



DEPARTMENT OF  
**COMPUTER  
SCIENCE**



# Me want cookie!

Towards automated and transparent  
data governance on the Web



Jesse Wright - [jesse.wright@cs.ox.ac.uk](mailto:jesse.wright@cs.ox.ac.uk)

“



Customers need to be given  
control of their own data

**Sir Tim Berners-Lee**

# Motivations



Solid



Web Agents



Why this do we care?



# To make **Solid** work with General Data Protection Regulation (**GDPR**) (EU) and the Data Protection Act (**DPA**) (UK)

- Solid detaches *data* from *Web applications* giving users control over their data and the applications they use
- Data subjects (in this case users) have *direct control* over how the data in their Personal Data Store is shared
- Once data is shared, applications still require *consent* from the data subject to be able to lawfully process personal data under GDPR & DPA; with consent explicitly given for the specific purpose(s) the data will be processed.
- RQ: Can users pre-define their consent and can ODRL be used in Solid to communicate terms of use agreements between Pod and Application?



*Is Automated Consent in Solid GDPR-Compliant?  
An Approach for Obtaining Valid Consent with the Solid Protocol*

Why this do we care? **Solid**



# Creating a future where semi-autonomous **web agents** can represent **legal entities** and perform **online interactions** on their **behalf**

## Identify

Identify **legal entities**, such as **individuals** or **organisations** on the Web.

## Discover

**Discover** other **agents** representing an entity from their Web identity.

## Data Usage

**Describe**, and **agree to**, any **usage controls** associated with data they exchange – to share protected data while articulating the recipient's legal or moral obligations.

## Provenance

Describe the **origin** and **provenance** of data they exchange - so systems can identify which external claims to believe for a given task, based on the agent's internal trust model.

## Unambiguity

**Unambiguously** describe **ground truths** they send, and **agreements** they make, using a formal representation.

## Serendipity

**Contextualise** a task which may be ambiguous or poorly defined, such that interacting agents can introduce new solution spaces or negotiating actors in a serendipitous manner.







Why this do we care? **Web Agents**



# Creating a future where semi-autonomous web agents can represent legal entities and perform online interactions on their behalf



*Here's Charlie!  
Realising the  
Semantic Web  
vision of Agents in  
the age of LLMs*

	Risk / Internal Modelling	Exchange
Integrity / Provenance		 
Data Tagging & Terms of Use		 

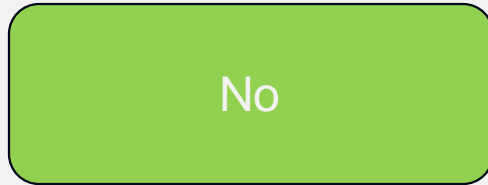
# Cookies







- Browser cookies are a mature, widely used and well-understood Web technology
- Makes a good target use case for academia, regulators and industry can ‘battle-test’ and mature technologies for semi-automated data governance.



- Pre-selected options
- Deceptive button colours
- Complex navigation
- Misleading labels
- Manipulative language



- Collect Consent
- Manage Consent
- Share Consent
- Prove Consent

## Want a cookie? 🍪

We and selected third parties use cookies or similar technologies as specified in the [cookie policy](#).

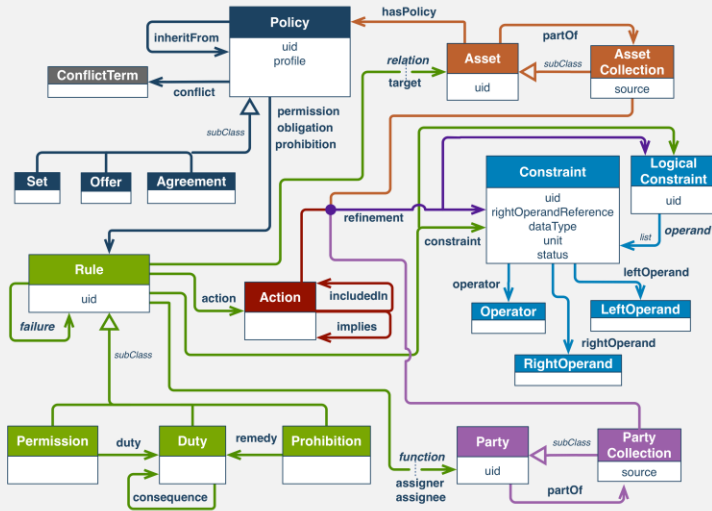
You can consent to the use of such technologies by using the "Accept" button.

