KU LEUVEN CiTiP

CENTRE FOR IT & IP LAW

imec

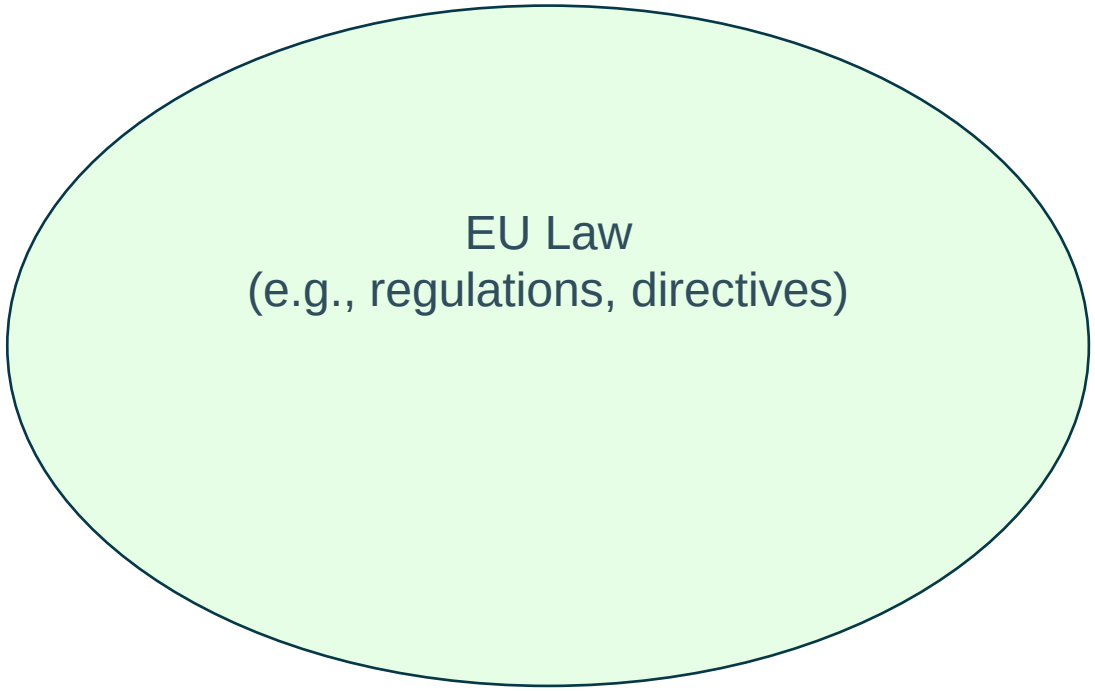
Principles of the EU digital single market legislation applicable to data spaces

European Language Data Space

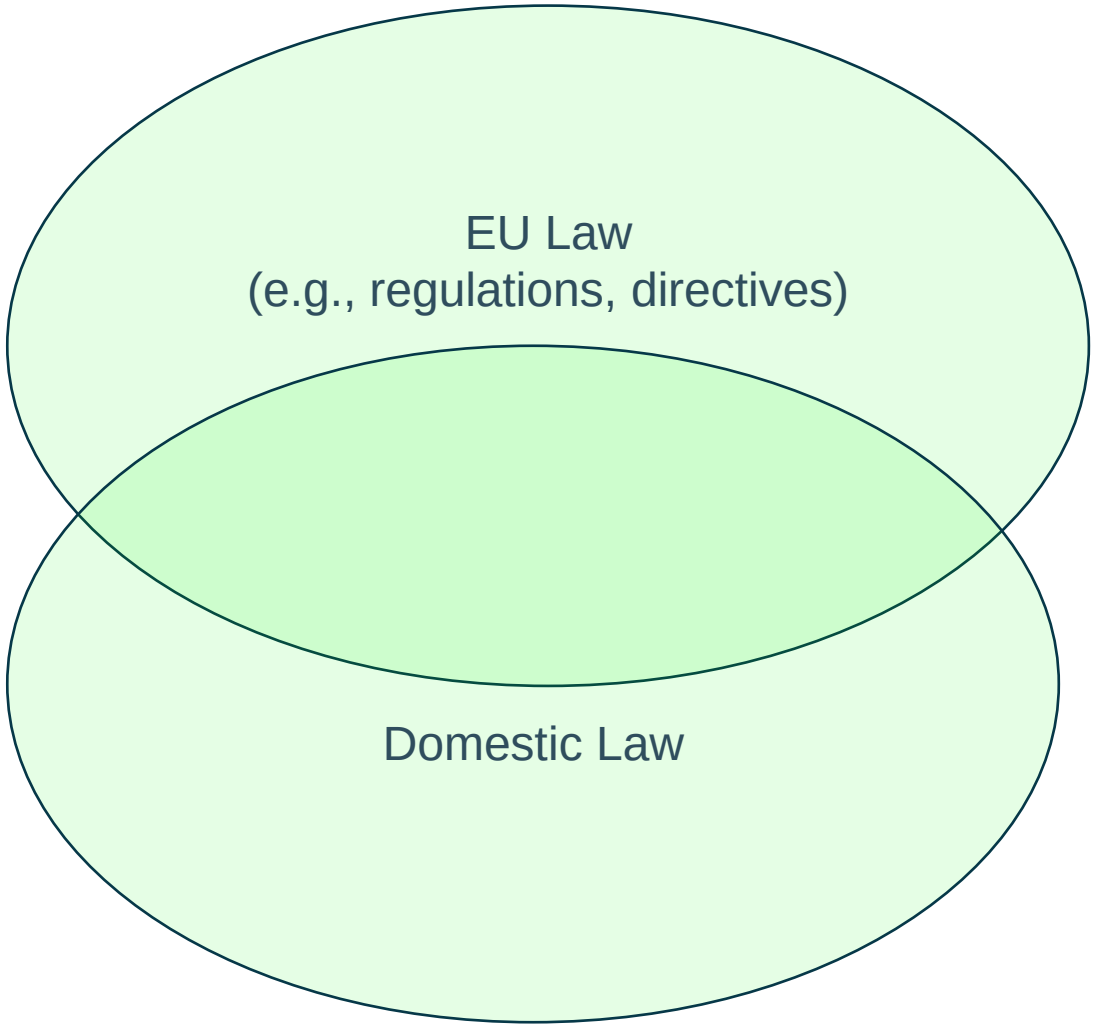
LDS Technology Workshop
Legislation and regulations for data spaces: an environment for the development of a European Data Market

Prof. Dr. Thomas Margoni
Research Professor of Intellectual Property Law
Centre for IP & IT Law (CiTiP)
Faculty of Law – University of Leuven (KUL)

