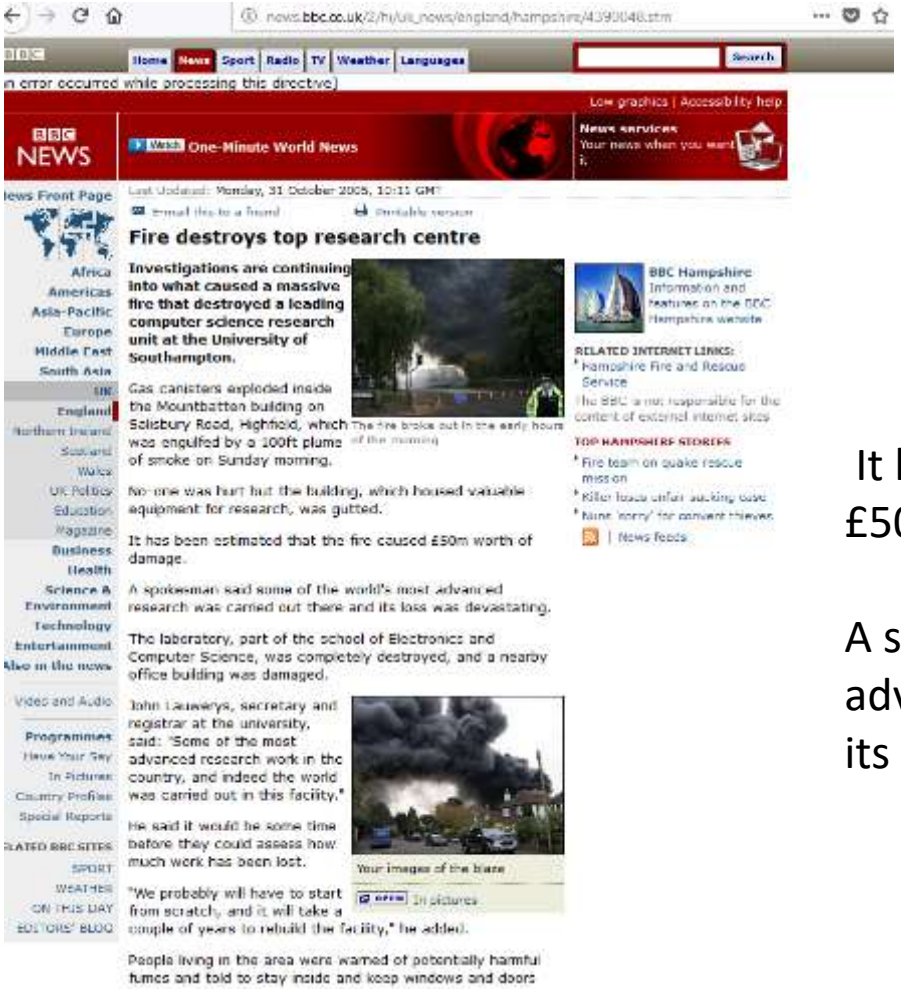
Does this excite you?

Why is RDM important?



BBC NEWS

One-Minute World News

Fire destroys top research centre

Investigations are continuing into what caused a massive fire that destroyed a leading computer science research unit at the University of Southampton.

Gas canisters exploded inside the Mountbatten building on Salisbury Road, Highfield, which was engulfed by a 100ft plume of smoke on Sunday morning.

No one was hurt but the building, which housed valuable equipment for research, was gutted.

It has been estimated that the fire caused £50m worth of damage.

A spokesman said some of the world's most advanced research was carried out there and its loss was devastating.

The laboratory, part of the school of Electronics and Computer Science, was completely destroyed, and a nearby office building was damaged.

John Lauwerys, secretary and registrar at the university, said: "Some of the most advanced research work in the country, and indeed the world was carried out in this facility."

He said it would be some time before they could assess how much work has been lost.

"We probably will have to start from scratch, and it will take a couple of years to rebuild the facility," he added.

People living in the area were warned of potentially harmful fumes and told to stay inside and keep windows and doors

It has been estimated that the fire caused £50m worth of damage.

A spokesman said some of the world's most advanced research was carried out there and its loss was devastating.

ESB liability for UCC campus flooding overturned by Court of Appeal

Liability for legal costs estimated at several million euro yet to be decided

Tue, Mar 20, 2018, 17:49

Mary Carolan

The Court of Appeal ruled that damage to UCC buildings arose "from a natural event", the ESB did not cause the flooding of UCC's buildings and it had no legal duty to avoid unnecessary flooding. Photograph: Matt Kavanagh

The Court of Appeal, in a significant judgment, has ruled the Electricity Supply Board (ESB) is not liable for extensive flood damage to buildings on the campus of University College Cork (UCC) after floods affected large parts of Cork city in late 2009.

The three judge court, in a lengthy judgment on Tuesday, overturned a Hi Court finding the ESB was 60 per cent liable in respect of flooding and warnings.

Outline

What is “research” data?

What’s your data management plan?

What is your data organisation style?

Some technical information (formats etc.)

Metadata and keeping track of your data

GDPR, security and safety

UCC

For me a win is:

1. **You become active managers of your data**
2. **You go away with a concrete plan of what you are going to do**

Question: Does good research **REQUIRE** good data management?

"If a cluttered desk is a sign of a cluttered mind, of what, then, is an empty desk a sign?"

