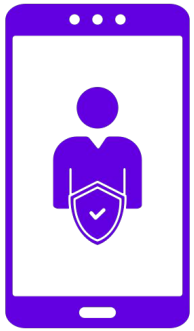
MADTECH MENTAL MODELS

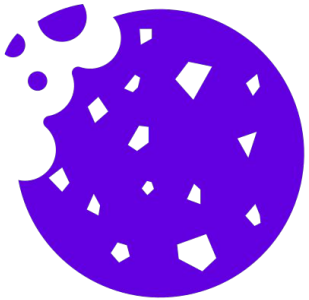
DATA CLEAN ROOMS

Data Clean Rooms are secure environments where two (or more) parties can upload their user data to collaborate on mutually agreed upon use cases without actually sharing the data or compromising the privacy of their users.

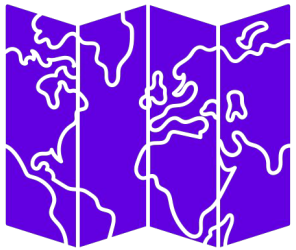
Why are they a thing now?



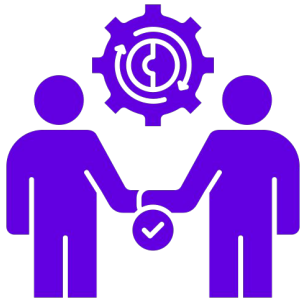
Data privacy regulation & consumer concern



