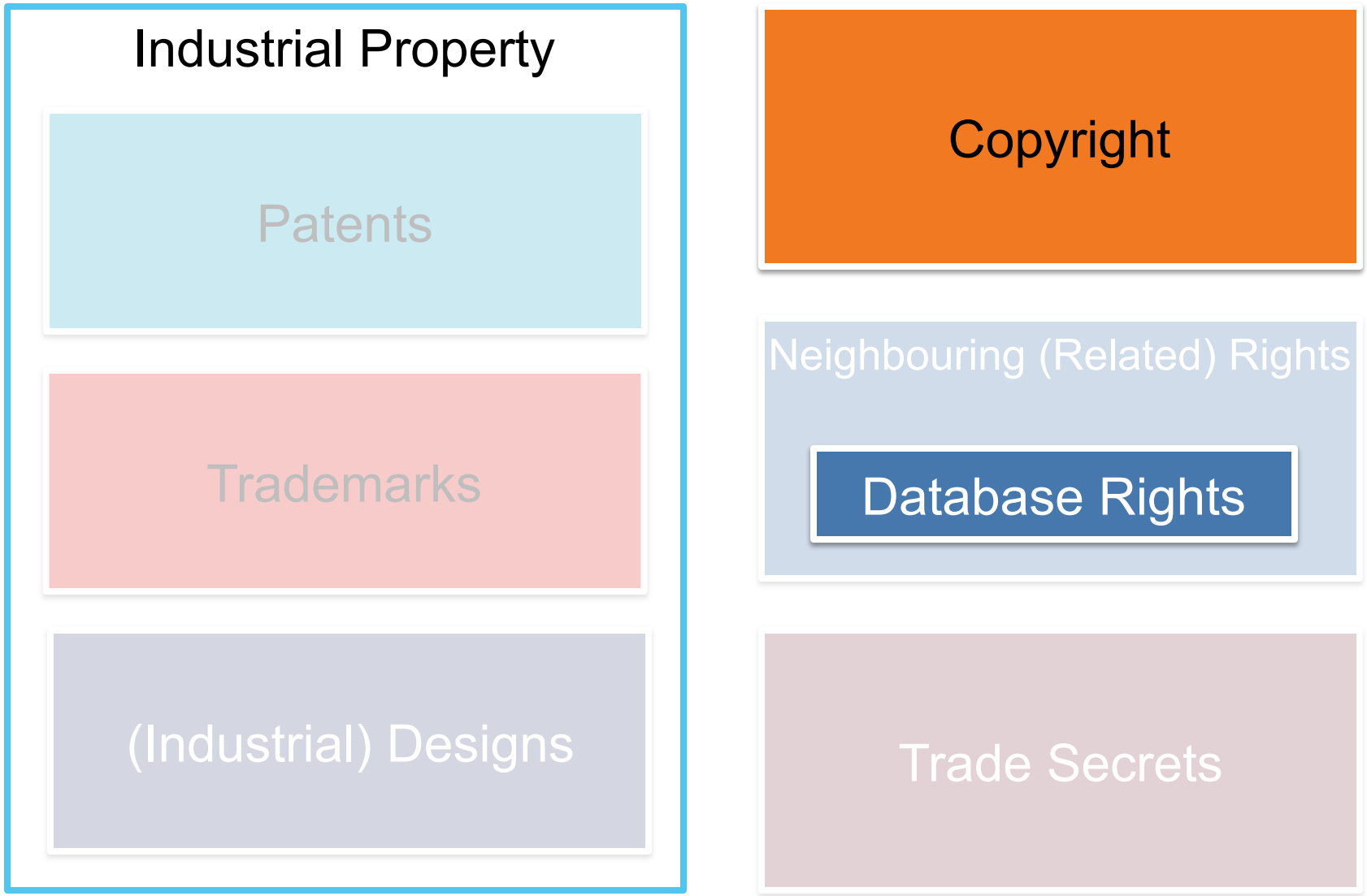
Xerox machine

The Internet



E-research Infrastructures

Intellectual property rights



The history of copyright

- Berne Convention for the protection of literary and artistic works 1886
- Agreement on Trade-Related Aspects of Intellectual Property Rights 1994
- World Intellectual Property Organisation's Copyright Treaty 1996
- Directive 2001/29/EC of 22 May 2001 on the harmonization of certain aspects of copyright in the information society
- Directive 93/98/EEC harmonizing the term of protection of copyright
- Directive 2009/24/EC on the legal protection of computer programs
- National implementations (CDPA, UrhG, Code de la propriété intellectuelle…)

