

Privacy & Innovation

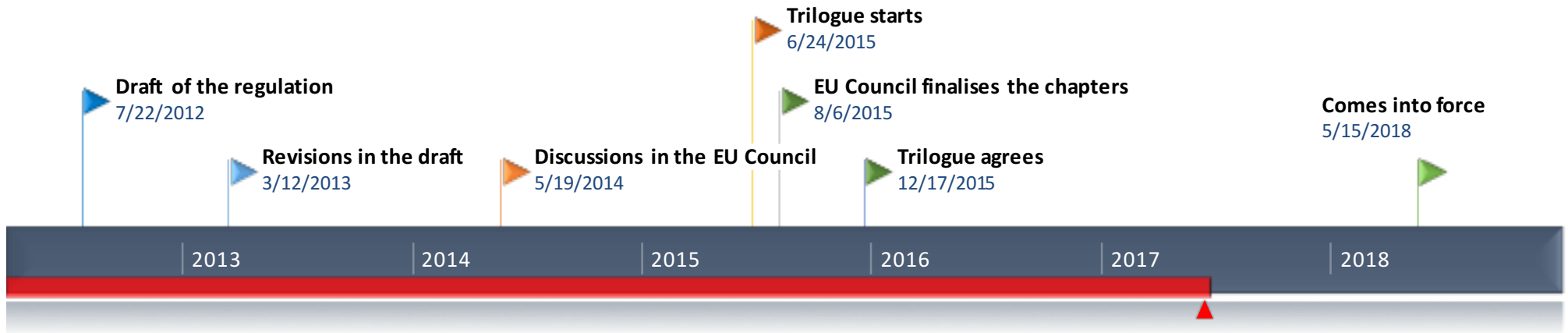
Sabrina Kirrane, Wirtschaftsuniversität Wien



Horizon 2020
European Union funding
for Research & Innovation

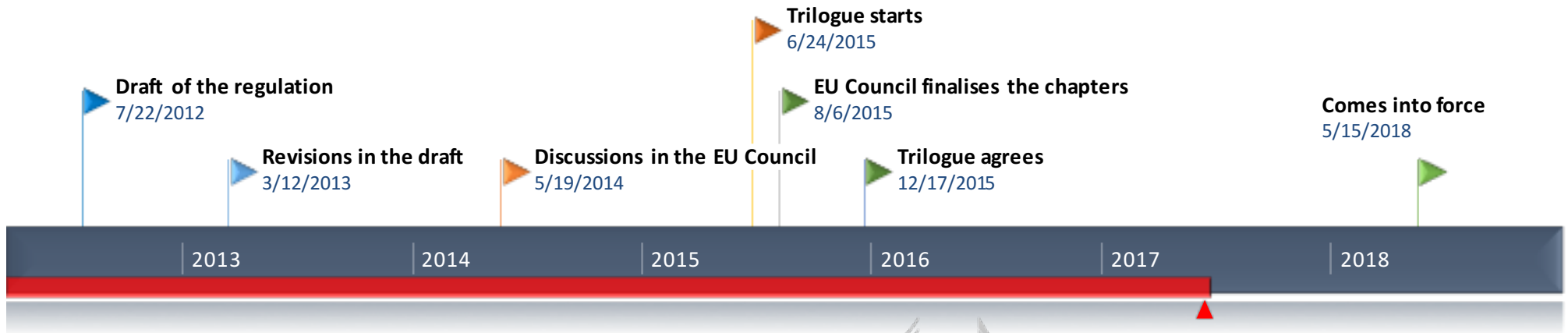


The General Data Protection Regulation





The General Data Protection Regulation



Horizon 2020
European Union funding
for Research & Innovation





Impact for data subjects?

1 Data to take away

I can get back the data I provided to an organisation or online-service and transmit those to other ones (social networks, Internet service provider, online streaming supplier, etc)



2 Better transparency

I know what is done with my data and it's easier for me to exercise my rights.



3 Child protection

Online services must obtain the parents' consent before registering any child under 16 or less if provided by national laws.



4 One-stop-shop

In case of problems with how my data is handled, I can contact my national data protection authority, whatever the country where the organisation is processing my data.