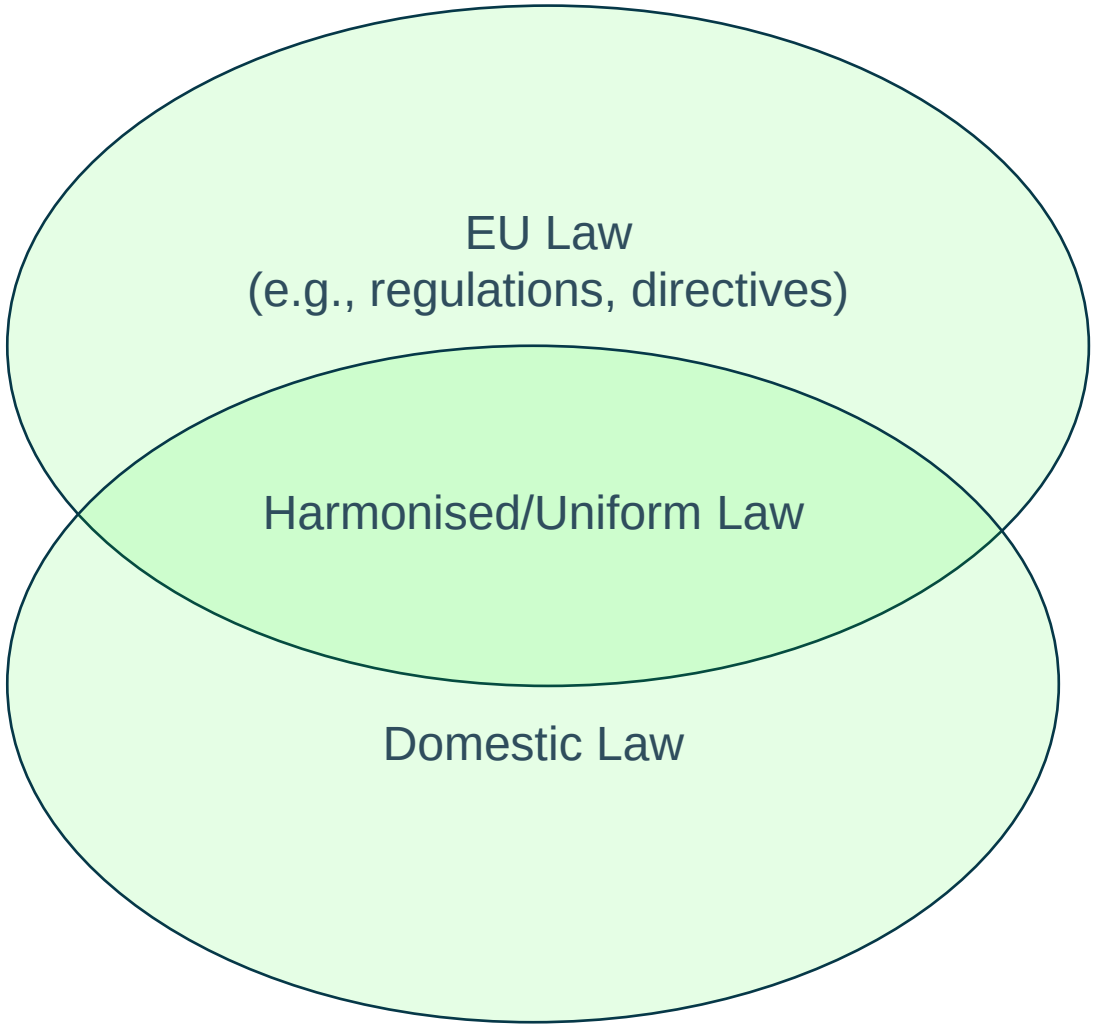
General legal framework

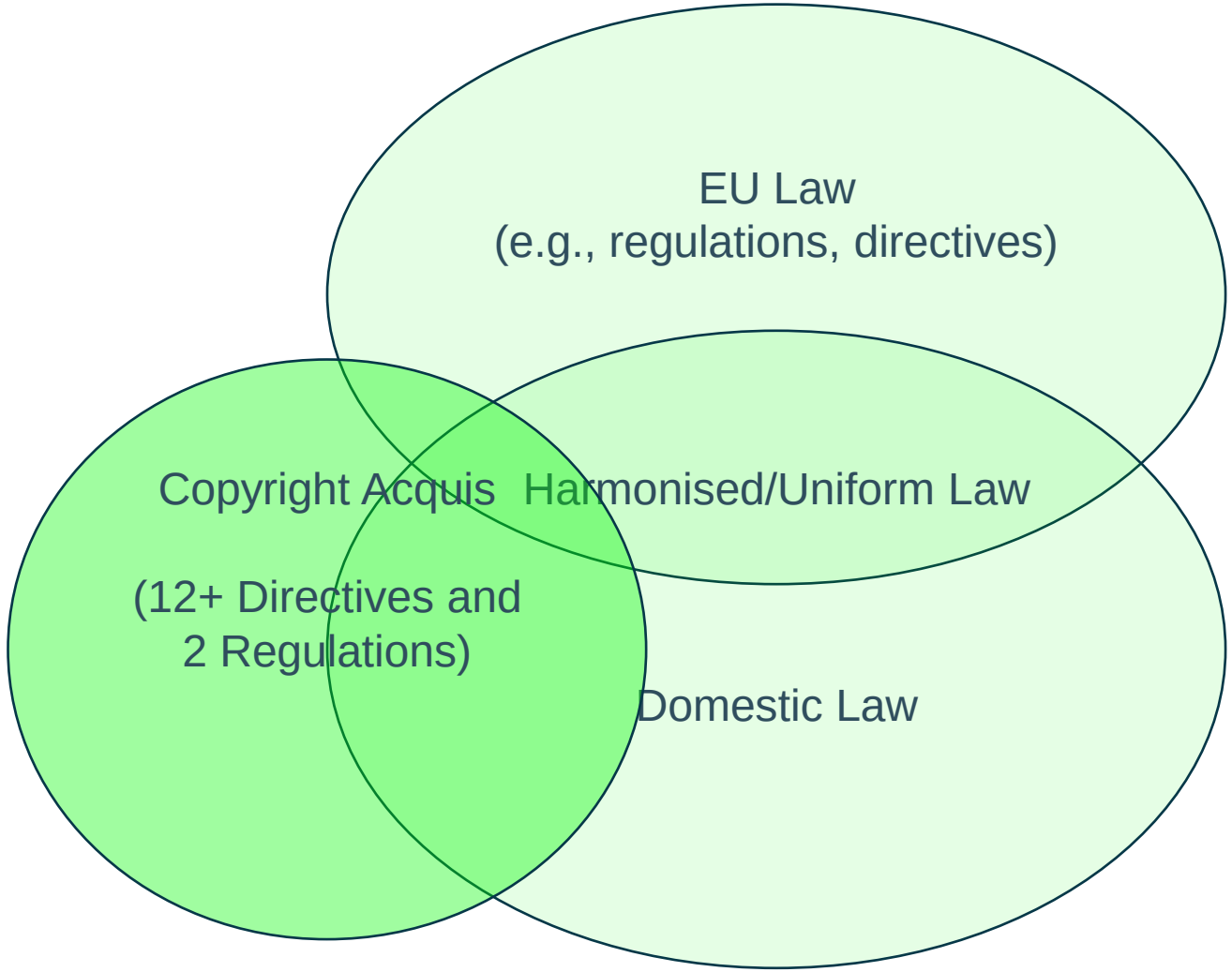


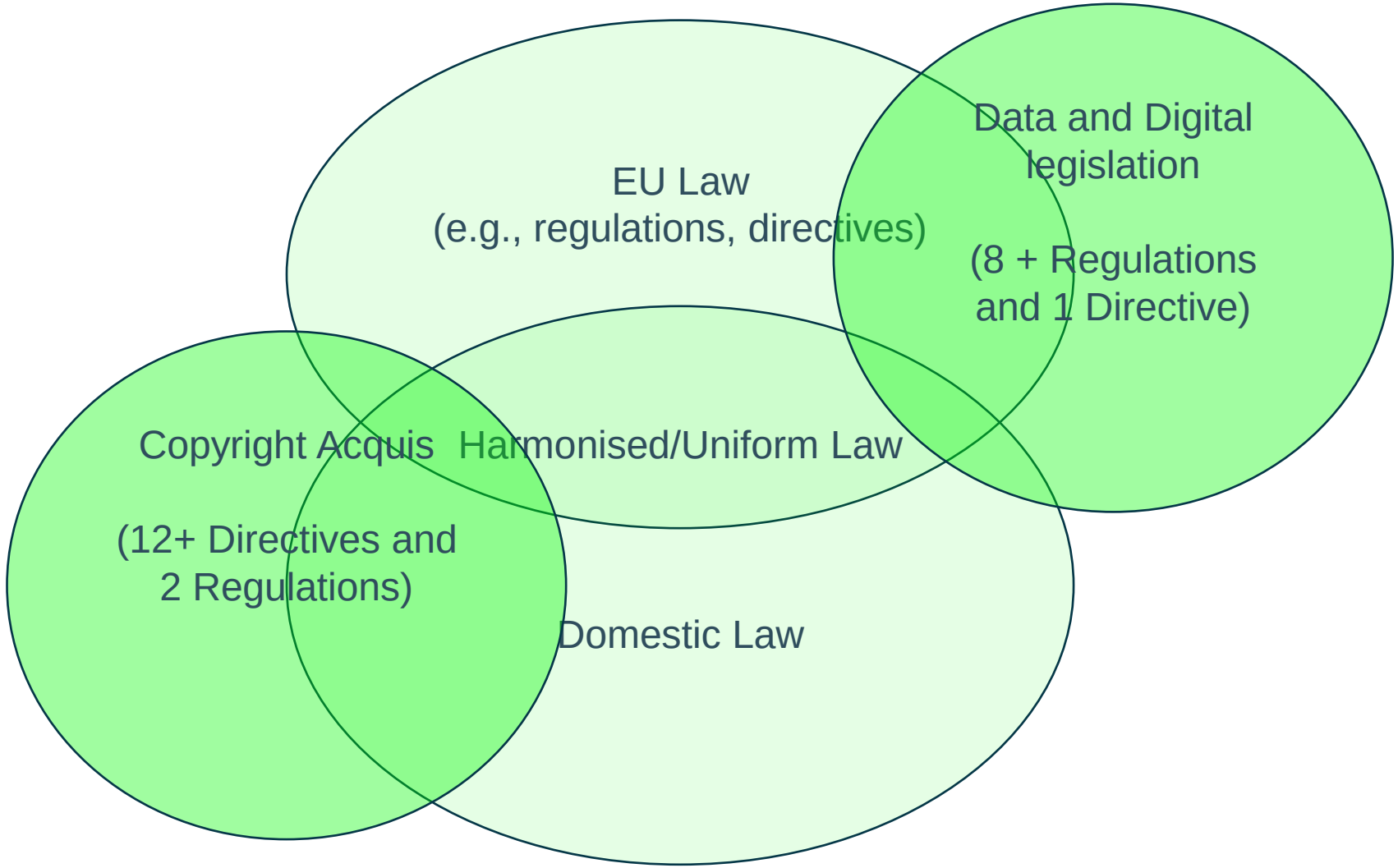
General legal framework

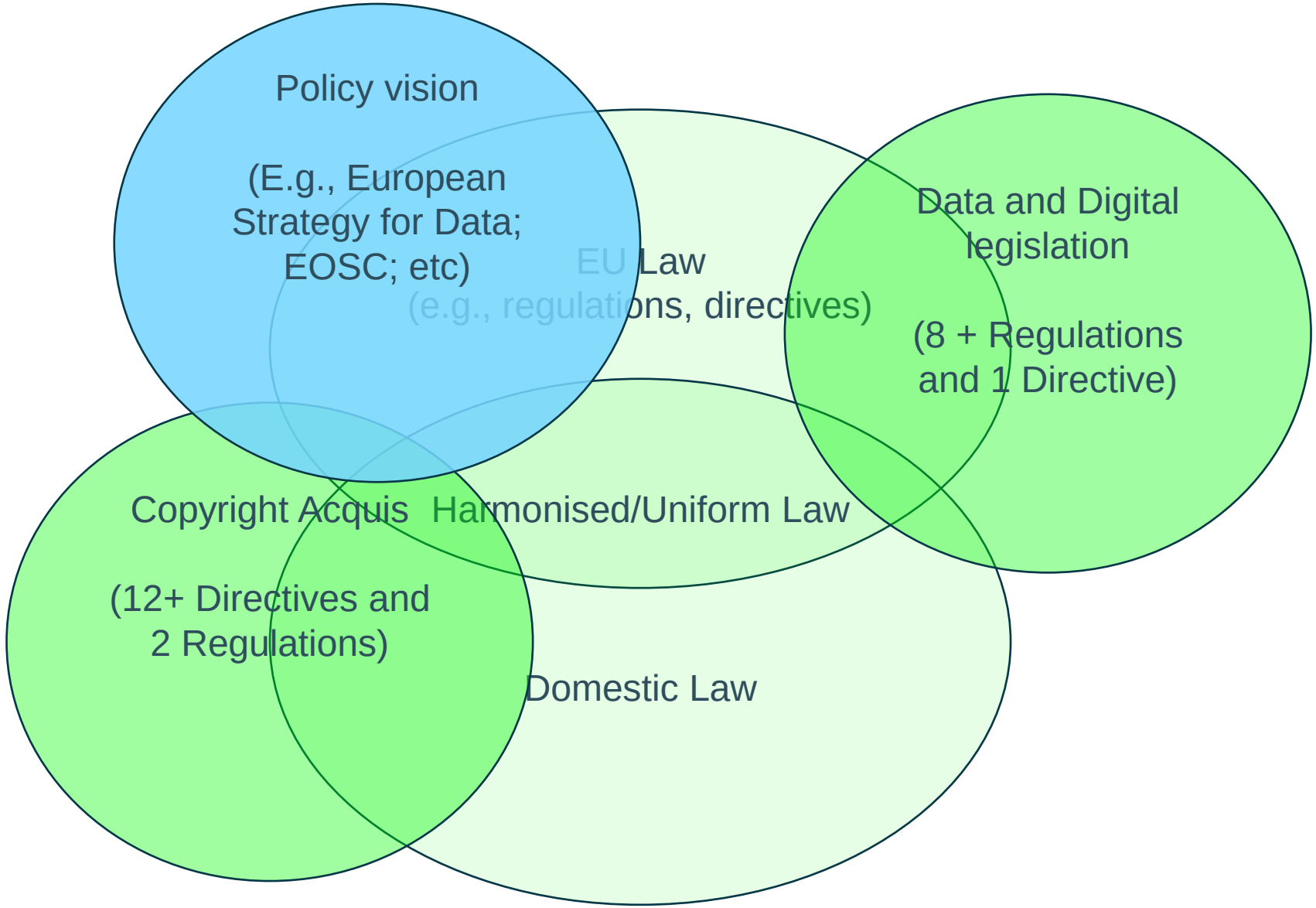


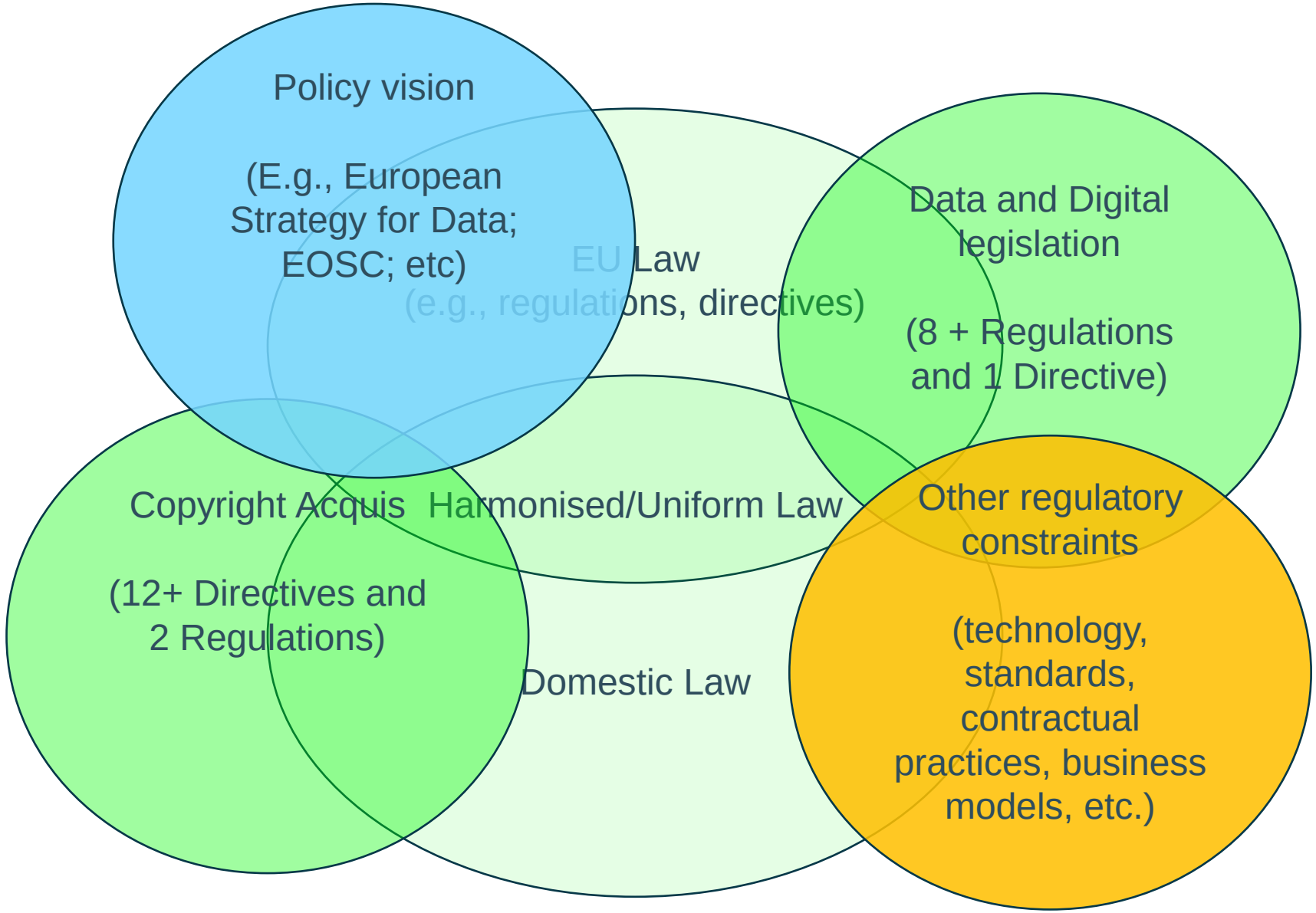
General legal framework

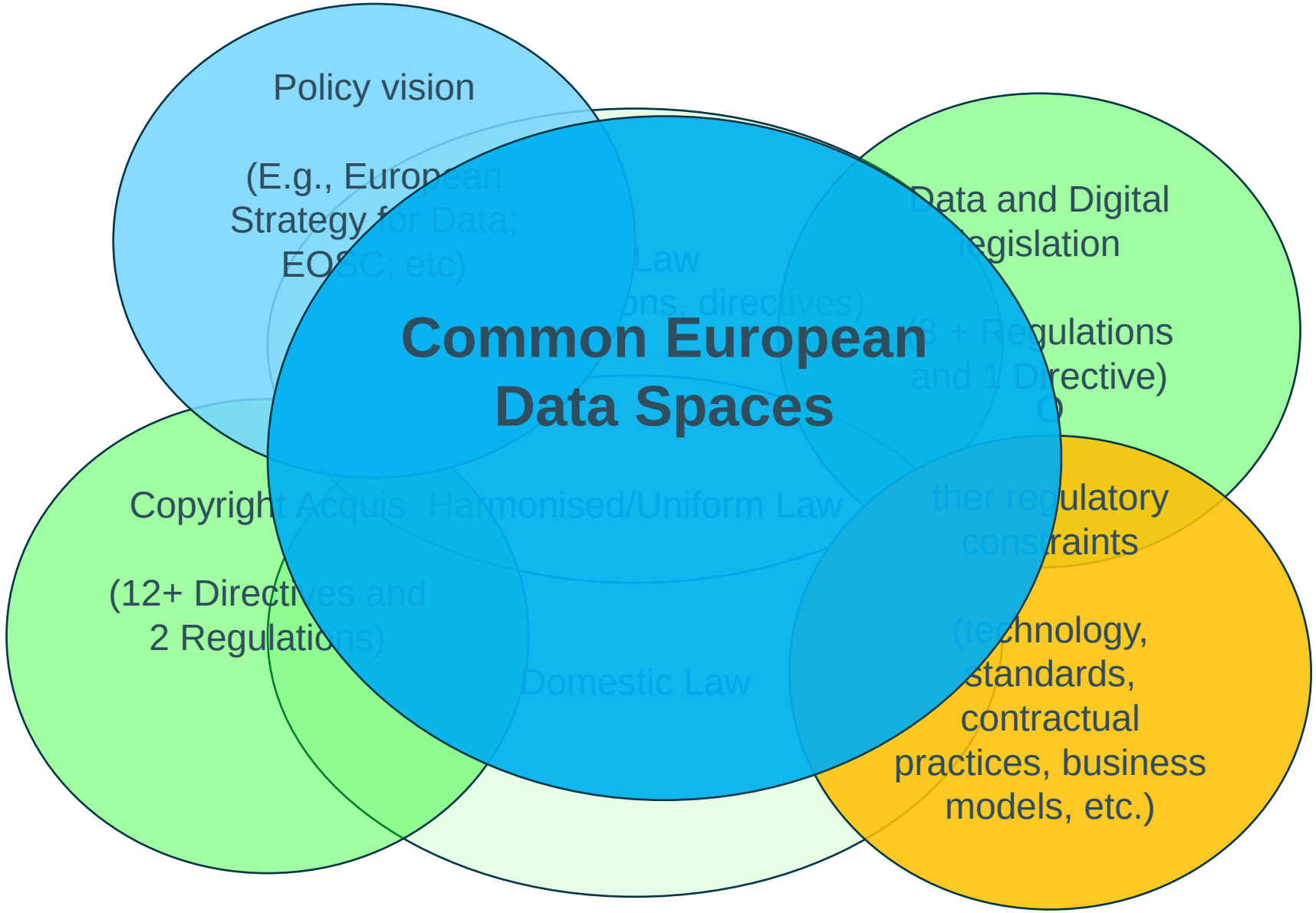



EU Law
(e.g., regulations, directives)
Copyright Acquis
(12+ Directives and
2 Regulations)
Harmonised/Uniform Law