Decline of traditional 3P identifiers

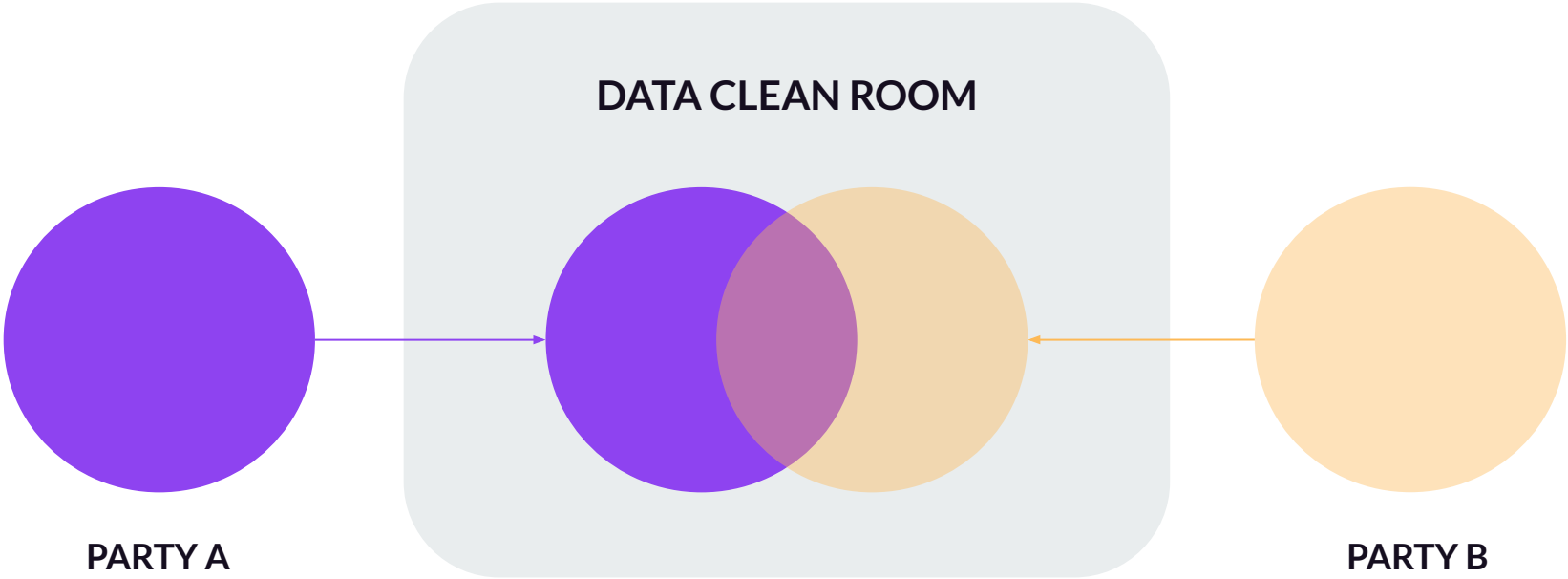


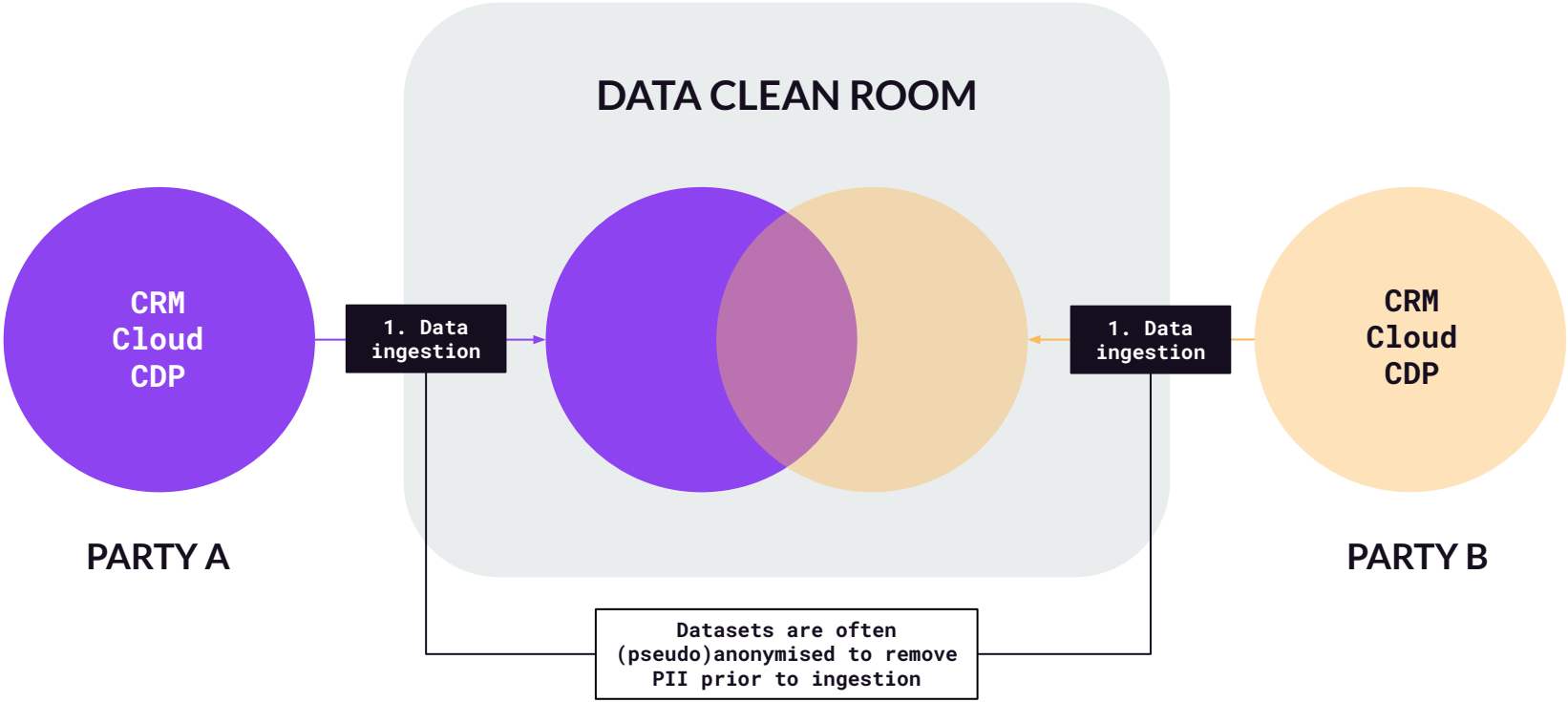
Increased media & data fragmentation

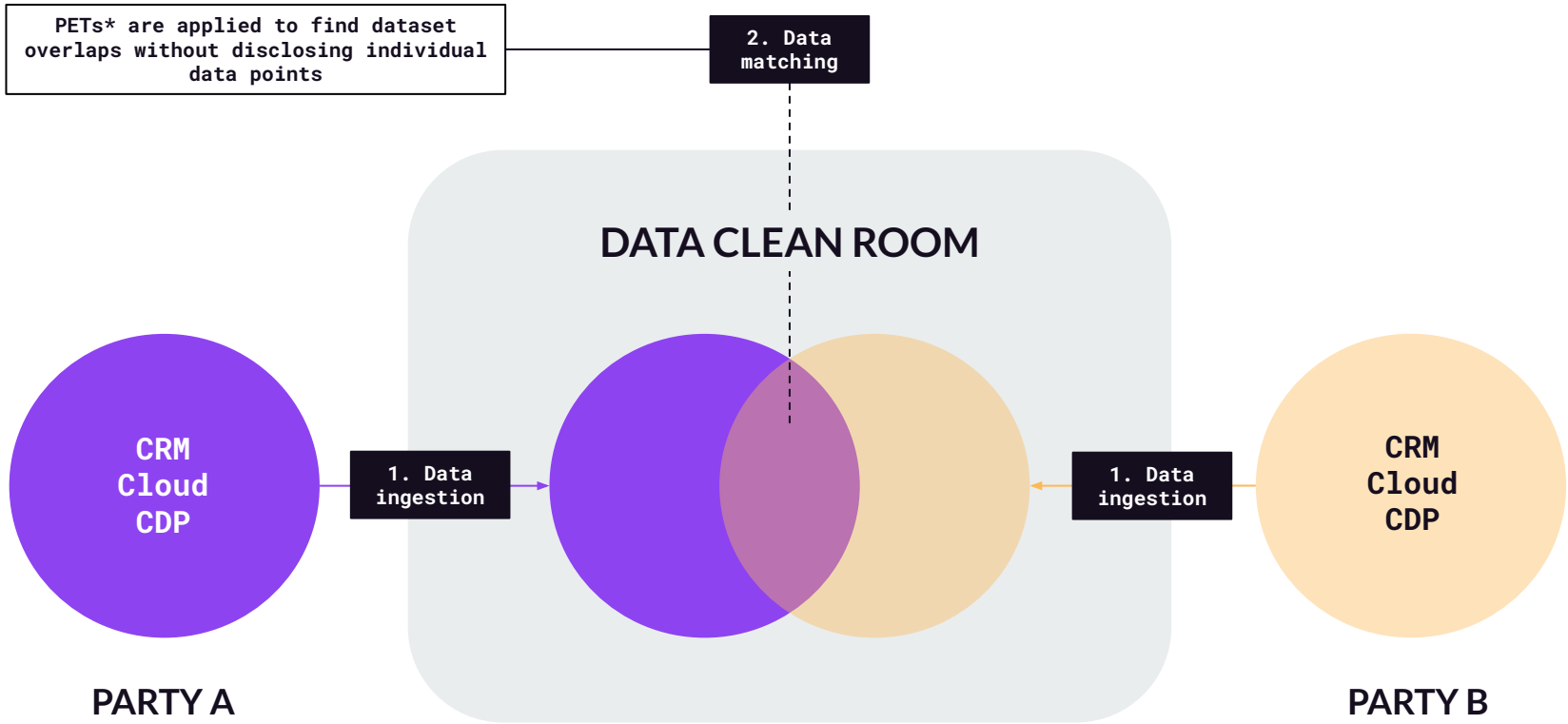


Increased demand for 1P data collaboration

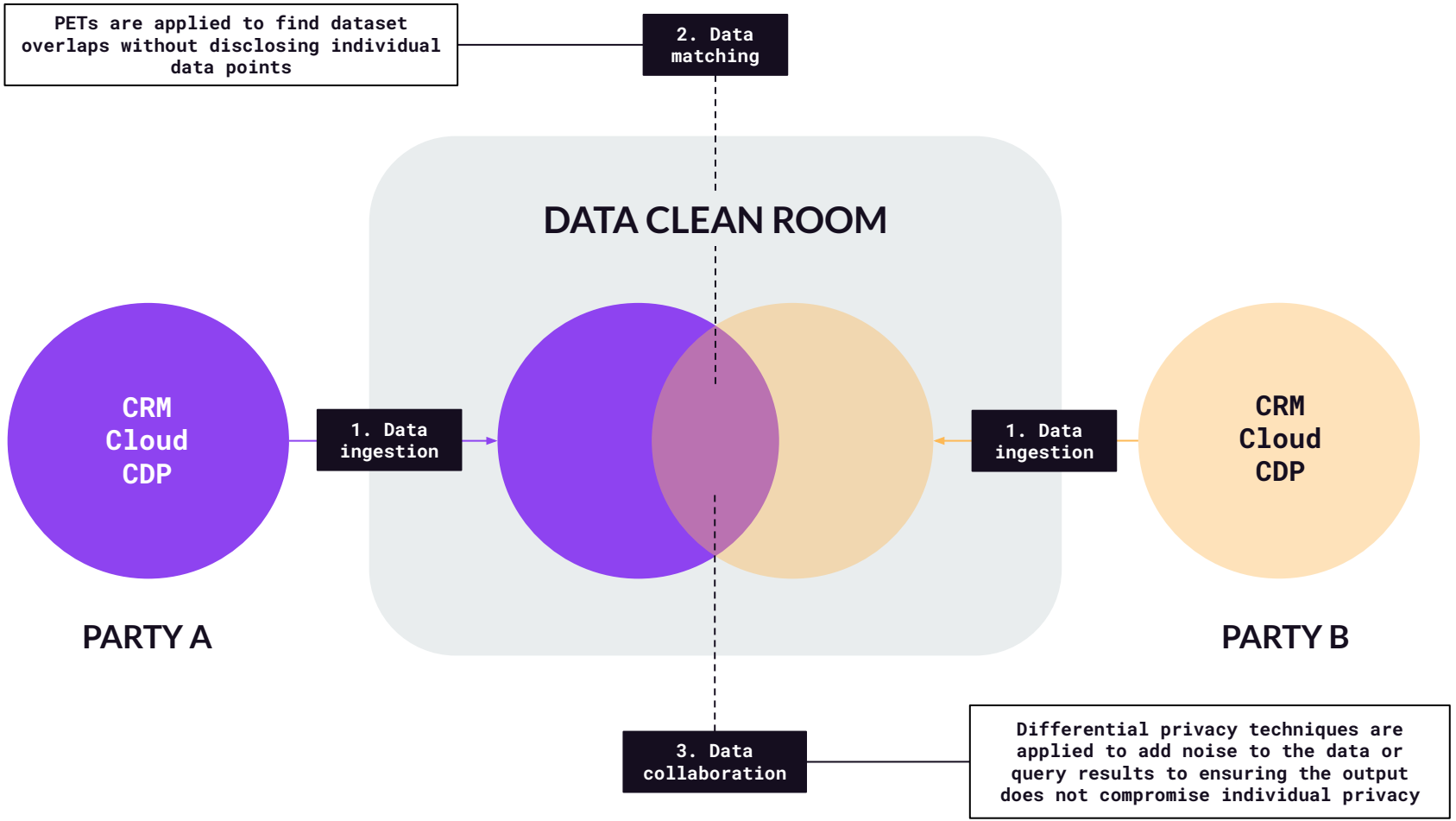
How they work
(on a high level)

