Understanding critical data literacy beyond data skills

Javiera Atenas, Leo Havemann, Cristian Timmermann, Caroline Kuhn

#Datapraxis

Conducting research with data about people is a privilege, not a right, thus, everyone managing research data, open data, or any type of data provided by people, including librarians and researchers, needs to build critical and ethical data literacies

- Data literacy is normally understood as a set of abilities to read, understand, create, and communicate data as information. Much like literacy as a general concept, data literacy focuses on the competencies involved in working with data.

- However, data literacy can be also understood a mean to participate in the (datafied) society, thus the skills needed to work with data go beyond technicalities and have a strong social component, ergo, need to be grounded on the overarching principles of data ethics.

- We suggest librarians and researchers get familiarised with a set of data skills that may help them work with data at management and research level, while being aware of the potential impact of data on individuals and the society, thus, handling data within an ethical and critical framework.

Basic data skills for librarians and researchers

Data storytelling

refers to the skills involved in telling the stories arising from data analysis

Data visualisation

refers to the ability to portray the results of analysis using graphic and visual techniques to aid communication

Data publication

refers to making the data available for any kind of use in data portals or repositories

Data collection

refers to the ability to collect data using different techniques, including surveys, interviews, questionnaires or retrieving data from different portals and sources including data mining.

Data curation

refers to the capacity of organising and maintaining data from storage and organisation in personal systems to create databases of records

Data analysis

refers to reviewing and studying the data to find ways to solve a problem or obtain results

Atenas, Havemann, Kuhn & Timmermann

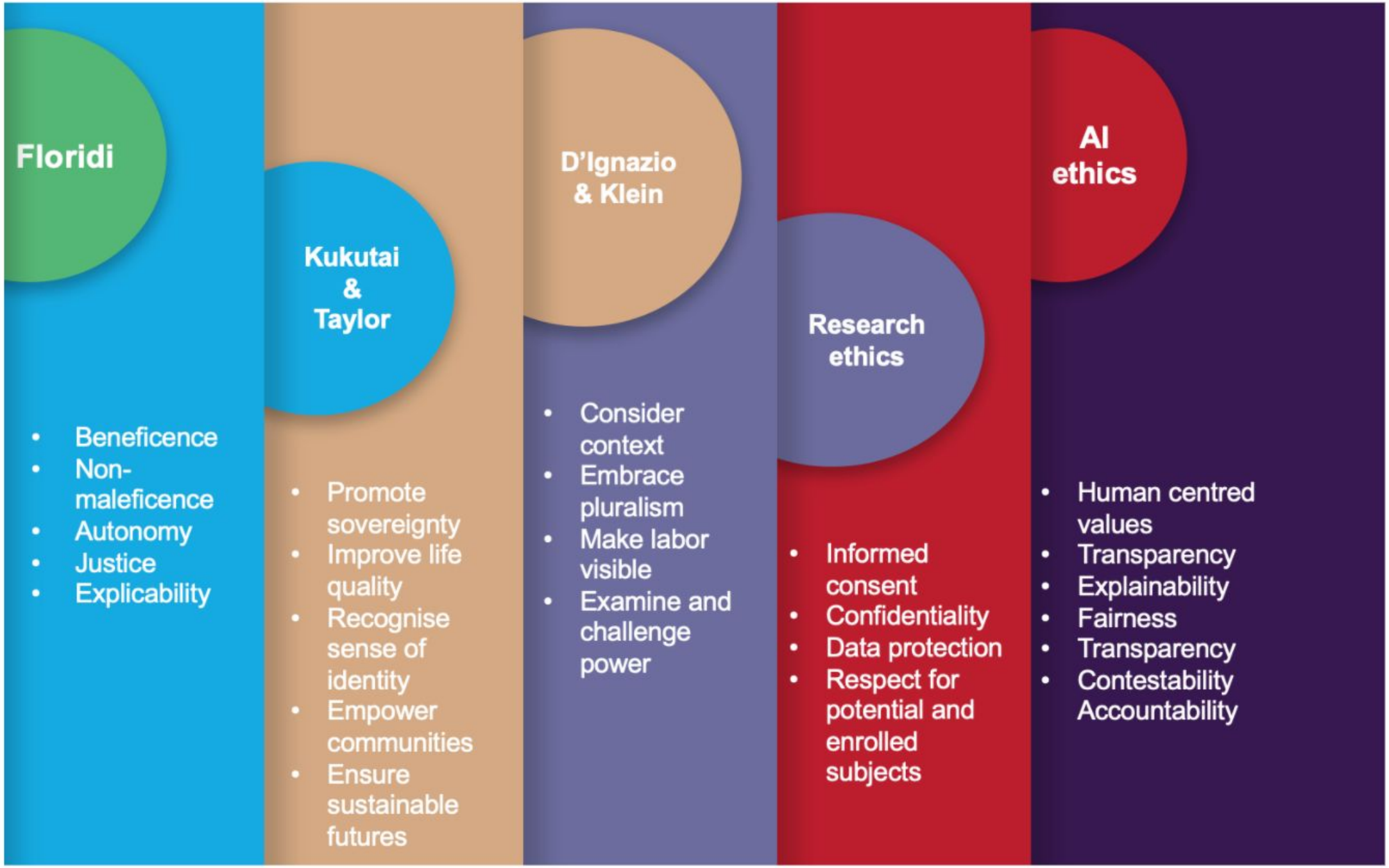
Why data ethics?