Domestic Law


EU Law
(e.g., regulations, directives)
Data and Digital legislation
(8 + Regulations and 1 Directive)
Copyright Acquis
(12+ Directives and 2 Regulations)
Harmonised/Uniform Law
Domestic Law

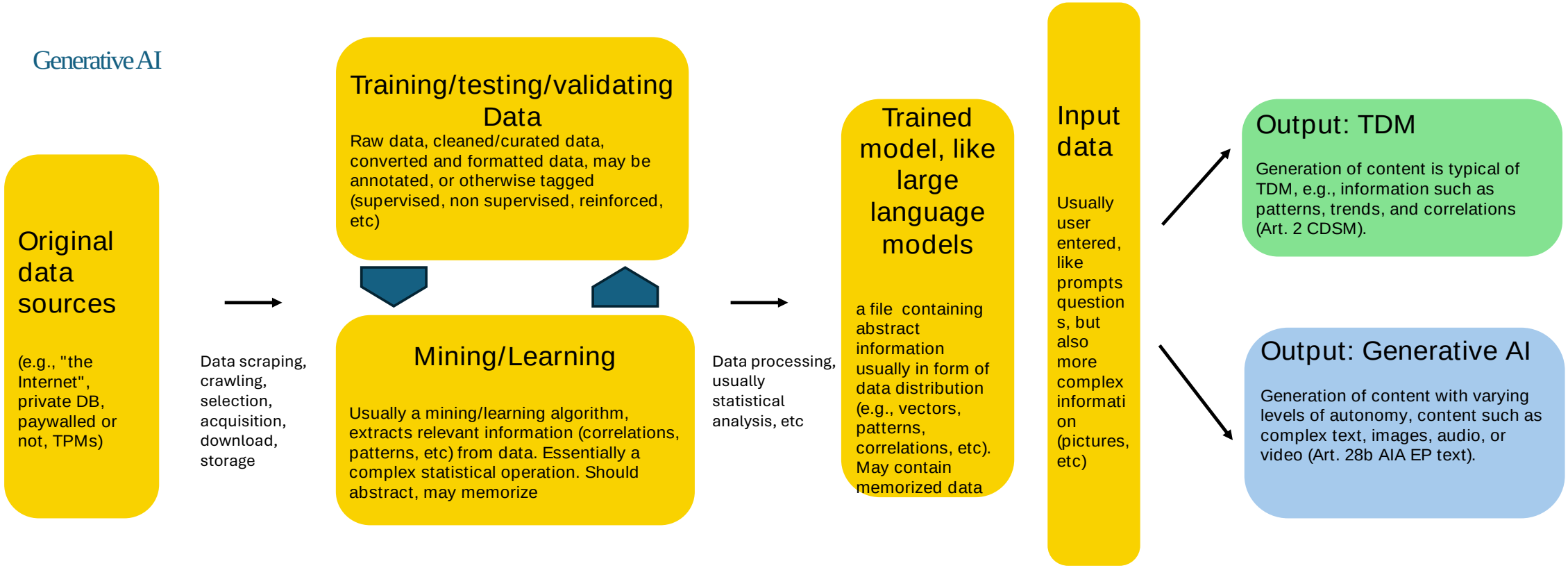

Policy vision
(E.g., European Strategy for Data; EOSC; etc)
EU Law
(e.g., regulations, directives)
Data and Digital legislation
(8 + Regulations and 1 Directive)
Copyright Acquis
(12+ Directives and 2 Regulations)
Harmonised/Uniform Law
Domestic Law


Policy vision
(E.g., European Strategy for Data; EOSC; etc)
EU Law
(e.g., regulations, directives)
Data and Digital legislation
(8 + Regulations and 1 Directive)
Copyright Acquis
(12+ Directives and 2 Regulations)
Harmonised/Uniform Law
Domestic Law
Other regulatory constraints
(technology, standards, contractual practices, business models, etc.)