5 Bigger sanctions

When infringing the regulation, the organisation at fault can be fined up to 20 000 000 € or 4% of its annual worldwide turnover.



6 Right to be forgotten

I can ask search engines to delist a web page that affects my privacy negatively or ask a website to erase an information, under certain circumstances.





Impact on data driven operations & Innovation?

Data Market



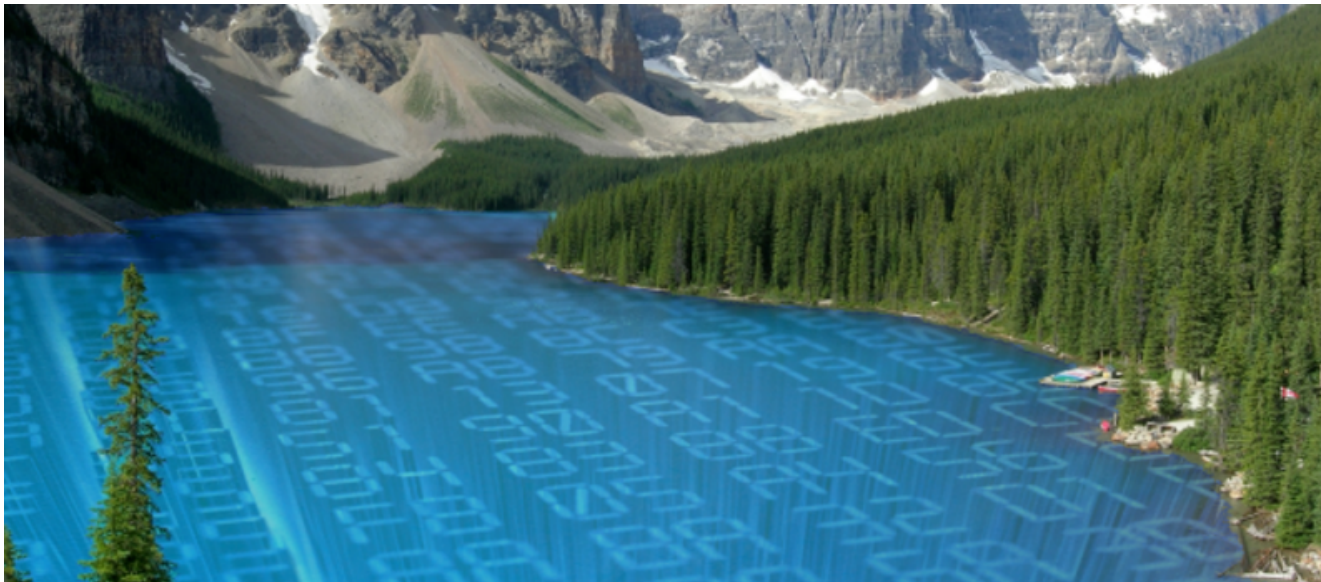
<http://themerkle.com/slur-io/>

Data Vault



<http://www.miamidatavault.com/>

Data Lake



<https://solutionsreview.com/data-integration/the-emergence-of-data-lake-pros-and-cons/>



Innovation via Anonymisation!

4.5.2016

EN

Official Journal of the European Union

L 119/1

I

(Legislative acts)

REGULATIONS

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 27 April 2016
on the protection of natural persons with regard to the processing of personal data and on the free
movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 16 thereof,

Having regard to the proposal from the European Commission,

After transmission of the draft legislative act to the national parliaments,

Having regard to the opinion of the European Economic and Social Committee ⁽¹⁾,

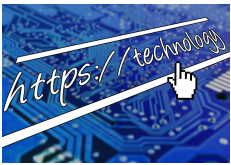
Having regard to the opinion of the Committee of the Regions ⁽²⁾,



Which \mathcal{K} should
we use for
 \mathcal{K} -Anonymity?

The GDPR does not apply to anonymous data where the data subject is no longer identifiable.