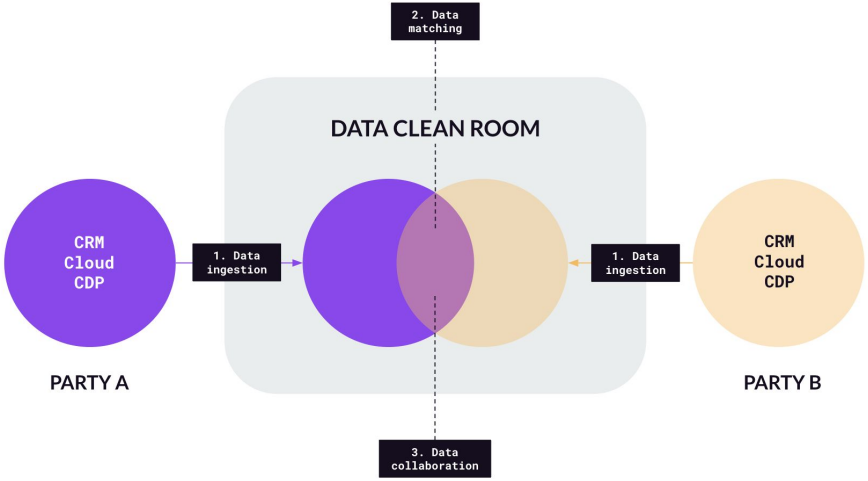

DATA CLEAN ROOM
PARTY A
PARTY B


DATA CLEAN ROOM
CRM
Cloud
CDP
1. Data
ingestion
1. Data
ingestion
CRM
Cloud
CDP
PARTY A
PARTY B
Datasets are often
(pseudo)anonymised to remove
PII prior to ingestion



*PETs = Privacy enhancing technologies


PETs are applied to find dataset overlaps without disclosing individual data points
2. Data matching
DATA CLEAN ROOM
CRM Cloud CDP
1. Data ingestion
1. Data ingestion
CRM Cloud CDP
PARTY A
PARTY B
3. Data collaboration
Differential privacy techniques are applied to add noise to the data or query results to ensuring the output does not compromise individual privacy