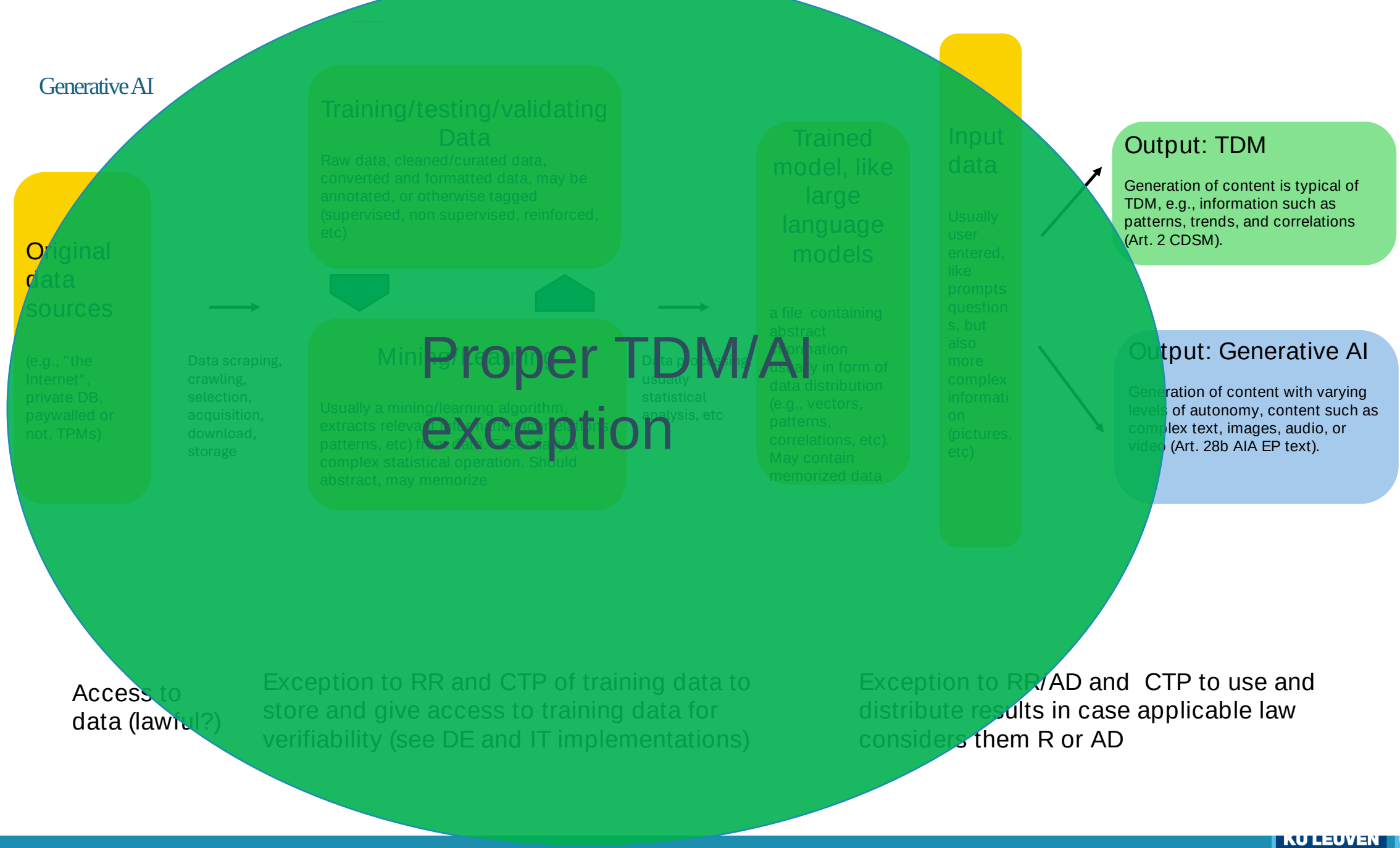

Policy vision
(E.g., European Strategy for Data; EOSC; etc)
Data and Digital legislation
(8 + Regulations and 1 Directive)
Copyright Acquis
(12+ Directives and 2 Regulations)
Harmonised/Uniform Law
Domestic Law
ther regulatory constraints
(technology, standards, contractual practices, business models, etc.)
Common European Data Spaces

A practical example
Mine language sources for model training

- Or any other content that could potentially be protected by copyright and or related rights to copyright (e.g., certain databases)



Access to data (lawful?)	Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)	Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD


Generative AI
Original data sources
(e.g., "the Internet", private DB, paywalled or not, TPMs)
Data scraping, crawling, selection, acquisition, download, storage
Training/testing/validating Data
Raw data, cleaned/curated data, converted and formatted data, may be annotated, or otherwise tagged (supervised, non supervised, reinforced, etc)
Mining/learning/extracting
Usually a mining/learning algorithm, extracts relevant information (features, patterns, etc) from data. Usually a complex statistical operation. Should abstract, may memorize
Data processing usually statistical analysis, etc
Trained model, like large language models
a file containing abstract information usually in form of data distribution (e.g., vectors, patterns, correlations, etc). May contain memorized data
Input data
Usually user entered, like prompts questions, but also more complex information (pictures, etc)
Output: TDM
Generation of content is typical of TDM, e.g., information such as patterns, trends, and correlations (Art. 2 CDSM).
Output: Generative AI
Generation of content with varying levels of autonomy, content such as complex text, images, audio, or video (Art. 28b AIA EP text).
Proper TDM/AI exception
Access to data (lawful?)
Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)
Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD

Art 5(1) ISD temporary copies

Trained model, like large language models

a file containing abstract information usually in form of data distribution (e.g., vectors, patterns, correlations, etc). May contain memorized data

Input data

Usually user entered, like prompts questions, but also more complex information (pictures, etc)

Output: TDM

Generation of content is typical of TDM, e.g., information such as patterns, trends, and correlations (Art. 2 CDSM).

Output: Generative AI

Generation of content with varying levels of autonomy, content such as complex text, images, audio, or video (Art. 28b AIA EP text).

Access to data (lawful?)	Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)	Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD

Generative AI

Art 3&4 CDSM (TDM exceptions)

Access to data (lawful?)	Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)	Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD

Access to data (lawful?)

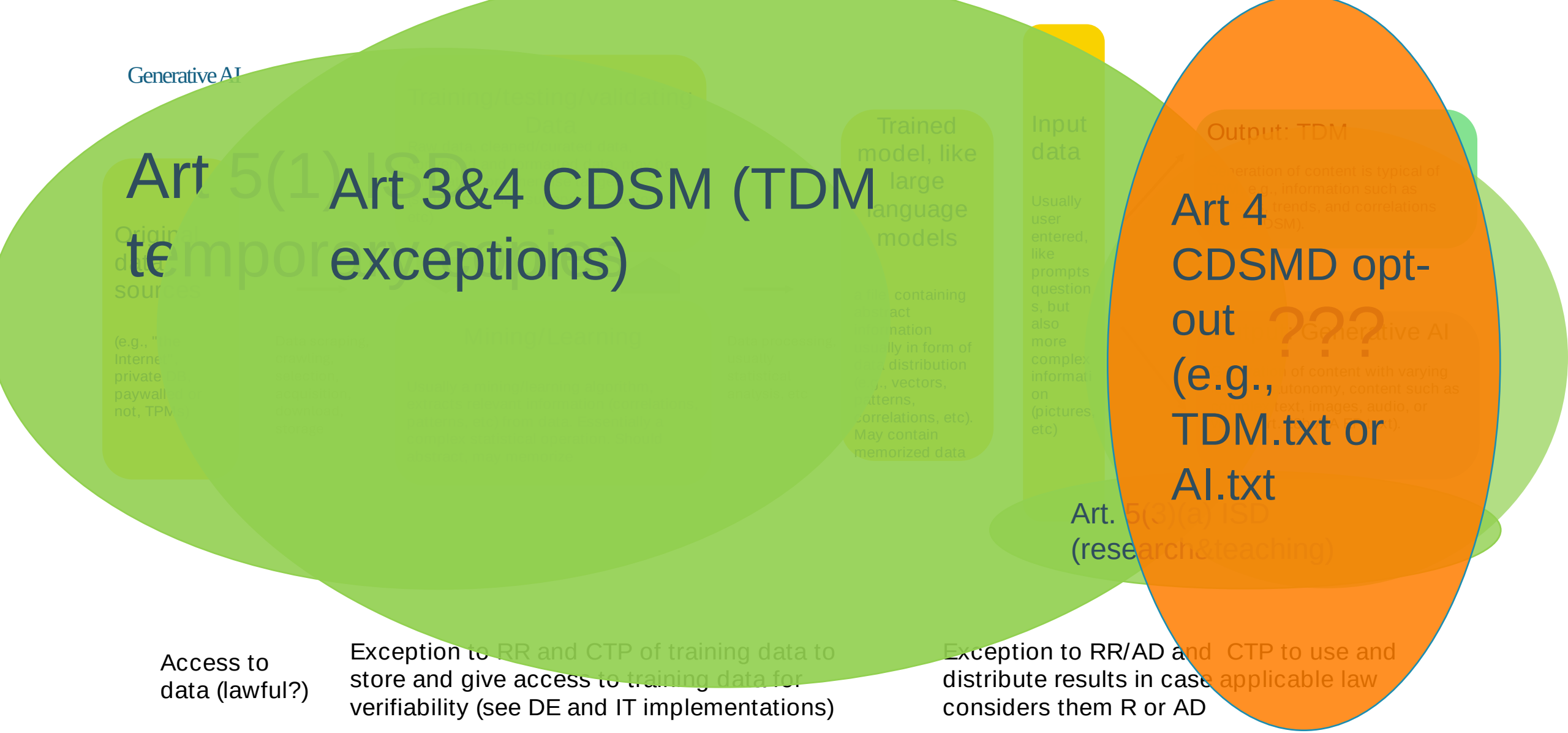
Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)