Accept all cookies

Manage cookies

... but ...

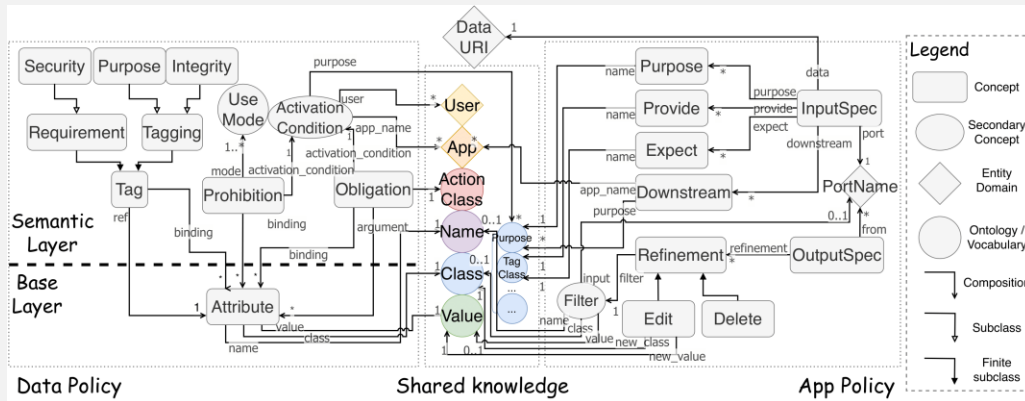
- **ODRL** is a W3C recommendation for the expression of policies over digital assets.



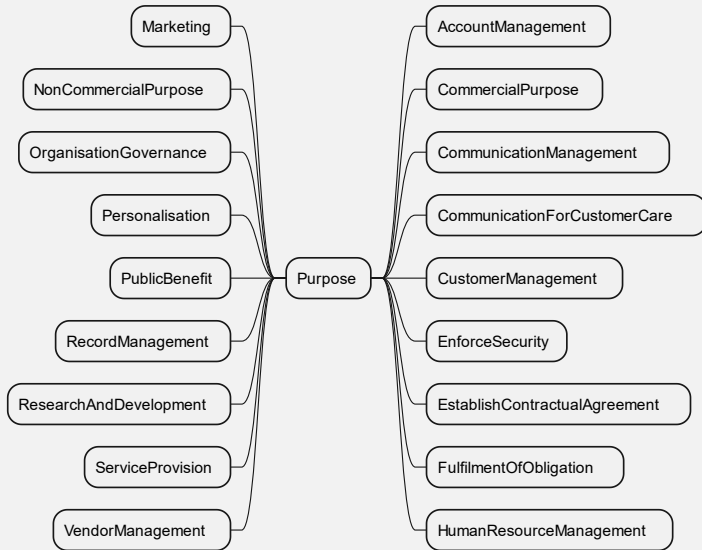
# Policy Languages



- **DToU** supports policy checking across applications and data providers, enabling users to decisions on application authorisation

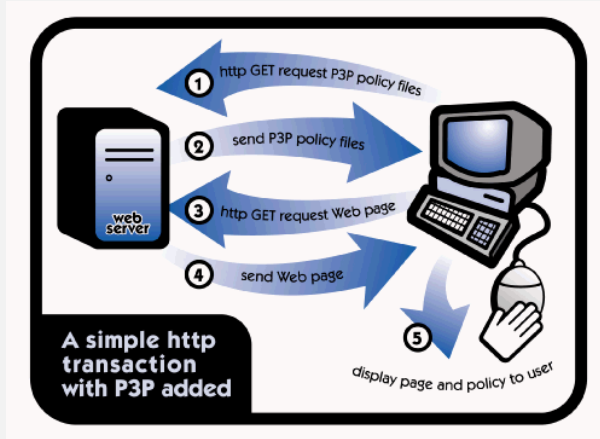


- **DPV** is a community-based specification, for the expression of metadata related to the processing of personal and non-personal data, based on legal requirements





- 2002 W3C spec
- Websites declared data practises in XML
- Browsers could block parts of the website accordingly
- Was never widely adopted due to lack of incentives or regulation



