

# A law literature review of the Internet of Things

---

Identifying common themes, debates, and  
gaps via Natural Language Processing (NLP)

Christof Koolen

EPIP Conference – Madrid  
September 8-10, 2021



# Agenda

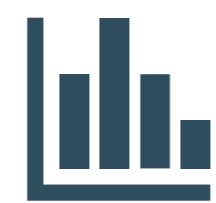
---



PROJECT  
BACKGROUND



IOT AND NLP



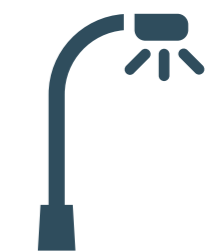
TEXT  
STATISTICS



CLUSTERING  
RESULTS



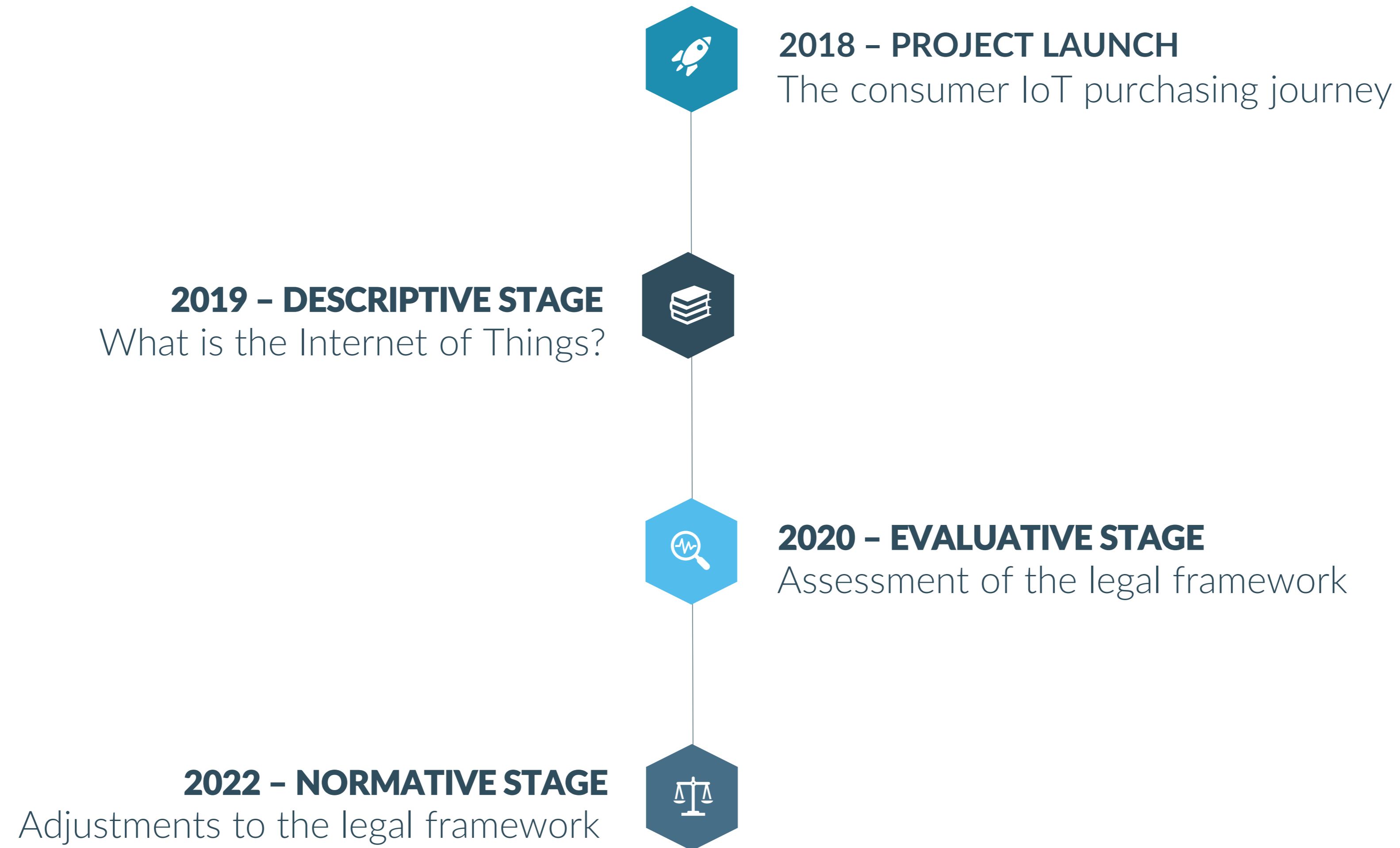
PUBLICATION  
TRENDS



AUTHOR  
SPOTLIGHT

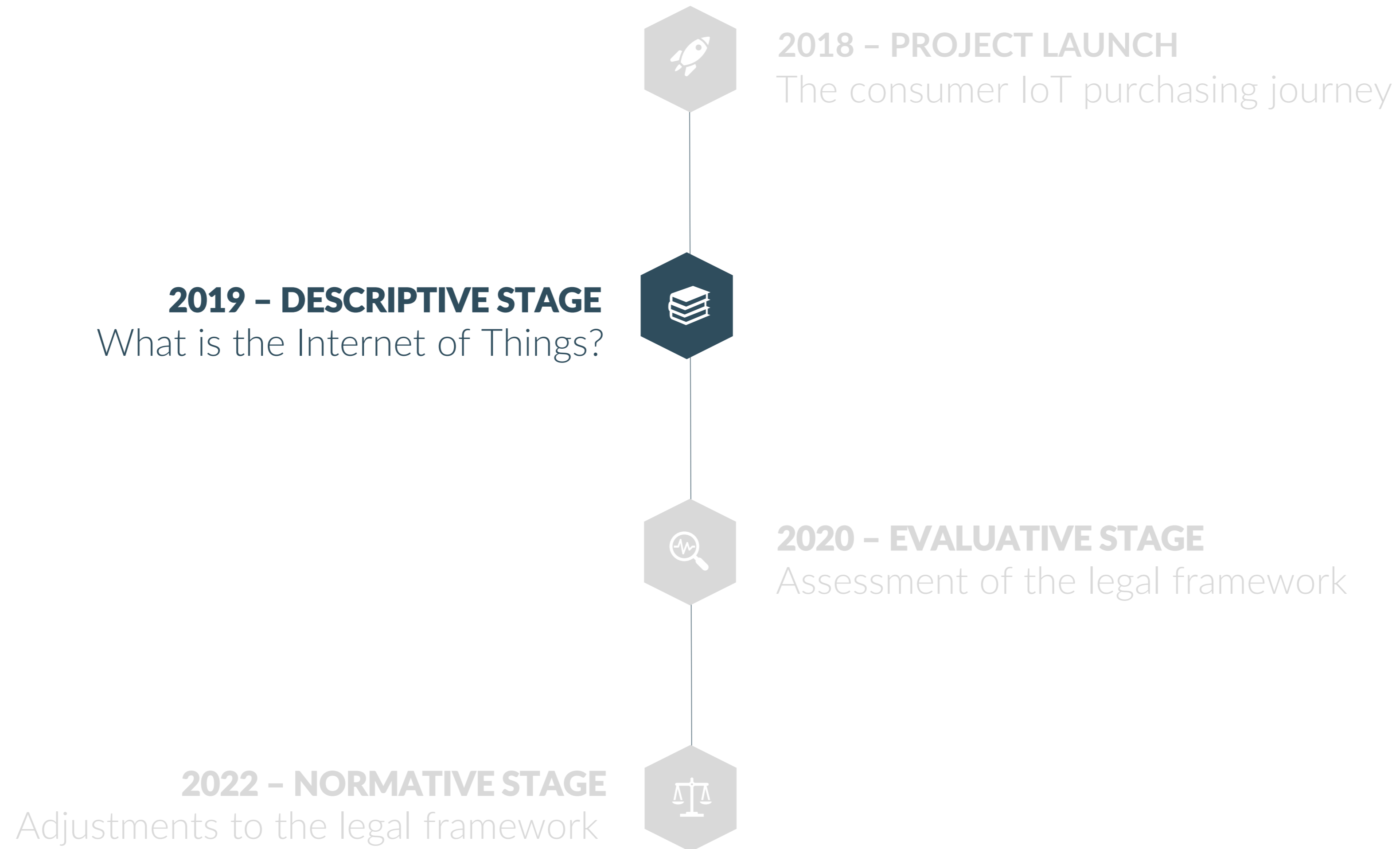
# Project background

Securing consumers' rights and interests in the era of smart technology and the IoT