- (26) The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

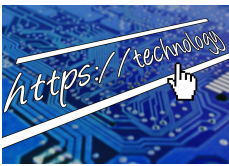


\mathcal{K} -anonymity

- A record cannot be distinguished from at least \mathcal{K} -1 others
- Approach
 - **Suppression** certain values of the attributes are replaced by an asterisk
 - **Generalization** individual values of attributes are replaced by with a broader category

A 3-anonymous patient table

Zipcode	Age	Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
4790*	≥ 40	Flu
4790*	≥ 40	Heart Disease
4790*	≥ 40	Cancer
476**	3*	Heart Disease
476**	3*	Cancer
476**	3*	Cancer



Is \mathcal{K} -Anonymity enough?

Homogeneity Attack

Bob	
Zipcode	Age
47678	27

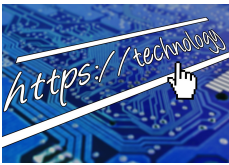
A 3-anonymous patient table

Zipcode	Age	Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
4790*	≥40	Flu
4790*	≥40	Heart Disease
4790*	≥40	Cancer
476**	3*	Heart Disease
476**	3*	Cancer
476**	3*	Cancer

Background Knowledge Attack

Carl	
Zipcode	Age
47673	36

\mathcal{K} -anonymity has deficiencies when sensitive values in an equivalence class lack diversity or the attacker has background knowledge



Is \mathcal{K} -Anonymity & \mathcal{L} -Diversity enough?

- Each equivalence class has at least \mathcal{L} well-represented sensitive values

Similarity Attack

Bob	
Zip	Age
47678	27

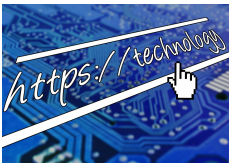
A 3-diverse patient table

Zipcode	Age	Salary	Disease
476**	2*	20K	Gastric Ulcer
476**	2*	30K	Gastritis
476**	2*	40K	Stomach Cancer
4790*	≥ 40	50K	Gastritis
4790*	≥ 40	100K	Flu
4790*	≥ 40	70K	Bronchitis
476**	3*	60K	Bronchitis
476**	3*	80K	Pneumonia
476**	3*	90K	Stomach Cancer

Conclusion

- Bob's salary is between [20k,40k].
- Bob has some stomach-related disease.

\mathcal{L} -diversity does not consider the semantic meanings of the sensitive values



Is \mathcal{K} -Anonymity, \mathcal{L} -Diversity & \mathcal{J} -Closeness enough?

- Distribution of sensitive attributes within each quasi identifier group should be “close” to their distribution in the entire original database

Background Knowledge Attack

Bob	
Zip	Age
47678	27

Conclusion

- Bob could have Flu, Heart Disease or Cancer!

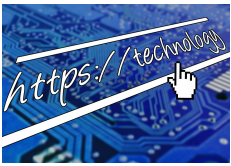
A completely generalised table

Age	Zipcode	Gender	Disease
*	*	*	Flu
*	*	*	Heart Disease
*	*	*	Cancer
.
.
.
*	*	*	Gastritis

A released table

Age	Zipcode	Gender	Disease
2*	476**	Male	Flu
2*	476**	Male	Heart Disease
2*	476**	Male	Cancer
.
.

There are other anonymisation approaches however it is getting harder and harder to guarantee anonymity!



Is \mathcal{K} -Anonymity, \mathcal{L} -Diversity & \mathcal{J} -Closeness enough?



Which \mathcal{K} should we use for \mathcal{K} -Anonymity?

- A layered approach to anonymisation may be needed
- Even then \mathcal{K} , \mathcal{L} & \mathcal{J} are highly dependent on the data
- Also, there is a tradeoff between anonymisation and utility

Considering that it is getting harder and harder to guarantee anonymity, *what is the alternative?*



Innovation via consent & transparency?

4.5.2016

EN

Official Journal of the European Union

L 119/1

I

(Legislative acts)

REGULATIONS

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 27 April 2016

on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 16 thereof,

Having regard to the proposal from the European Commission,

After transmission of the draft legislative act to the national parliaments,

Having regard to the opinion of the European Economic and Social Committee (1),

Having regard to the opinion of the Committee of the Regions (2),





Innovation via **consent** & transparency?