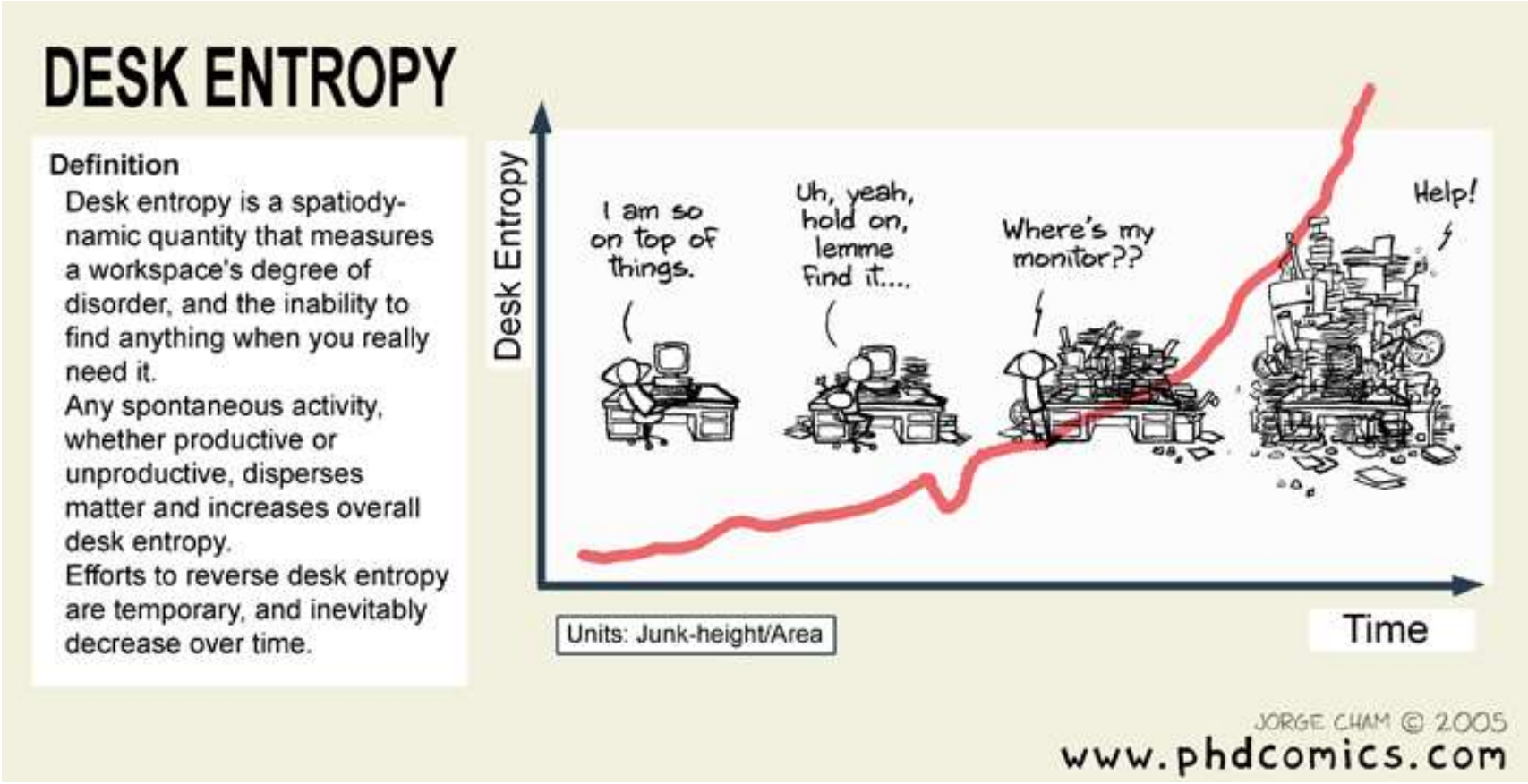
Laurence J. Peter*

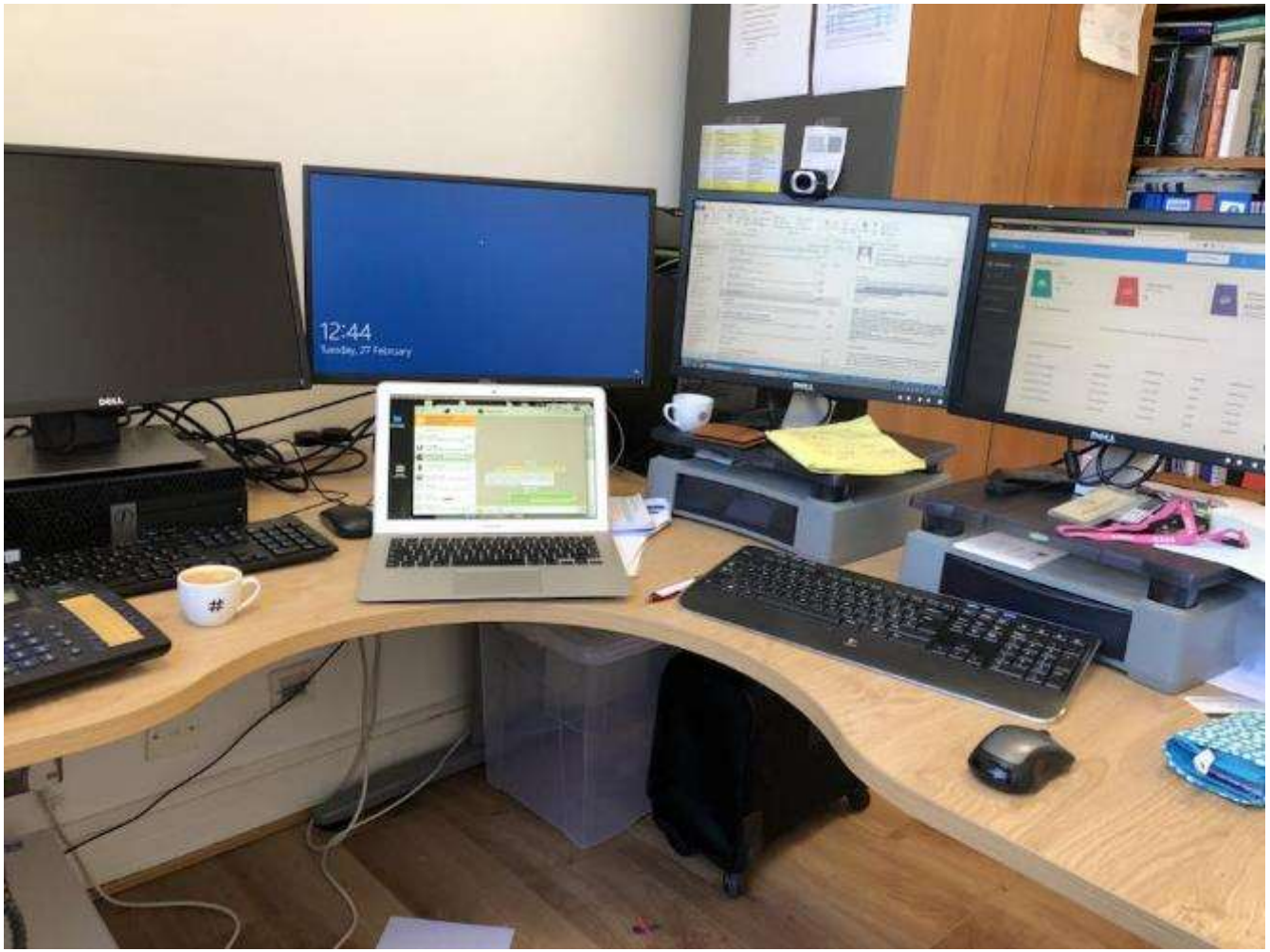
"Data is the fabric of the modern world: just like we walk down pavements, so we trace routes through data and build knowledge and products out of it?"

Ben Goldacre

THE FILE ROOM AT THE CHAOS THEORY & ENTROPY INSTITUTE.

Desk Entropy- where is the data?




12:44
Tuesday, 27 February

Activity

You have completed your post-doctoral study with flying colours and published a couple of papers to disseminate your research results. Your papers have been cited widely in the research literature by others who have built upon your findings. However, three years later a researcher has accused you of having falsified the data.

Do you think you would be able to prove that you had done the work as described? If so, how?

What would you need to prove that you haven't falsified the data?

What should you have done throughout your research study to be able to prove now that you had done the work as described?

SECTION 1: WHAT IS DATA?

"Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly".

"Data management should be woven into every course in science, as one of the foundations of knowledge".

'Editorial: Data's Shameful Neglect' (10 September 2009) in Nature 461, p. 145, doi:10.1038/461145a. Published online 9 September 2009; corrected 23 September 2009.

Conduct a data inventory

- What types of data do you have?
- Who needs to be able to access your data?
- Who can access your data?
- Where is it stored?
- How is it filed?

What is research data?

"Research data are collected, observed, or created, for the purposes of analysis to produce and validate original research results."

Edinburgh, Mantra

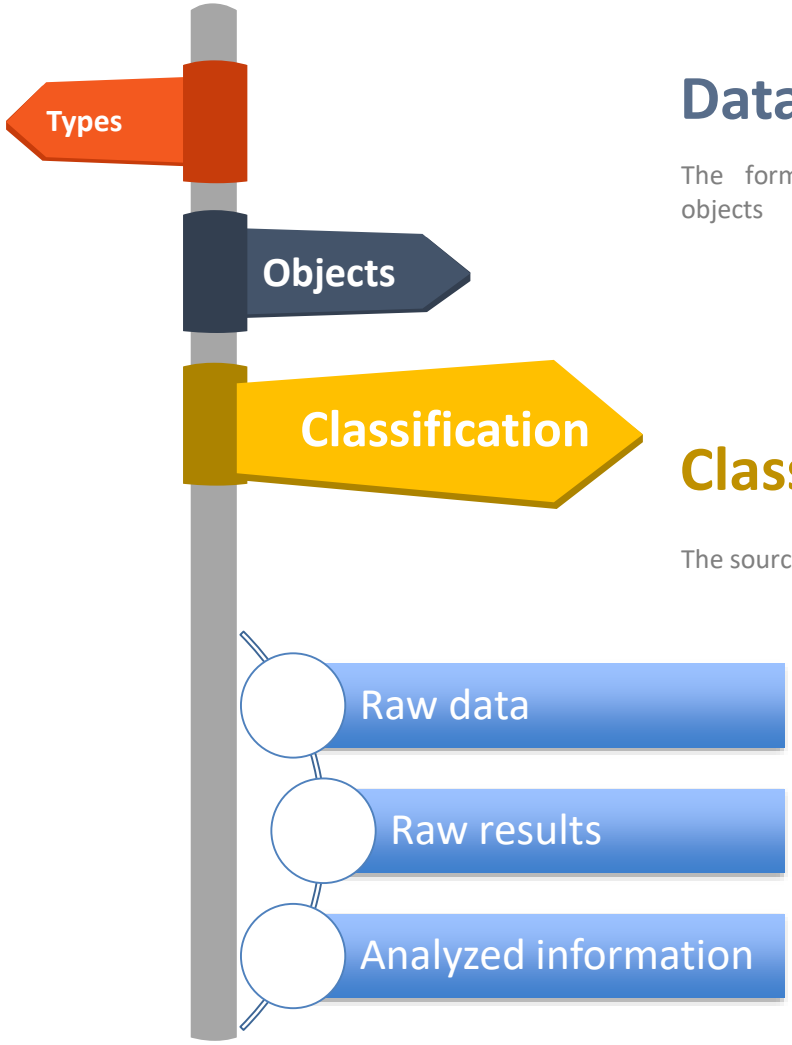
"The *evidence* on which academic researchers build their analytic or other work"

HEFCE

1. Different to information
2. Lowest level of abstraction when creating new knowledge
3. Distinct from commercial data (sales, marketing etc)

Data types

The format



Data Objects

The format plus the physical objects

Classification

The source or “how” of data

Context is everything

Mary Ryan

Context is everything

Feminist scholarship

historians

Mary Ryan (14 December 1873 – 16 June 1961) was the first woman in Ireland or Great Britain to be a professor at a university. She was the Professor of Romance Languages at University College Cork in 1910.

1°, 2° data

Primary Data

Direct – you control what is collected

Secondary Data

Readily available, large samples

BRAINSTORM

What are the advantages and disadvantages of primary and secondary data?

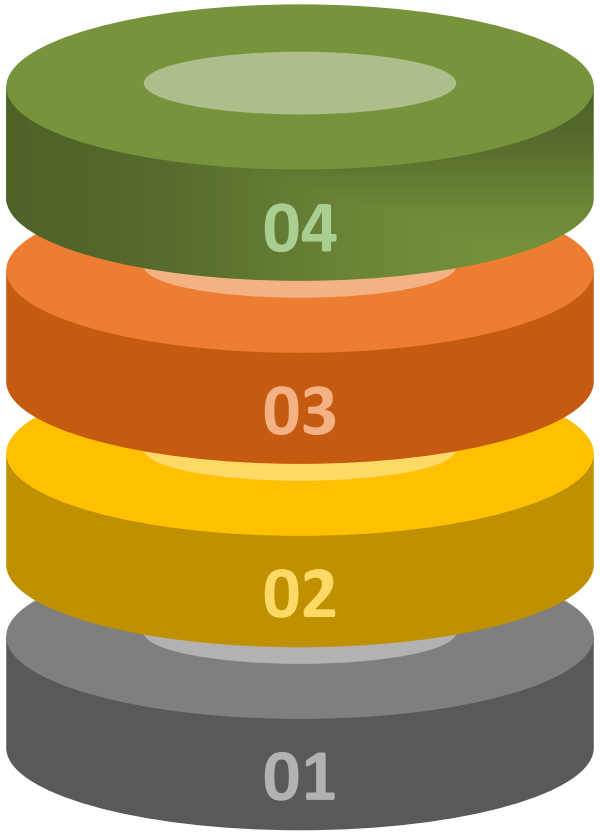
Types of data

Text

Plain text files
Flat files such as EMBL
MS Word
Portable Document Format (PDF)
Rich Text Format (RTF)
Hyper-Text Markup Language (HTML)
Extensible Markup Language (XML)

Multimedia

JPEG
TIFF
GIF
Dicom
MPEG
Quicktime
Bitmap
PNG



Numerical

SPSS
Stata
MS Excel
SAS
Flat files: fixed field format files, delimited files
Hierarchical files

models

3D
Statistical
Similitude
Macroeconomic
Causal

5: software, 6: Discipline specific 7: instrument specific

Research data objects

Research data (traditional and electronic research) may comprise the following objects:

- Documents (text, Word), spreadsheets
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, films
- Test responses
- Slides, artefacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research
- Statistical or other data files
- Database contents (video, audio, text, images)
- Models, algorithms, scripts
- Contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- Methodologies and workflows
- Standard operating procedures and protocols

SECTION 2: STAGES OF RDM

What is research data management?