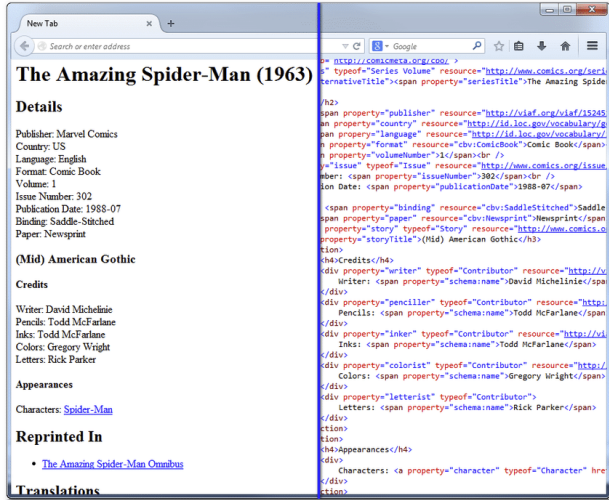


- On / off signal that browsers send to websites
- Has gained traction thanks to backing by California Consumer Privacy Act (CCPA)
- Similar attempt Do Not Track (DNT) which failed without enforcement



**Vision**

- Embed machine-readable terms-of-use requests in cookie consent dialogues
- Standardise HTML attributes for browser extensions to automate cookie preference management



- Use a “Data-Policy” header to signal user consent for data processing

```
Request
Pretty Raw Hex Hackvector
1 GET / HTTP/2
2 Host: www.professionallyevil.com
3 Dnt: 1
4 Upgrade-Insecure-Requests: 1
5 User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64)
  AppleWebKit/537.36 (KHTML, like Gecko) Chrome/103.0.0.0 Safari/537.36
6 Accept:
  text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/
  webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.9
7 Sec-Fetch-Site: none
8 Sec-Fetch-Mode: navigate
9 Sec-Fetch-User: ?1
10 Sec-Fetch-Dest: document
11 Sec-Ch-Ua: ".Not/A)Brand";v="99", "Google Chrome";v="103",
  "Chromium";v="103"
12 Sec-Ch-Ua-Mobile: ?0
13 Sec-Ch-Ua-Platform: "Windows"
14 Accept-Encoding: gzip, deflate
15 Accept-Language: en-US,en;q=0.9
16
17
```



- Browsers and websites participate in a dialogue to establish the terms-of-use agreement best suited to the user



- More **expressive** than Global Privacy Control
- Allows the *client* to specify the terms-of-use they agree to rather than blocking parts of the websites with P3P policies they dislike



**Call for action**



Joint work between **academia, industry** and **regulators** is required to:



- Align this proposal with EU and UK *regulation*
- Apply regulatory *incentives* or *pressures* for adoption
- Align with existing Enterprise Data Governance solutions
- Reduce engineering, legal and compliance costs for industry

**Call for action**

## Calls to *legal* and *policy* experts:

- Ensure compatibility between regulatory frameworks and our software architectures
- Call the European Data Protection Board (EDPB) to help make these terms-of-use **legally binding Data Sharing Agreements (DSAs)** or **lawful consent for data processing by data subjects** under **GDPR**.
- Same call Information Commissioners Office (ICO) in w.r.t. **UK Data Protection Act 2018**.



Call on *industry* to:

- Co-design solutions that will reduce friction with internal data governance architectures
- Joint Research Centres (JRCs) with experience implementing mechanised data governance to validate the proposal



## Benefits to users:

- Better UX - no invasive popups
- More autonomy over personal data

## Benefits to implementors:

- Possible “safe harbour” provisions
- Automated compliance verification
- Ease of implementation
- More users (?)





## Benefits to regulators:

- Automated compliance enforcement
- Reduction in possible dark patterns undermining regulatory intentions

# Acknowledgements



DEPARTMENT OF  
**COMPUTER  
SCIENCE**



**GHENT  
UNIVERSITY**



OXFORD  
**MARTIN**  
SCHOOL



Jesse Wright is funded by the Department of Computer Science, University of Oxford



Beatriz Esteves is funded by SolidLab Vlaanderen (Flemish Government, EWI and RRF project VV023/10)



Rui Zhao is funded by the Ethical Web and Data Architecture in the Age of AI (EWADA) project, whose funds come from Oxford Martin School, University of Oxford.



Full Paper

Short Paper



Thank you

