Big Data and the application of anonymization techniques
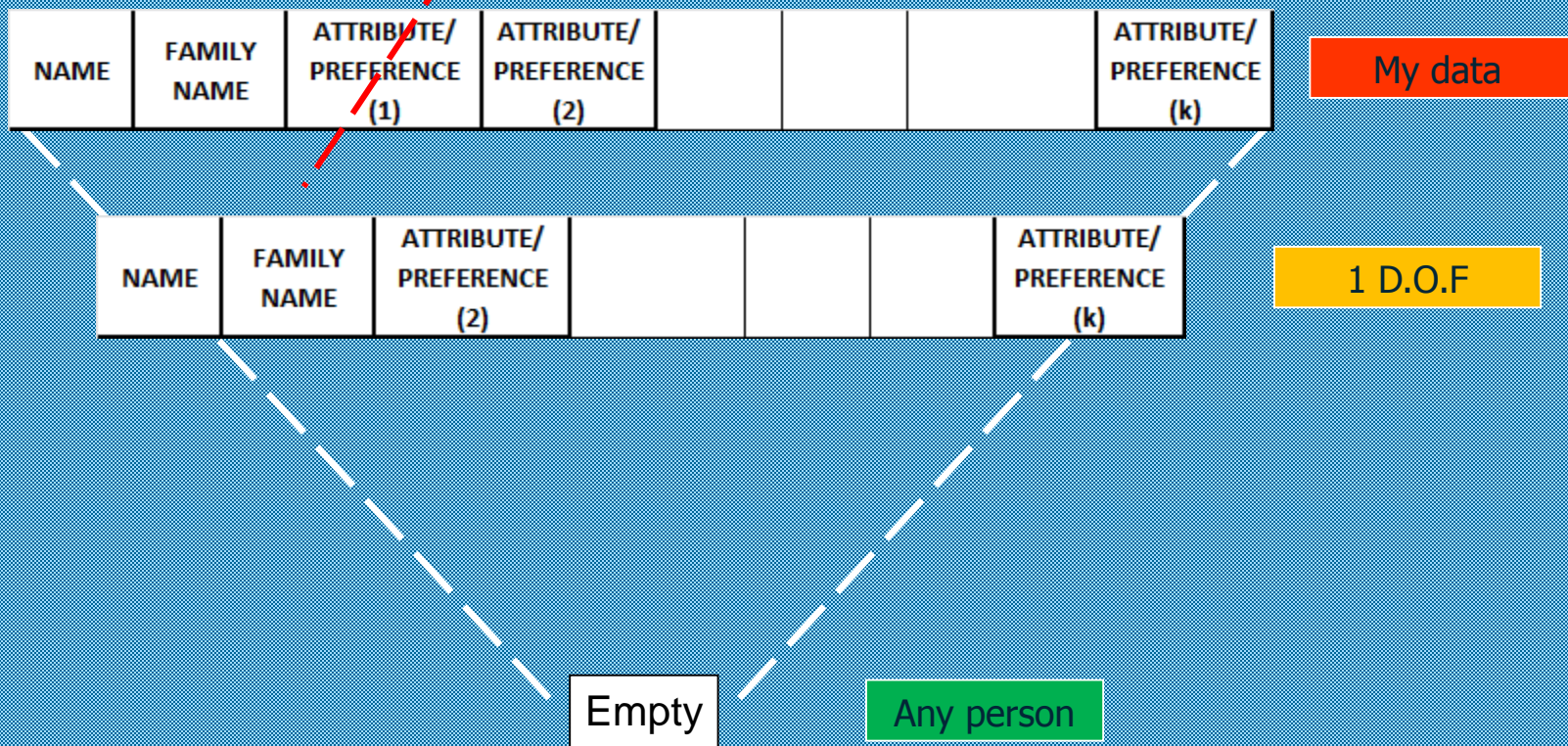
Annual Privacy Forum 2015

