DSI Data Day – July 4, 2018

Ethical Challenges of Data Research

Markus Christen, UZH Digital Society Initiative

Overview

In this presentation, we will discuss three topics:

1) **The Research Data Life Cycle and its guiding Ethical Principles**

2) **Additional challenges through Big Data**

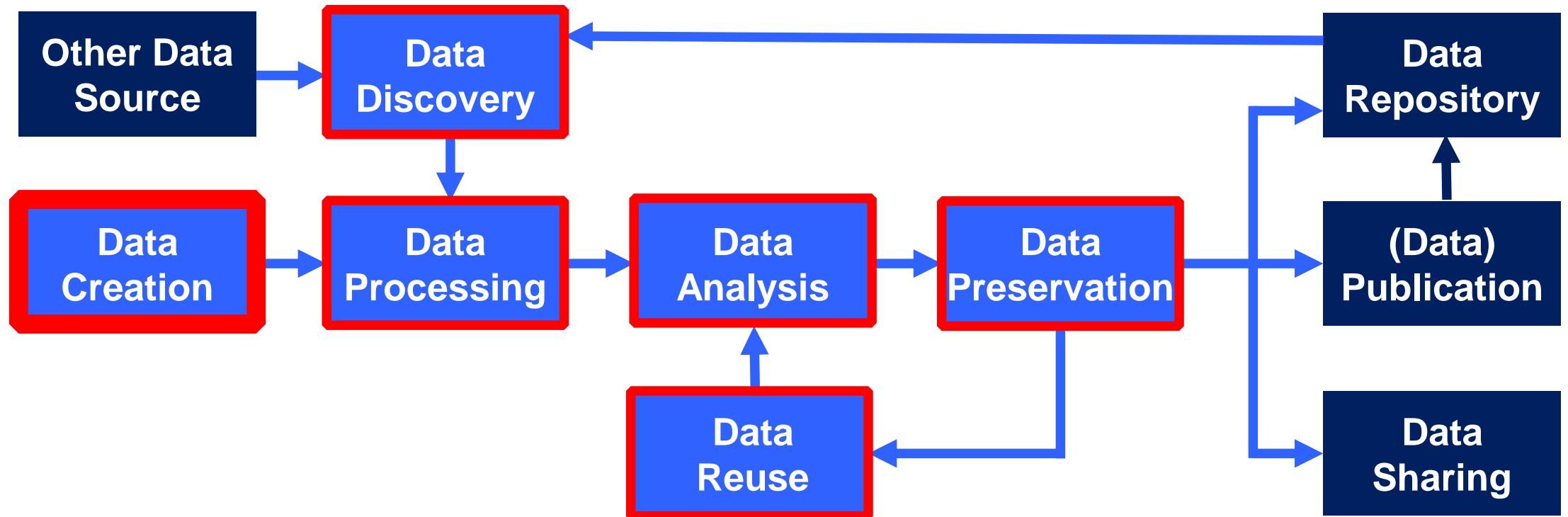3) **Sketch of a framework for data ethics in research**