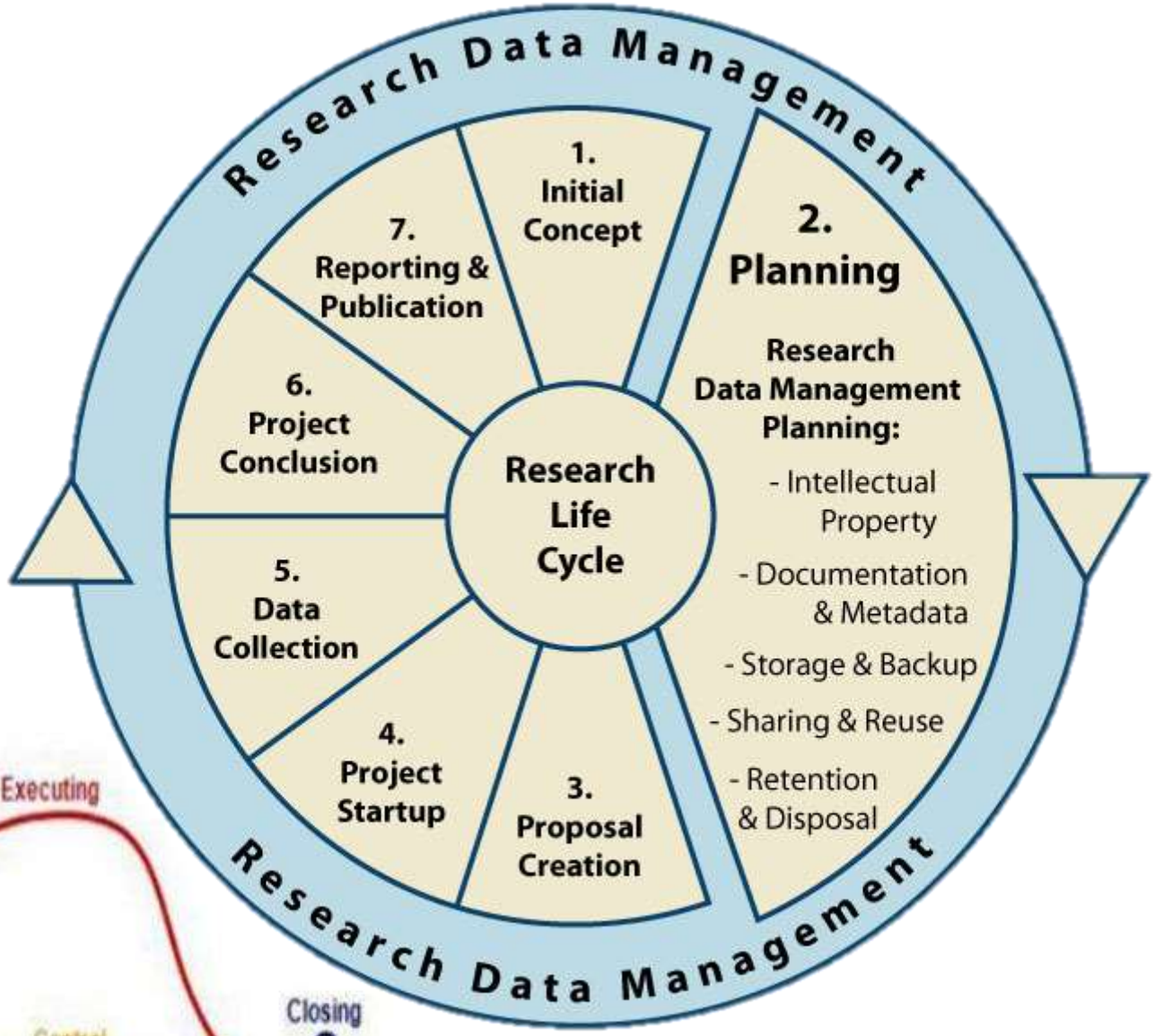
“Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information."

Digital Curation Centre (2011)
Making the Case for Research Data Management
http://www.dcc.ac.uk/sites/default/files/documents/publications/Making%20the%20case.pdf

Stages of RDM

- Planning
- Operational (housekeeping)
  - Documentation
  - naming, versioning etc
  - Storage and back up
- Post-project

Data management as a discipline has been around since 1950's. It took a big leap in 1970s when electronic media emerged, another one in the 00's with cloud storage. Even more recently GDPR and big data have brought data management to the fore.


Research Data Management
1. Initial Concept
2. Planning
Research Data Management Planning:
- Intellectual Property
- Documentation & Metadata
- Storage & Backup
- Sharing & Reuse
- Retention & Disposal
3. Proposal Creation
4. Project Startup
5. Data Collection
6. Project Conclusion
7. Reporting & Publication
Research Life Cycle
Research Data Management


Level of Process Interaction
Executing
Planning
Initiating
Control
Closing
Start
TIME
Finish

Why plan data management

Compliance:
Helps researchers meet the requirements of the various codes of practice
Meets the requirements of funders
Meets the requirements of publishers

Efficiency:
Improves management of data and the research process.
Encourages systematic documentation and descriptions of the research data.
Provides guidelines and procedures ensuring consistency.

Security:
Safeguards against data loss.
Ensures confidentiality and ethical compliance.
Guarantees legal compliance to intellectual property rights such as copyright.

Access:
Allows researchers to validate and verify published results.
Enables collaborative research opportunities thereby increasing the potential scale and scope.
Prevents duplication of research within a particular field.
Allows data sharing and future use when the data is preserved in retrievable formats.
Increases citations for the researcher.

Quality:
Allows for data replication or reproducibility.
Increases the accuracy or reliability of the data.
Ensures research data integrity.

IUA Policy on RI

Data storage and retention good practice

The definition of 'data' for these purposes encompasses the methodology used to obtain results, the actual research results and the analysis and interpretations by the researchers.

Primary responsibility for observing good practice in the use, storage and retention of data sits with the individual researcher, supported by the institution, and should follow the principles below:

Data should be recorded in a **clear and accurate format**. Particular attenton should be paid to the completeness, integrity and security of these records.

Data should be stored in *secure and accessible* form and must be held for a length of time and a minimum period of five years from the date of publication is normally recommended, depending on the type of data,

in the absence of an institutional records management and retention policy. Data should be organised in a manner that allows **ready verification** either in paper or electronic format.

Original data should be **authenticated**, in order to protect the university (or other research performing institution) and researcher against allegations of falsification of data. Research data and records may be discoverable in the event of litigation. This means that the research data and records may be accessed by the university (or other research performing institution) and its legal advisers, to determine their relevance to any litigation process.

IUA National Policy statement on Research Integrity

-Primary responsibility sits with the individual researcher.

-completeness, integrity and security of records.

-Secure and accessible.

-Five years from the date of publication

-Original data should be authenticated

-Research data and records may be discoverable in the event of litigation

OECD types of misconduct by scholars

Core "Research Misconduct"	Research practice misconduct
Fabrication of data **F**alsification of data **P**lagiarism FFP normally includes: Selectively excluding data from analysis Misinterpreting data to obtain desired results (including inappropriate use of statistical methods) Doctoring images in publications Producing false data or results under pressure from a sponsor	Using inappropriate (e.g., harmful or dangerous) research methods Poor research design Experimental, analytical, computational errors Violation of human subject protocols Abuse of laboratory animals
Data-related misconduct Not preserving primary data Bad data management, storage Withholding data from the scientific community NB: The above applies to physical research materials as well	**Publication-related misconduct** Claiming undeserved authorship Denying authorship to contributors Artificially proliferating publications ("salami-slicing") Failure to correct the publication record
Personal misconduct Inappropriate personal behaviour, harassment Inadequate leadership, mentoring, counselling of students Insensitivity to social or cultural norms	**Financial, and other misconduct** Peer review abuse e.g., non-disclosure of conflict of interest, unfairly holding up a rival's publication Misrepresenting credentials or publication record Misuse of research funds for unauthorised purchases or for personal gain Making an unsubstantiated or malicious misconduct allegation