4.Aggregated output syndication

Insights

- Data environments
- MadTech platforms
- Analytics platforms
- BI tools

Activation

- DSPs/SSPs
- Walled gardens
- Retail media networks
- Publishers

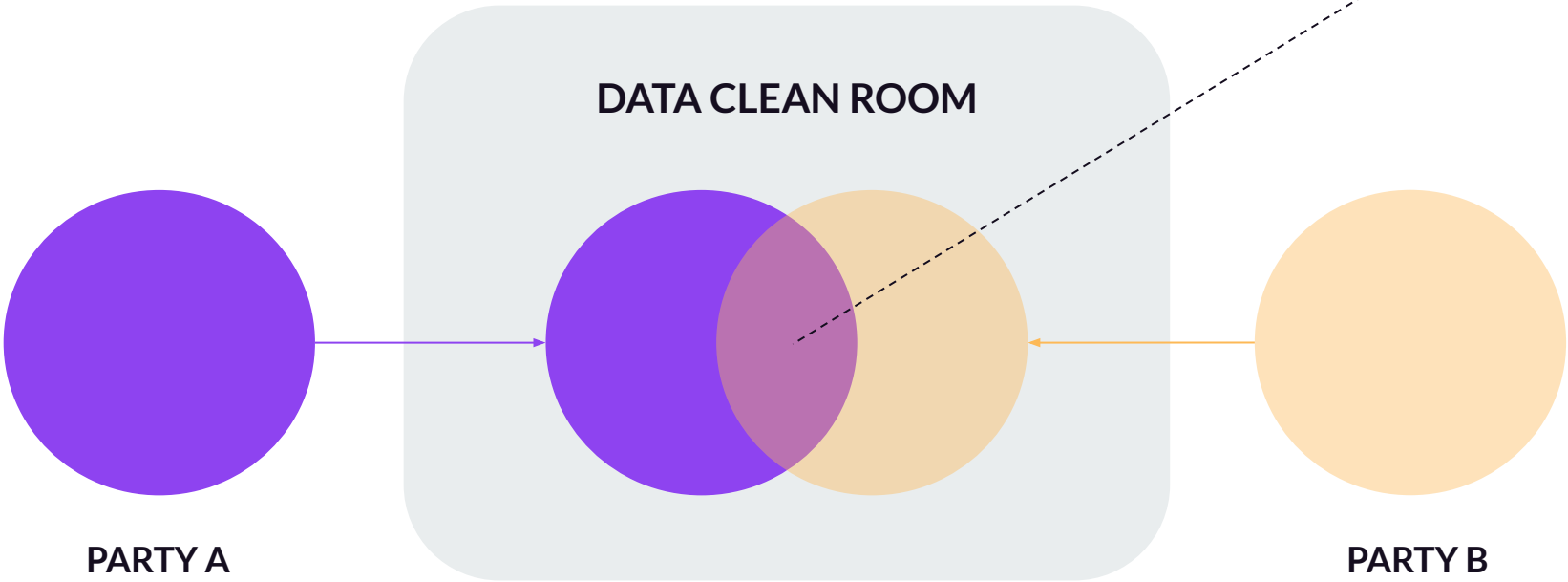
Measurement

- Measurement providers
- Analytics platforms
- BI tools

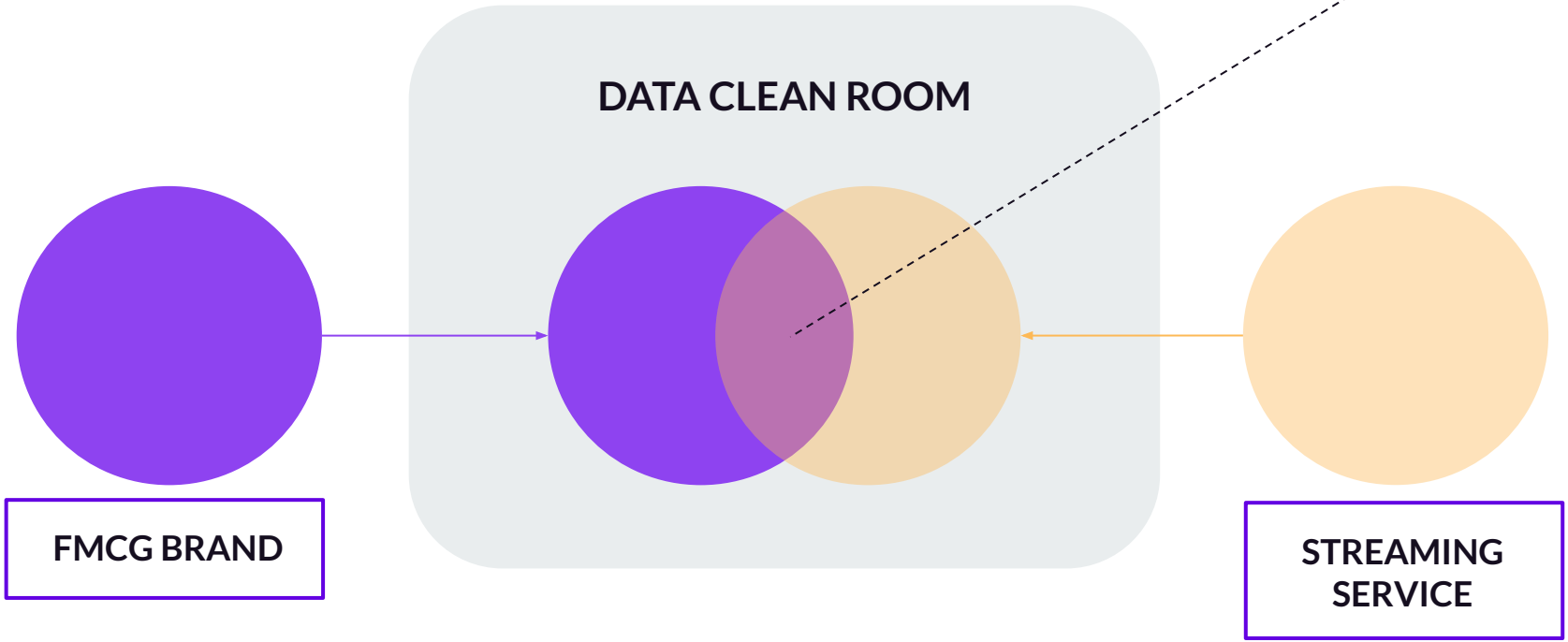
Diving a bit deeper

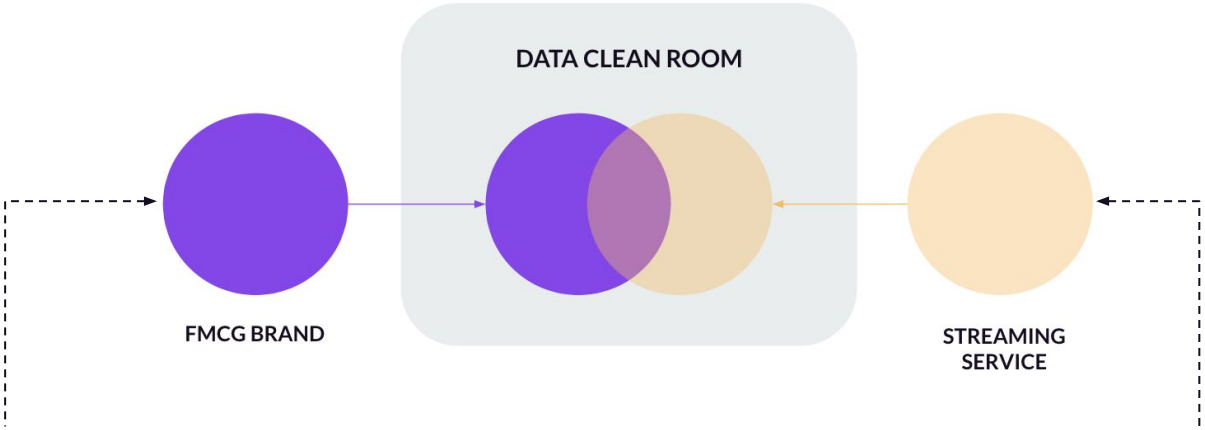
What data matching looks like in a Data Clean Room (*think venn diagrams…*)

What does this actually look like?



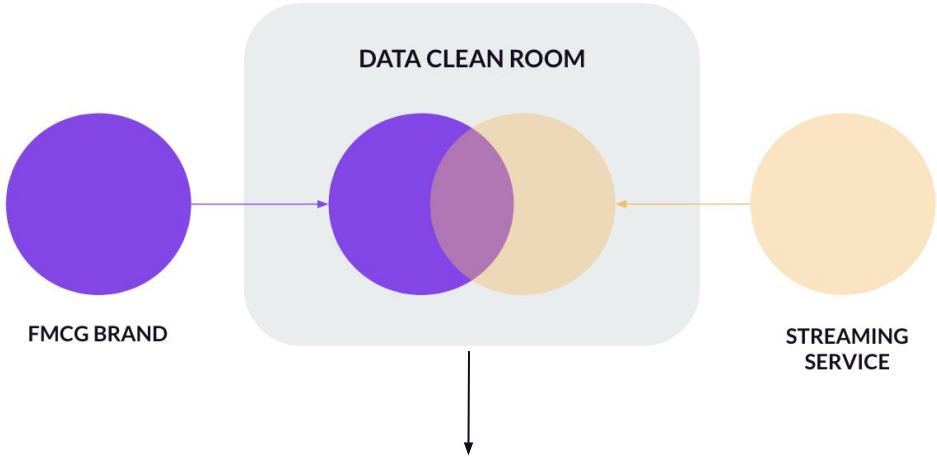
What does this actually look like?