Copyright

- Many issues around the legal use or reuse of data stem from copyright
- So, what is it?
- What is protected by copyright?
- What rights are reserved (by the owner)?
- For how long does the legal protection last?
- Are there any exceptions?

- Detailed answers are, of course, dependent on jurisdiction!

What does copyright protect?

- Scientific, literary and artistic works, *regardless of their value*
- Copyright does **not** protect ideas, raw facts, mathematical formulae
  - So scientific measurements of natural phenomena are not copyright
- A key measure is originality
  - personal imprint; selection and arrangement; labour, skill and judgement; the author's own intellectual creation
- Also covers
  - derivative works – translations, adaptations
  - compilations and collections
- Copyright is automatic
  - You don't have to claim it, you just get it when you create something

What does copyright grant you?

- Typically *economic* and *moral* rights
- If you own copyright in *X*, you reserve the economic right to
  - make copies of it
  - make it available to the public (eg. via publication or upload)
  - make derivative works
- and the moral right to
  - assert paternity of it (right of attribution)
  - disclose it (right of first publication)
  - have it respected (right of protection against distortion)

How long does copyright last?

- In the EU and in the US:
  - 70 years after the death of the author
- Other countries
  - min. 50 years after the death of the author

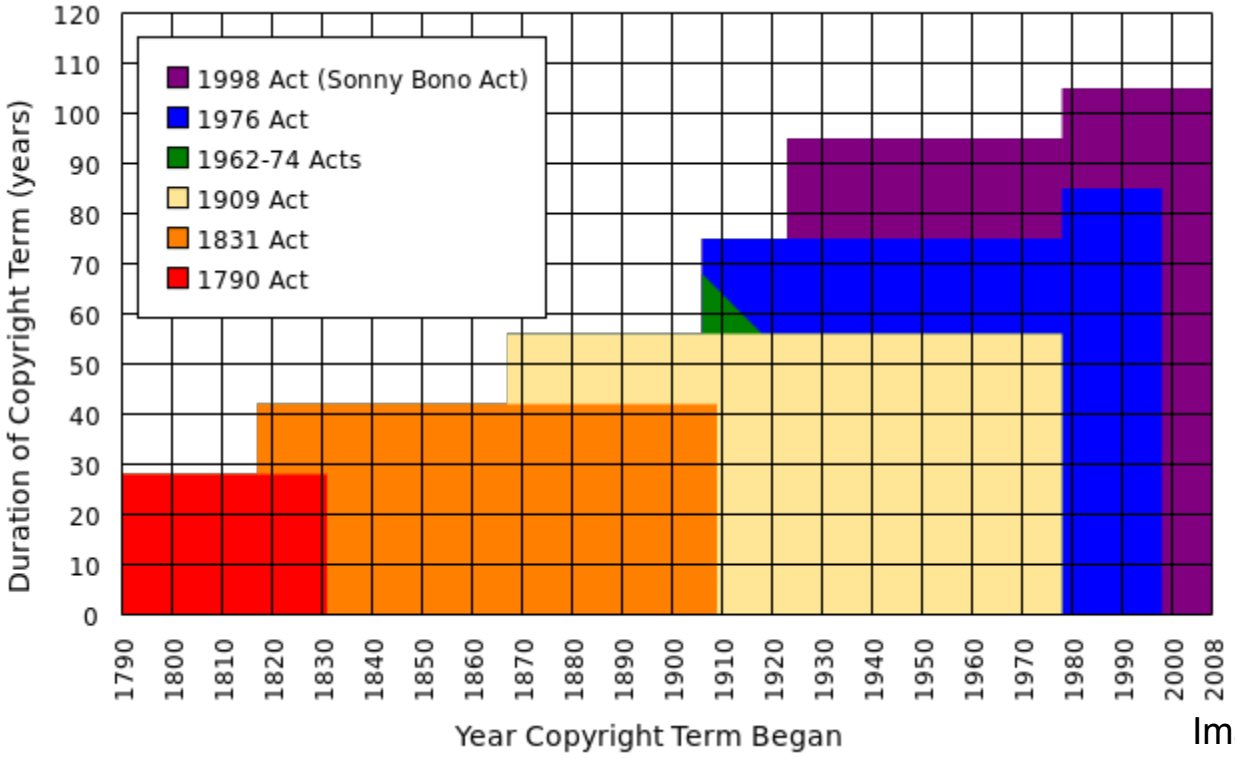


Image © Tom W. Bell

Are there any exceptions?

- You bet!
- Scope of exceptions varies significantly across jurisdictions
- Often cover things like
  - Non-commercial research (can anyone define non-commercial? Or research?)
  - Private copies
  - Text and data mining
  - Fair use (in the US)

> Fair dealing with a literary, dramatic, musical or artistic work for the purposes of research for a non-commercial purpose does not infringe any copyright in the work provided that it is accompanied by a sufficient acknowledgement.
>
> s. 29 CDPA

Database right

- A comparatively new idea in European law
  - EU Directive of 11 March 1996 on the legal protection of databases
- National implementations are usually word-for-word
- Does not apply to companies from countries which do not provide for an equivalent level of protection for EU databases