SECTION 2.1.1 DATA MANAGEMENT PLANS

Funding policy and DMP

H2020

- http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

NIH

- https://researchethics.od.nih.gov/DataManagement1.aspx

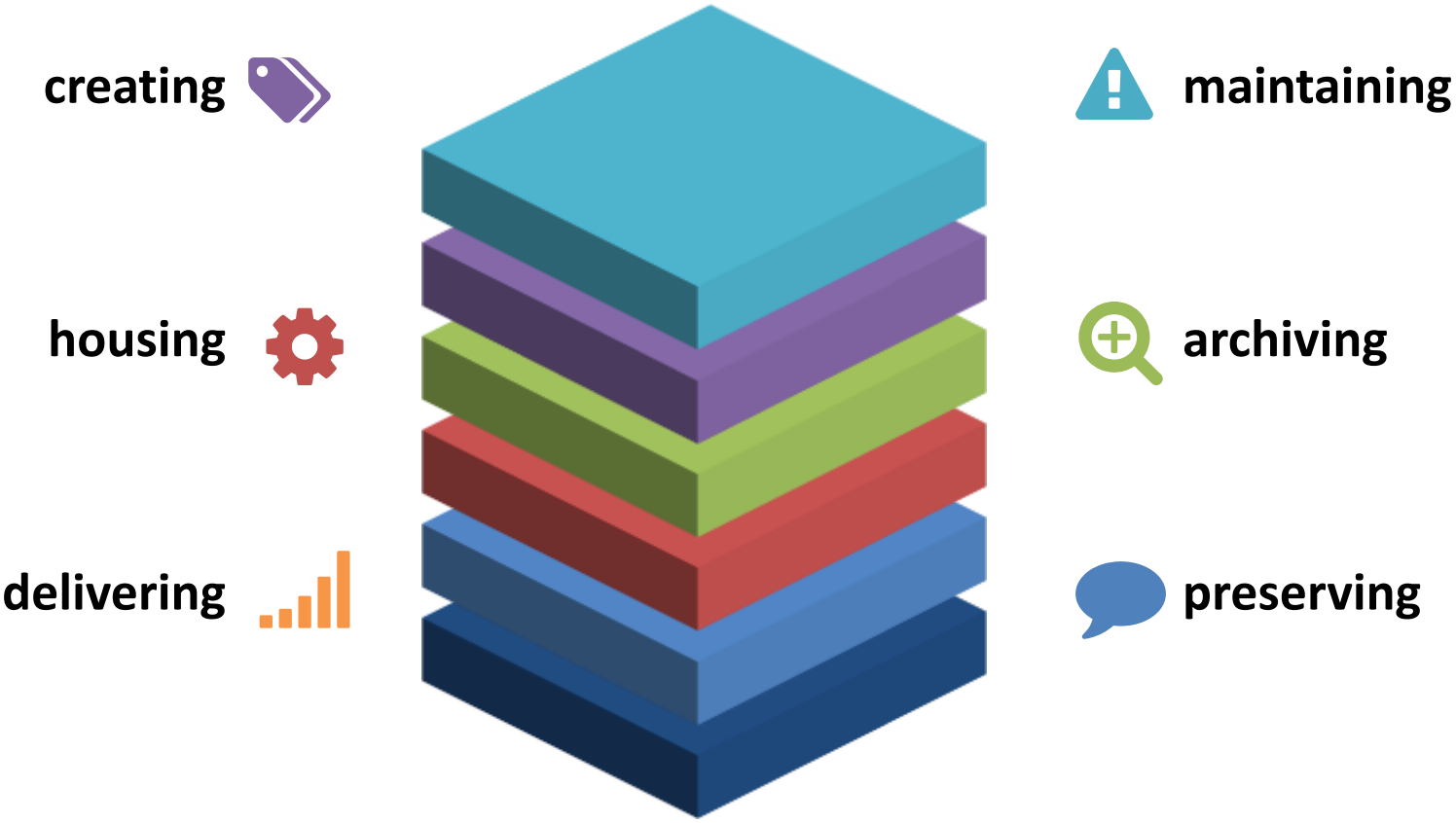
HRB

- http://www.hrb.ie/funding/policies-and-principles/open-research/

UCC

- https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiYtpXLwoXcAhXKCMAKHXWJDSsQFggvMAA&url=https%3A%2F%2Fwww.ucc.ie%2Fen%2Fmedia%2Fresearch%2Fresearchatucc%2Fpoliciesdocuments%2FResearchDataManagementPolicy.docx&usg=AOvVaw0uOjCceqdD_RGC_vaFrP9Z

Considerations for a DMP



DMP checklist

From digital curation center:
http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP-checklist-flyer.pdf

Or from Swedish National Data Service:

https://snd.gu.se/sites/snd.gu.se/files/Checklist%20Data%20Management%20Plan_2017-10-16.pdf

Many others exist, some discipline specific, some suitable for multi-partner projects

Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	**Questions to consider:** - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	**Questions to consider:** - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) policy? - Does your funder have a Research Data Management policy? - Are there any formal standards that you will adopt? **Guidance:** List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.

	to them here.
Data Collection	
What data will you collect or create?	**Questions to consider:** - What type, format and volume of data? - Do your chosen formats and software enable sharing and long-term access to the data? - Are there any existing data that you can reuse? **Guidance:** Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.
How will the data be collected or created?	**Questions to Consider:** - What standards or methodologies will you use? - How will you structure and name your folders and files? - How will you handle versioning? - What quality assurance processes will you adopt? **Guidance:** Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning

	for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.
Documentation and Metadata	

Documentation and Metadata	
What documentation and metadata will accompany the data?	**Questions to consider:** - What information is needed for the data to be to be read and interpreted in the future? - How will you capture / create this documentation and metadata? - What metadata standards will you use and why? **Guidance:** Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.
Ethics and Legal Compliance	
How will you manage any ethical issues?	**Questions to consider:** - Have you gained consent for data preservation and sharing? - How will you protect the identity of participants if required? e.g. via anonymisation - How will sensitive data be handled to ensure it is stored and transferred securely? **Guidance:** Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and reused.
How will you manage copyright and Intellectual Property Rights (IPR) issues?	**Questions to consider:** - Who owns the data? - How will the data be licensed for reuse? - Are there any restrictions on the reuse of third-party data? - Will data sharing be postponed / restricted e.g. to publish or seek patents? **Guidance:** State who will own the copyright and IPR of any data that you will collect or create, along with the licence(s) for its use and reuse. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.

Storage and Backup	
How will the data be stored and backed up during the research?	**Questions to consider:** - Do you have sufficient storage or will you need to include charges for additional services? - How will the data be backed up? - Who will be responsible for backup and recovery? - How will the data be recovered in the event of an incident? **Guidance:** State how often the data will be backed up and to which locations. How many copies are being made? Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage provided by university IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes. If you choose to use a third-party service, you

	should ensure that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.
How will you manage access and security?	**Questions to consider:** - What are the risks to data security and how will these be managed? - How will you control access to keep the data secure? - How will you ensure that collaborators can access your data securely? - If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems? **Guidance:** If your data is confidential (e.g. personal data not already in the public domain, confidential information or trade secrets), you should outline any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.
Selection and Preservation	

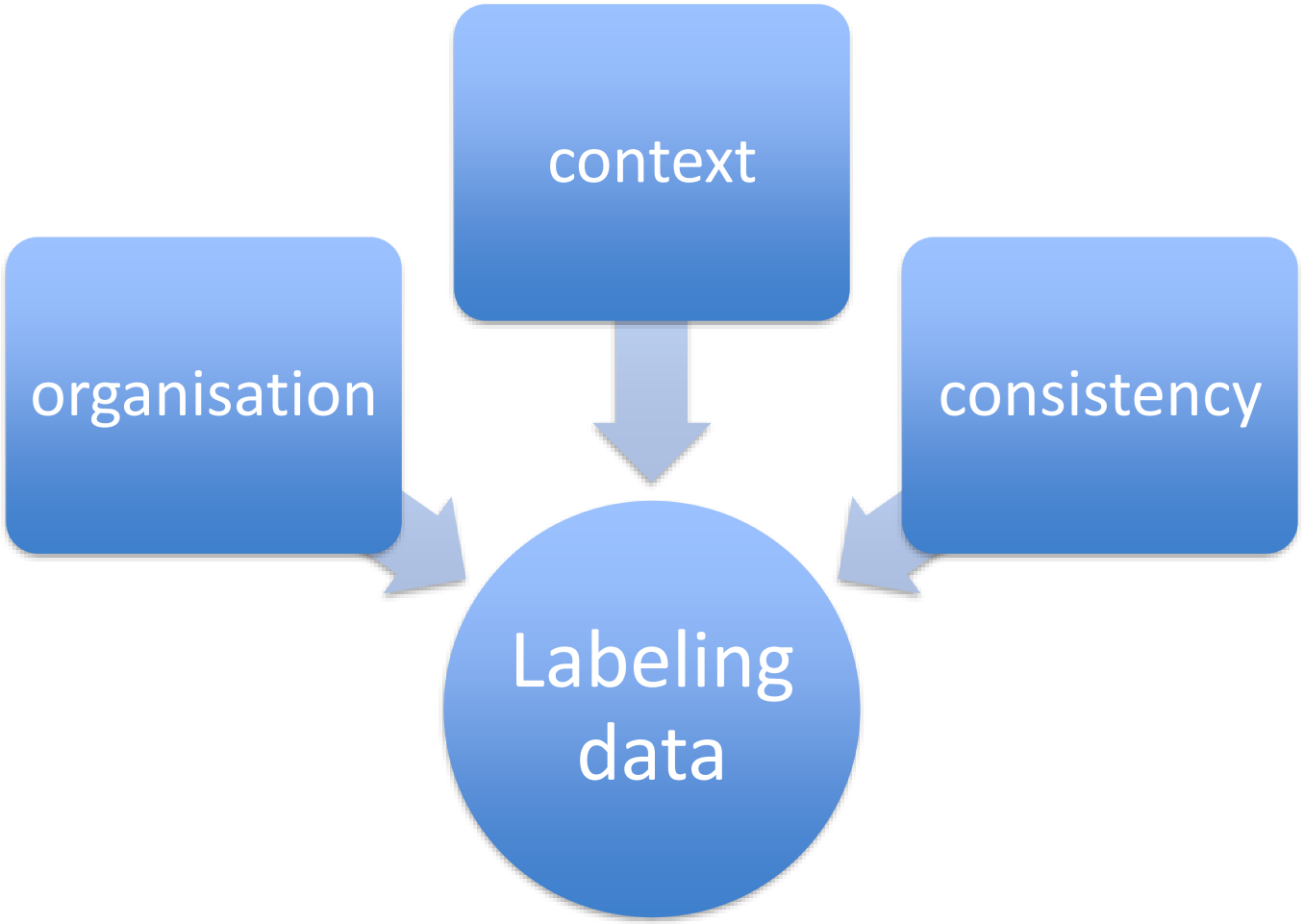
Selection and Preservation	
Which data should be retained, shared, and/or preserved?	**Questions to consider:** - What data must be retained/destroyed for contractual, legal, or regulatory purposes? - How will you decide what other data to keep? - What are the foreseeable research uses for the data? - How long will the data be retained and preserved? - **Guidance:** Consider how the data may be reused e.g. to validate your research findings, conduct new studies, or for teaching. Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential reuse value, what is economically viable to keep, and any additional effort required to prepare the data for data sharing and preservation. Remember to consider any additional effort required to prepare the data for sharing and preservation, such as changing file formats.
What is the long-term preservation plan for the dataset?	**Questions to consider:** - Where e.g. in which repository or archive will the data be held? - What costs if any will your selected data repository or archive charge? - Have you costed in time and effort to prepare the data for sharing / preservation? **Guidance:** Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.
Data Sharing	
How will you share the data?	**Questions to consider:** - How will potential users find out about your data? - With whom will you share the data, and under what conditions? - Will you share data via a repository, handle requests directly or use another mechanism? - When will you make the data available? - Will you pursue getting a persistent identifier for your data? **Guidance:** Consider where, how, and to whom data with acknowledged long-term value should be made available. The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing. Consider how people might acknowledge the reuse of your data.