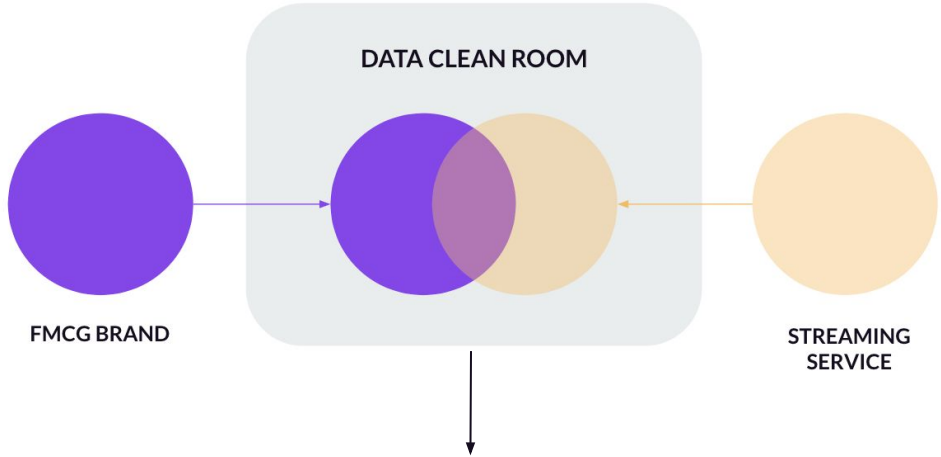


Encrypted ID	Age Group	Food Cohort
id_xx1	25-34	Fresh Food Lovers
id_xx8	18-24	Sweet Tooths
id_xx3	18-24	Fresh Food Lovers
id_xx5	25-34	Craft Beer Enthusiasts
id_xx7	35-44	Sweet Tooths

Encrypted ID	Primary Device	Genre Cohort
id_xx2	TV	Documentary Lovers
id_xx5	Desktop	Blockbuster Buffs
id_xx1	TV	Award Winners
id_xx3	Tablet	Blockbuster Buffs
id_xx9	Smartphone	Animation Aficionados

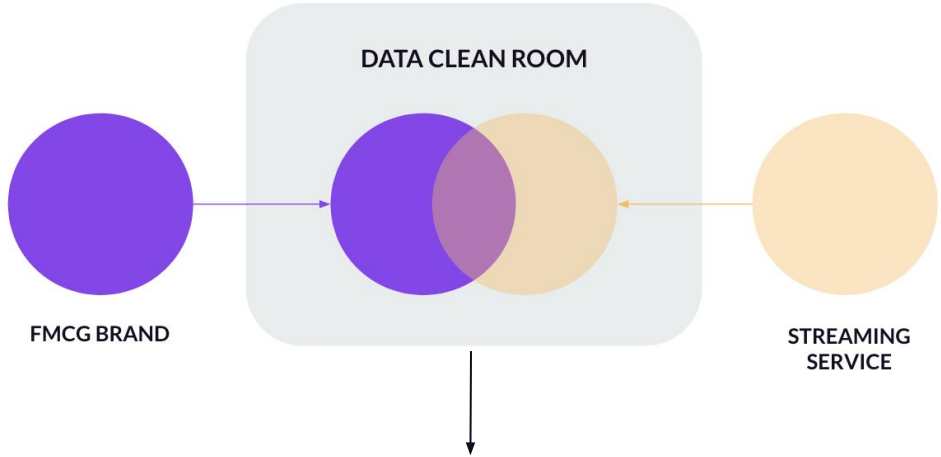


Food Cohort	Genre Cohort	Sum/Count
Sweet Tooths	Blockbuster Buffs	1174
Fresh Food Lovers	Documentary Lovers	374
Craft Beer Enthusiasts	Award Winning Classics	284
Sweet Tooths	Animation Aficionados	732
Fresh Food Lovers	Award Winning Classics	48



Food Cohort	Genre Cohort	Sum/Count
Sweet Tooths	Blockbuster Buffs	1174
Fresh Food Lovers	Documentary Lovers	374
Craft Beer Enthusiasts	Award Winning Classics	284
Sweet Tooths	Animation Aficionados	732
Fresh Food Lovers	Award Winning Classics	48

User-level data goes in, but doesn't come out



Food Cohort	Genre Cohort	Sum/Count
Sweet Tooths	Blockbuster Buffs	1174
Fresh Food Lovers	Documentary Lovers	374
Craft Beer Enthusiasts	Award Winning Classics	284
Sweet Tooths	Animation Aficionados	732
~~Fresh Food Lovers~~	~~Award Winning Classics~~	~~48~~

User-level data goes in, but doesn't come out

Redacted due to users being below minimum threshold

While INNER JOIN queries (venn diagram overlap) are the most common DCR matching function, other queries are also used depending on use case and what's permitted to ensure collaboration is done in way that preserves data privacy.

Conceptualising
Data Clean Room use cases

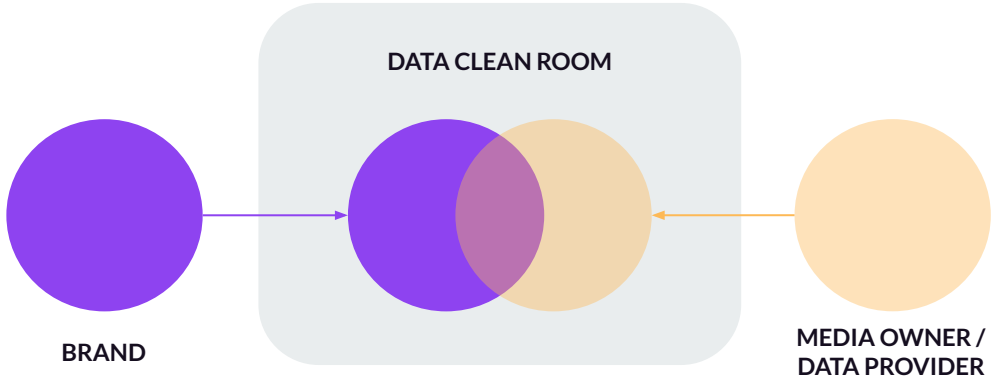
Data Clean Rooms allow one party to use another party's data to reveal new data points about their own users.

All use cases are ultimately unlocked by this concept.

Audience addressability and activation

Brands want to identify existing or new audiences for targeting and/or measurement at scale.

Brands use media owner or data provider data to reveal more attributes about their users.