# Project background

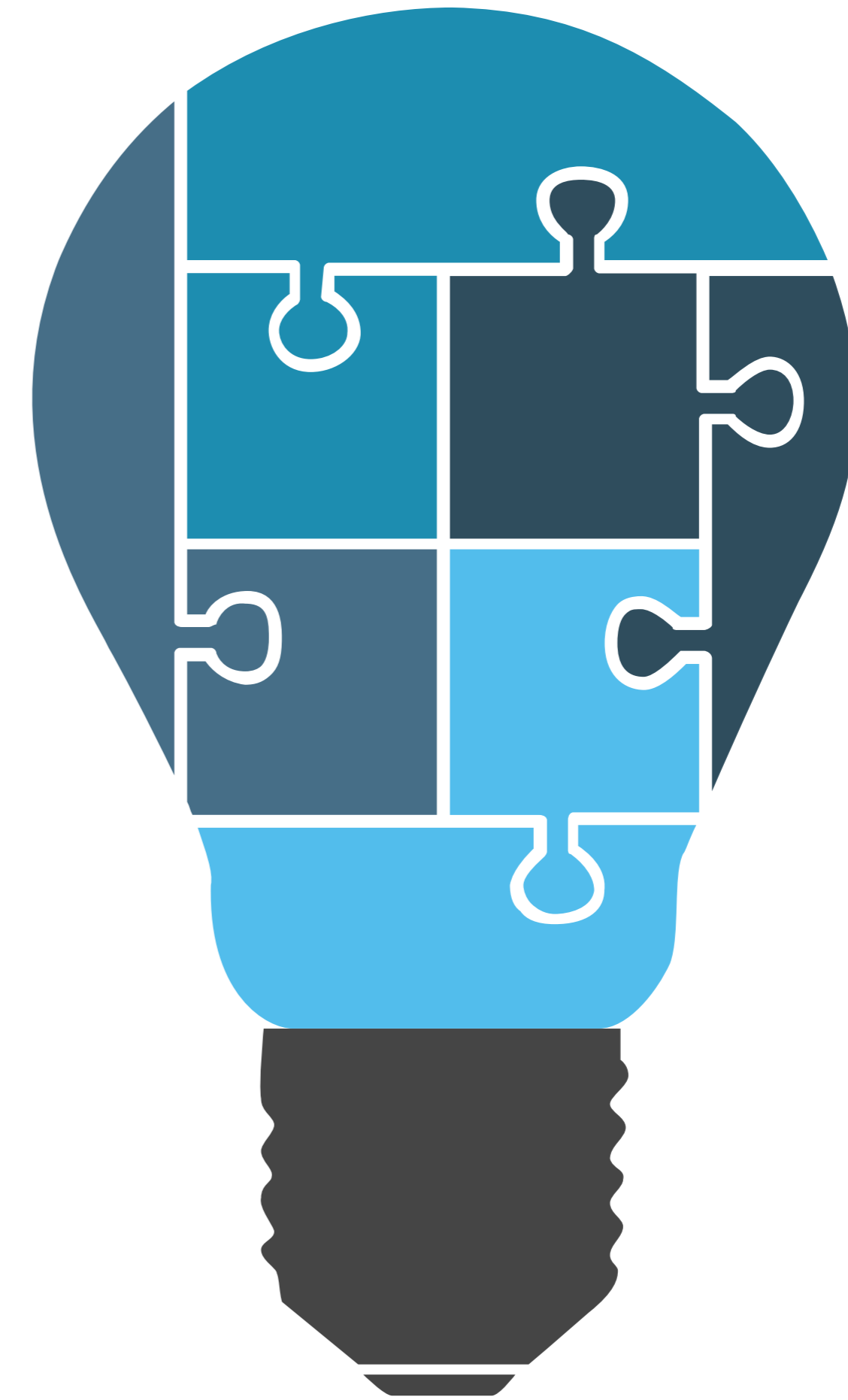
Situating the NLP project within the PhD



# IoT and NLP

Why document clustering?

- Nascent stages of IoT development
- Growing topic of conversation
- Vast and unwieldy body of literature
- No comprehensive law literature review



# IoT and NLP

Expected output of the NLP exercise



CLUSTER  
TOPICS



WEIGHT PER  
TOPIC



PUBLICATION  
TRENDS



LEADING LEGAL  
SCHOLARS



MOST  
AUTHORITATIVE  
JOURNAL ARTICLES

# Text statistics

Descriptive information of the dataset

304



**Journal articles** that matched with the search queries for the phrases Internet of Things, IoT, or smart device (excl. mentions in passing).

404



**Authors** who were identified within the dataset.

95



**Law reviews** that contained articles covering the IoT. The initial list contained 130 top publication venues for law and technology journals as identified by Google Scholar and W&L Law Journal Rankings.