Data Sharing	
How will you share the data?	**Questions to consider:** - How will potential users find out about your data? - With whom will you share the data, and under what conditions? - Will you share data via a repository, handle requests directly or use another mechanism? - When will you make the data available? - Will you pursue getting a persistent identifier for your data? **Guidance:** Consider where, how, and to whom data with acknowledged long-term value should be made available. The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing. Consider how people might acknowledge the reuse of your data.
Are any restrictions on data sharing required?	**Questions to consider:** - What action will you take to overcome or minimise restrictions? - For how long do you need exclusive use of the data and why? - Will a data sharing agreement (or equivalent) be required? **Guidance:** Outline any expected difficulties in sharing data with acknowledged long-term value, along with causes and possible measures to overcome these. Restrictions may be due to confidentiality, lack of consent agreements or IPR, for example. Consider whether a non-disclosure agreement would give sufficient protection for confidential data.

Responsibilities and Resources	
Who will be responsible for data management?	**Questions to consider:** - Who is responsible for implementing the DMP, and ensuring it is reviewed and revised? - Who will be responsible for each data management activity? - How will responsibilities be split across partner sites in collaborative research projects? - Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners? **Guidance:** Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.
What resources will you require to deliver your plan?	**Questions to consider:** - Is additional specialist expertise (or training for existing staff) required? - Do you require hardware or software which is additional or exceptional to existing institutional provision? - Will charges be applied by data repositories? **Guidance:** Carefully consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.

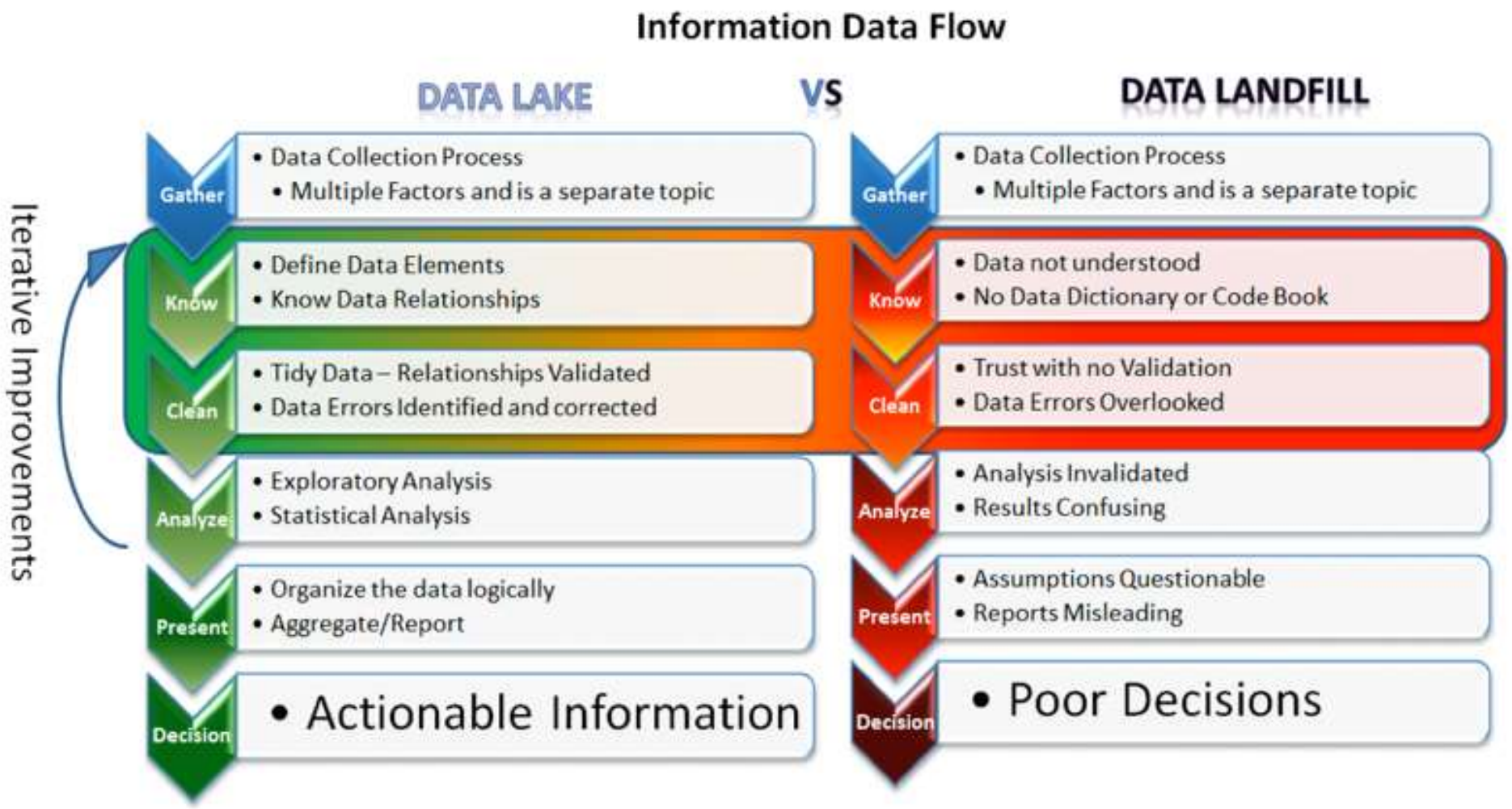
Example DMP

Administrative information	Name and ID of project, project description, funding body, PI &ID, Name of Data contact person, related policies, date of 1st version, dates of updates
Data collection	Description of data (type, format, volume), existing datasets, methodology notes, data structural conventions – names and versions
Metadata	Info needed for future interpretation, standards
Ethics and legal	Details of consent, participant ID protection, secure storage of sensitive data, copyright and IP rights, storage (physical location, back ups, who is responsible). Security (risk management, access arrangements), sharing policy.
Responsibilities & resources	For preservation, hardware, software requirements


context
organisation
consistency
Labeling data

Test your research memory

- What did you have for breakfast this morning?
- What did you have for dinner on Sunday 17th June?
- What did you have for dinner on 17th June 2011?


Information Data Flow
DATA LAKE
VS
DATA LANDFILL
Iterative Improvements
Gather
• Data Collection Process
• Multiple Factors and is a separate topic
Know
• Define Data Elements
• Know Data Relationships
Clean
• Tidy Data – Relationships Validated
• Data Errors Identified and corrected
Analyze
• Exploratory Analysis
• Statistical Analysis
Present
• Organize the data logically
• Aggregate/Report
Decision
• Actionable Information
Gather
• Data Collection Process
• Multiple Factors and is a separate topic
Know
• Data not understood
• No Data Dictionary or Code Book
Clean
• Trust with no Validation
• Data Errors Overlooked
Analyze
• Analysis Invalidated
• Results Confusing
Present
• Assumptions Questionable
• Reports Misleading
Decision
• Poor Decisions

Why Create a file structure system:

Organization -important for future access and retrieval

Simplifies your workflow in managing files

Data files are easier to locate and browse

Eases data sharing: clear organization is intuitive to team members and colleagues

Data files are distinguishable from each other within and across folders

Document your system and use it consistently!

Good practices for organizing data files

First: define the types of data and file formats for the research

Be Clear, Concise, Consistent, Correct, and Conformant

Choose a meaningful directory hierarchy/naming convention which includes important contextual information

Document your system and use it consistently -choose a naming convention and ensure that the rules are followed systematically by always including the same information in the same order

