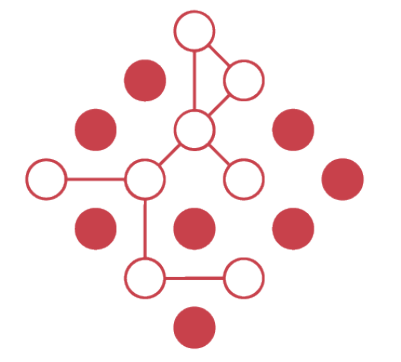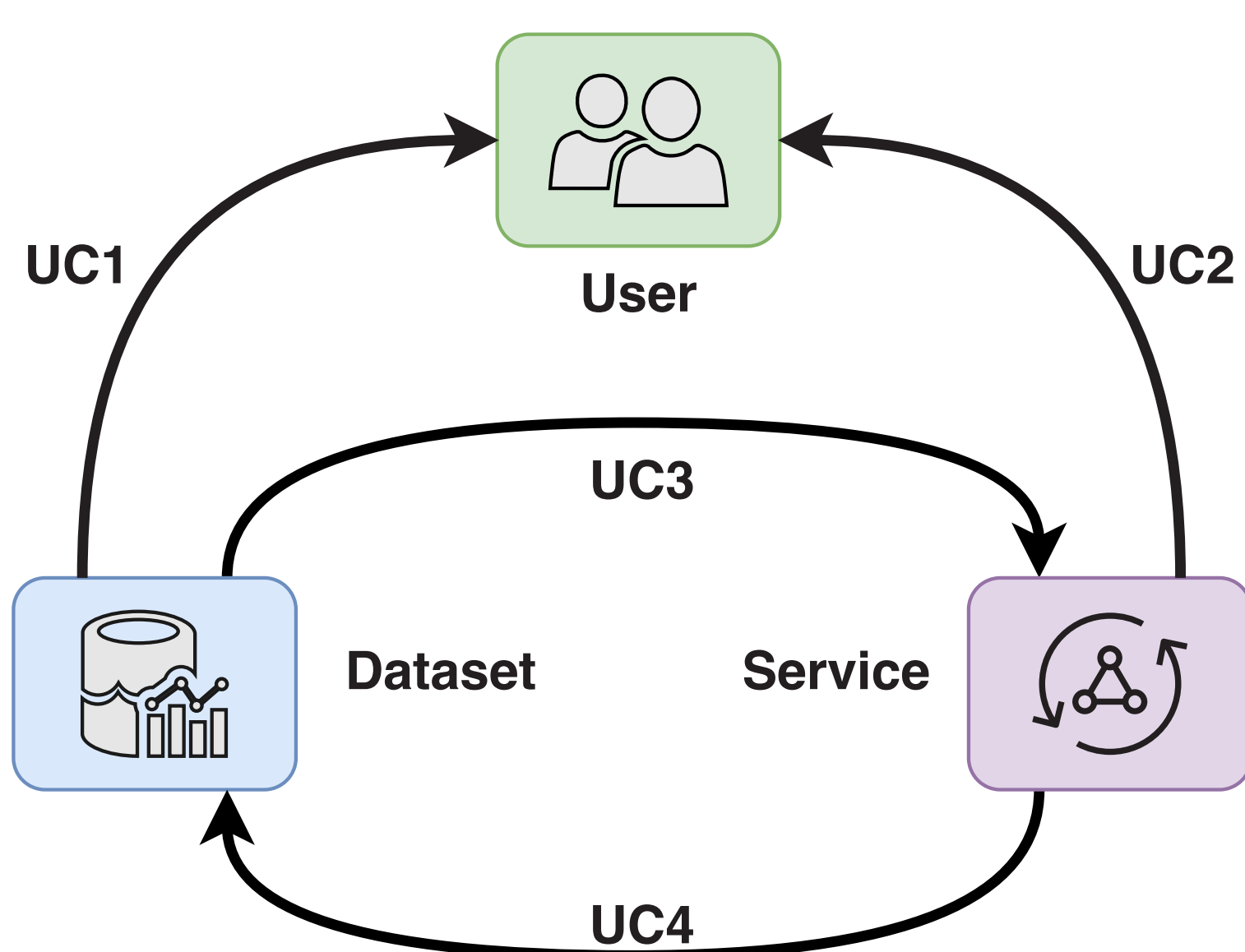# Evaluating Tripartite Recommendations in Data Markets