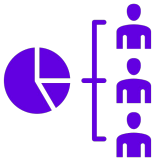


This enables/informs audience use cases including but not limited to:

Discovery

Targeting

Expansion/LAL

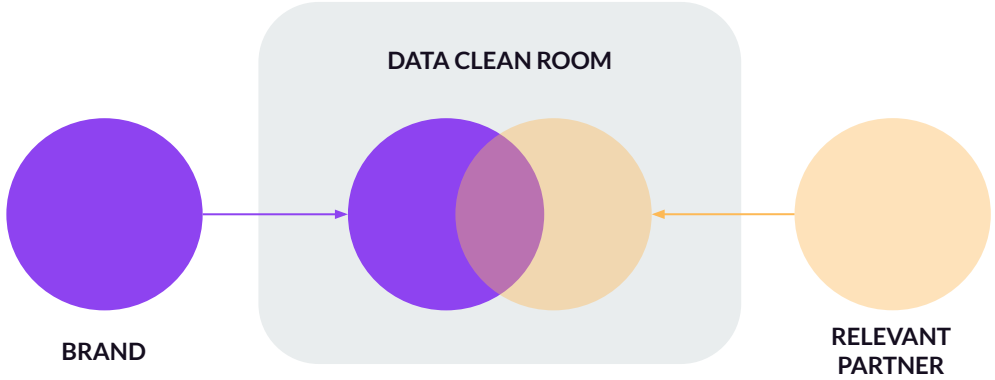


Segmentation

Data enrichment and insight generation

Brands want to enrich their 1PD to learn more about customers to improve product dev and messaging

Brands use partner data to enrich their existing customer data with additional data points.

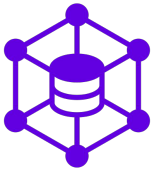


This enables/informs use cases including but not limited to:

Identity resolution*

Customer insights

Audience enrichment



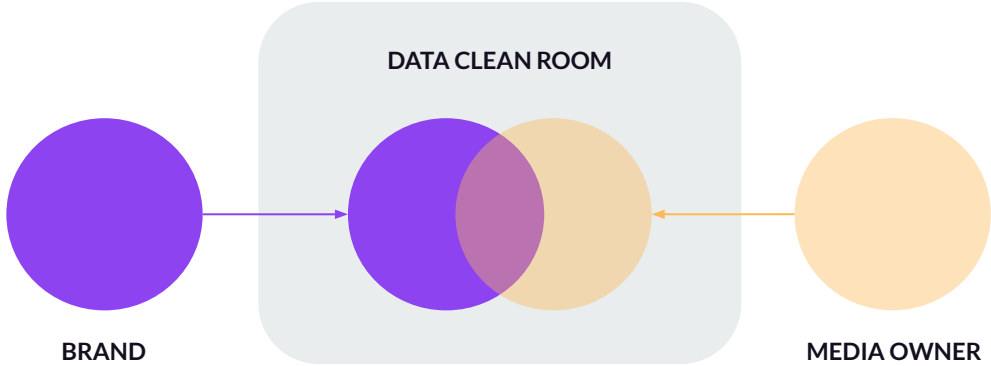
Audience modeling

*Requires an ID graph

Attribution, Measurement, Optimisation

Brands want to better understand the returns on their ad spend.

Brands use media owner data to reveal data points (e.g. exposure, conversion) needed to quantify campaign effectiveness.



This enables/informs use cases including but not limited to:

Reach and frequency

Audience validation

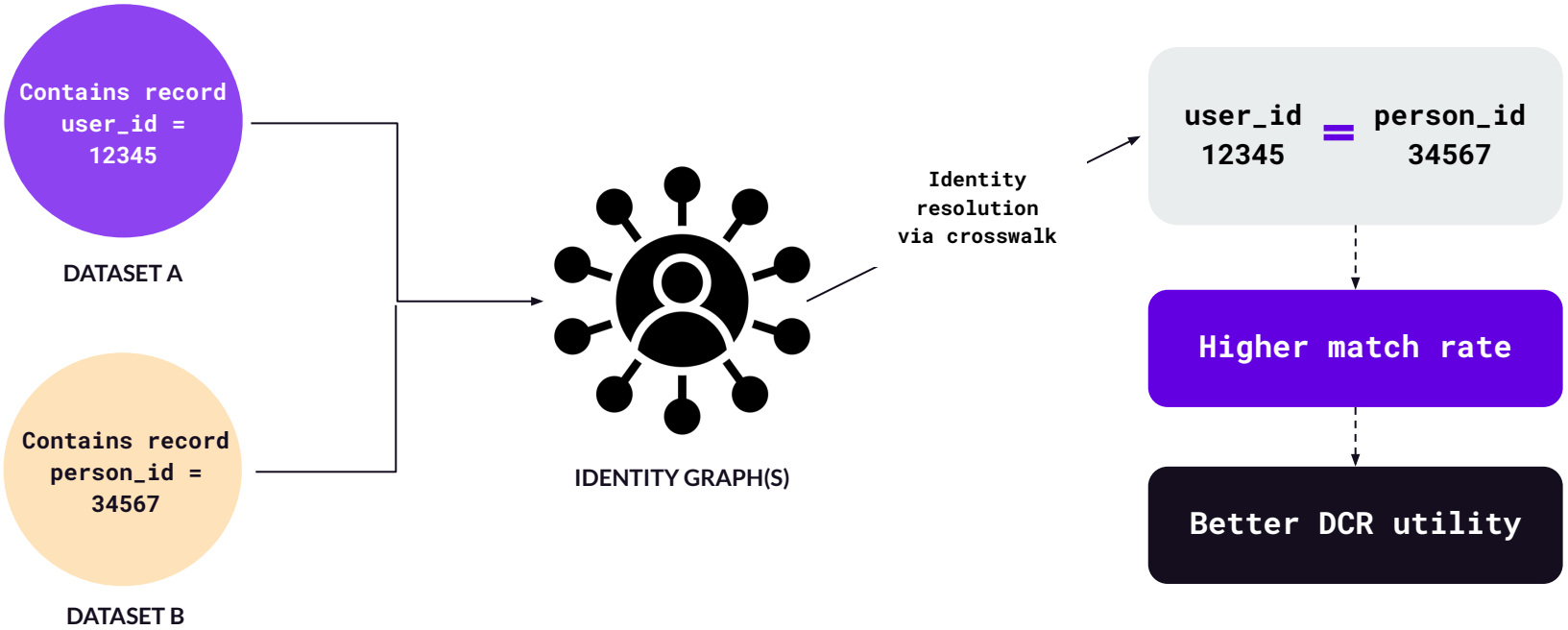
Incrementality measurement

Attribution analysis/modeling

Predictive modeling

The role of Identity Graphs in Data Clean Rooms

Data clean rooms use identity graphs to improve match rates between parties in scenarios where the datasets don't share the same identifier.



User data models in Data Clean Rooms

For more advanced use cases, data scientists can apply models to Data Clean Rooms in the following ways:

BUILD

Use the matched dataset as data to train a new model which can then be deployed within the DCR

IMPORT

Bring an existing, pre-trained model into the DCR for deployment against the data inside

Most of these use cases aren't new, but DCRs provide users with access to 1P data sources with whom to collaborate that were previously practically impossible.