8.334



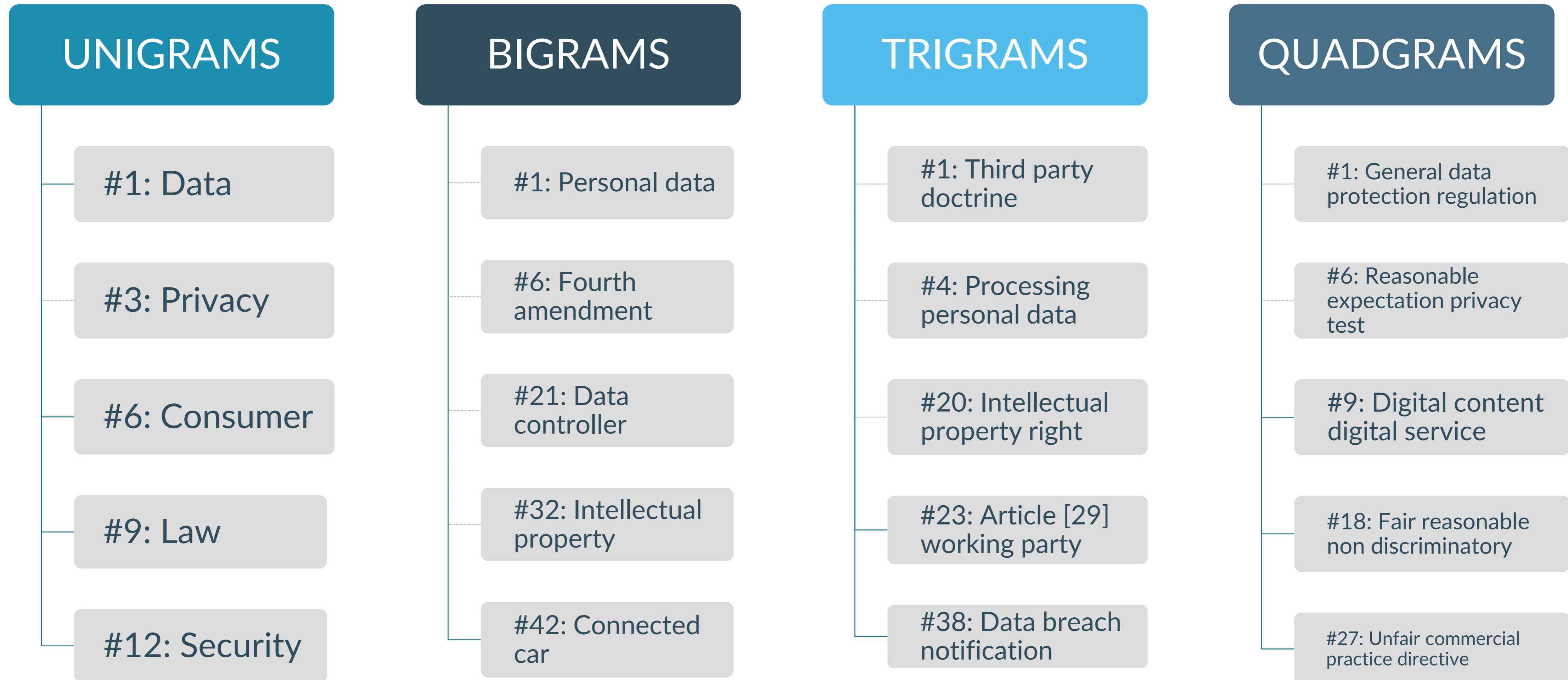
**The average word count** per article, without including references.