The Research Data Life Cycle and its guiding Ethical Principles

Research Data Life Cycle

Guiding Principles & Tools

Fairness

Societal value

Participant benefits

Valid methods and questions

Informed consent

Rules of crediting

Autonomy

Anonymity/ confidentiality

Review processes

Integrity / Responsibility
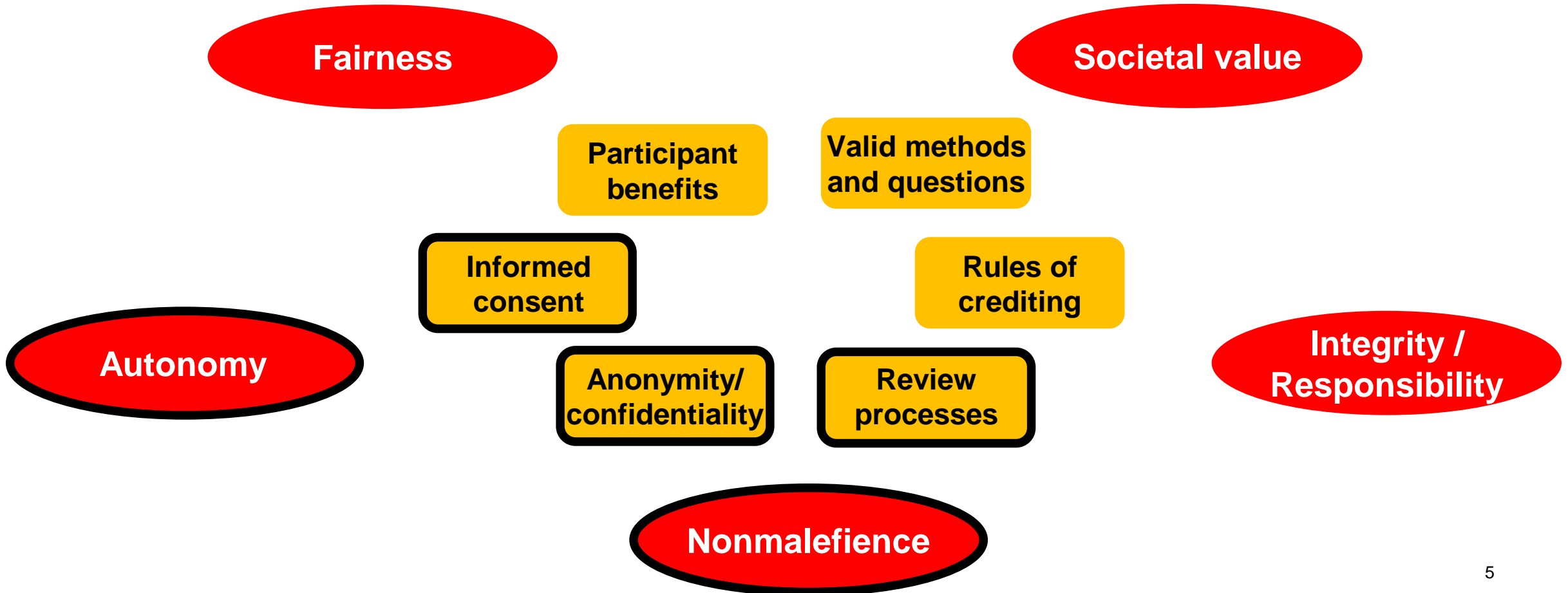
Nonmalefience

Additional challenges through Big Data

Additional challenges through Big Data?

Big Data

Size & complexity of data

Sophisticated analysis tools

Ethics

Volume

Variety

Velocity

Veracity

Surveillance

Privacy

Falling behind

Truthfulness

Pattern discovery

Categorization

Decision-making

Machine learning

Exploitation

Discrimination

Responsibility

Autonomy
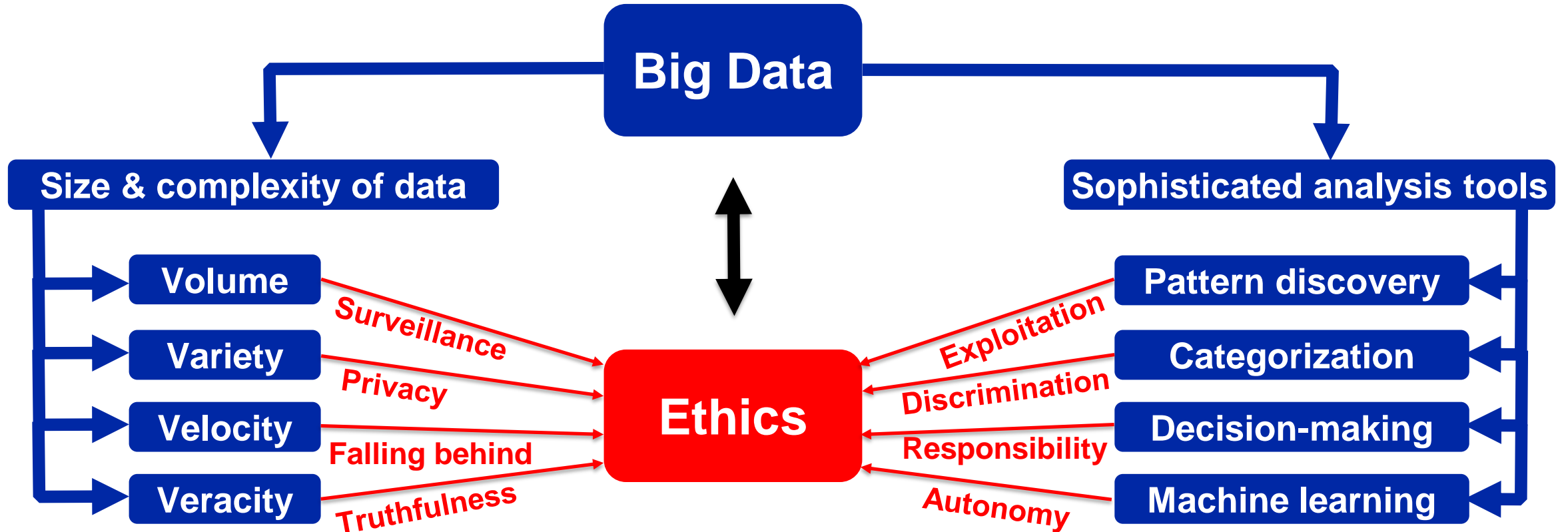
Big Data and contextual integrity (of data)