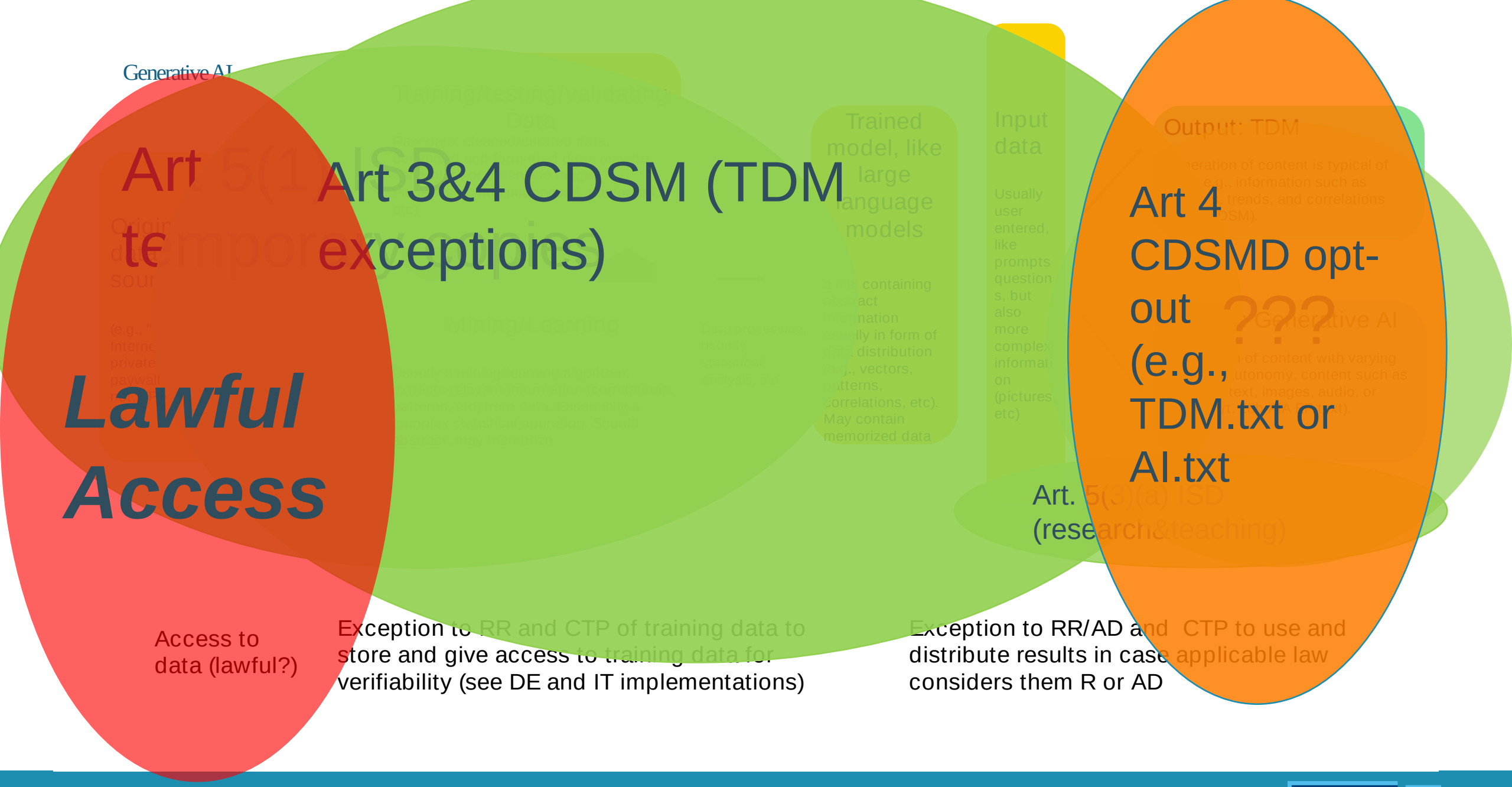
Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD

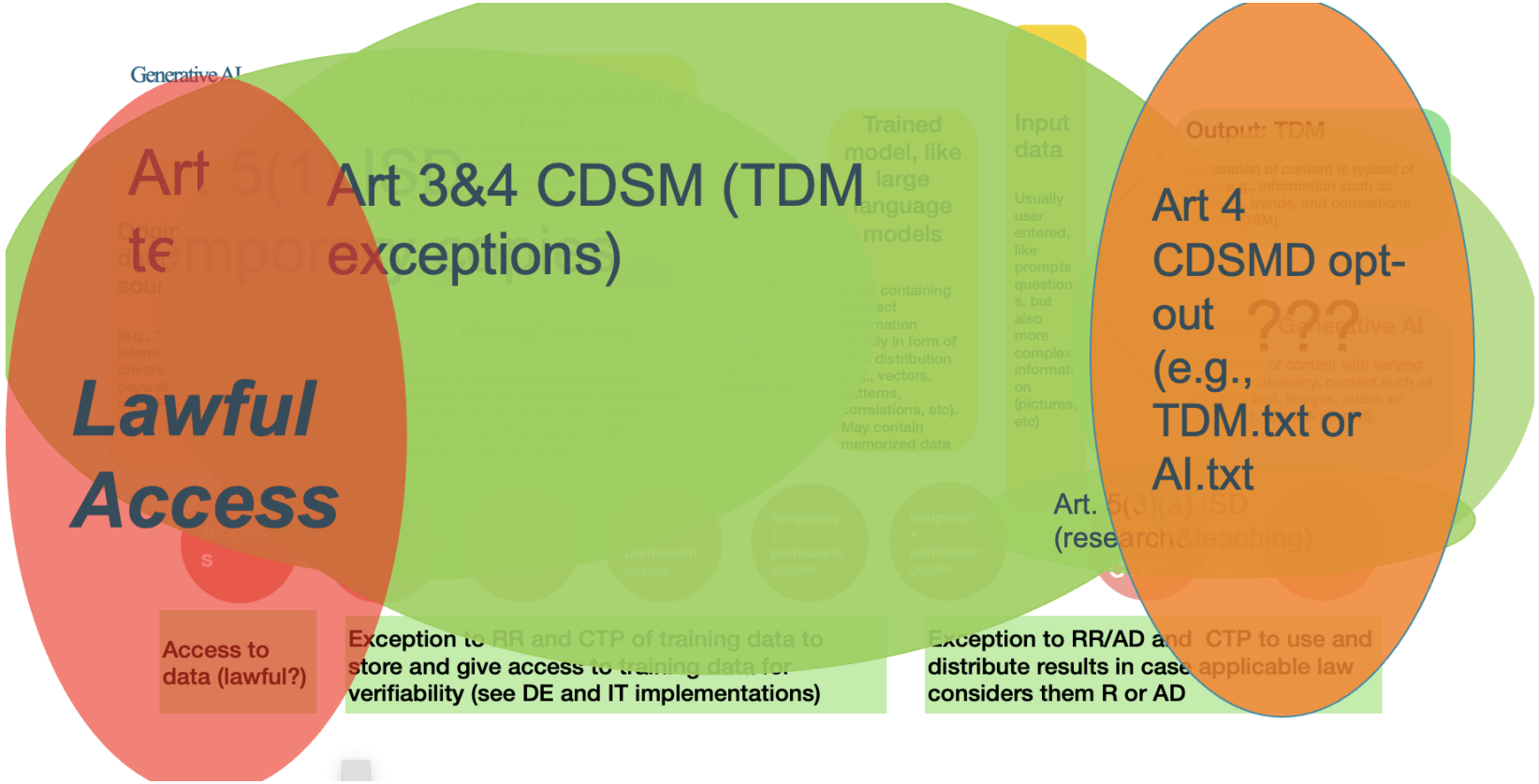
Access to data (lawful?)