4.5.2016 EN Official Journal of the European Union L 119/1

I

(Legislative acts)

REGULATIONS

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 27 April 2016

on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 16 thereof,

Having regard to the proposal from the European Commission,

After transmission of the draft legislative act to the national parliaments,

Having regard to the opinion of the European Economic and Social Committee ⁽¹⁾,

Having regard to the opinion of the Committee of the Regions ⁽²⁾,

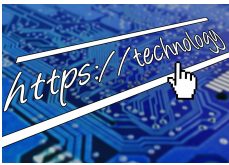


Can we adopt a soft opt-out approach?

What level of consent is specific enough?


Affirmative action, freely given, specific, informed and unambiguous indication

- (32) Consent should be given by a clear **affirmative act** establishing a **freely given, specific, informed and unambiguous indication** of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement. This could include ticking a box when visiting an internet website, choosing technical settings for information society services or another statement or conduct which clearly indicates in this context the data subject's acceptance of the proposed processing of his or her personal data. Silence, pre-ticked boxes or inactivity should not therefore constitute consent. Consent should cover all processing activities carried out for the same purpose or purposes. When the processing has multiple purposes, consent should be given for all of them. If the data subject's consent is to be given following a request by electronic means, the request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided.

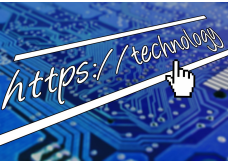


Innovation via **consent** & transparency?

- Condition of use
 - Consent will not be valid if the individual has no choice but to give their consent
- Opt-In
 - Research shows that users spend almost no time browsing the text of the agreement before clicking the box
 - It is far from clear that the opt-in model achieves the goal of informed
- Opt-Out
 - Facilitates innovation while still safeguarding autonomy
 - Suitably prominent opt-out box might help to establish that clicking the button was a positive indication of consent




Can we adopt a soft opt-out approach?



Innovation via **consent** & transparency?

- We need a **non-intrusive** means to obtain consent for new usages
- When it comes to data driven services we **may not know the type of consent** we need a priori
- At this stage it is difficult to say **what constitutes specific enough**



What level of consent is specific enough?



Innovation via consent & transparency?

Official Journal of the European Union

L 119/1

I

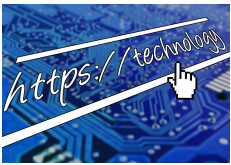
(Legislative acts)

-Personal data that is corrected, used, consulted or otherwise processed
-time limits should be established
-rectification or deletion

Do we have to delete data from logs & backups?