3.289.125



**Unique features (i.e. word combinations)** identified within a corpus that contains 2.533.675 words.

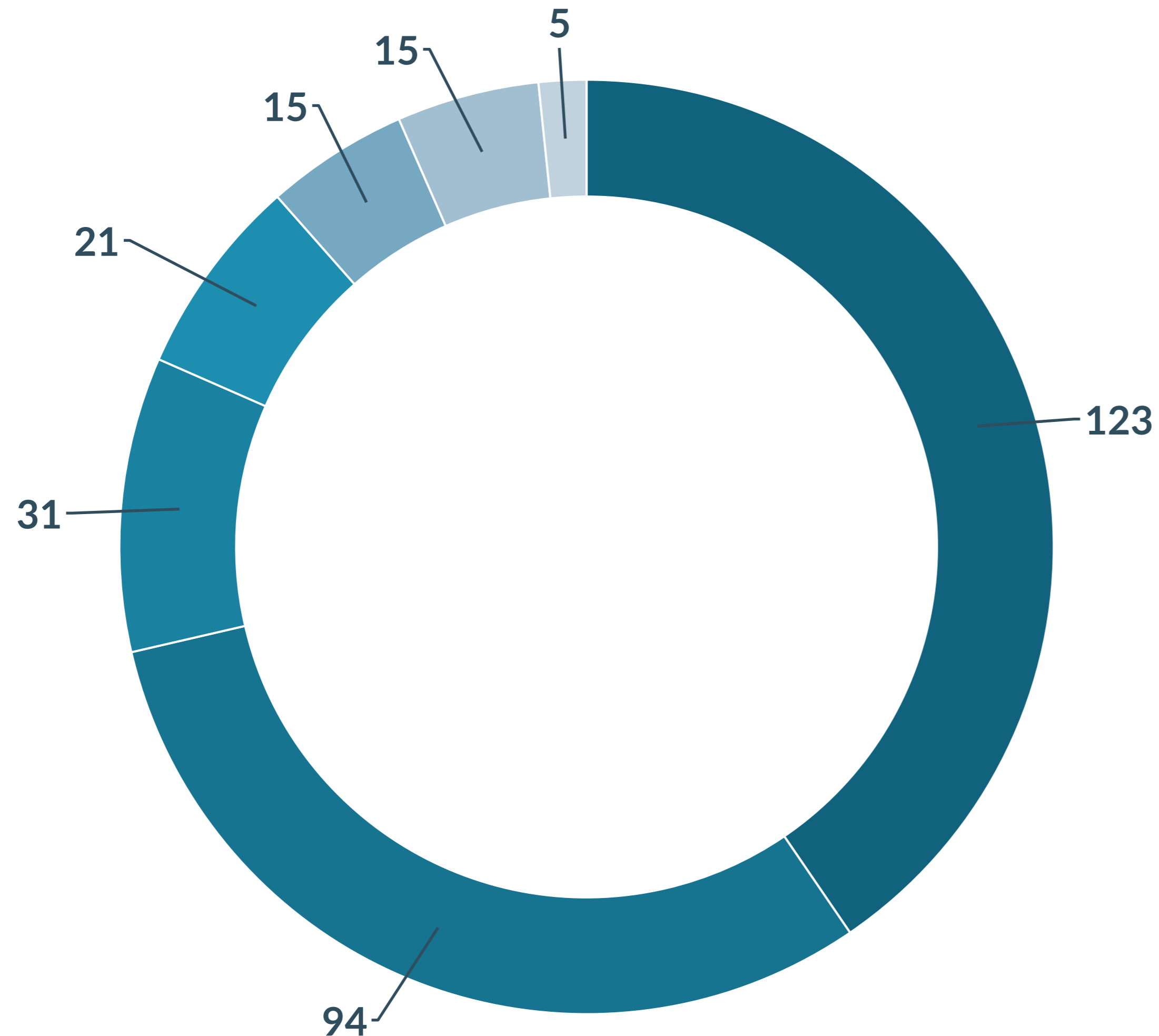
# Text statistics





# Clustering results – General

How are the journal articles distributed across  $k = 7$  topics?

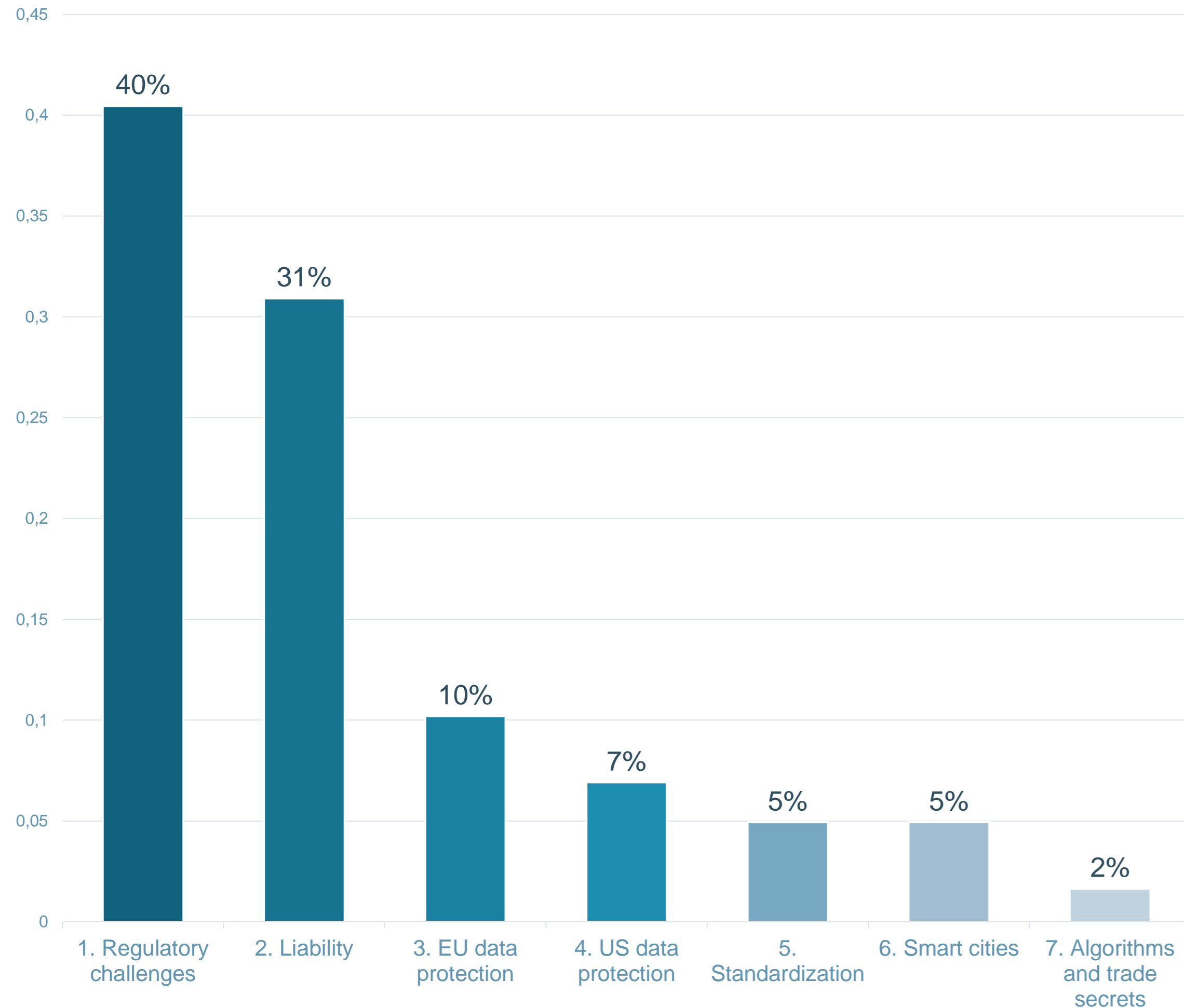


- 1. Regulatory challenges (T6)
- 2. Liability (T1)
- 3. EU data protection (T0)
- 4. US data protection (T4)
- 5. Standardization (T2)
- 6. Smart cities (T3)
- 7. Algorithms and trade secrets (T5)