  - i.e. the United States
- Database right protects
  - a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.
  - qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents

Database rights & science

- a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.
- qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents

- i.e. scientific measurements are not covered by intellectual property rights, *but their collection and organisation may be*

Database rights…

- prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of the database
- prevent repeated and systematic extraction and/or re-utilization of insubstantial parts of the contents of the database implying acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database
- for 15 years after each “substantial investment”
  - Potentially, forever!

Licensing

- **License**, *verb,* to give someone official permission
- **Licence**, *noun*, an official document that gives you permission
- A licence is a way of allowing people to use your stuff subject to certain conditions
  - It's a promise not to sue
- Licence ≠ transfer (sale ≠ rental)
- Licences may reserve some rights to the owner, or none
- Rights that can be licensed
  - Copyright
  - Database rights

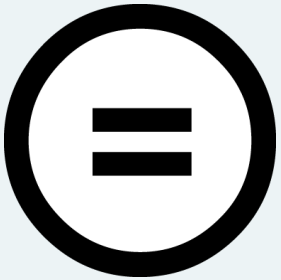
Using licences

- If you want to use someone's data, you must agree to the terms of the licence under which those data are released
- If you can't find a licence, then either
  - (a) the data are public domain and you can do what you like, or
  - (b) you do not have permission to use the data
  - ***You need to ensure you know which is which!***
    - Ignorantia juris non excusat!
- If you have created data or a database and wish to share, you should choose a licence under which to release them
  - ***Only make sure you have the right to do so!***
    - Don't try to sell the Eiffel Tower – it's not yours!

Creative Commons

- CC Foundation started in 2002
- CC is a public licence suite + CC0 (= rights waiver)
- Versions: 1.0, 2.0, 2.5, 3.0; 4.0 launched in November 2013
  - nearly 50 ported (national) versions
  - no ported versions of CC 4.0
- Three layers: machine-readable, lawyer-readable, human-readable
- Used e.g. by
  - Al Jazeera, Flickr, Google, Open Course Ware, Public Library of Science, Radiohead, Wikipedia, whitehouse.gov