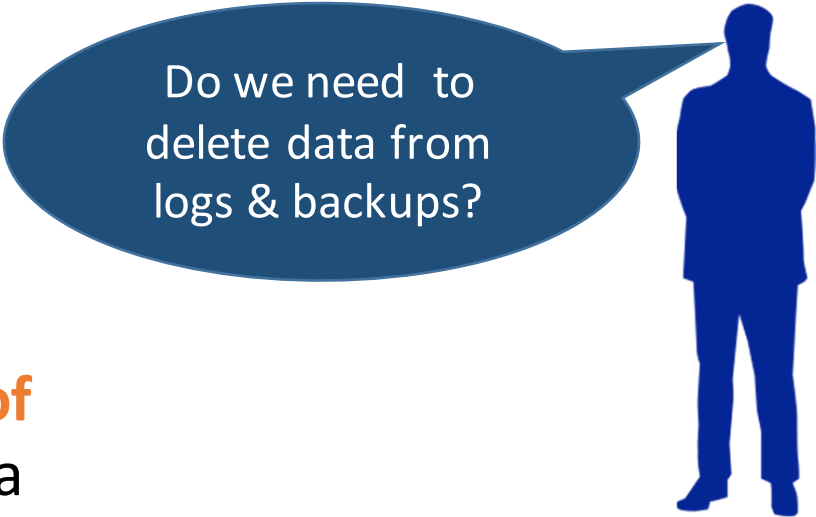


- (39) Any processing of personal data should be lawful and fair. It should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed. The principle of transparency requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used. That principle concerns, in particular, information to the data subjects on the identity of the controller and the purposes of the processing and further information to ensure fair and transparent processing in respect of the natural persons concerned and their right to obtain confirmation and communication of personal data concerning them which are being processed. Natural persons should be made aware of risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing. In particular, the specific purposes for which personal data are processed should be explicit and legitimate and determined at the time of the collection of the personal data. The personal data should be adequate, relevant and limited to what is necessary for the purposes for which they are processed. This requires, in particular, ensuring that the period for which the personal data are stored is limited to a strict minimum. Personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means. In order to ensure that the personal data are not kept longer than necessary, time limits should be established by the controller for erasure or for a periodic review. Every reasonable step should be taken to ensure that personal data which are inaccurate are rectified or deleted. Personal data should be processed in a manner that ensures appropriate security and confidentiality of the personal data, including for preventing unauthorised access to or use of personal data and the equipment used for the processing.



Innovation via consent & **transparency**?

- Time limits will involve a **major cultural change**
- In the case of rectification it may suffice to **update data in the line of business application(s)** and enter a new record in the log indicating that the data was updated at the request of the data subject
- It might be possible to use **cryptographic delete**



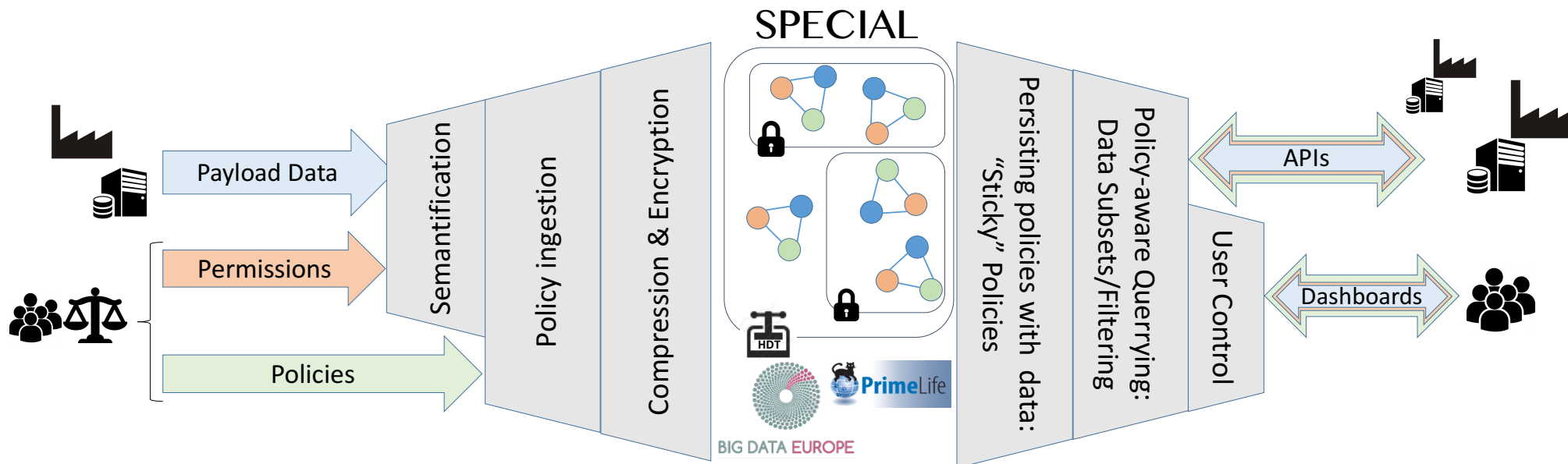
Do we need to delete data from logs & backups?

In the Era of Big Data how can we support consent and transparency at scale?



SPECIAL

Scalable Policy-aware Linked Data arChitecture for prIvacy, trAnsparency and complIance





SPECIAL

Scalable Policy-aware Linked Data Architecture for privacy, transparency and compliance

Technical/Scientific contact

Sabrina Kirrane

Vienna University of Economics
and Business

sabrina.kirrane@wu.ac.at

Administrative contact

Philippe Rohou

ERCIM W3C

philippe.rohou@ercim.eu