- **Contextual integrity** is inspired by the idea of *spheres of justice* (Walzer 1983): societies consist of different social spheres each defined by different types of goods, distribution mechanisms of goods as well as relevant values and principles that justify allocations etc.

- The major **ethical challenge** is to prevent the domination of a single good, distribution mechanism, principle etc. *across spheres*.

- "Translating" this idea to the **information sphere** (Nissenbaum 2004) means that the type of information that is revealed and the flows between different spheres have to be *appropriate for the context*.
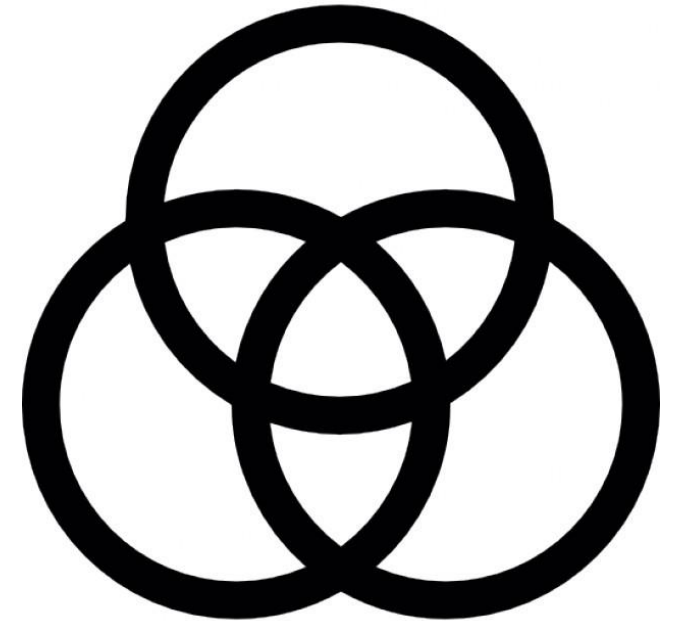
Big Data challenges for autonomy & consent

When research data adds up to "Big Data", the following problems may be intensified:

- When individuals provide multiple data sources through research, they are often in a position of **informational asymmetry**: they are not aware of the informational links between social spheres that are generated in this way.

- In the context of Big Data, the amount of information extracted from data might **exceed ex-ante expectations** of both users and researchers.

- The current orientation on autonomy and harm-reduction puts the focus on the individual and **conceals the moral obligations of the other players** involved in (Big) data research.

Sketch of a framework for data ethics in research

Reframing (some) of the guiding values

In the following, it is proposed to reframe three of the five values presented in the beginning:

Autonomy: Users ought to be aware of how their data records are used in order to promote *their* values and gain control over privacy-related choices.

Responsibility: Users (both researchers and data providing research subjects) should be held responsible and accountable for the ways in which they use their personal information and the information about other people. If some subjects are wronged, it must be possible to attribute personal responsibility for the wrongs in question.

Fairness: The benefits of knowledge and information ought to be fairly apportioned to all participants in interactions, so as to rule out inequality of opportunity and exploitation by some at the expense of others.

New tools for enabling autonomy

- Enable research participants to **gain awareness on what guides their choices** (privacy preferences), e.g. through a privacy preferences self-assessment tool that will provide a value profile that outlines the privacy preferences of participants with respect to their participation in research or data donation.

- Provide information (to participants and researchers) on **what they potentially may disclose** when providing certain types of data. This may include a security issues taxonomy; i.e. forensic and security assessment of relevant risks when using the platform.

The goal is to shift away the focus from (mere) informed consent towards empowering research participants and data donators.

New tools for improving responsibility

- Ensure longer-term relations between participants and researchers through an infrastructure (social network) that allows for **bidirectional relations** (e.g., for suggesting new research questions by participants; citizen science).

- **Empower the researcher** both regarding legal / ethical requirements and technical instruments (e.g. for data anonymization) for doing responsible research with personal data. This may include profile anonymisation tools, including masking and synthetic data methods used in statistical disclosure control (micro-aggregation, noise addition, etc.).