- Training in basic data skills, though essential, often leaves ethics aside or at best, implied.
- The need for improved data literacy and data ethics is universal
- Sustainable ethical principles to guide decision-making in data practices, in education and beyond, are urgently needed.
- Using an ethical framework to enable the critical understanding of the wide spectrum of data issues in the context of HE, can therefore support educators, librarians and researchers in assessing their own practices, and foster participatory and collaborative learning and research activities, co-creating knowledge for social transformation.
- It is important to note that each step of the data cycle requires ethical principles in order to put people and communities first.

Data ethics - beyond the tickbox

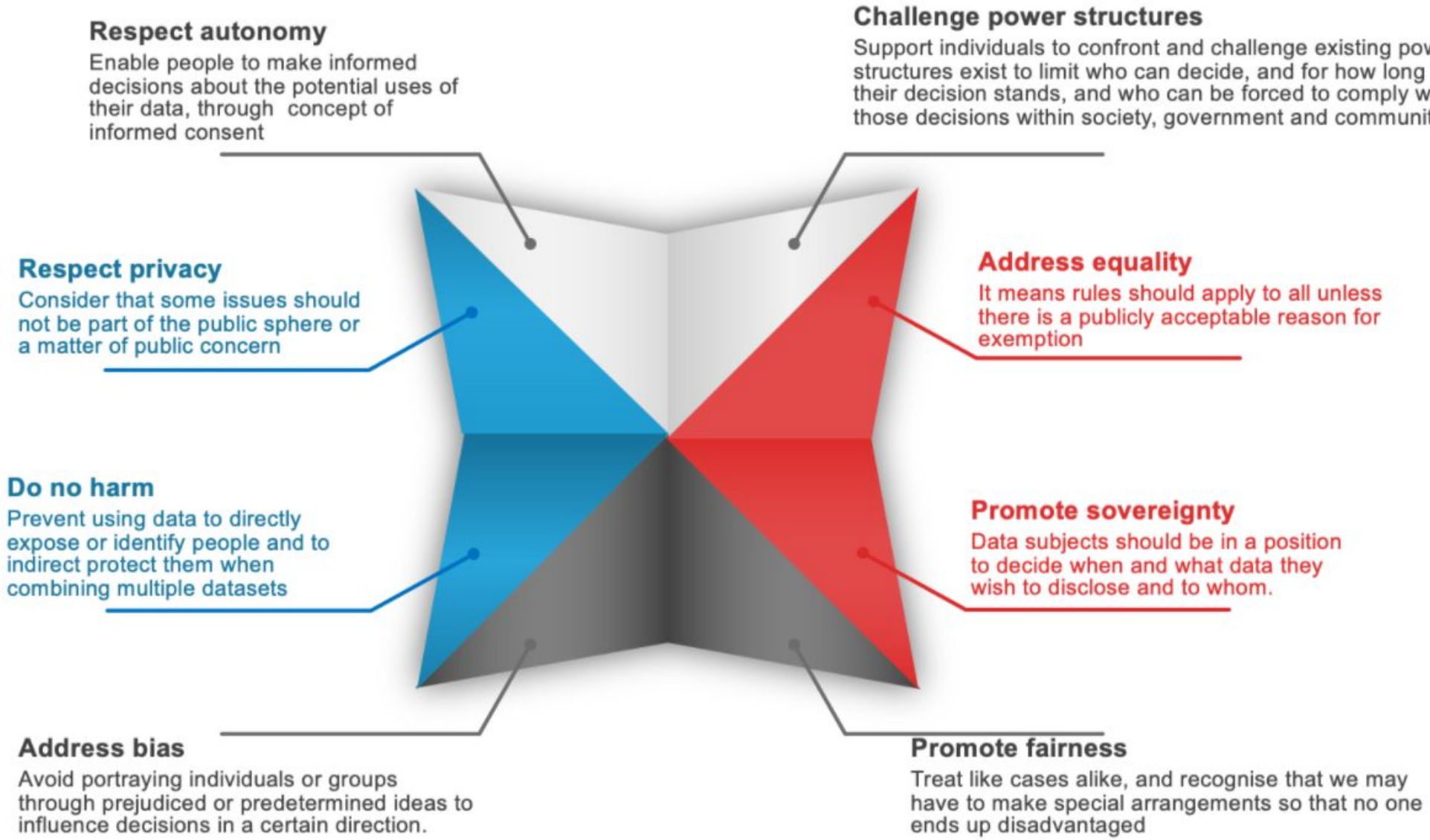
- With datafication and algorithmic automation of decisionmaking we are seeing increased calls for "**ethics training**" for those who will collect, process or exploit our personal data.
- But it is far from clear what exactly ethics training might cover - quite **superficial** (tickbox) or **narrow** approaches are common - posing the danger of "*ethicswashing*".
- Instead we propose a view of ethics as core to, and embedded across, data lifecycles: "**ethics as method**", inseparable from **criticality.**

