
THE DATA CARDS PLAYBOOK • RESOURCE

OFTEn Framework

Participatory activities for purposeful, transparent,
and people-centric dataset documentation

pair-code.github.io/datacardsplaybook ↗

#datacardsplaybook



01 ASK • FRAMEWORK

#datacardsplaybook

OFTEn Framework

What is Knowledge Acquisition?

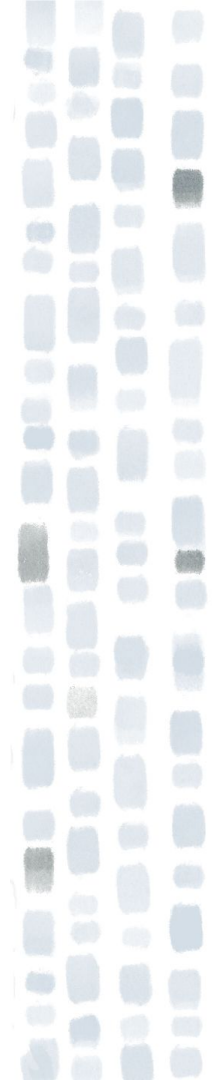
KNOWLEDGE ACQUISITION (KA)

—

is the process of extracting, structuring and organizing knowledge from one source, usually human experts, so it can be used in software such as an Expert Systems (ES).

OFTEn is a knowledge acquisition framework designed to elicit information about any dataset, that is important for transparency in reporting.

OFTEn can be used on any dataset to identify and evaluate content that's important to your agents.



OFTEn represents general stages in a dataset's lifecycle

ORIGINS

Early stages of a dataset's lifecycle when decisions to create a dataset are made

FACTUALS

Actual data collection processes and raw outputs

TRANSFORMATIONS

Raw data is transformed into a usable form through operations like filtering, validating, parsing, formatting, and cleaning

EXPERIENCE

Dataset is tested, benchmarked, or deployed in practice (experimental, production, or research)

n = 1

Actual samples from the dataset - or vignettes - representing normal datapoints and outliers



ORIGINS

Examples of Questions

Various planning activities such as defining requirements, design decisions, collection or sourcing methods, and deciding policies which dictate final outcome.

Who collected it?

Why was it collected?

What is it about?

How, where & when was it collected?

When was it updated? **Who** maintains it?

...

Resulting Themes

Authorship

Motivations

Intended applications

Licenses

Versions

Sources

Collection methods

Errata

Accountable parties

Example

[Credit Card Fraud Detection, Kaggle](#)

Who created this dataset?

Publisher:

Machine Learning Group,
ULC

Publisher Type:

Academic - University
Industry - Financial

FACTUALS

Examples of Questions

Statistical and other factual attributes that describe the dataset, deviations from the original plan, and any pre-wrangling analysis.

How many instances and features?

What are the labels in this dataset? **Who** labelled it?

What is the structure and format?

How should an example be read or interpreted?

Where is the dataset hosted?

...

Resulting Themes

Number of instances

Number of features

Number of labels

Source of labels

Source of data

Breakdown of subgroups

Shape of features

Description of features

Missing or duplicates

Inclusion criterion

Example

[Credit Card Fraud Detection, Kaggle](#)

Who does this dataset represent?

Represented Population:
September 2013 European Cardholders

Geographic Distribution:
Europe-only

TRANSFORMATIONS

—

Examples of Questions

Resulting Themes

Example

Summaries of labeling, annotation, or validation tasks. Inter-rater adjudication processes.

Feature engineering and modifications made to handle privacy, security, or PII.

How was this dataset cleaned or verified?

How was this data annotated, validated, or rated? **Who** did this?

Which features correlate with each other?

What are the known sensitive variables?

What metrics were used to define dataset quality?

...

Rating or annotation

Filtering

Processing

Validation

Statistical properties

Synthetic features

Handling PII

Sensitive variables

Impact on fairness

Skews or biases

[Credit Card Fraud Detection, Kaggle](#)

How was identity handled?

Privacy:

Due to confidentiality issues, background information is not provided.

Feature Description:

All features transformed with PCA; with the exception of “Time” and “Amount.”

EXPERIENCE

—

Examples of Questions

Using the data for specific tasks, undergoing access training, making modifications to suit the task, acquiring results and comparing to other similar datasets, and noting any expected/unexpected behaviors.