Be Clear, Concise, Consistent, Correct, and Conformant

When consistency goes wrong

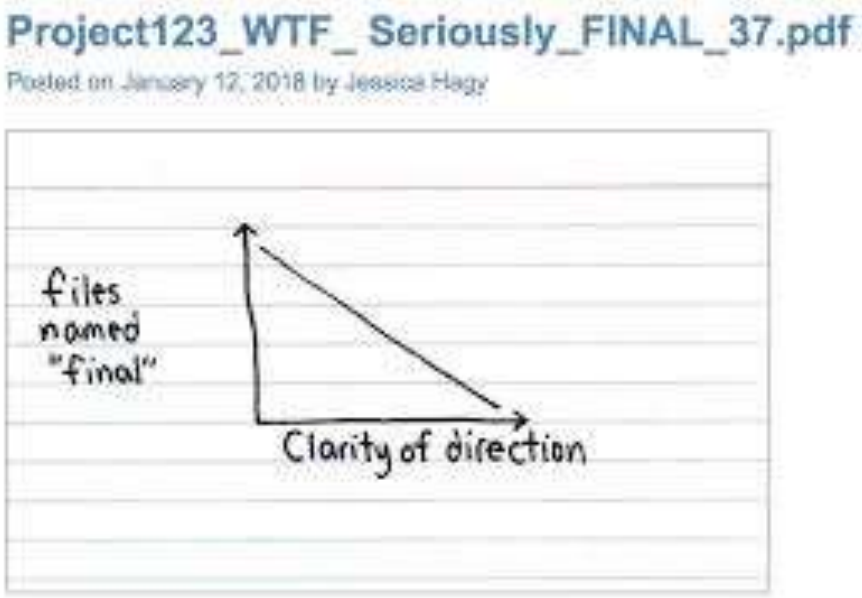
Doc.Final

Doc. Final 2

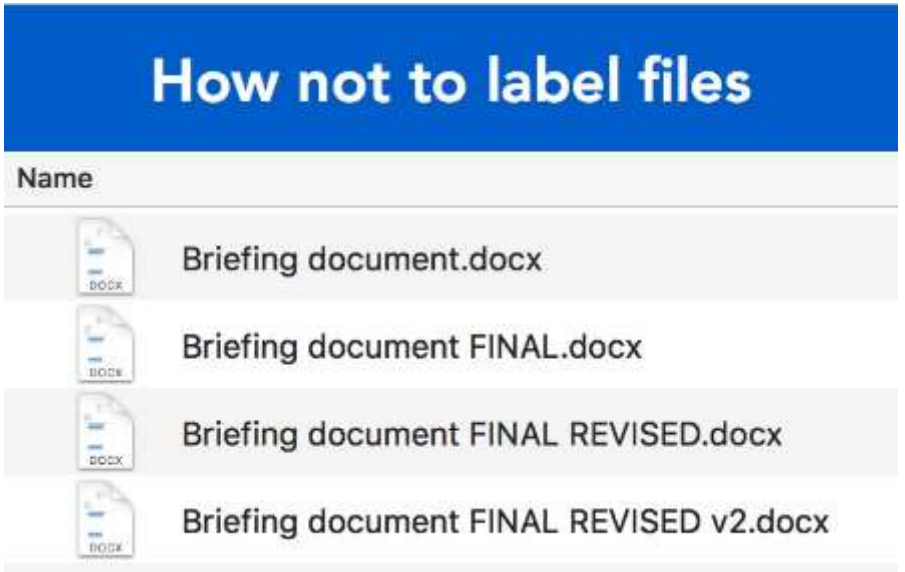
Doc. Definitive final

Doc. Copy of final

Doc. Copy



Naming files

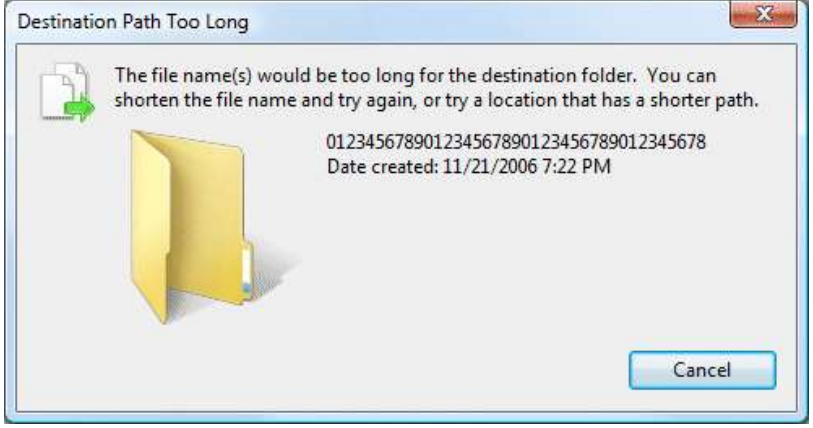


- Distinguishable within folder
- Distinguishable between authors
- Easy to locate (by author and others)
- Logical sequence
- Cannot be accidently deleted
- Versioned
- Can be moved and retain context

The good, bad and ugly of file names

- Intereview1.doc
- 2346254jkfdjli8jhrjhqg4hjt.jpg
- Myfile_mylocation_date_version_contains_firstversion_team name_usedfor.xls
- P1_WP1_theme_ver.docx
- Joannes_file_final

Some ideas



- 25 characters
- No special characters e.g. & ; * % $ £] { ! @ /
- Number order files only if using leading zeros : e.g., 001, 002, 003, etc. will order files up to 999
- _ not .
- Descriptive independent of location
- Consistent dates (if used)
- Assume Joanne = JOANNE = joanne in naming but not always for sorting
- Use extensions where possible (often default)
- Ordinal number for major changes, decimal for revisions 1.1, 1.2, 2.0 etc.

Bulk renaming tools

- Bulk renaming = batch renaming = mass file renaming
- Especially useful when instrument assigns file names (eg camera)

- Avoids accidental overwriting
- Saves time
- Can replace unsuitable characters ($%^*)

Batch renaming tools Windows:

- Adobe Bridge http://ist.mit.edu/adobe-creative-cloud
- Ant Renamer: http://www.antp.be/software/renamer
- Bulk Rename Utility: http://www.bulkrenameutility.co.uk/
- ImageMagick: http://www.imagemagick.org
- PSRenamer: http://www.powersurgepub.com/products/psrenamer.html
- RenameIT: http://sourceforge.net/prpjects/renameit

Mac:

- Adobe Bridge :http://ist.mit.edu/adobe-creative-cloud
- ImageMagick: http://www.imagemagick.org/
- Name Changer: http://web.mac.com/mickeyroberson/MRR_Software/NameChanger.html
- PSRenamer: http://www.powersurgepub.com/products/psrenamer.html
- Renamer4Mac : http://renamer4mac.com/
- Name Mangler: http://manytricks.com/namemangler/

Linux

- GNOME Commander: http://www.nongnu.org/gcmd/
- GPRename: http://gprename.sourceforge.net/

Unix: the use of the grep command to search for regular expressions

versioning

- Include in DMP
- Record every change
- Always keep raw data
- Discard obsolete versions
- Consider
  - Multiple users
  - Multiple sites/locations
  - Synchronisation across storage

Discarding obsolete files

We've updated our Privacy and Cookies Policy

BBC Sign in News Sport Weather Shop Earth Travel

NEWS

Home Video World UK Business Tech Science Stories Entertainment & Arts

TODAY'S NEWS IN VERTICAL VIDEO

DOWNLOAD THE APP

Technology

Nasa sells shuttle PCs without wiping secret data

8 December 2010 Share

US space agency Nasa has been left red-faced after selling off computers without ensuring that highly sensitive data had been removed.

An internal investigation found 10 cases where PCs were sold despite failing data removal procedures.

Another four PCs - which were about to be sold - were found to contain data restricted under arms control rules.

Nasa is selling of hundreds of PCs used on its Space Shuttle programme

The computers were being sold off as Nasa winds down its space shuttle operations.

Write down 3 things you are going to do to improve your data management in the next 2 weeks.

1.

2.

3.

Name: _____

(internal) post address _____

STORAGE AND SECURITY

What to consider

- What are your options?
- Backups?
- Being proactive
- Passwords and safety
- Encryption
- Safe destruction of obsolete data

Where to store data

- Local
  - portable or
  - Integrated
- NAS
- Cloud

Cloud storage

	OneDrive	Dropbox	Google Drive	Box	Amazon Cloud Drive
File size restrictions?	10GB	10GB with website, none with Dropbox apps	5TB	250MB for free plan, 5GB for paid personal plan	2GB*
Free storage?	5GB**	2GB	15GB	10GB	No***
Can I earn extra free storage?	No**	Yes	No	No	No
Paid plans	$2/month for 50GB**	$10/month for 1TB	$2/month 100GB, $10/month for 1TB	$10/month for 100GB	$12/year for unlimited photos, $60/year for unlimited files
OSes supported	Windows, Mac, Android, iOS, Windows Phone	Windows, Mac, Linux, Android, iOS, Windows Phone, BlackBerry, Kindle Fire	Windows, Mac, Android, iOS	Windows, Mac, Android, iOS, Windows Phone, BlackBerry	Windows, Mac, Android, iOS, Kindle Fire

UCC – Research Data Store

- Anytime, any place
- Staff and student (@ucc.ie address)
- <=1TB – FREE!!
- PI requests to Eoghan Ó Carragáin :Library – research data store
- Shared – per research group, incl students
- Install receiver on machine and click on it on desktop
- Access the data and analyse the data (e.g. spss)

Post-project data management

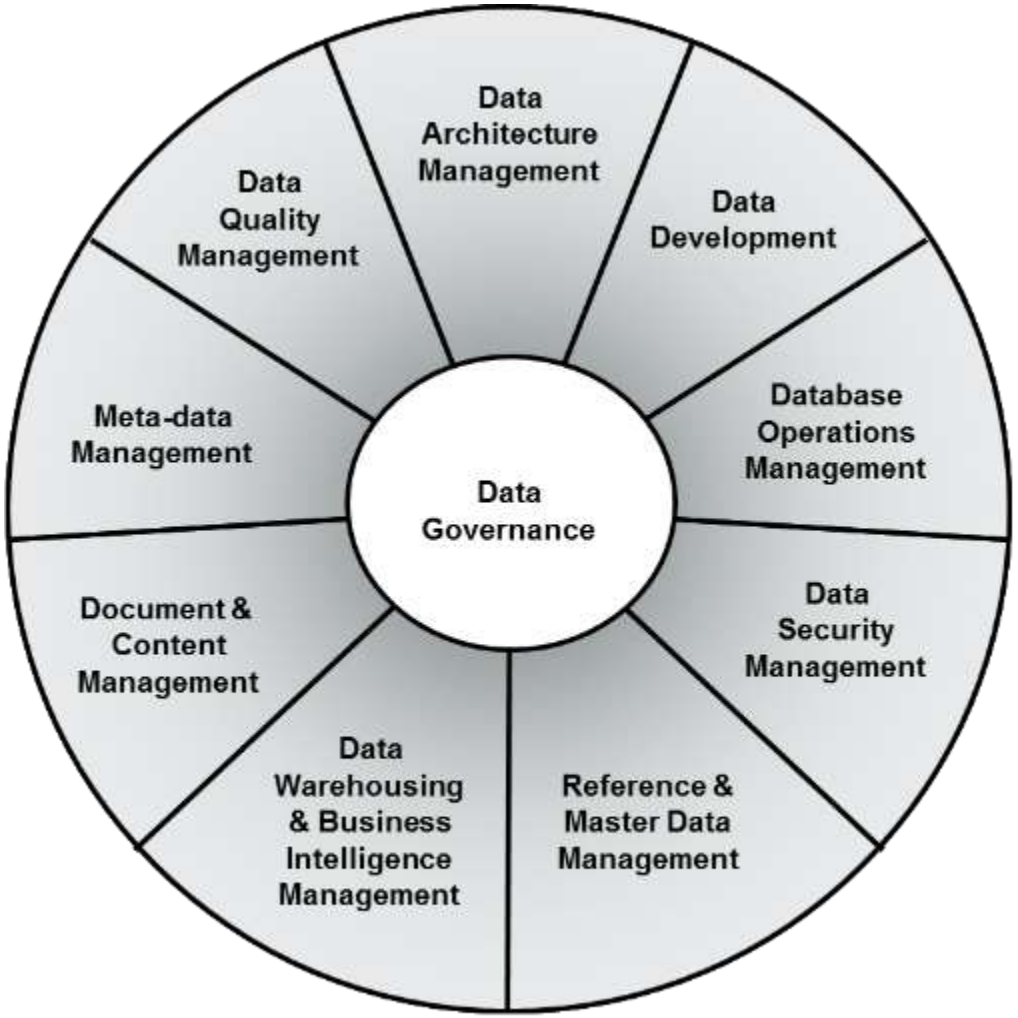
- Correspondence (electronic mail and paper-based correspondence)
- Project files
- Grant applications
- Ethics applications
- Technical reports
- Technical appendices
- Research reports
- Research publications
- Master lists
- Signed consent forms
- Social media communications such as blogs, wikis, tweets etc.