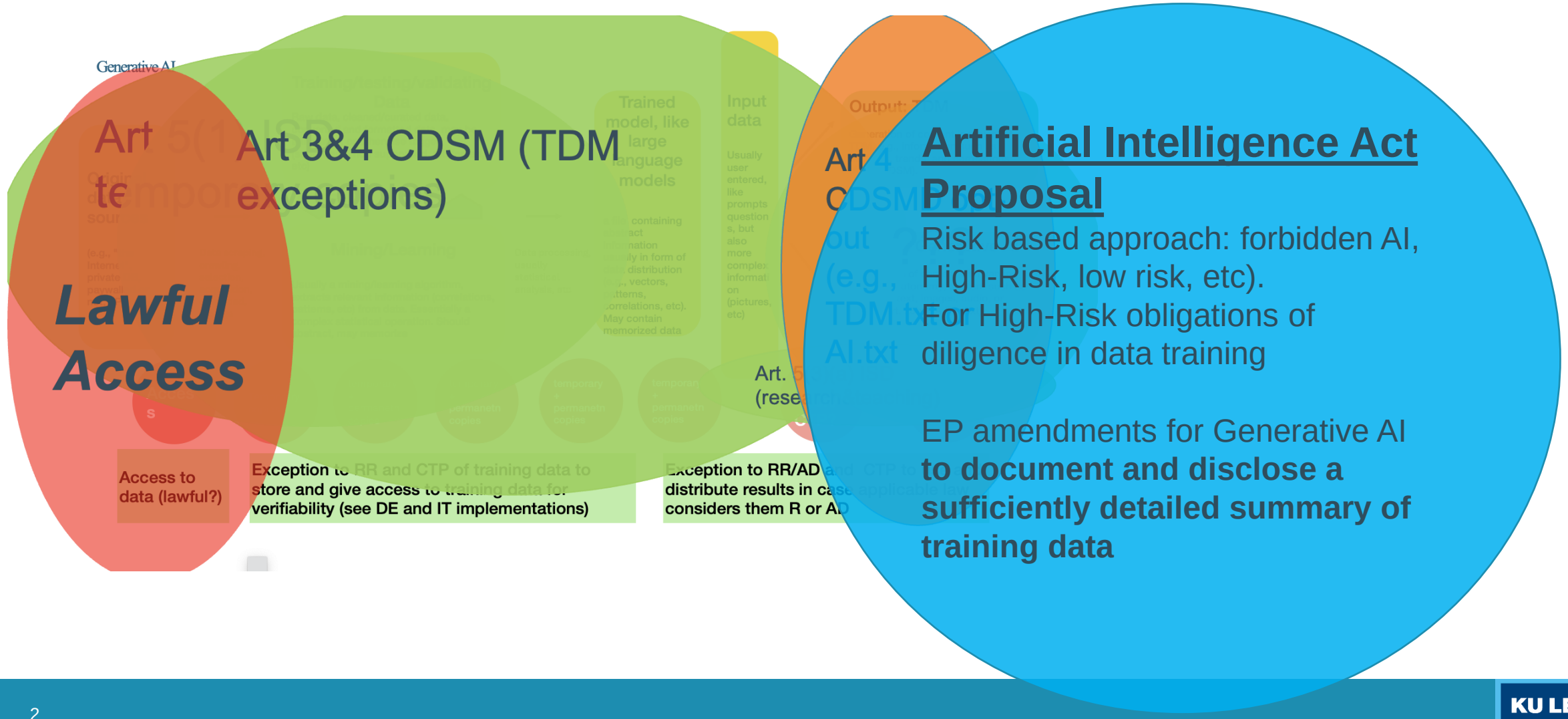
Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)

Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD

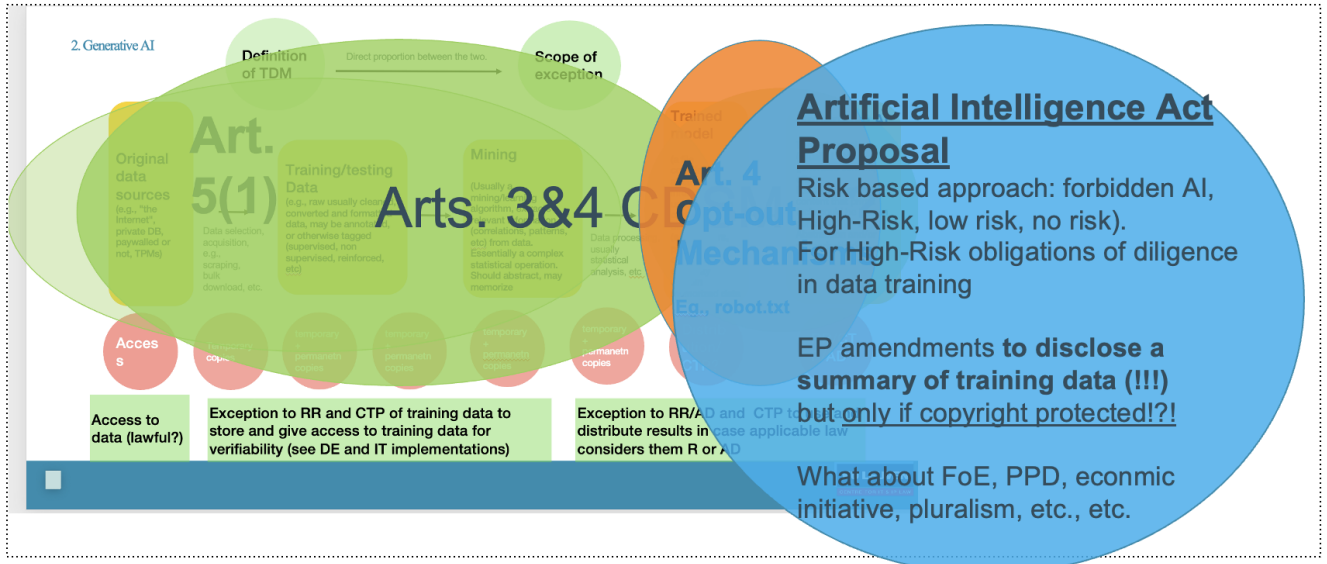

Generative AI
Art
te
Art 3&4 CDSM (TDM exceptions)
Art 4 CDSMD opt-out (e.g., TDM.txt or AI.txt
???
Art. 5(3)(a) ISD (research&teaching)
Access to data (lawful?)
Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)
Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD


Art 3&4 CDSM (TDM exceptions)
Art 4 CDSMD opt-out (e.g., TDM.txt or AI.txt
Lawful Access
Access to data (lawful?)
Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)
Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD

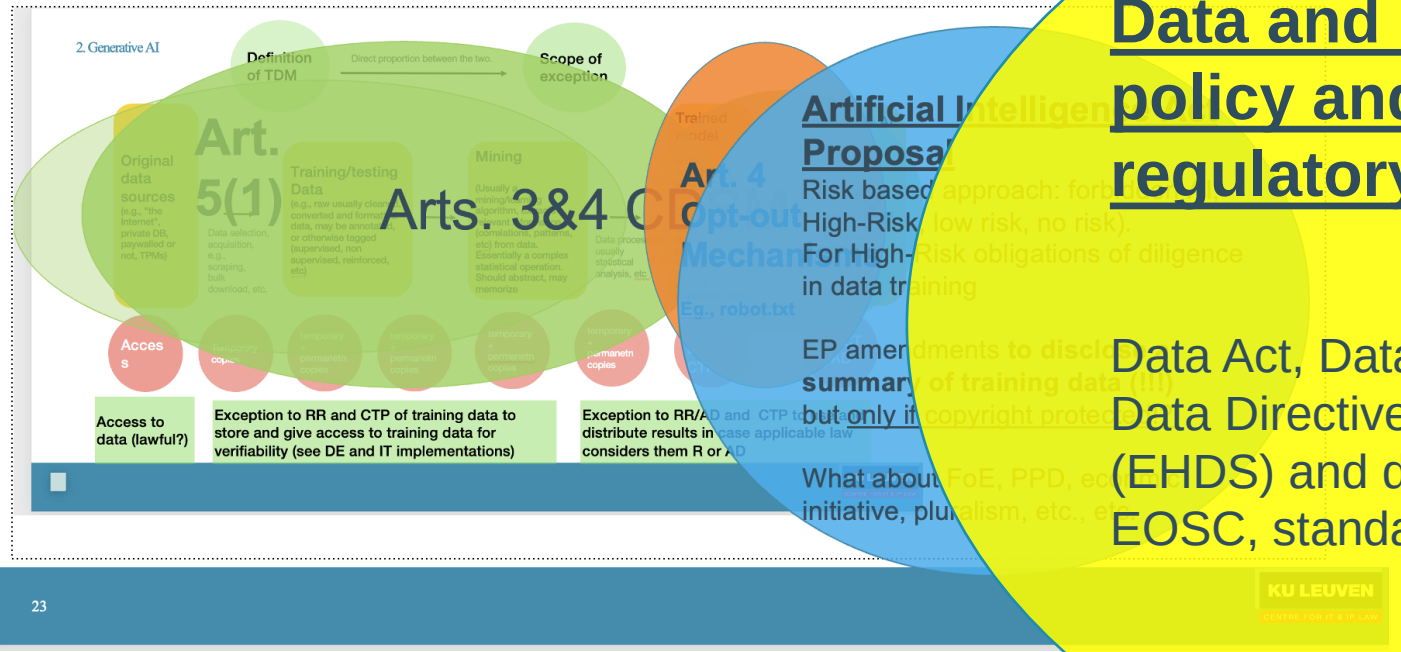

Art
te
Art 3&4 CDSM (TDM exceptions)
Lawful Access
Art 4 CDSMD opt-out (e.g., TDM.txt or AI.txt
???
Access to data (lawful?)
Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)
Exception to RR/AD and CTP to use and distribute results in case applicable law considers them R or AD