7-8 October, Luxembourg

Giuseppe D'Acquisto

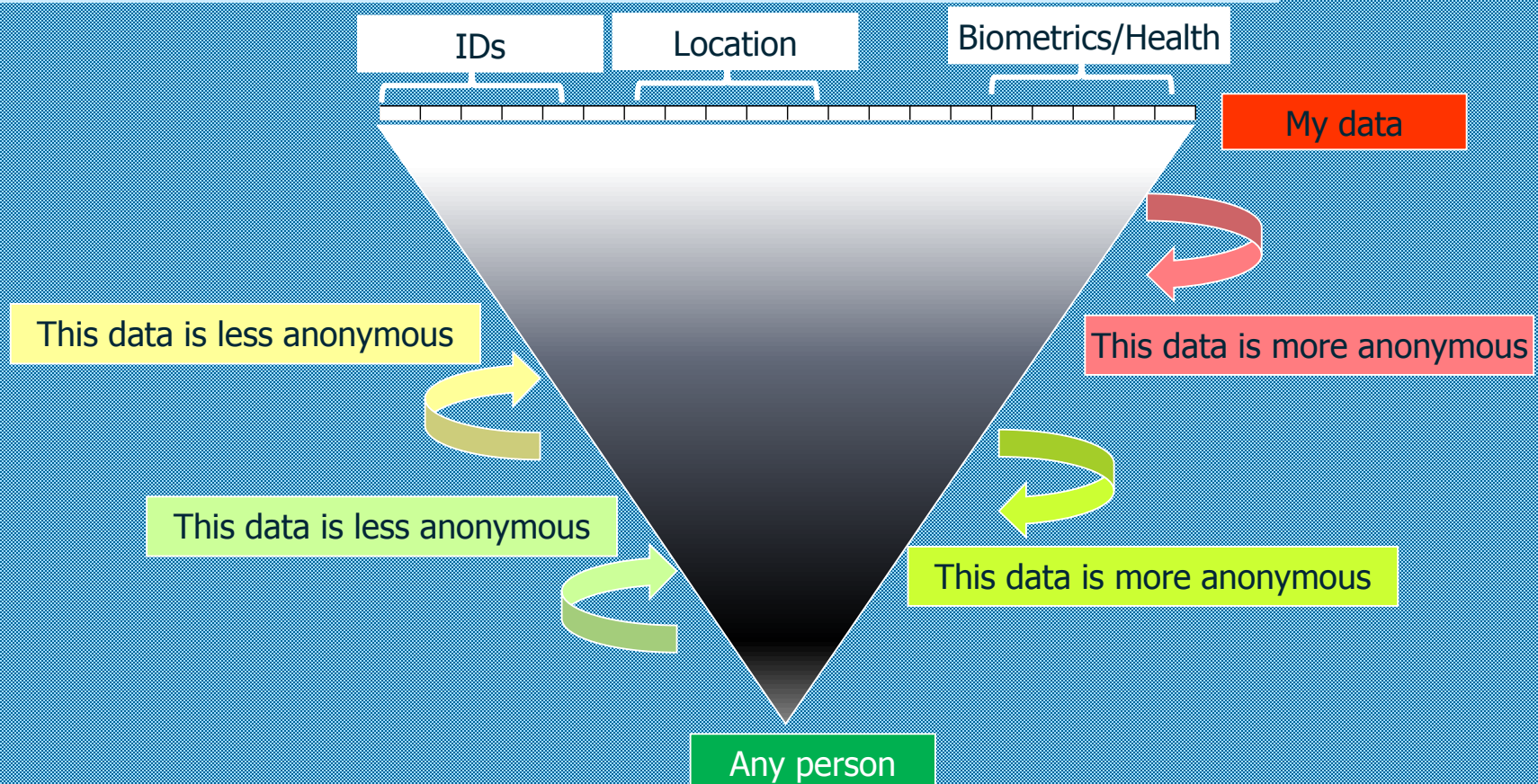*Garante per la protezione dei dati personali*

Annual
Privacy
Forum

The concept of anonymization

NAME	FAMILY NAME	ATTRIBUTE/ PREFERENCE (1)	ATTRIBUTE/ PREFERENCE (2)				ATTRIBUTE/ PREFERENCE (k)

My data

NAME	FAMILY NAME	ATTRIBUTE/ PREFERENCE (2)				ATTRIBUTE/ PREFERENCE (k)

1 D.O.F

Empty

Any person

Anonymization is a relative concept

IDs	Location	Biometrics/Health

My data

This data is less anonymous

This data is more anonymous

This data is less anonymous

This data is more anonymous

Any person

Anonymization is absolute from legal perspective

<table>
<tr><td>

**This
is
personal
data**

🔴
</td><td>

**This
is not
personal
data**

🟢
</td></tr>
</table>

The anonymization approach in the U.S.