What is your succession plan?

DATA MANAGEMENT PROFESSIONALISATION

DAMA

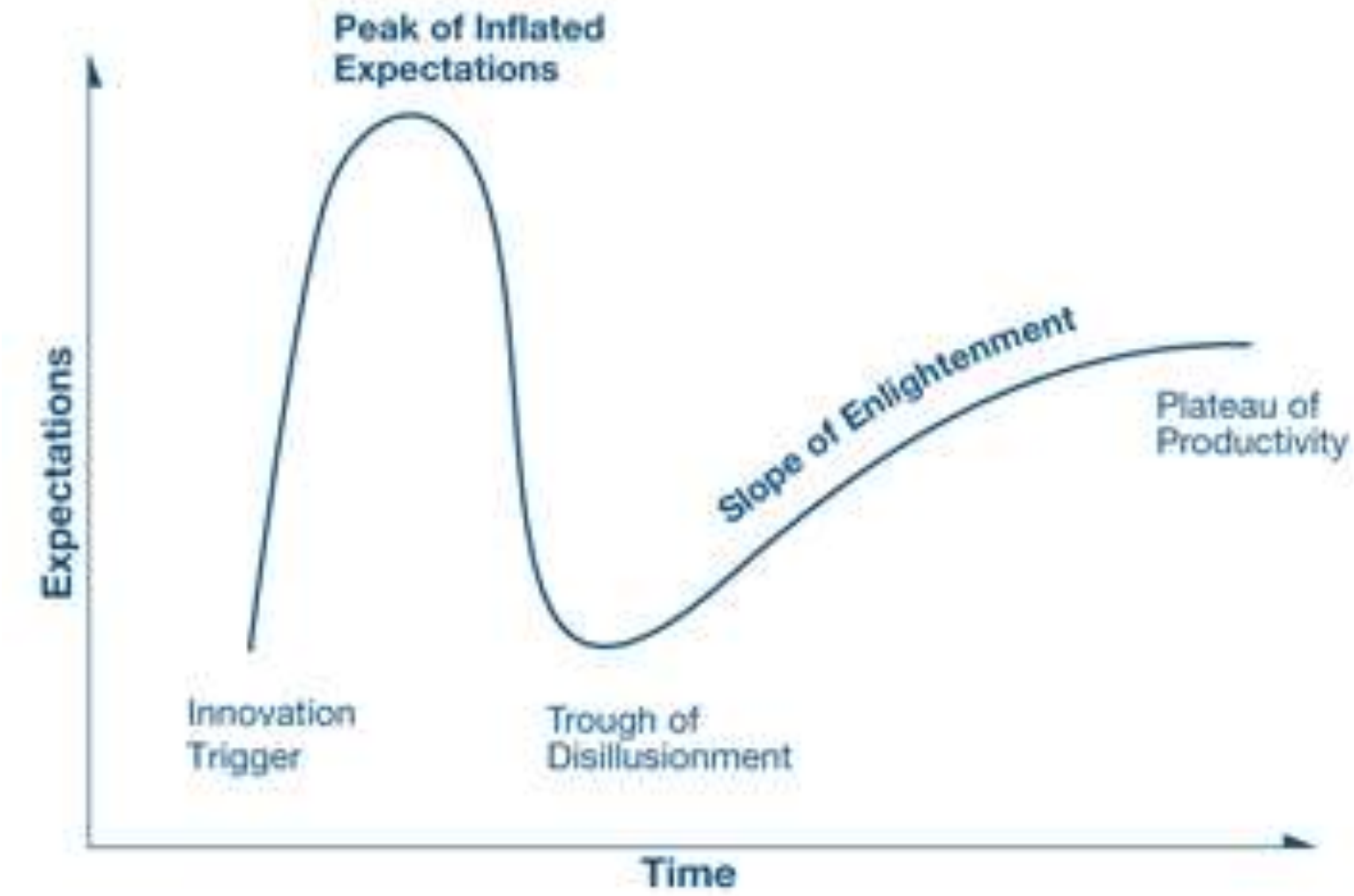
DMBOK



	Associate	Practitioner	Master	Fellow
DAMA Membership	Central or Chapter	Central or Chapter	Central or Chapter	Central or Chapter
Experience	6 months → 2 years	>= 3,5 years	>= 10 years	>= 20 years
		Add 2 years if no tertiary education	Add 5 years if no tertiary education	
Education	Degree/ Diploma/ Certification	Degree/ Diploma/ Certification	Degree/ Diploma/ Certification	Degree/ Diploma/ Certification
DMBoK Knowledge	Knowledge and understanding of DMBoK Principles	Strong knowledge and understanding of DMBoK Principles	Advanced knowledge and understanding of DMBoK Principles	Expert knowledge and understanding of DMBoK Principles
Practical Implementation		Involved/ team member in practical implementation of DMBoK	3+ years managing DM programs or projects 2+ years experience in developing or implementing DM business processes	5+ years managing DM programs or projects 10+ years experience in developing or implementing DM business processes
Contributing to the DM Profession		Conference Speaker Member or Advisor of Chapter Board (or other similar Association) In-house training, education	← + Webinar/ Workshop Presenter Published Blogs, White Papers, Articles, Books	← + Contribution to CDMP (exams or training material) Globally recognised thought leadership International recognition of achievement in Data & Information Management Contribution to the CDMP exams, the CDMP training material and/ or DAMA I Bodies of Knowledge
Number of Exams	1	3	3	CDMP Master
Exams Required	DM Fundamental	DM Advanced 2 electives	DM Advanced 2 electives Case Study	
Exam Pass Rate	60%	70%	80%	
Substitutions may be applied to (see https://www.dama.org/content/cdmp-practitioner for details)	N/A	Data Operations/ Database Administration Data Warehousing/ Business Intelligence & Analytics	Data Operations/ Database Administration Data Warehousing/ Business Intelligence & Analytics	
Adjacent Knowledge Area Certifications (one only)		e.g.: TOGAF®; Zachman Certification; etc	e.g.: TOGAF®; Zachman Certification; etc	
Industry Specialisation Certifications (one only)		TBC	TBC	
Re-certification	Move to Practitioner	3 yearly intervals Contribute a minimum of 3 exam questions per exam topic per 3 year interval Update CPD's online	3 yearly intervals A minimum of 3 exam questions per exam topic annually Update CPD's online	Continued Contribution
How to get there	Apply & Take exam	Convert (if necessary), Apply &	Apply, Take Exams & Supply Case	Nominated by CDMP Master or

TRENDS

Gartner Hype cycle



Data SWOT 2018?

- IT in institution responsibility for cloud data
- Data breaches increasing at 40% pa (particularly ransomware)
- GDPR fines?
- Increasing sophisticated data analytics (IQ boost for data management)
- Data continue exponential growth, storage to slow though –increased need for data hygiene