Art
te
Art 3&4 CDSM (TDM exceptions)
Lawful Access
Art
CDSM
out
(e.g.,
TDM.txt
AI.txt
Access to data (lawful?)
Exception to RR and CTP of training data to store and give access to training data for verifiability (see DE and IT implementations)
Exception to RR/AD
distribute results in ca
considers them R or A
Artificial Intelligence Act Proposal
Risk based approach: forbidden AI, High-Risk, low risk, etc).
For High-Risk obligations of diligence in data training
EP amendments for Generative AI to document and disclose a sufficiently detailed summary of training data

TDM & Generative AI



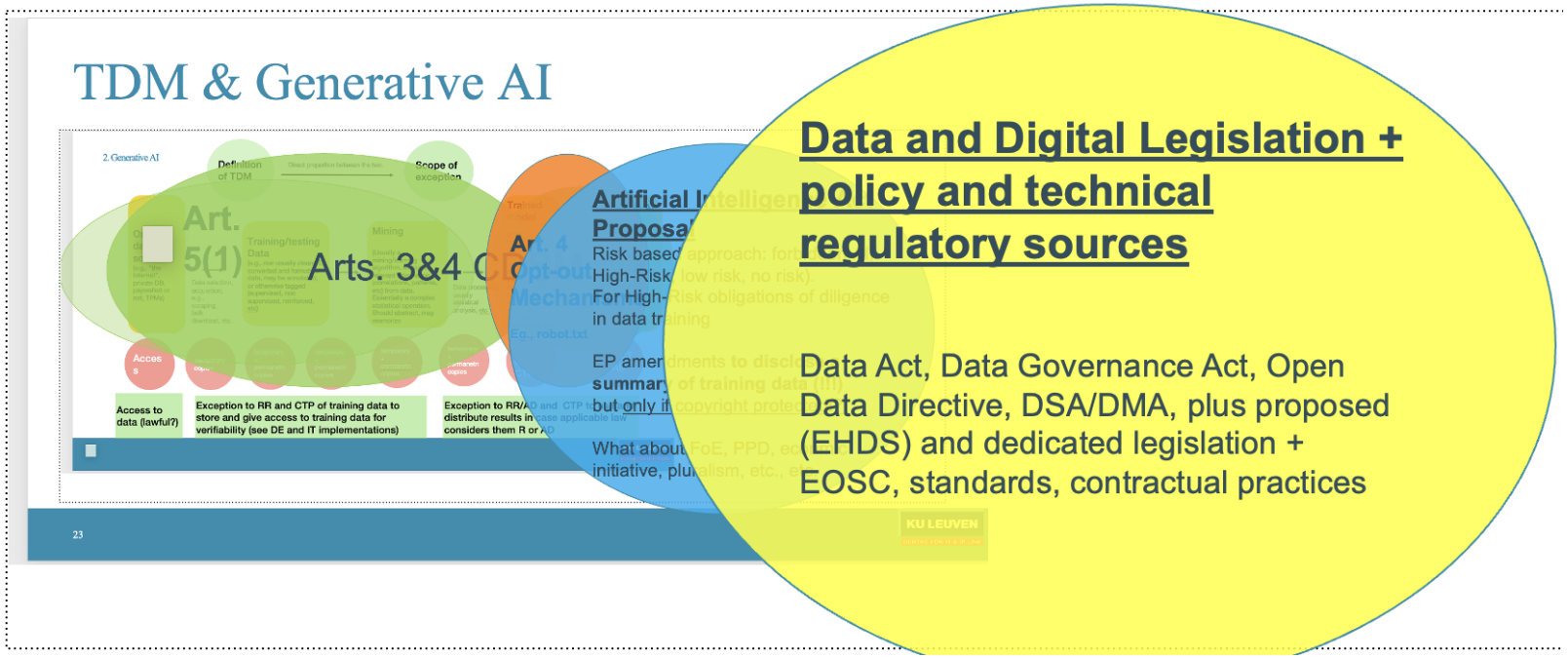
TDM & Generative AI



Data and Digital Legislation + policy and technical regulatory sources

Data Act, Data Governance Act, Open Data Directive, DSA/DMA, plus proposed (EHDS) and dedicated legislation + EOSC, standards, contractual practices

3. Common European Data Spaces (CEDS)




TDM & Generative AI
Data and Digital Legislation + policy and technical regulatory sources
Common European Data Spaces

Some examples of the impact on “data” of DDL

Non personal data access and portability:

- user of IoT has right to access (as co-generator) IoT data for free and to ask data holder to transfer data to designated third party including for commercial purposes (but no to develop directly competing product, yes for secondary markets, repair, additional services)

- Actual positive **B2G obligation to give access to privately held datasets** when request comes from PSB (including ROs) in cases of special need (e.g, climate, health emergencies, etc).

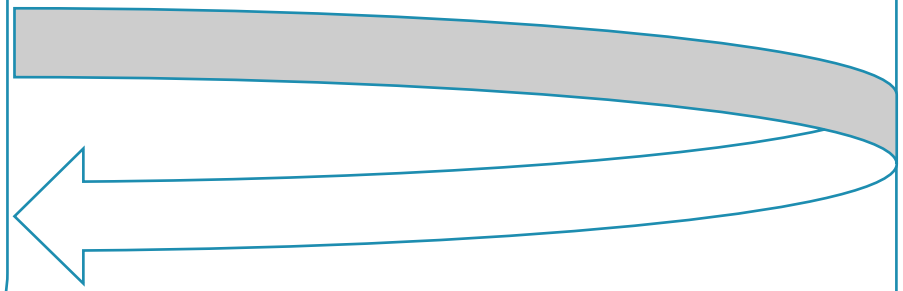
- **Right to switch** in cloud and hedge

- **No SGDR** in IoT data

Data property (e.g, copyright, trade secrets, etc)

Ownership and/or exclusive control *de jure* or *de facto*, freedom of contract

Data access, use and portability rights (IoT, B2C, B2B, B2G, etc.)



Technical Protection Measures Art. 11 DA, e.g., prohibition of circumvention, injunctive relief and damages against users and third parties

Data governance

No to SGDR or any other rights to use IoT data, introduce limitation To freedom of contract, etc.

Yes to Independent administrative authorities, data altruism, fairness in data transactions and value allocation.

KU LEUVEN CiTiP

CENTRE FOR IT & IP LAW

imec

Aim: To support the creation of common data spaces that collectively provide a data sovereign, interoperable and trustworthy environment for data sharing to enable data re-use within and across sectors, fully respecting EU values and supporting the European economy and society.

DATA SPACES SUPPORT CENTRE

The Data Spaces Support Centre receives funding from the European Union Digital Europe Programme under grant agreement no 101083412.

Funded by the European Union

Further readings

- Margoni, Thomas; Strowel, Alain; 2024. **Contractual freedom and fairness in EU data sharing agreements**, in Research Handbook on Intellectual Property Licensing
- Margoni, et al, **Data property, data governance and Common European Data Spaces** in Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht; 2023, https://zenodo.org/record/7906945
- Margoni, Kretschmer, **A deeper look into the EU TDM exceptions: harmonisation, data ownership and the future of technology**, in GRUR Int., 2022, https://doi.org/10.1093/grurint/ikac054
- Ducuing, Margoni, et al, **White paper on Data Act proposal, CiTiP Working paper 2022,** https://lirias.kuleuven.be/retrieve/682728

KU LEUVEN CiTiP

CENTRE FOR IT & IP LAW

imec

Thank you!

thomas.margoni@kuleuven.be

KU Leuven Centre for IT & IP Law (CiTiP) - imec

http://www.law.kuleuven.be/citip