Bringing the “Clean” to Data Clean Rooms

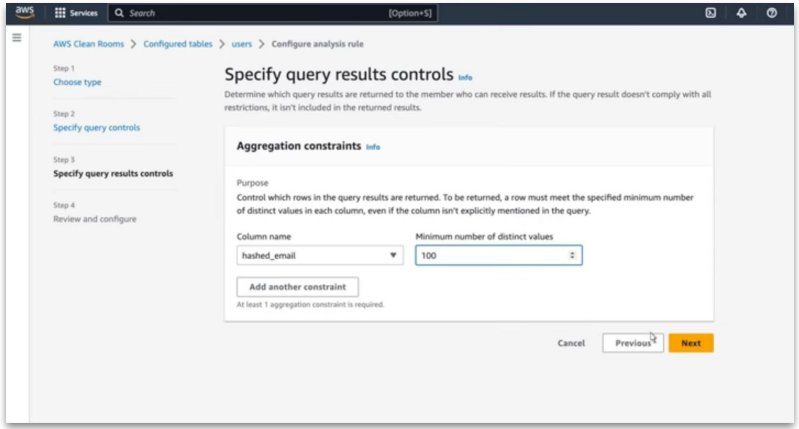
Privacy Enhancing Technologies (PETs)

Data clean rooms employ a broad suite of PETs to ensure that data analysis and collaboration between parties can occur in a secure, privacy-compliant manner. Here are some of the PETs commonly used by DCRs:

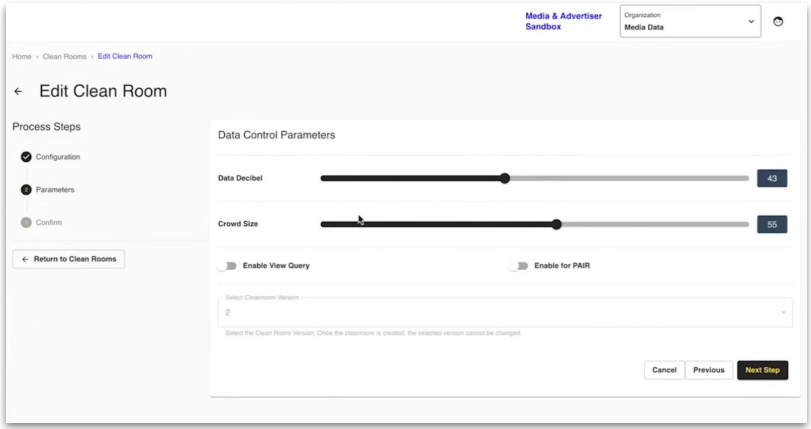
1. **Data Anonymisation** - Removes or alters personal IDs so data cannot be used to re-identify individuals
2. **Data Pseudonymisation** - Replaces personal IDs with fake/pseudo IDs to facilitate analysis without exposing personal identities
3. **Differential Privacy** - Adds noise to the data or query results to prevent identification of individuals from dataset
4. **Homomorphic Encryption** - Technique allowing computations to be performed on encrypted data without needing to decrypt first
5. **Secure Multi-Party Computation (SMPC)** - Allows parties to contribute data to a shared analysis/model without revealing individual datasets
6. **Zero-Knowledge Proofs** - Cryptographic method that is used to verify the accuracy of data or computations without exposing the underlying data

Application of PETs can be calibrated (to a degree)

The degree to which PETs are applied in a given DCR instance can be adjusted based on the requirements of the users and the agreements in place between data-sharing parties.



AWS Data Clean Room



Habu Data Clean Room (on Snowflake)