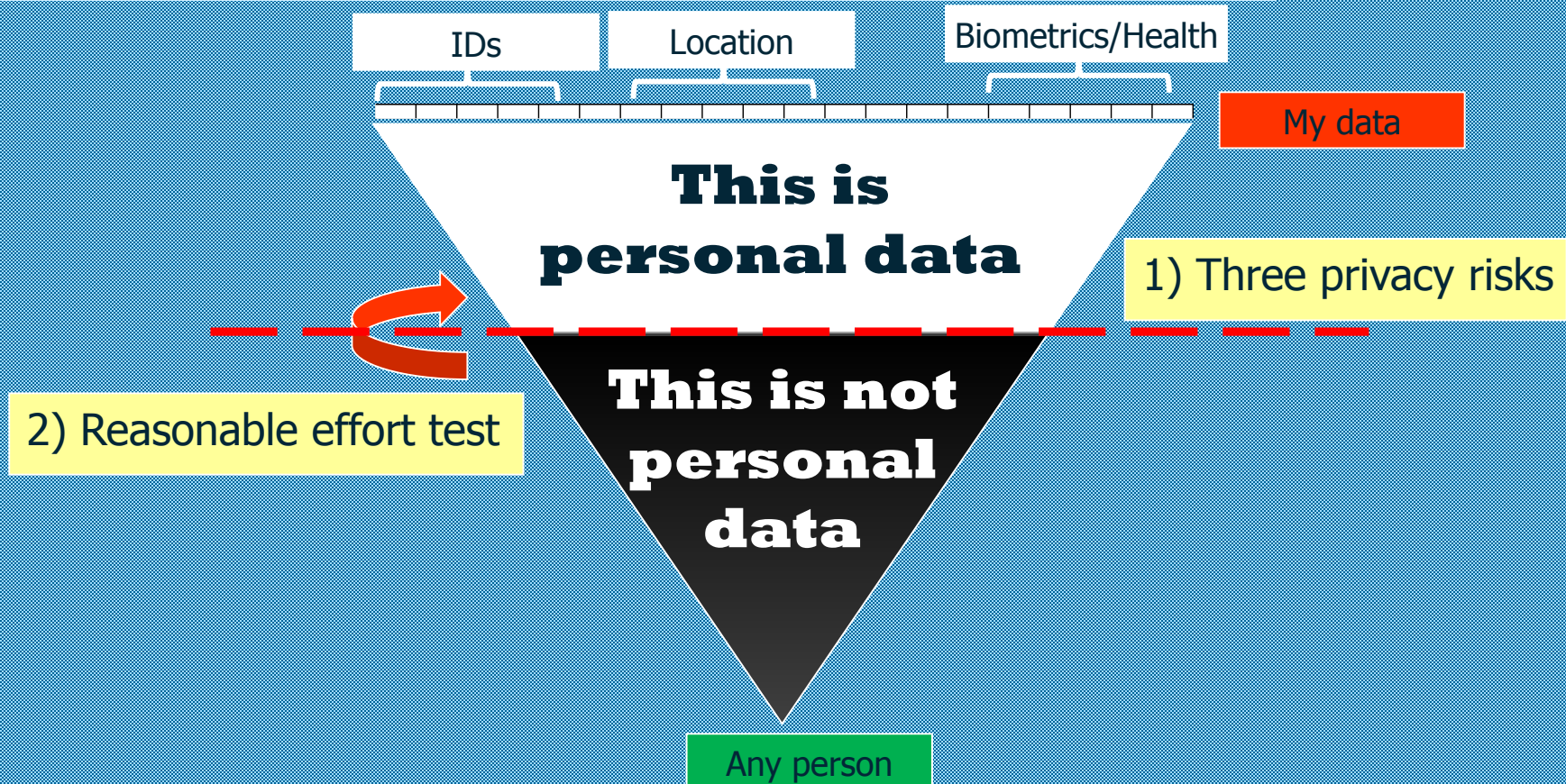**Health Insurance Portability and Accountability Act of 1996 (HIPAA)**

NAME	FAMILY NAME	ATTRIBUTE/ PREFERENCE (1)	ATTRIBUTE/ PREFERENCE (2)				ATTRIBUTE/ PREFERENCE (k)