Who has used the dataset?

What kinds of tasks has it been used for?

What were the methods, results, or errors?

How should it be used and not be used?

How many times has it been used?

...

Resulting Themes

Intended performance

Alternate applications

Unexpected performance

Caveats

Insights

Experiences

Stories

Use & use case evaluation

Example

[Credit Card Fraud Detection, Kaggle](#)

Who has used this dataset in the past?

Known use cases - industry application:
Unknown

Known use cases - research and academia:
Widely used in academia for sample problems to train fraud detection classifiers

n = 1 (Samples)

—

Examples of Questions

In-and-out-of distribution datapoints, demonstrates noteworthy data points with specific attributes, and where applicable, model outcomes on them.

What does an actual example in this dataset look like?

What does a prototypical example look like?

Why is it typical or atypical?

What is the expected result of a task on this data point?

Resulting Themes

Examples or links to typical examples and outliers

Examples that yield false positives or false negatives

Examples that demonstrate handling of null or zero feature values

Example

[Credit Card Fraud Detection, Kaggle](#)

Time: 0

V1: -1.3598071336738

V3: -0.0727811733098497

V4: 2.53634673796914

V5: 1.37815522427443

...

Amount: 149.62

Class: 0

	Who	What	When	Where	Why	How
<p>ORIGINS</p> <p>–</p>	Who is the publisher of the dataset?	<p>What challenges should the dataset have addressed?</p> <p>What are the incentives for data professions, like labellers, practitioners, researchers?</p>	When was this dataset created? Launched?	Where did the funding come from?	Why was this dataset created? What was the process prior?	<p>How were the methods decided?</p> <p>How many parties were involved?</p>
<p>FACTUALS</p> <p>–</p>	Who is the data about? Who is responsible for the data? Who are the labellers?	What are the subgroups in the data that can affect outcomes?	What time period does the data represent? When does the data perform abnormally?	Where can the dataset be accessed? Where was the data collected or model created?	Why were the reported metrics chosen?	How many unique labels exist in the dataset? How were these generated?
<p>TRANSFORMATIONS</p> <p>–</p>	How was PII handled in this dataset?	<p>What methods were used to clean or verify this dataset?</p> <p>What categories are labelled in all?</p>	When should features be engineered and how? Do these need to be updated?	<p>Do location features correlate with any other sensitive features?</p> <p>Where is the data safe to use?</p>	Why were the chosen transformations applied to the dataset?	How are biases handled in the data?
<p>EXPERIENCE</p> <p>–</p>	<p>Can outcomes from this dataset be used to identify individuals?</p> <p>Who can use this dataset? Are there any trainings required?</p>	<p>In practice, what tasks has data been used for?</p> <p>What were the methods, results, or errors discovered when the dataset was used?</p>	Under what circumstances or use cases, and when would this dataset not be effective?	<p>Where in the world is this dataset accessible?</p> <p>Where has it been used?</p>	Why is the expected representation of the dataset different from the observed representation?	<p>In practice, how many times has this dataset been used?</p> <p>How expensive is the data in various parts of the world?</p>
<p>n = 1 (Samples)</p> <p>–</p>	<p>Is the datapoint typical or atypical?</p> <p>How do models behave here?</p>	<p>What is the size of the datapoint?</p> <p>What's the consent, redaction, and withdrawal process to intervene in a datapoint?</p>	When does the outcome on a datapoint change? Show examples through counterfactuals?	<p>What factors are baked into the datapoint?</p> <p>What are the risks involved if things go wrong with predictions?</p>	<p>Why is this image datapoint cropped a certain way?</p> <p>Why are certain categories not populated in this datapoint?</p>	<p>How does this datapoint relate to a real-world input?</p> <p>How does the outcome relate to a real-world output?</p>

Try it yourself

Frame a question using OFTE n.

Rows represent life cycle stages - O,F,T,E, and n.
Columns represent prompts for questions.
Frame a question about “Factuals,” starting with “Where.”

	Who	What	When	WHERE	Why	How
ORIGINS	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]
FACTUALS	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]
TRANSFORMATIONS	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]
EXPERIENCE	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]
N(1) EXAMPLE	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]





The [Data Cards Playbook](#) [↗] by [Google Research](#) [↗] is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. You are free to share and adapt this work under the [appropriate license terms](#) [↗].