- **Empower the participant** with the ability to verify how safe is the anonymization performed by the data collector/researcher.

The goal is to provide both the infrastructure and tools for stable relations between researchers and participants as a prerequisite for responsible research.

New tools for promoting fairness

- Provide a **broader set of utilities** (not only monetary compensation) like visualizing the contribution of research participants, e.g. through donated data, to certain scientific results.

- Create **novel types of interactions** (using, e.g., co-private protocols, Domingo-Ferrer 2011, and, more generally, co-utile protocols, Domingo-Ferrer et al. 2015) that allow collaborative contribution to a common good (like ensuring each other's privacy).

- Provide **anti-discrimination tools**, i.e. models and protocols of data acquisition and analysis for quantifying the risk of discriminatory decisions as a (possibly unwanted) consequence of data profiling and data mining.

The goal is to demonstrate that contributing to research is based on a fair exchange and mutual respect of the involved parties.

An envisaged data research ethics infrastructure (1)

The Digital Society Initiative of the University of Zurich is currently conceptualizing a Swiss Infrastructure for Responsible Online Research (SIRON) that should include some of those tools.
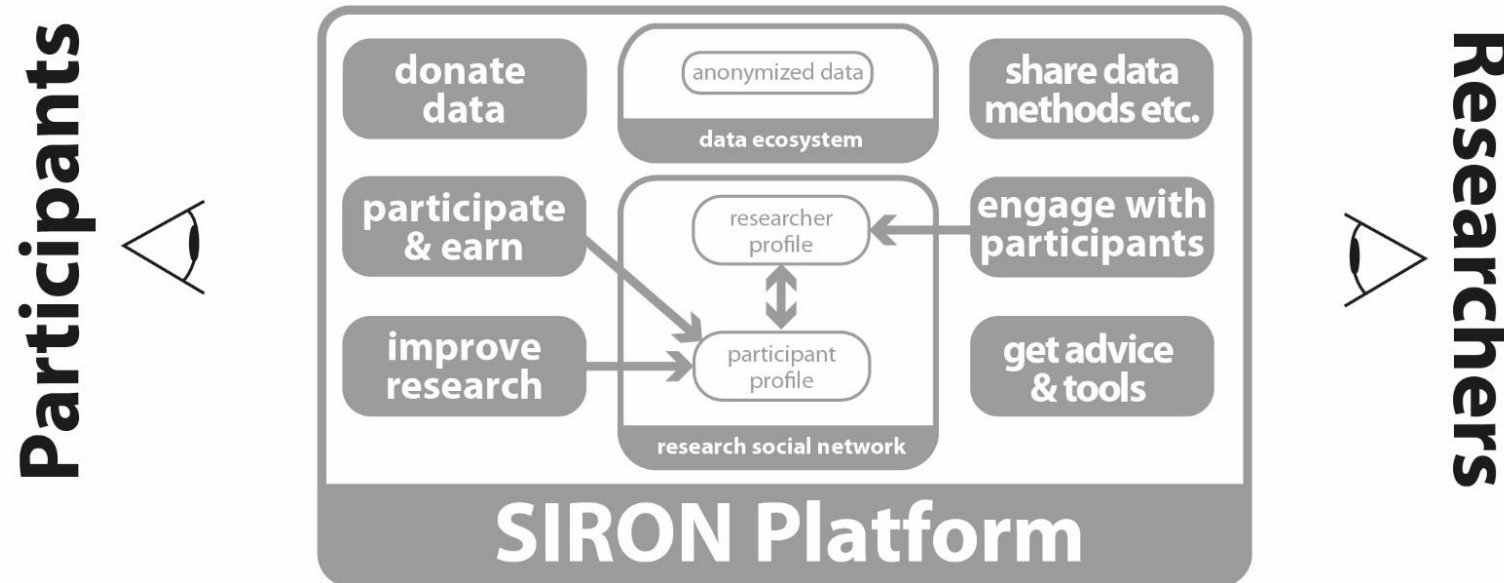
An envisaged data research ethics infrastructure (2)

In collaboration with the Delegate for Data Protection of the UZH, the DSI is currently conceptualizing a "One-Stop-Shop" for Data Protection & Data Ethics, where researchers…:

… obtain an integral assessment on the data protection and ethical issues their project may rise including documentation;

… get advice on how to handle potential problems (by providing best practice examples, guidelines, digital tools, etc.);

… are directed to the responsible body (data protection officer, faculty-internal ethics board, KEK) if needed.