This is personal data **This is not personal data**

The anonymization approach in the WP29 Opinion

IDs	Location	Biometrics/Health

This is personal data

My data

1) Three privacy risks

2) Reasonable effort test

This is not personal data

Any person

Engineering may not be enough

IDs

Location

Biometrics/Health

My data

After engineering

This is personal data

This is not personal data

Gap to fill with policies

Any person

Safeguards

| First processing in compliance | IDs | Location | Biometrics/Health |

This is personal data

My data

After Anonymization

This is not personal data

If the data is in the personal sphere (the device), then art. 5(3) applies

Any person
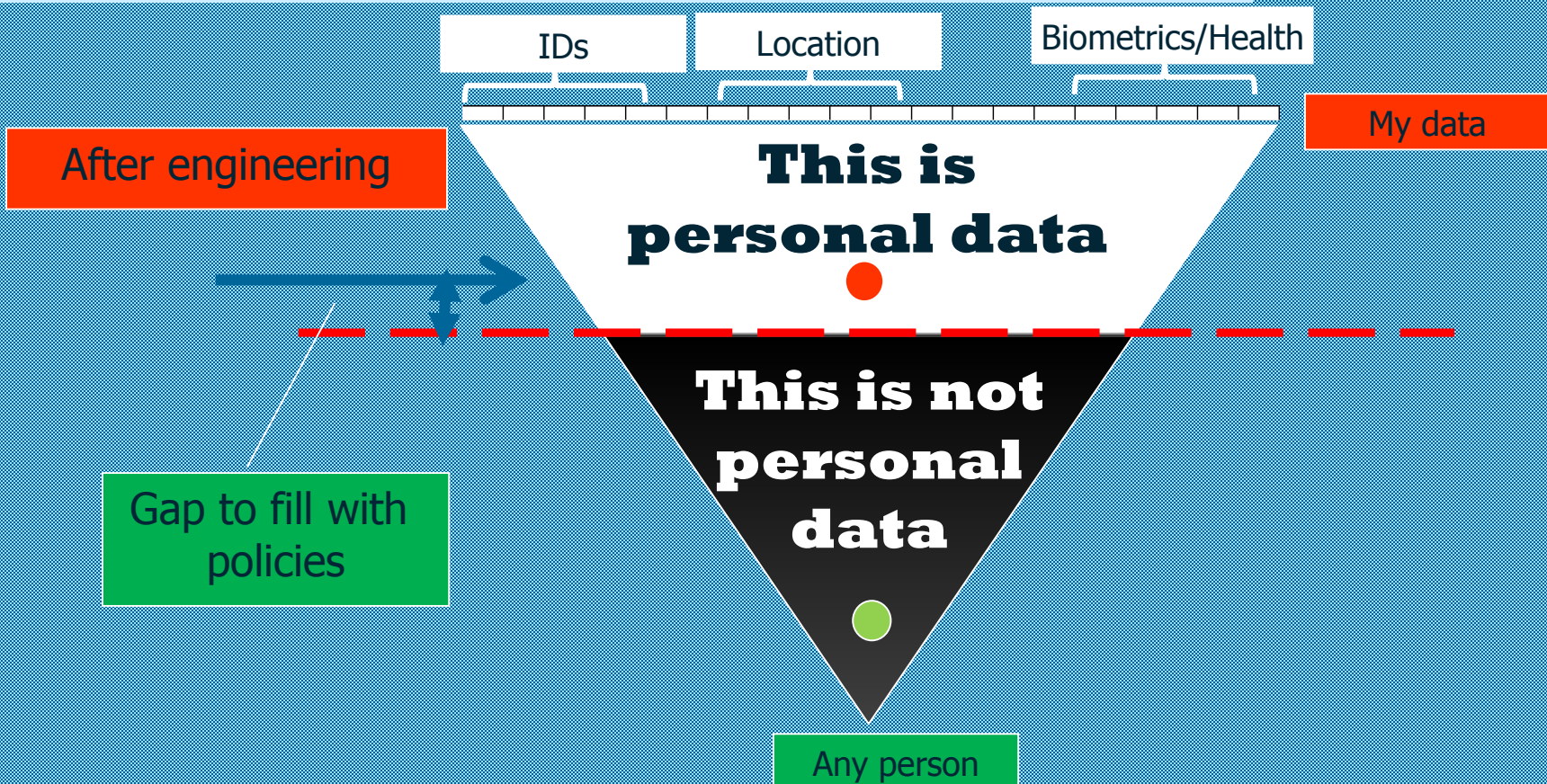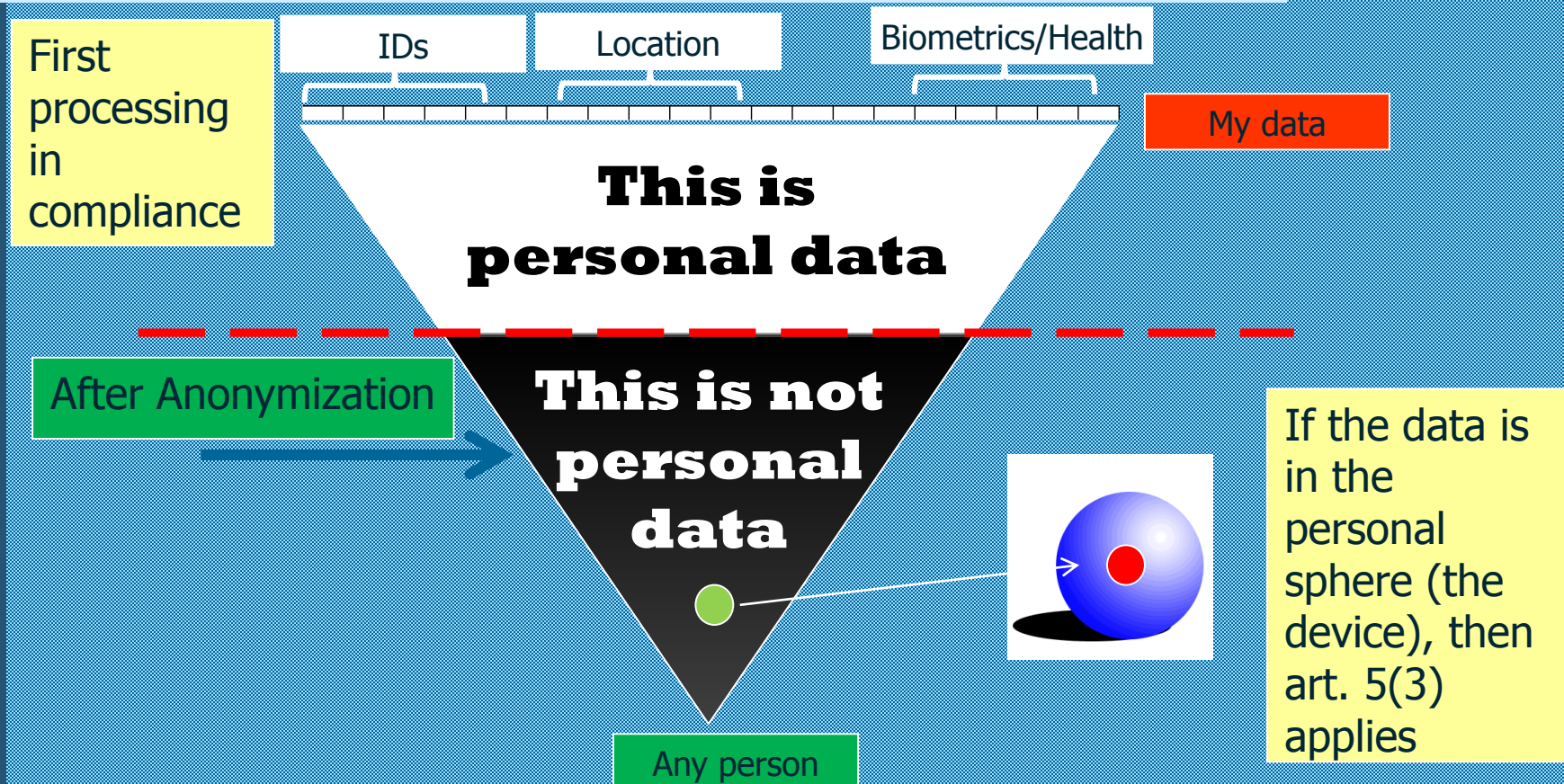
Additional safeguards

