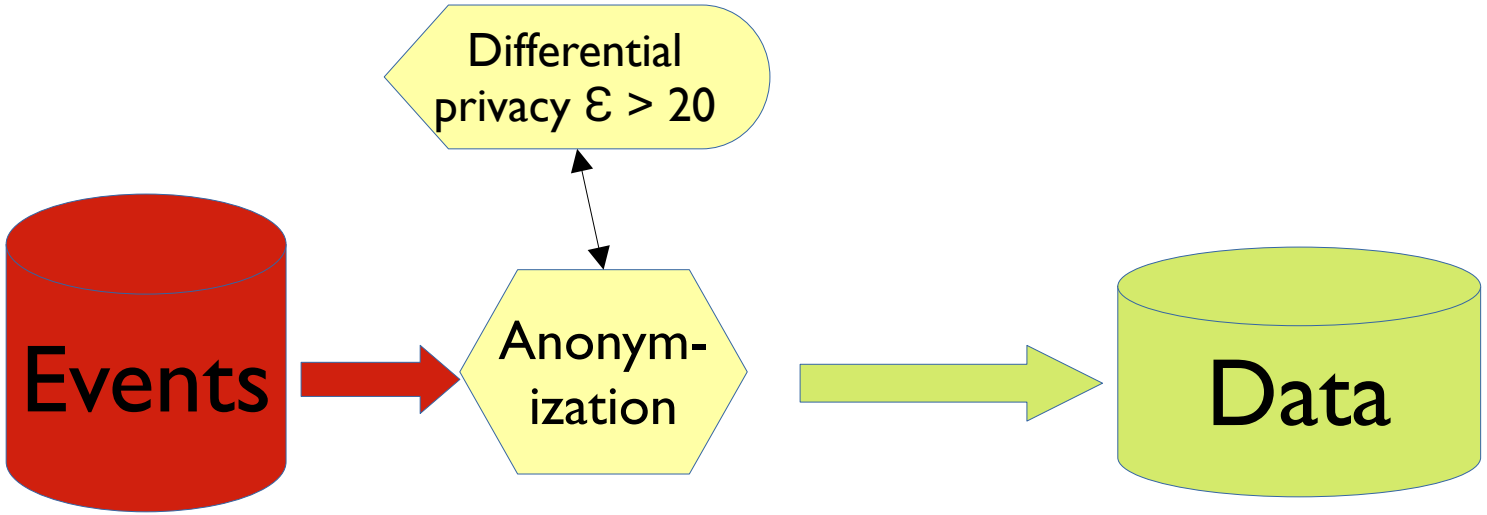
ERCIM
European Research Consortium
for Informatics and Mathematics

Privacy and legal issues in Big Data

Rigo Wenning
ERCIM/W3C Legal counsel
rigo.wenning@ercim.eu rigo@w3.org @rigo@mamot.fr