Creative Commons aspects

	BY Attribution	You must: identify the creator(s) of the Licensed Material (…); retain a copyright notice; retain a notice that refers to this Public License (and) to the disclaimer of warranties; retain a URI or hyperlink to the Licensed Material if reasonably practicable; indicate if You have modified the Licensed Material (…)
	ND No Derivatives	Adapted Material means material (…) that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner (…)
	NC Non Commercial	not primarily intended for or directed towards commercial advantage or monetary compensation. (…) exchange of the Licensed Material for other material subject to Copyright and Similar Rights by digital file-sharing or similar means is NonCommercial provided there is no payment of monetary compensation in connection with the exchange.
	SA ShareAlike	if You produce and Share Adapted Material, (…) the License You apply must be a Creative Commons license with the same License Elements, this version or later, or a Compatible License.

Licensing and Open Data

- Strong movement towards Open Access to research data
  - Budapest Open Access Initiative Statement 2002
  - OECD Principles 2007
  - RCUK Common Principles on Data Policy 2011
  - G8 Open Data Charter 2013
- If you create research data funded by public money, you will need to make special arguments *not* to release them openly
- Can still protect your intellectual investment through “open-access compatible” licensing

Openness and Creative Commons

Other open public licences

- Open Database Licence (ODbL)
- Open Data Commons Attribution Licence (OCL)
- GNU Free Documentation Licence (for software documentation)
- Open Game Licence
- Free Art Licence (for artistic works)
- Open Source is a similar, but different movement (software whose source code is available for inspection and modification)

Copyright vs public domain

Public Domain

Non-copyrightable
Ideas, raw facts, 'works of nature', official documents

Copyright-expired
> 70 years post mortem auctoris

Copyright waived

Copyrighted Works

Orphan Works

Open Access

Personal data

- Personal data are subject to different laws **as well as** copyright and/or database right
- Personal data are
  - "...any information relating to an identified or identifiable person..."
    - art. 2 (a) Directive 95/46/EC
  - i.e. there is a data ***subject*** as well as a data creator
- Personal *sensitive* data relate to
  - race, religion, health, political opinions, sexual life, criminal history
- In research terms…
  - medical records, medical images, interviews, samples of speech, names, addresses…

Working with personal data

- *Consent* of the data subject is a key issue
  - Getting it (but what if they're dead?)
  - Recording it (consent needs to be specific; what about unforeseen future uses?)
- *Anonymisation* may be an acceptable way to manage reuse
  - But how easy is anonymisation for, eg, medical images?
  - And how easy is it to anonymise against all possible "mashups"?
- Anyone working with personal data falls under the Data Protection Act
  - ***You need to be aware of your responsibilities!***
    - Ignorantia juris non excusat, again!