Underpinning ethics frameworks
Atenas, Havemann, Kuhn and Timmermann

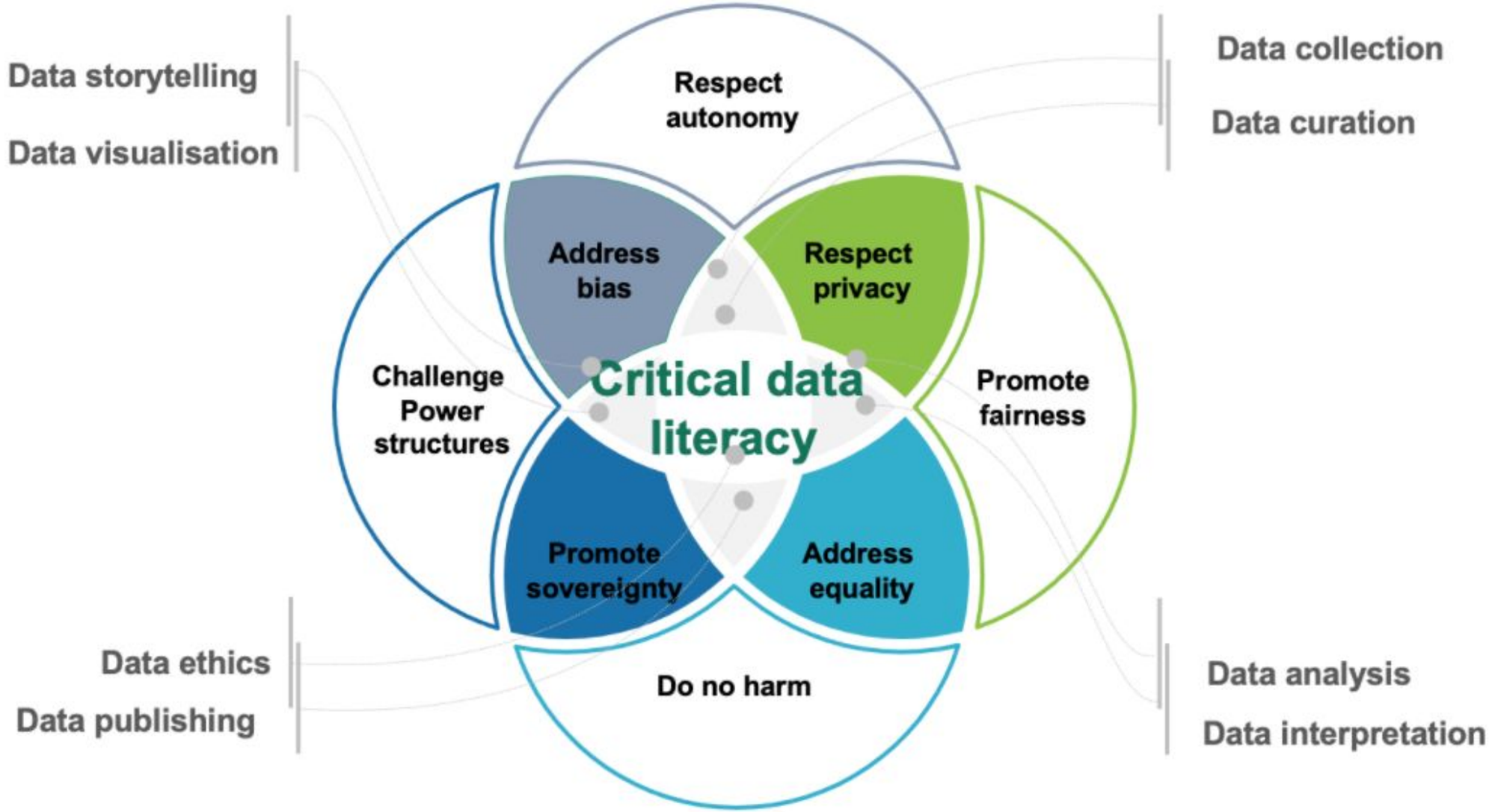


A framework on data ethics to develop critical data literacy

Atenas, Havemann, Kuhn and Timmermann



Critical data literacy model



Atenas, Havemann, Kuhn and Timmermann

Atenas, Havemann, Kuhn and Timmermann

Atenas, Havemann, Kuhn and Timmermann

Core technical data skills

Skills/Level	Basic	Intermediate	Proficient	Advanced
Data curation skills	The capacity of organising datasets in simple folders	The capacity of identifyng datasets from different sources and organise them using simple database systems	The capacity of using data management tools for data curation, organising files in order to facilitate retrieval by descring it through controlled vocabularies	The capacity of developing databases to automate the process of organising merge datasets, embedding standarised metadata into the files to facilitate information retrieval
Data management skills	The capacity to identify and retrieve datasets from different sources	The capacity of identify and retrieve datasets from different portals in different formats and organise them using a standard format	The capacity of extract, filter and compare data from different data sources creating a new datasets and organise them in data management systems	The capacity of identify, retrieve, select, filter and format data in different formats to creating databases in data management systems
Data Mining skills	The capacity of extracting data from different sources published in different formats to create new datasets	The capacity of extracting and collecting data and analyse it for trends, patterns, summaries, and meaning using basic statistic modelling	The capacity to manage and process large data sets both using relational and non-relational databases	The capacity of proficiently understand Natural Language Processing (NLP) Machine Learning and Deep Learning Algorithm to analyse large amounts of data using programming and statistics languages
Data visualisation skills	The capacity of create simple graphics to explain data using the right kind chart to showcase data	The capacity of developing simple infographics portraying data in a clear and straightforward manner	The capacity of using statistical modelling software to design complex charts	The capacity of use data visualisation software techniques to present their findings using complex statistical modelling