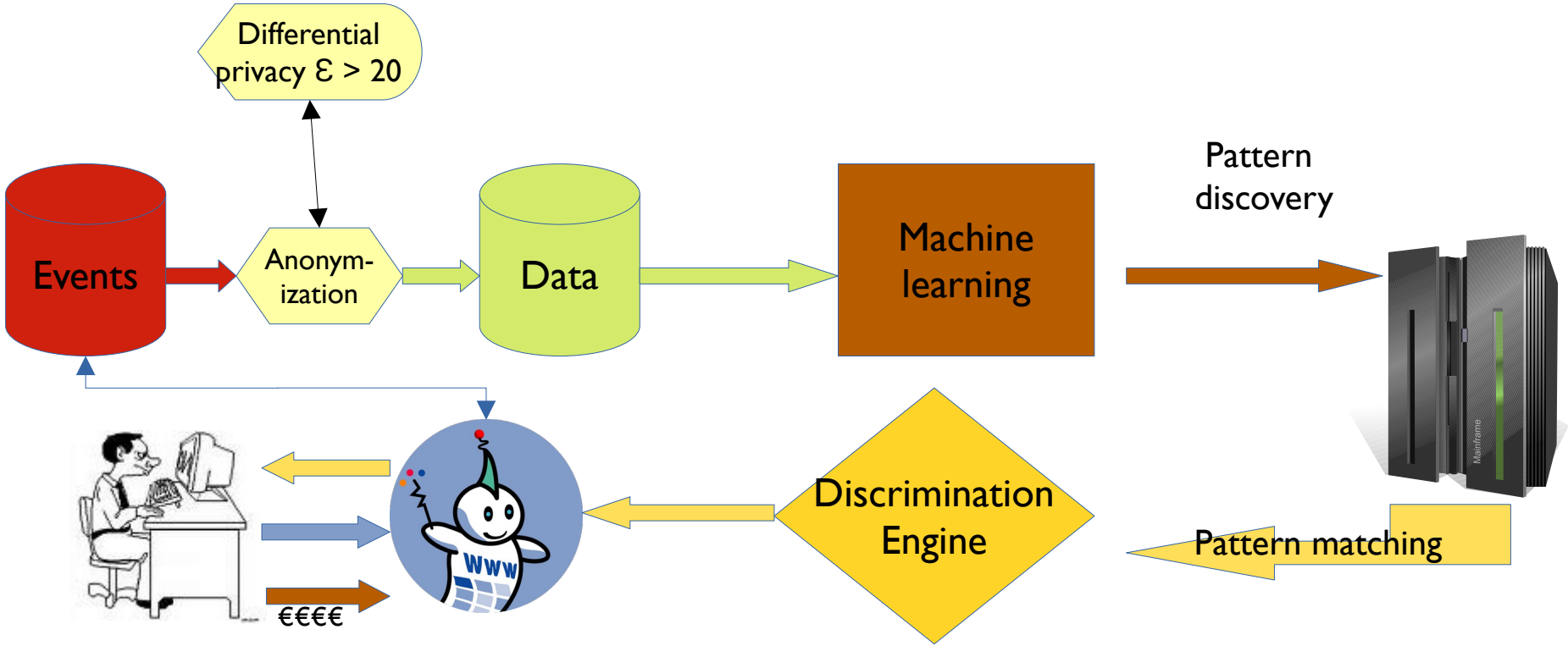
ERCIM
European Research Consortium for Informatics and Mathematics

Escaping GDPR



ERCIM
European Research Consortium for Informatics and Mathematics

Discrimination without PII



ERCIM
European Research Consortium for Informatics and Mathematics

Legal implications

- Purpose limitation despite anonymization?
  - Limited support in GDPR
- Consent to anonymization?
  - Stretching GDPR to mask a hole
- AI regulation?
  - Not there yet, act NOW

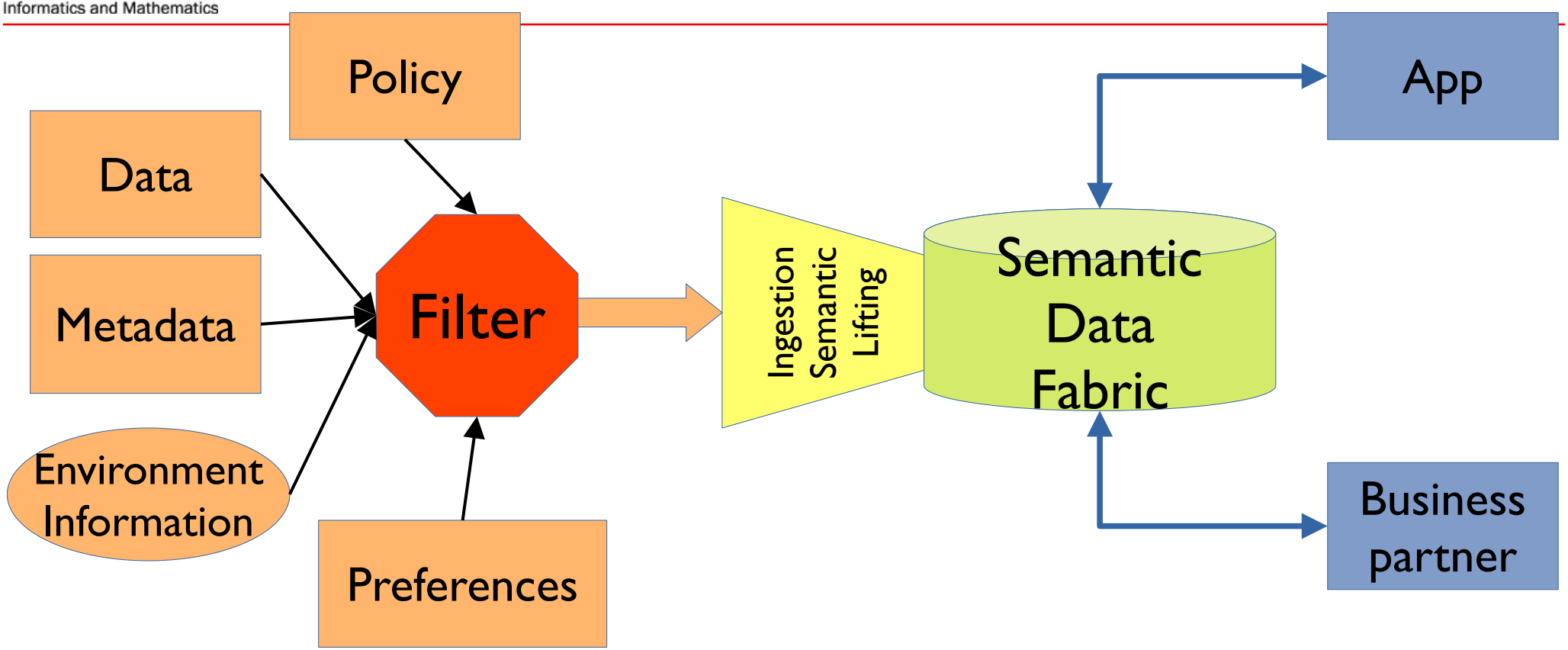
ERCIM
European Research Consortium for Informatics and Mathematics