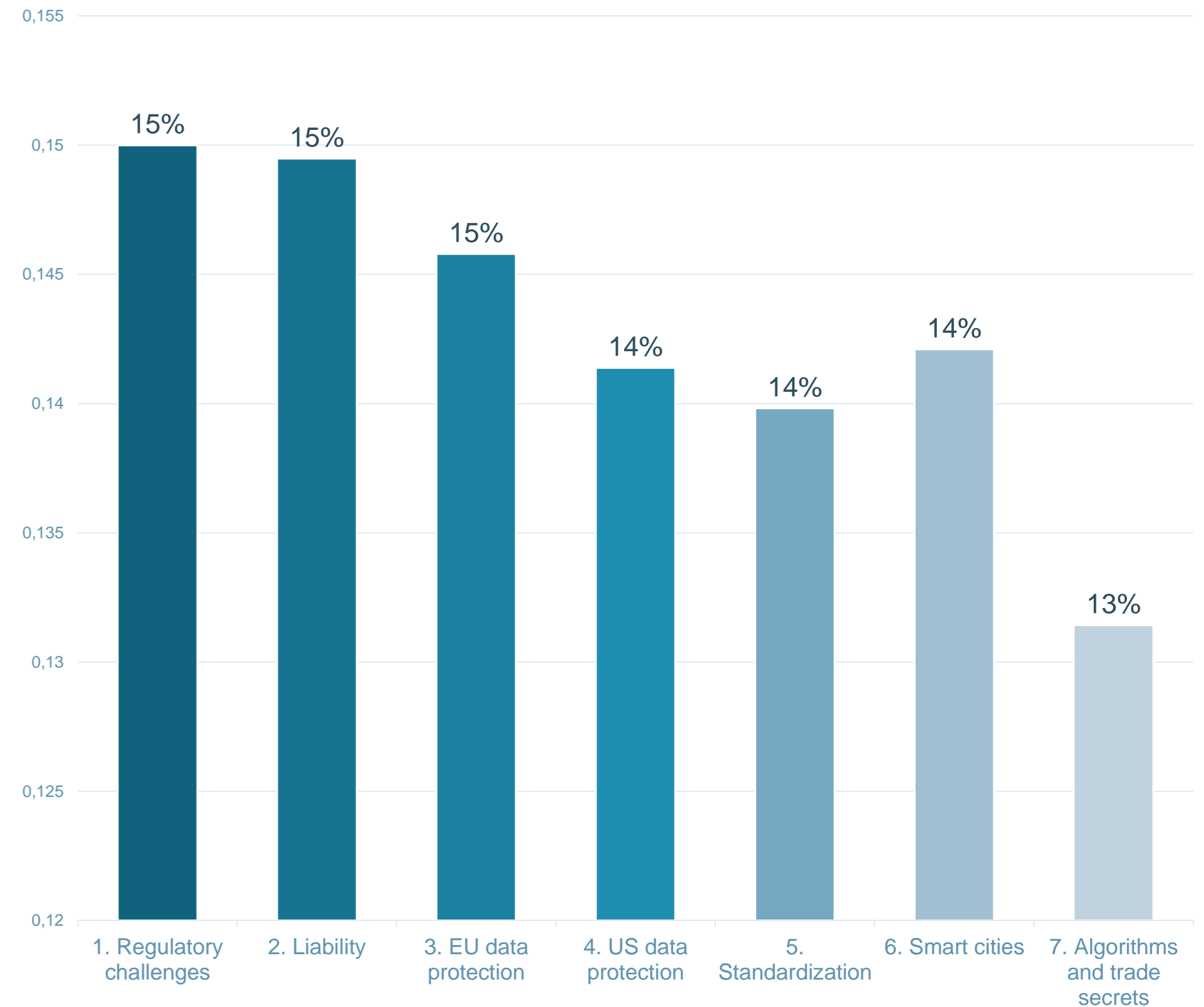
# Clustering results – General

Topic weight distribution – Absolute versus relative

## Absolute topic weight



## Relative topic weight



# Clustering results – Standardization

Examining what defines the cluster via TF-IDF

## T2 – STANDARDIZATION – TOP TERMS

Patent	Royalty	Competition	SSOs
SEP	Standard	Market	Huawei
FRAND	Pool	Holder	Standardization
Infringement	Court	Price	Data pool
Licensing	SEP holder	IoT	Antitrust