4 Vs

Volume

Variety

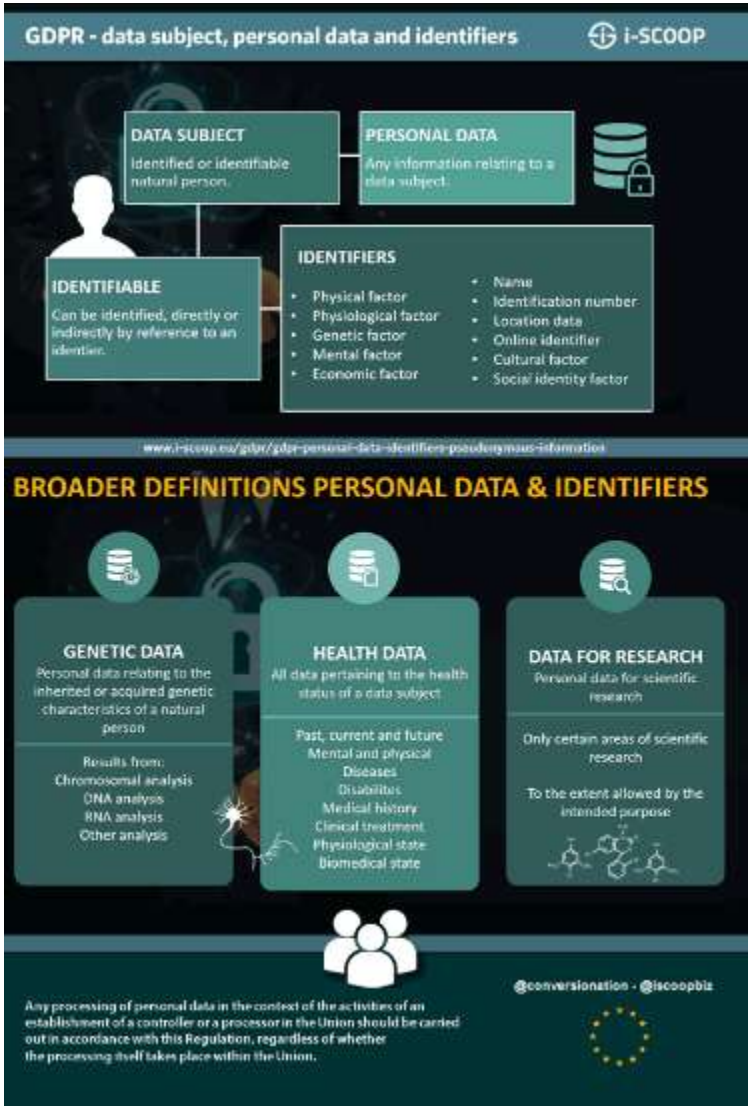
Velocity (of analysis)

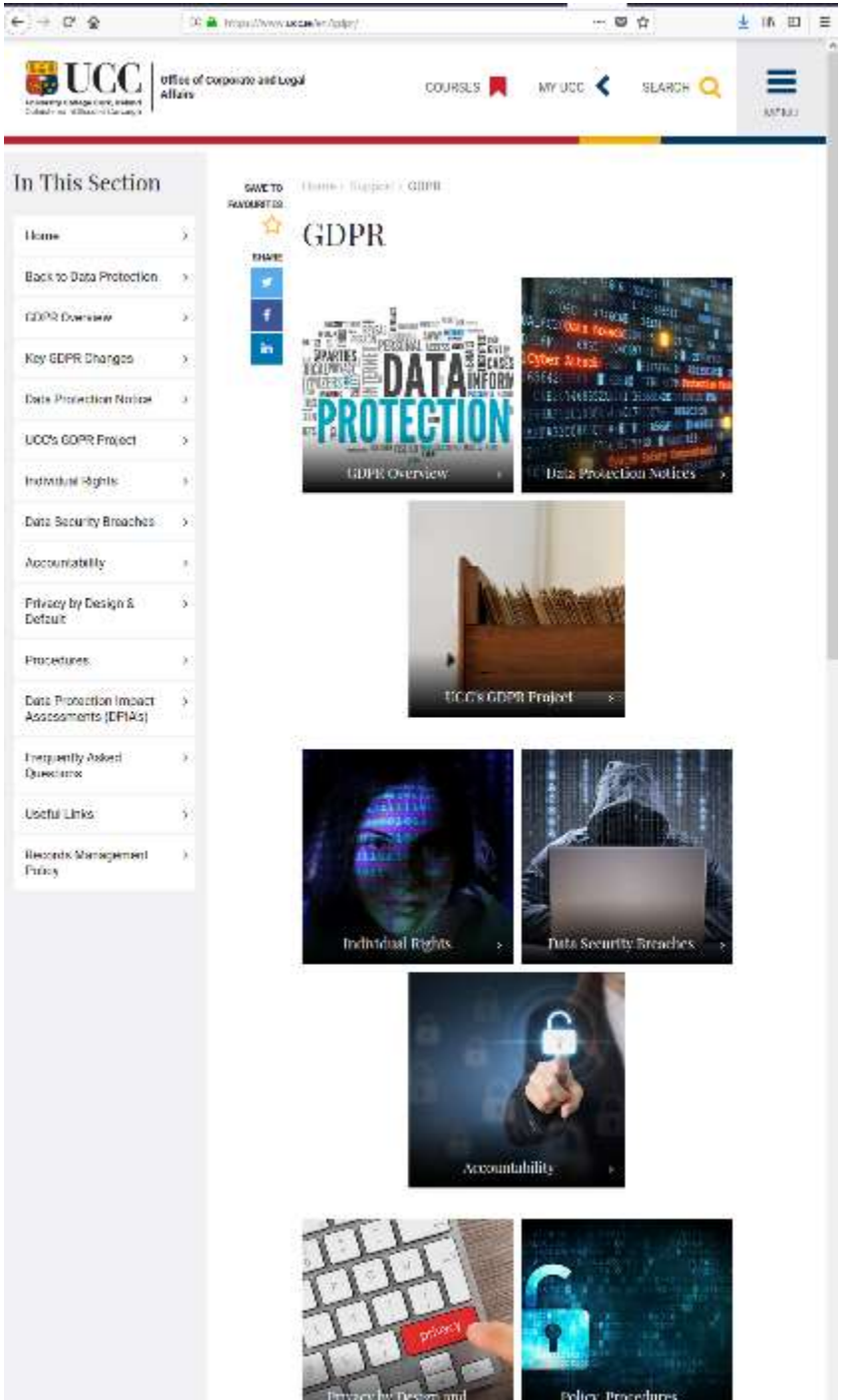
Veracity

Definition: extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

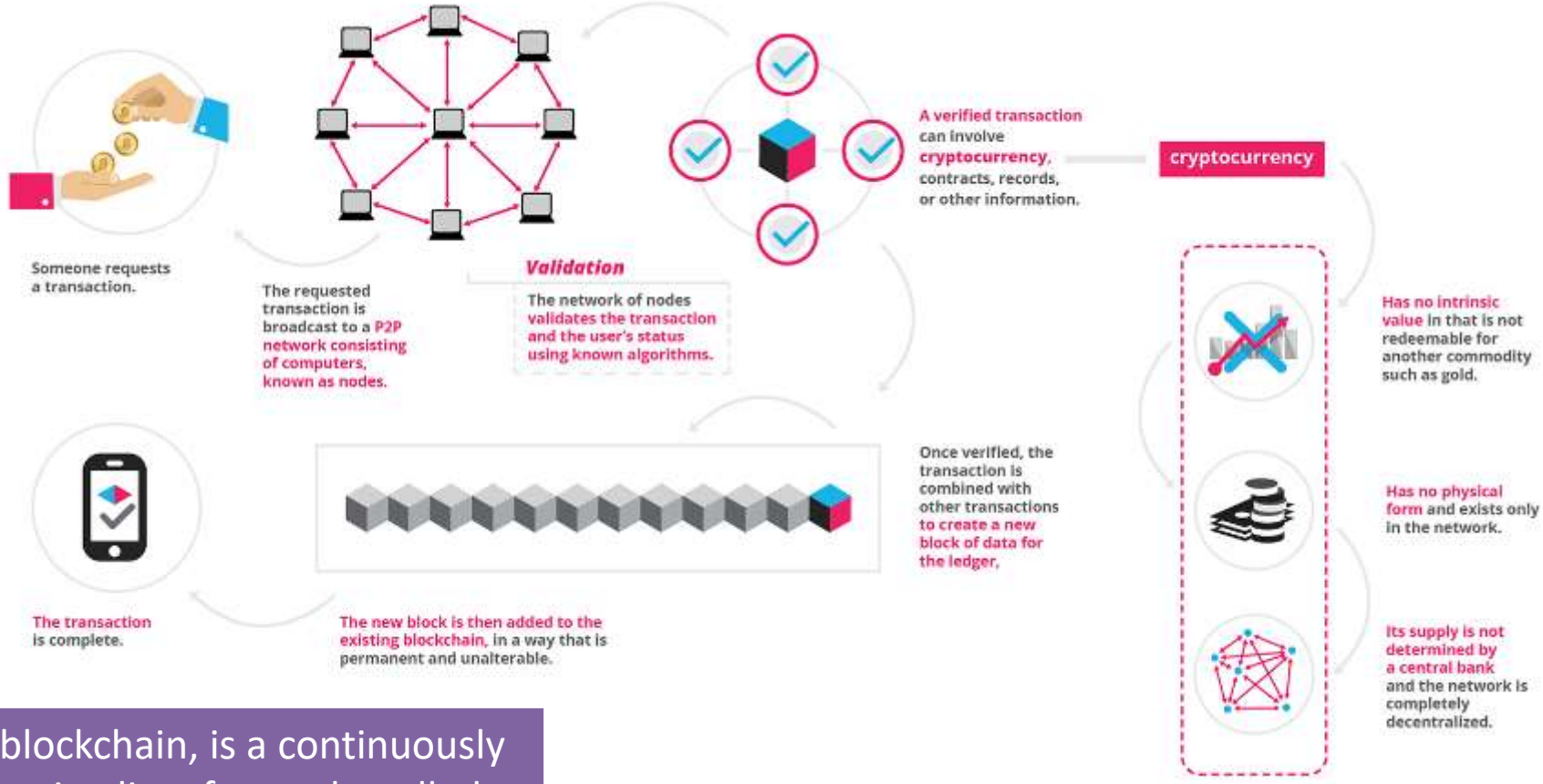
GDPR

http://gdprandyou.ie/gdpr-for-individuals/





Blockchain



A blockchain, is a continuously growing list of records, called blocks, which are linked and secured using cryptography

Storage v data

“The total amount of digital data generated in 2013 was about 3.5 zettabytes (that's 35 with 20 zeros following). By 2020, we'll be producing, even at a conservative estimate, 44 zettabytes of data annually.”

Mark Whitby, Seagate

By 2020, we can predict a minimum capacity gap of over six zettabytes – nearly double all the data that was produced in 2013.

Silicon is showing limitations, need to look at DNA, quantum computing, RRAM, HAMR

https://youtu.be/yR2wWQYiVKM

Data richness

https://youtu.be/-Gj93L2Qa6c