Dominik Kowald, Matthias Traub, Dieter Theiler, Heimo Gursch, Emanuel Lacic, Stefanie Lindstaedt, Roman Kern, Elisabeth Lex
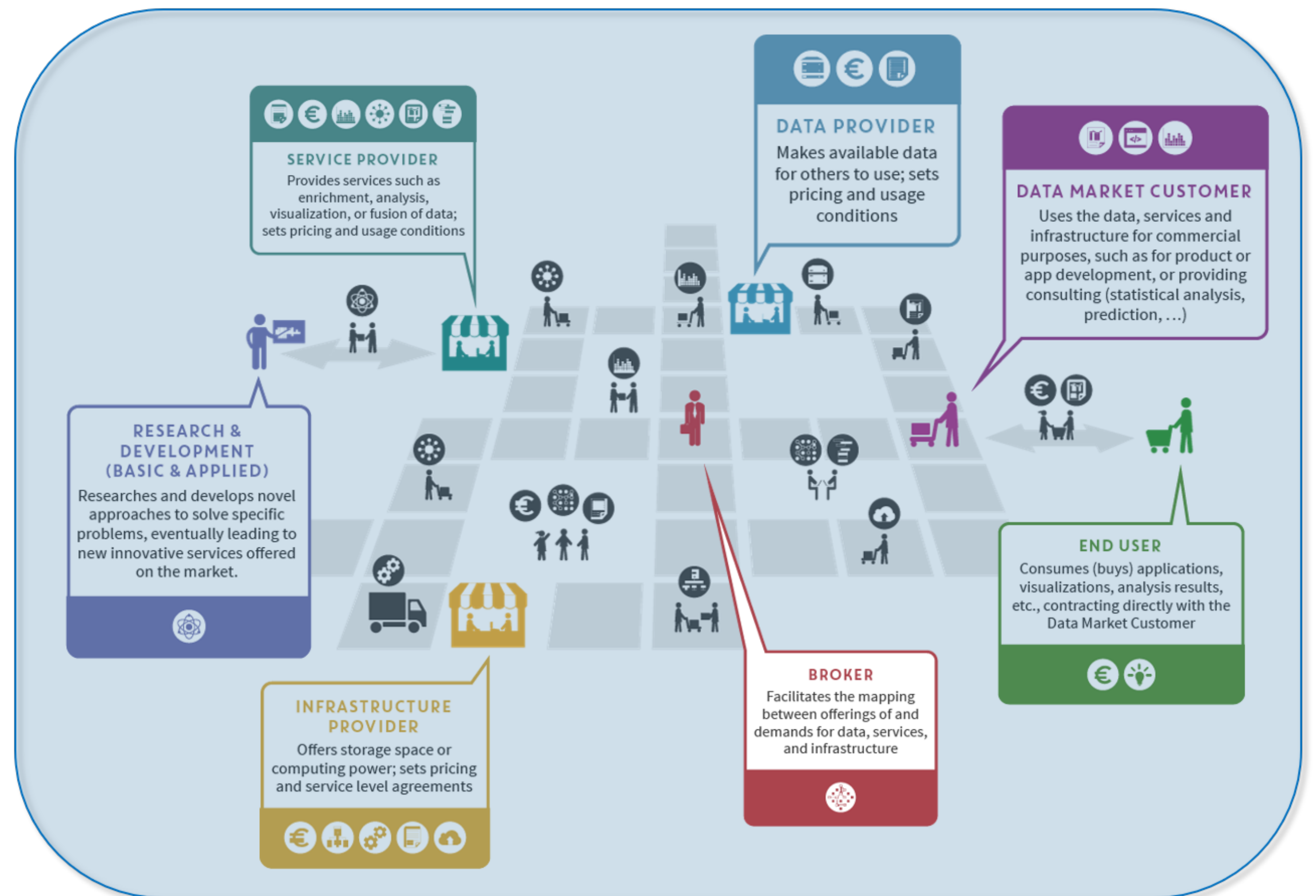Know-Center GmbH & Graz University of Technology (Austria)

Problem

- Most of the research in recommender systems is focused on the **bipartite relationship between users and items**.

- We extend this to the **tripartite relationship between users, datasets and services** that is present in data markets.

- There is **no open dataset and methodology** available to evaluate this kind of recommendations.

Meta Kaggle Dataset

- This dataset allows **simulating a real-world data market** by providing data of users, datasets, and services (i.e., kernels in Kaggle terms).

- To get interactions between datasets and services, we establish a **collaboration network** (i.e., set a link when a user has interacted with both).

| Feature	#
Users	6,108
Datasets	45
Services	3,334
User/dataset interactions	2,962
User/service interactions	18,593
Dataset/service interactions	95,249

Conclusion

- There is no **one-size-fits-all solution for recommendations in data markets**.

- There is a need for **more sophisticated algorithms** that also incorporate other sources, e.g., content-based filtering.

- For future work, we want to validate our evaluation results in an (online) DMA setting and investigate **further use cases for the Meta Kaggle dataset**.