# Clustering results – Standardization

Calculating which 5 articles are located nearest to the cluster centroid

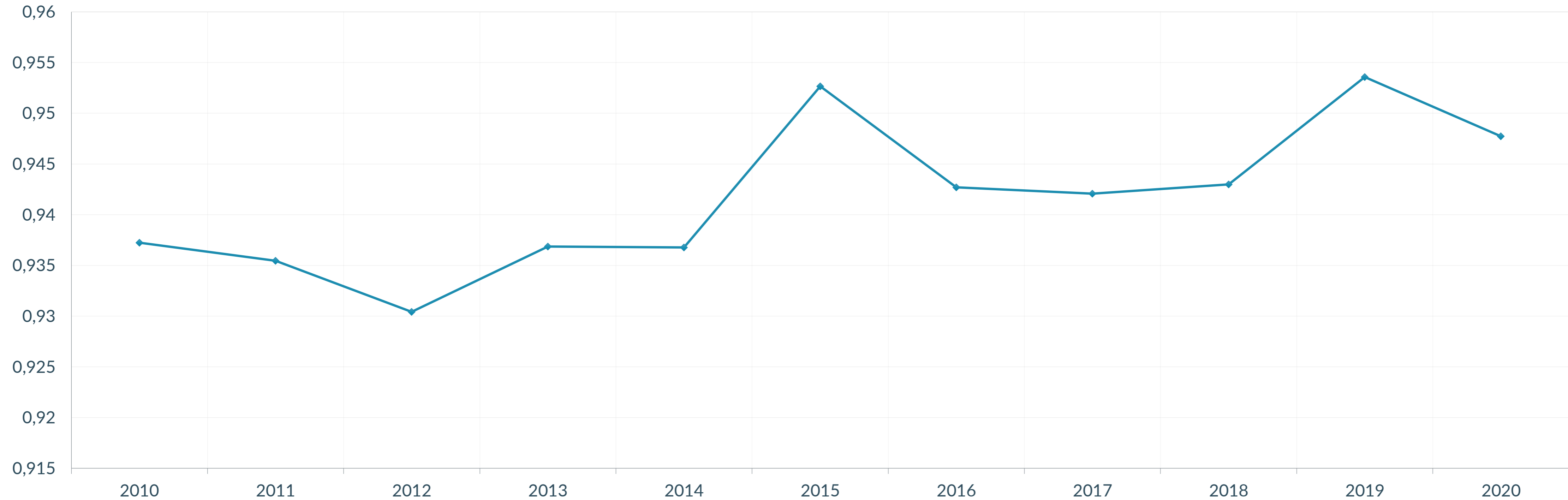
## T2 – STANDARDIZATION – TOP ARTICLES

1. Ashish Bharadwaj and Srajan Jain, “A comparative study of the evolving jurisprudence on standard essential patent licensing” [2020] 23 The Journal of World Intellectual Property 310-327.
2. Evelina Kurgonaitė, Pat Treacy and Edwin Bond, “Looking Back to the Future—Selective SEP Licensing Through a Competition Law Lens?” [2020] 11 Journal of European Competition Law & Practice 133-146.
3. Jonathan M. Barnett, “Antitrust overreach: undoing cooperative standardization in the digital economy” [2019] 25 Michigan Technology Law Review 163-238.
4. Rupprecht Podszun, “Standard Essential Patents and Antitrust Law in the Age of Standardisation and the Internet of Things: Shifting Paradigms” [2019] 50 IIC - International Review of Intellectual Property and Competition Law 720-745.
5. Beatriz Conde Gallego, Josef Drexler, “IoT Connectivity Standards: How Adaptive is the Current SEP Regulatory Framework?” [2019] 50 IIC - International Review of Intellectual Property and Competition Law 135-156.

# Publication trends – Standardization

Assessing the mean item-centroid distance on a yearly basis

## T2 – Standardization – Publication trends



2010s: "Smartphone wars"