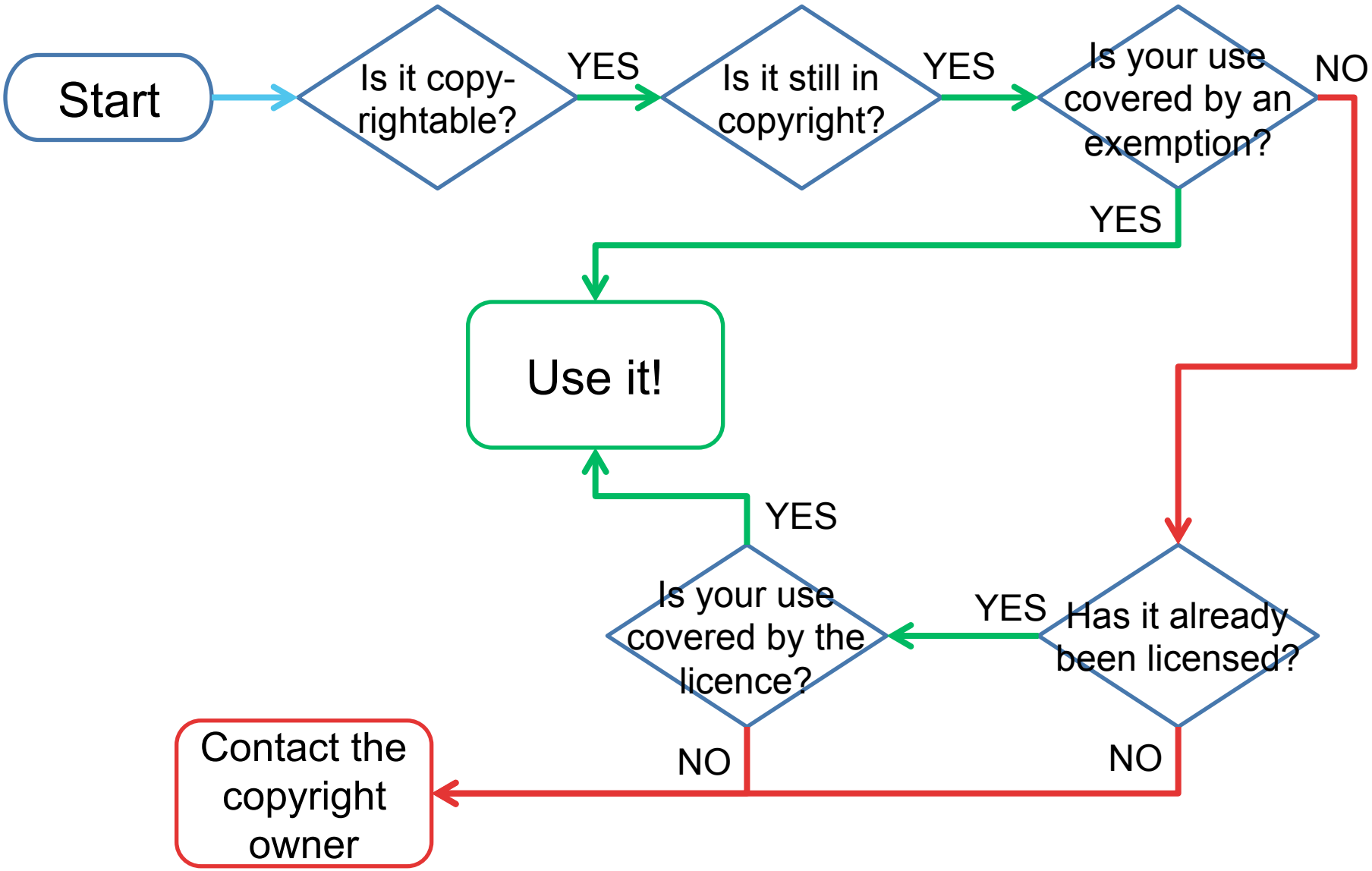
Ethical considerations in research data

- Personal data is one area which raises ethical as well as legal concerns
- Open Access raises other potential ethical concerns:
  - Unintended secondary uses and misappropriation
    - Your data may be used for purposes you didn't anticipate, or to bolster arguments you do not support
  - Dual use
    - E.g. data on epidemiological pathways will undoubtedly be a benefit for public health, but could be used for terrorism
  - Violations of privacy and confidentiality
    - E.g. interviews with criminal gang members, persecuted minorities
  - Unequal distribution of research results
    - If the data need supercomputers to process them, are they truly open for those of lower economic muscle?

Why the heck *should* you care?

- The era of open research data is here
- You will need to publish data as well as papers
- You will want to re-use other researchers' data in a professional capacity

Summary: how to use a dataset



Summary: how to deal with your data

- **Check**
  - Your institution's policy
  - Your obligations
  - Your funding body's requirements
- **Check**
  - Copyright and (maybe) IP rights
  - Database rights
  - Personal data
- **Choose**
  - A licence or waiver
    - Think about multiple licences, openness, interoperability
- **Contact**
  - A data intermediary
    - Data Controller, Data Librarian, Data Repository

Acknowledgements

- Includes material from
  - Legal Issues in Research Data Collection & Sharing
    - Paweł Kamocki, EUDAT. Retrieved May, 2014.
    - http://www.eudat.eu/training-materials-downloads