On the re-use of data

| IDs | Location | Biometrics/Health |

This is personal data

My data

This is not personal data

Anonymization as a compatible further purpose

- Non incompatibility of purposes
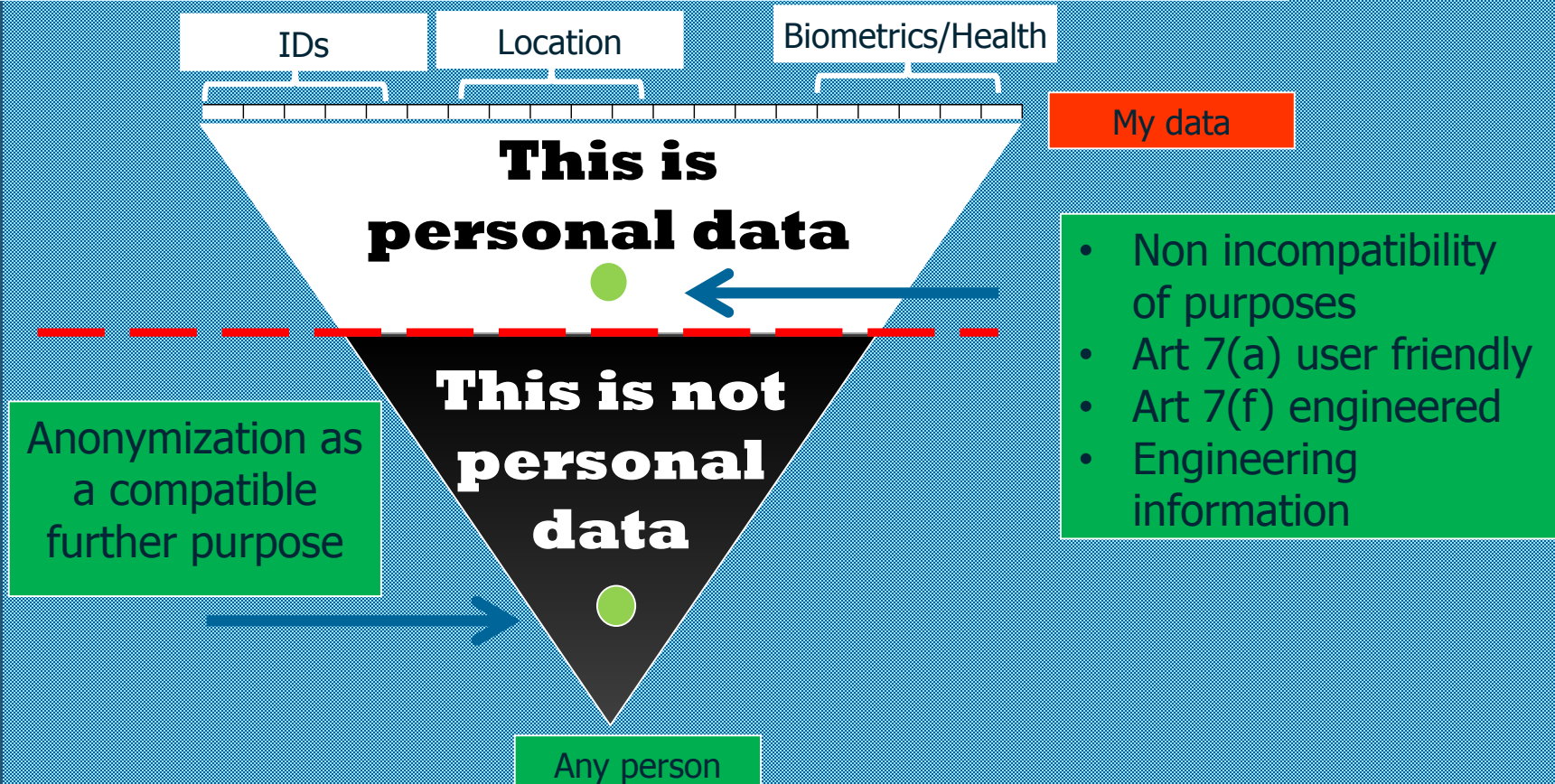- Art 7(a) user friendly
- Art 7(f) engineered
- Engineering information

Any person

Conclusions

- There is room for privacy principles also in Big Data

- New tools for safeguarding data subjects

  - Policy

  - Technology

  - Probability/Information theory

- The key is the capability to deal with complexity: anonymization is difficult (but not impossible)…

- …but, bad anonymization is very easy (AoL 2006 – Netflix 2009 - NY taxis 2014)

Thank you very much

g.dacquisto@gpdp.it