Data Market Austria

The Data Market Austria (DMA) is an initiative (FFG flagship project) to bring together various **stakeholders of Austria's data economy** (https://datamarket.at/):

ScaR Framework

The ScaR framework is a scalable recommendation-as-a-service framework based on microservices. It supports (i) multiple heterogeneous **data sources**, (ii) processing and considering **streaming data** immediately for the recommendation process, (iii) **scaling the recommender system** to be suitable for cloud-based environments, and (iv) combining (near) **real-time recommendation algorithms** with context-dependent data:
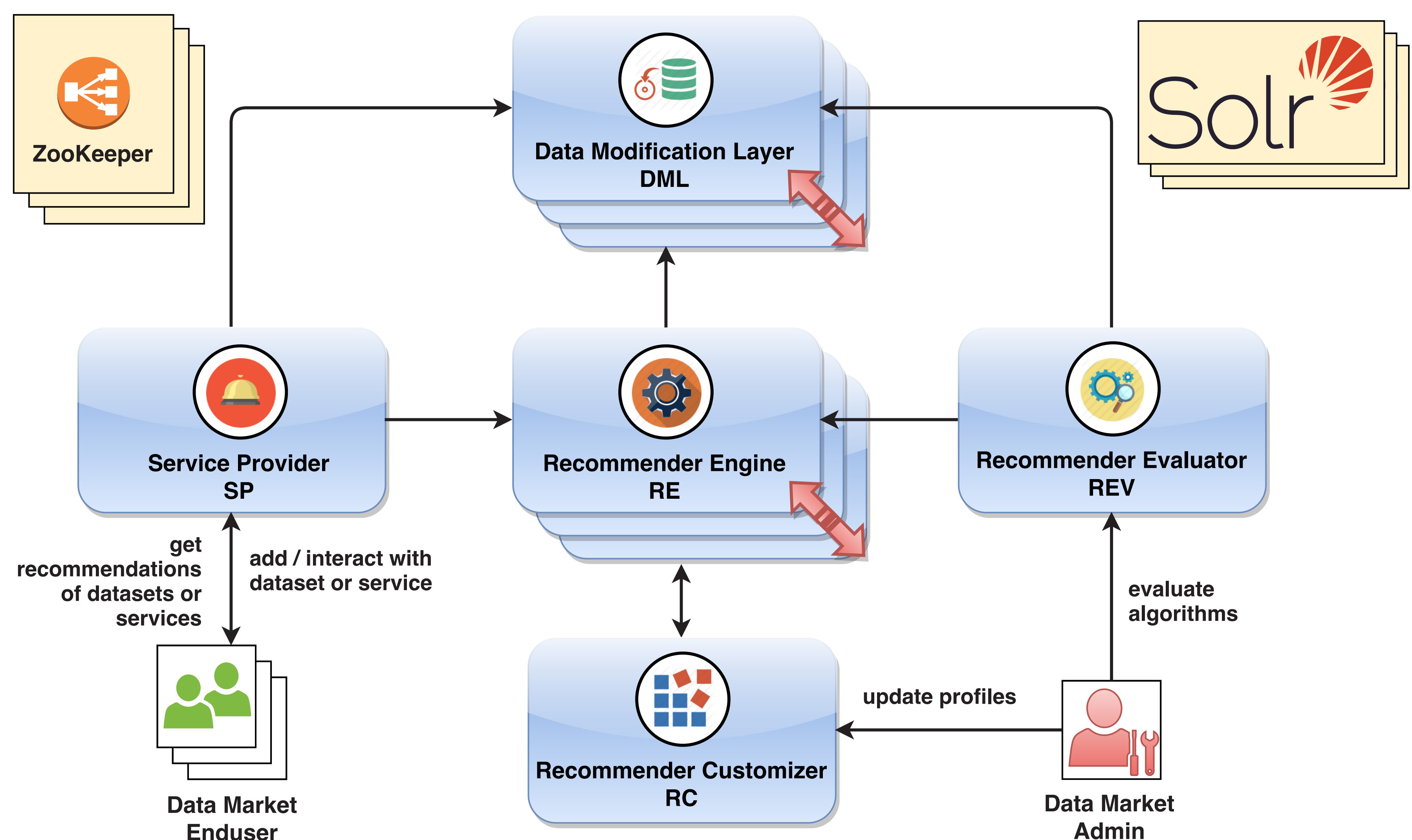
Evaluation

We evaluate a **MostPopular (MP)** and a user-based **Collaborative Filtering (CF)** algorithm for 4 recommendation use cases using the Meta Kaggle dataset:

Algorithm	P@1	F1@5	R@10	MRR@10	MAP@10	nDCG@10
UC1: MP	**0.823**	**0.470**	**0.717**	**0.217**	**0.597**	**0.729**
UC1: CF	0.705	0.431	0.611	0.192	0.484	0.635
UC2: MP	0.103	0.050	0.066	0.023	0.026	0.072
UC2: CF	**0.137**	**0.086**	**0.114**	**0.037**	**0.054**	**0.121**
UC3: MP	1.000	0.411	0.707	0.232	0.580	0.750
UC3: CF	**1.000**	**0.636**	**0.934**	**0.281**	**0.925**	**0.948**
UC4: MP	0.000	0.000	0.000	0.000	0.000	0.000
UC4: CF	**0.022**	**0.006**	**0.006**	**0.003**	**0.004**	**0.009**