Atenas, Havemann, Kuhn and Timmermann

Data literacy training activities				
Skills / Level	**Basic**	**Intermediate**	**Proficient**	**Advanced**
Critical thinking	Data can be used to foster basic understanding of critical thinking by showing different means to present it	Data can can be used to verify information from the media comparing how information published about the same issue is portayed by different outlets	Data can be used to analyse a wide range phenomena related and be used to write reports to propose solutions to problems	Data from different sources can be used to develop and present complex evidence-based arguments in a wide formats aiming at a diverse audiences
Data analysis skills	Open datasets can be used to gain understanding basics concepts of quantitative and qualitative research methods	Open data can be used to gain experience in learning basics of data analysis using quantitative and qualitative data analysis software	Open Data can be used build large data sets crossing information from various sources and using data cleaning tools before conducting analysis using different methods and reporting data insights using data storytelling techniques.	Data maining can be used to obtain insightful information through by performing complex data analysis using advanced statistical modelling techniques
Research skills	Data can help understanding the scientific method and become familiar with the concepts of quantitative and qualitative research methods	Data can be used to explain the data cycle from collection to publication including consent and data ethics	Open data can be used to replicate experiments following the research methods explained in literature asociated with such data or similar data and to report the findings using a wide range of techniques	Data ethics approached can be used to assess the quality of a study or a research project in regards with how it has been collected, analysed and published, thus it can be used to mitigate the potential harmful impact of data

The #datapraxis dream team

Javiera Atenas
University of Suffolk
@jatenas

Caroline Kuhn
Bath Spa University
@carolak

Leo Havemann
Open University & UCL
@leohavemann

Cristian Timmermann
University of Ulm
@c_timmermann

datapraxis.net