Moving the goal posts

- Initially, data protection protected "autonomy" and opinion building, thus protected democracy
- Came the internet and the use of DP as a consumer protection tool
- Came AI and we start to talk about data rights and commercial value chains.

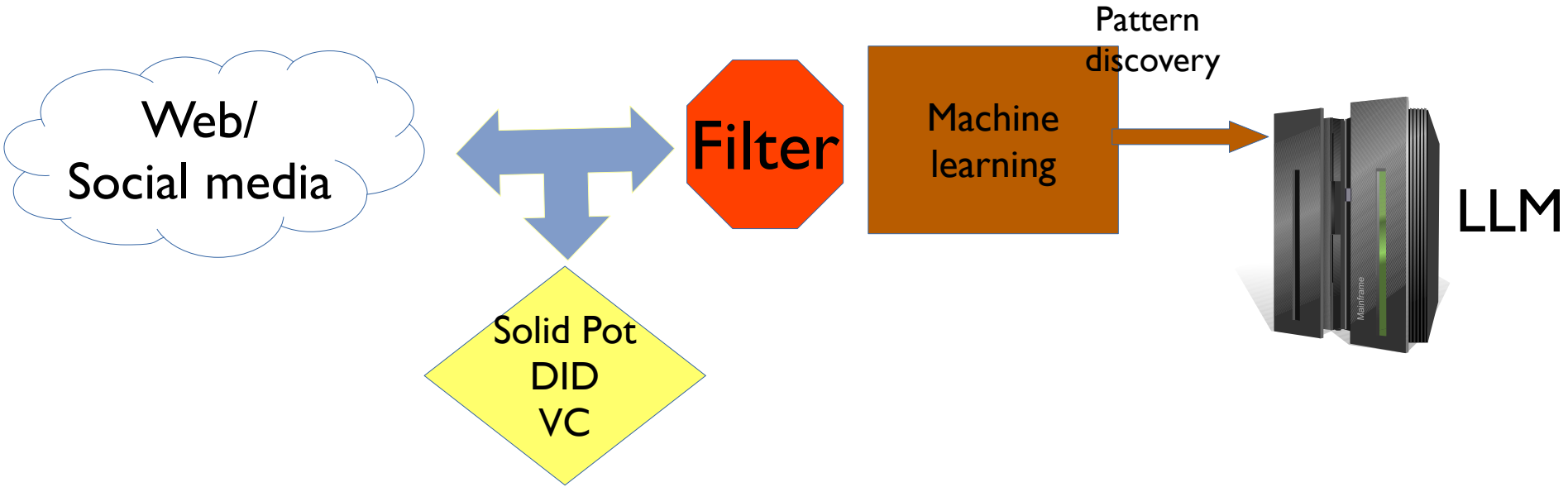
ERCIM
European Research Consortium for Informatics and Mathematics