Data Clean Room categories

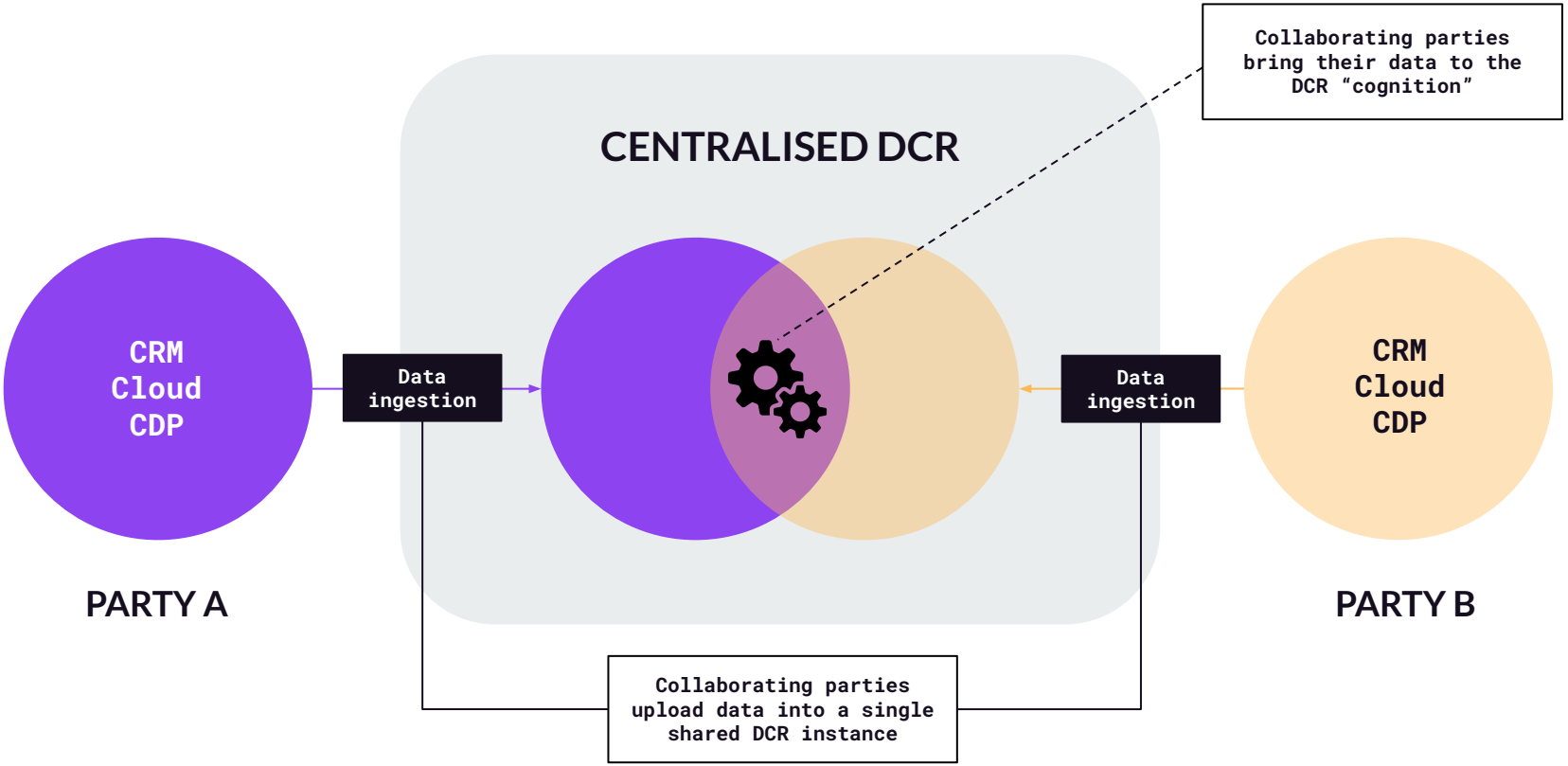
PURE PLAY

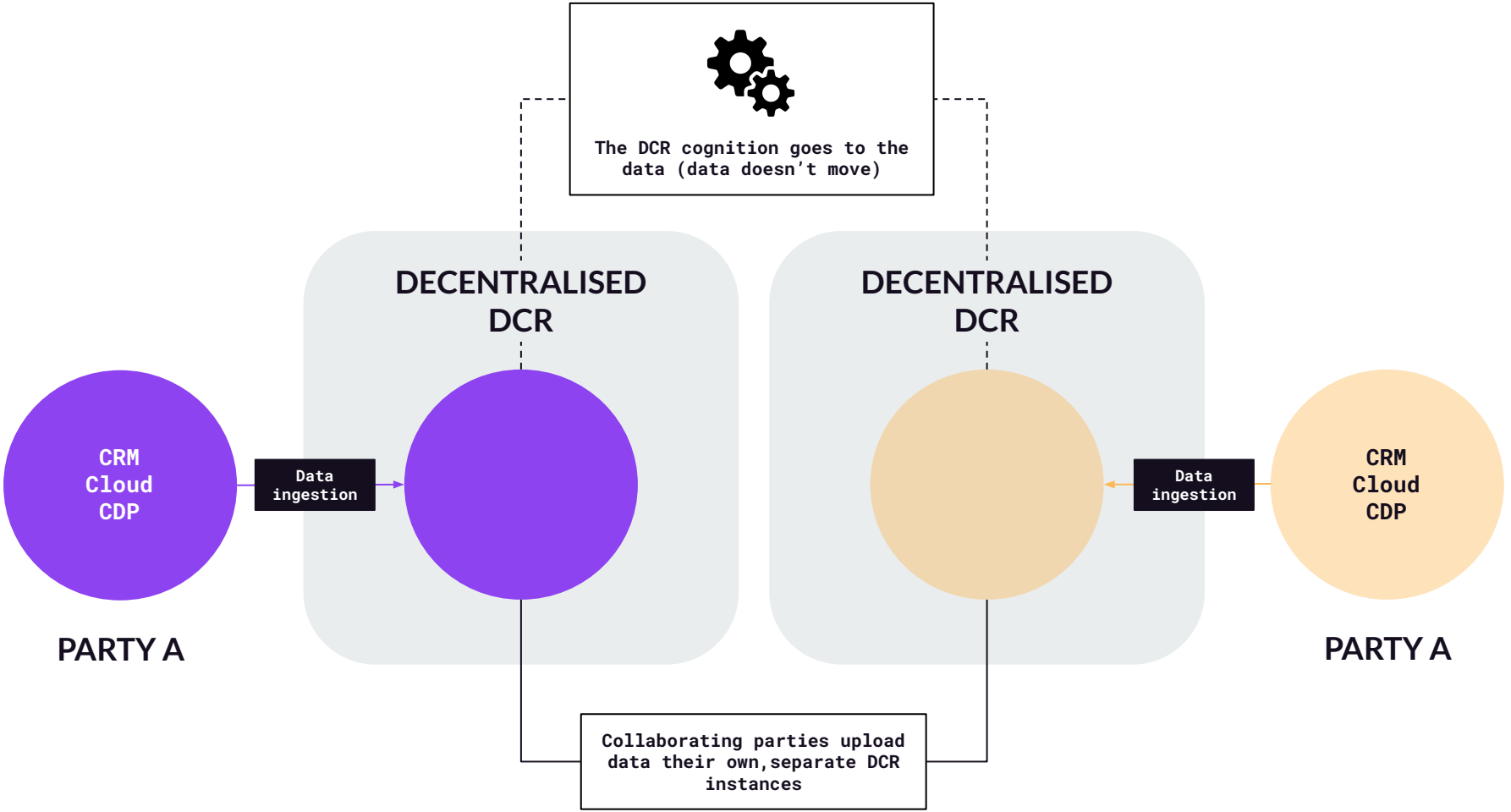
DATA PLATFORMS

MEDIA

Most are built on public cloud infrastructure (which can also be considered another DCR category)

Data Clean Rooms can be centralised or decentralised.

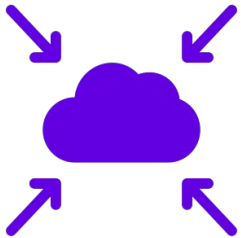

Collaborating parties bring their data to the DCR “cognition”
CENTRALISED DCR
CRM Cloud CDP
Data ingestion
Data ingestion
CRM Cloud CDP
PARTY A
PARTY B
Collaborating parties upload data into a single shared DCR instance


The DCR cognition goes to the data (data doesn't move)
DECENTRALISED DCR
DECENTRALISED DCR
CRM Cloud CDP
Data ingestion
Data ingestion
CRM Cloud CDP
PARTY A
PARTY A
Collaborating parties upload data their own,separate DCR instances

In most cases...

Decentralised

Zero/Minimal
data movement



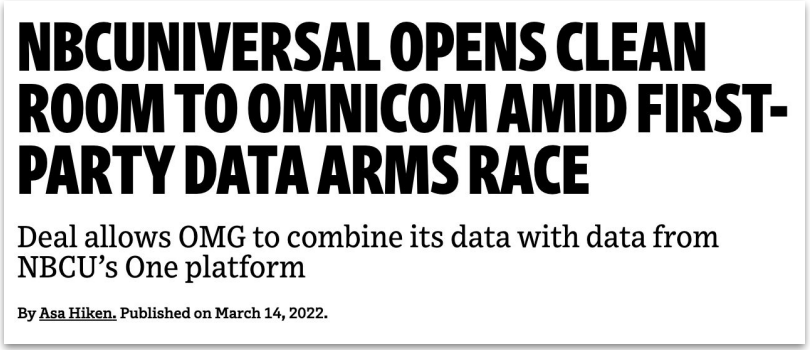
Centralised

Some degree of
data movement

A few more things

New revenue source for media owners

Just like DCRs provide brands with access to execute use cases with new sources of partner data, they also enable new revenue streams for media owners with scaled 1PD (independent of media inventory).



Platform or Product?

The answer is both, but it depends on the provider, the collaborators, and what the DCR is being used for.

- **Pure-play** (Habu, Infosum) - **Platform**
- **Media companies** (Disney, NBCU) - **Platform**
- **Data companies** (LiveRamp, Acxiom) - **Product**
- **Walled gardens** (Google, Amazon Ads) - **Product**
- **Data platforms** (Snowflake, Databricks) - **Product**
- **Managed clouds** (AWS, GCP, Azure) - **Product**

Thanks!

contact@skeletonkey.digital