2015: Huawei v. ZTE

2018: Unwired Planet

2020: Sisvel / Haier; Qualcomm

# Author spotlight – General

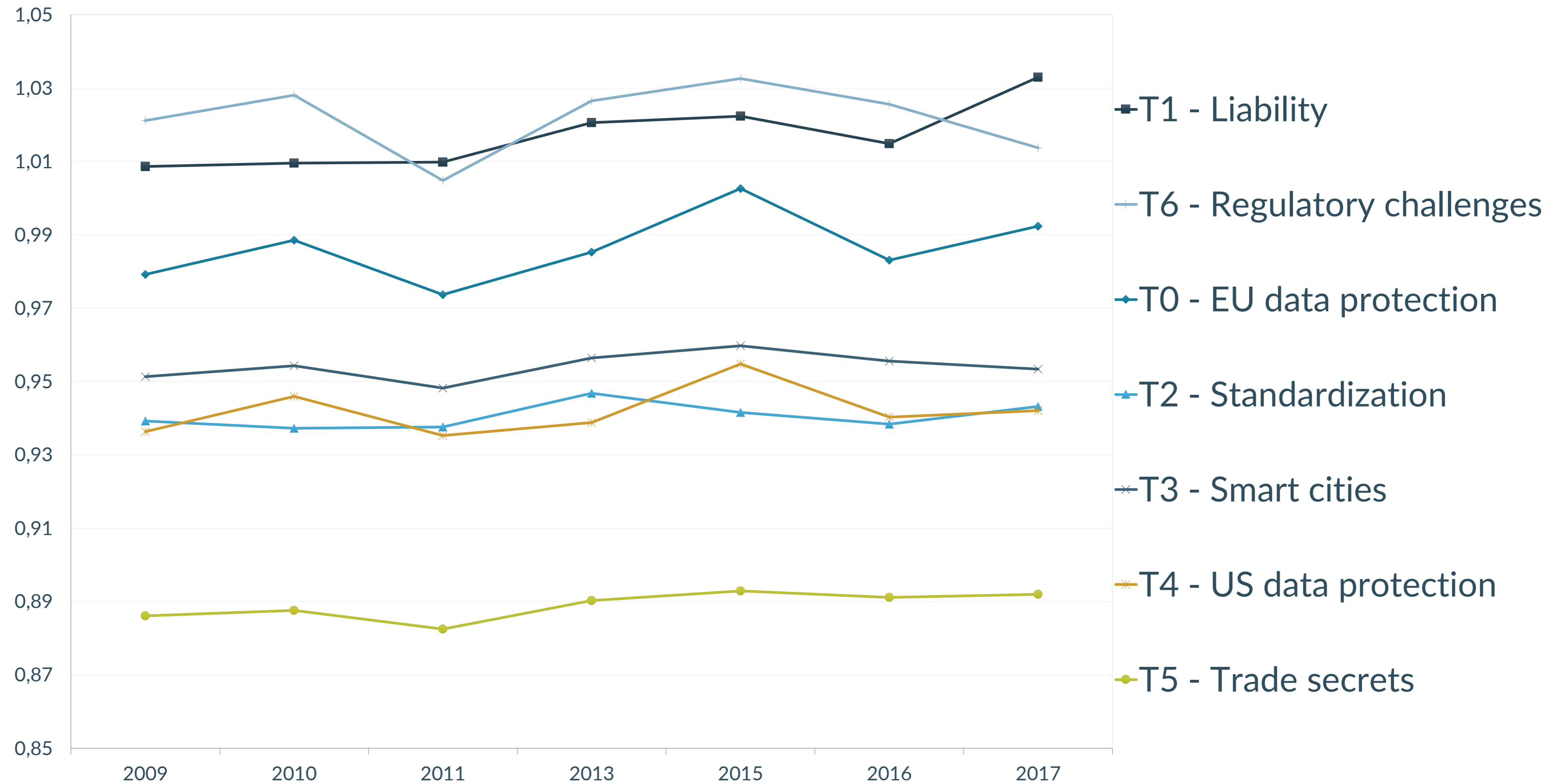
---

## MOST FREQUENTLY PUBLISHED AUTHORS

1. Rolf H. Weber	7 articles
2. Lachlan Urquhart	5 articles
3. Scott J. Shackelford	5 articles
4. Guido Noto La Diega	4 articles
5. Wolfgang Kerber	4 articles

# Author spotlight – Rolf H. Weber

Publishing trends from one author with reference to his own works



# Conclusion

What are the main takeaways?

---

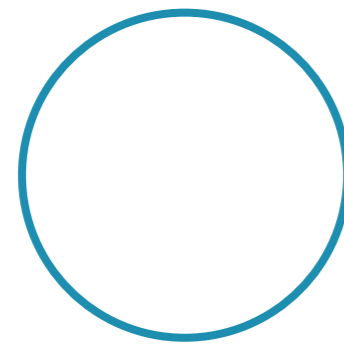
- **Natural Language Processing** used to good effect. Note, however, that it reduces the complexity of the puzzle, but that it remains a puzzle nonetheless.
- **Prior knowledge** of the topic is still required (e.g. naming topics, identifying the causes for trend changes).
- **Interpretation** of the results is not black or white. Often many possible solutions (e.g. number of topics).
- **Underfitting vs overfitting** considerations suggest that you want enough detail, but not too much.



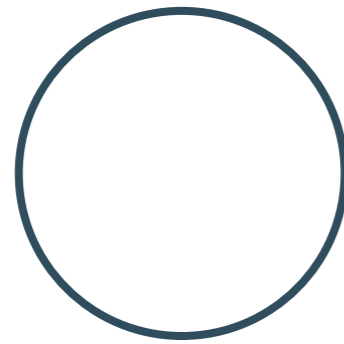
# Discussion

Food for thought

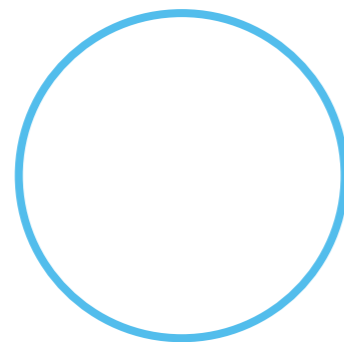
---



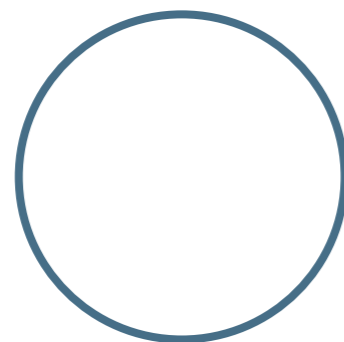
Thoughts on using **Natural Language Processing** for literature reviews?



Comments on the **methodology** and findings?



Is there room to **expand** the project?



Would **your own research** benefit from NLP?

# Questions or suggestions? Get in touch!

---

[christof.koolen@kuleuven.be](mailto:christof.koolen@kuleuven.be)  
[linkedin.com/in/christofkoolen](https://www.linkedin.com/in/christofkoolen)