Trapeze applications



ERCIM
European Research Consortium for Informatics and Mathematics

Applying Trapeze to public info



ERCIM
European Research Consortium for Informatics and Mathematics

Legal implications

- What is public information?
  - What is commons?
  - What is societal conversations not meant to be available for any purpose?
- What is allowed: Special rules in 2019/790/EU on copyright
- GDPR and publicly available information

ERCIM
European Research Consortium for Informatics and Mathematics

GDPR bombshell

- 2023-03-30: RITENUTO pertanto che nella situazione sopra delineata, il trattamento dei dati personali degli utenti, compresi i minori, e degli interessati i cui dati sono utilizzati dal servizio si ponga in violazione degli artt. 5, 6, 8, 13 e 25 del Regolamento; → preliminary prohibition of ChatGPT in Italy

ERCIM
European Research Consortium for Informatics and Mathematics

The box is open

- 2023-04-13 Creation of a ChatGPT Task force by EDPS
- An avalanche of journal articles & and reports with legal analysis on GDPR
- First timid voices on copyright implications (see Directive 2019/790/EU)

ERCIM
European Research Consortium for Informatics and Mathematics

GDPR issues:

1. lack of information for data subjects whose data is processed by OpenAI (Art. 13 GDPR),
2. a lack of legal basis for the data processing (Art. 6 (1) GDPR),
3. the incorrect processing of personal data because the information presented by ChatGPT does not always correspond to the actual data (Art. 5 (1) d GDPR),
4. a breach of the requirements of Art. 8 GDPR, according to which an age verification mechanism is required to protect underage users.

ERCIM
European Research Consortium for Informatics and Mathematics

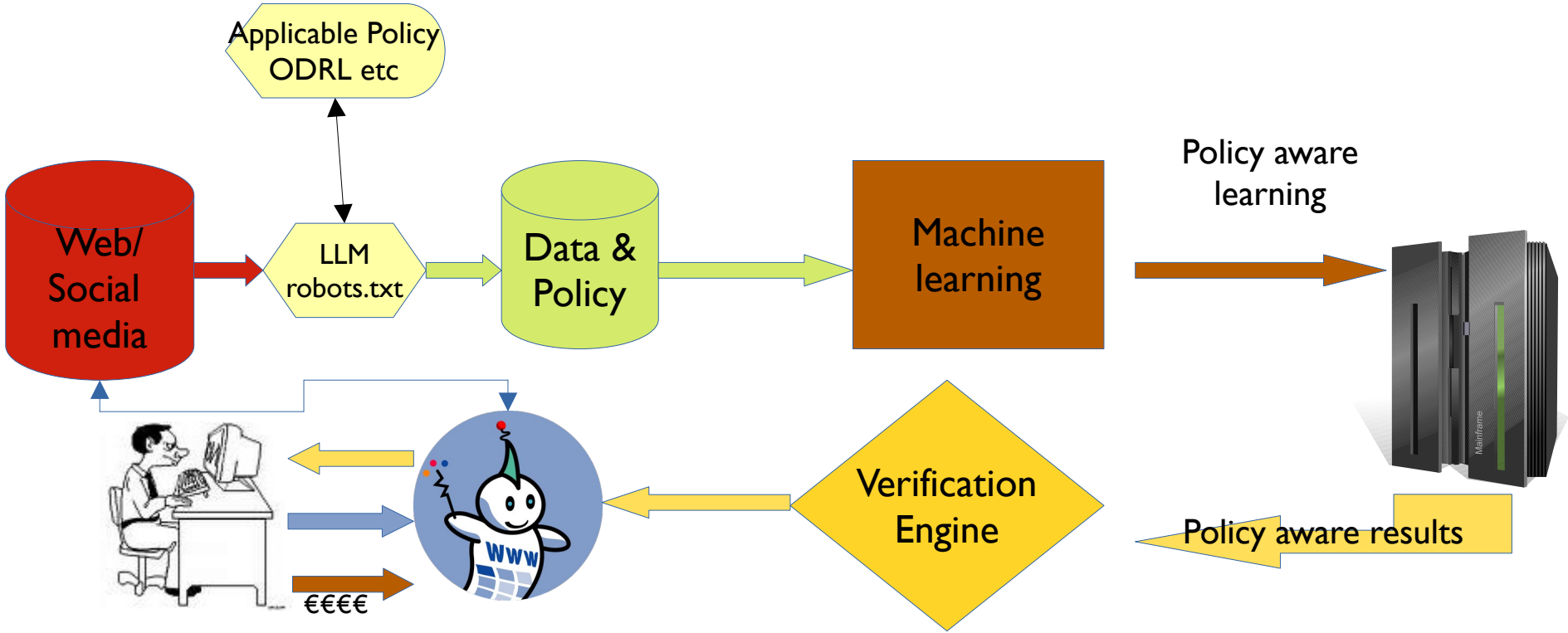
Copyright issues

- LLM may be seen as text & data mining (Art. 2(2))
  - ‘text and data mining’ means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations;
- LLM ingestion in the wild like ChatGPT may violate sui generis rights according to Art. 5-7 of Directive 96/9/EC
- The LLM itself may constitute a sui generis protected database
- Issues around the democratic discourse by overprotection of speech & misrepresentations at the same time

Rigo's Solutionism

ERCIM
European Research Consortium for Informatics and Mathematics

A new way of crawling for LLMs



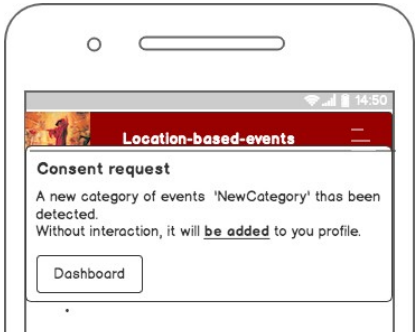
ERCIM
European Research Consortium for Informatics and Mathematics

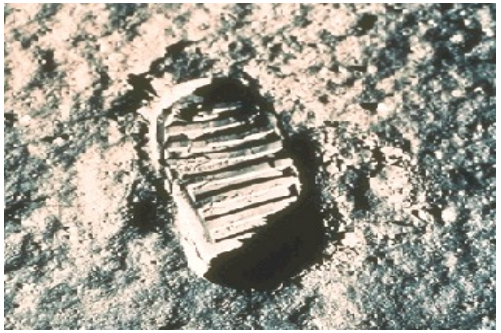
A new robots.txt approach for AI

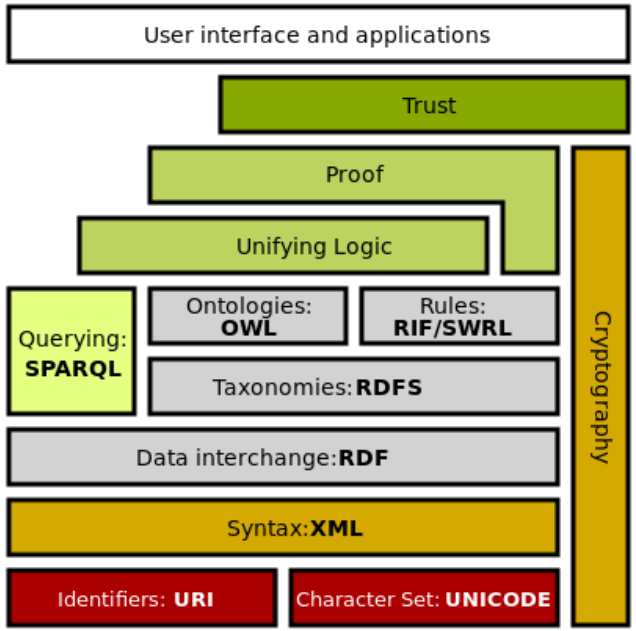
- More sophisticated expressions:
  - Access control expressions
  - Usage control expressions
  - Rights labeling
  - Revenue distribution
- A security for the creator of an LLM

ERCIM
European Research Consortium
for Informatics and Mathematics

Discussion







Rigo Wenning, rigo@w3.org