DOING GOOD WITH DATA
A Maturity Model for Evaluating Data Literacy Competencies in Non-Governmental Organizations

The presented data literacy maturity model (DLMM) is based on a master's thesis research project that contributes to the growing scholarly interest in exploring and describing data literacy. The study was set out to identify and outline data literacy competencies for non-governmental organizations (NGOs) in an initial exploration of representing the topic in a maturity model. The suggested DLMM aims at opening the discussion on data literacy and to stimulate increased awareness of what can be understood and expected from it. Throughout the study, action design research has been used to illuminate the question of how data literacy can be described in a maturity model for NGOs throughout three development iteration phases. The study draws on data from data practitioners of the Datenschule, an educational program of the Open Knowledge Foundation Germany that offers practical data training, as well as a partnering NGO. The analysis yielded a data literacy maturity model that evolved to describe eleven competencies to asses in organizations. The main contribution of the study is a data literacy maturity grid that is complemented by a self-assessment excel tool. The results reveal that a maturity model for describing data literacy is a helpful tool to raise awareness and educate on the topic, but should not be interpreted as an absolute evaluation tool. The model rather offers option values for better evaluating and planning data practices.

THE
DATA LITERACY MATURITY MODEL

Background

We live in times in which massive amounts of data are collected every day. Data streams emerge from various new sources, such as mobile phones, credit cards, televisions and city infrastructure. Within the course of the last years, data literacy turned out to be a meaningful priority to different groups. The emerging importance of data journalism or analytical skills on the modern job market show that we have to equip ourselves with new skills for the data era. While data practitioners from sciences and business are progressing in the world of data, social change organizations are still lagging behind when it comes to integrating data into their activities, despite the potential of using data to solve wicked social problems. To close this gap, we developed the data literacy maturity model.

.

Purpose

The data literacy maturity model will help you to better understand the required skills that are needed to kick-off your data projects, identify strengths and gaps and thus will empower you to plan your future data practice in accordance with your goals. It should more precisely help:

o To reflect on, analyze and record your existing competencies or the skills, abilities and knowledge connected to data handling skills.

o To better understand the type of behaviors that might be expected when working with data. With practical advice and suggestions on how you might continue to develop your data practice.

o To set out a personal development plan to identify the activities you feel are important to develop your data - practice.
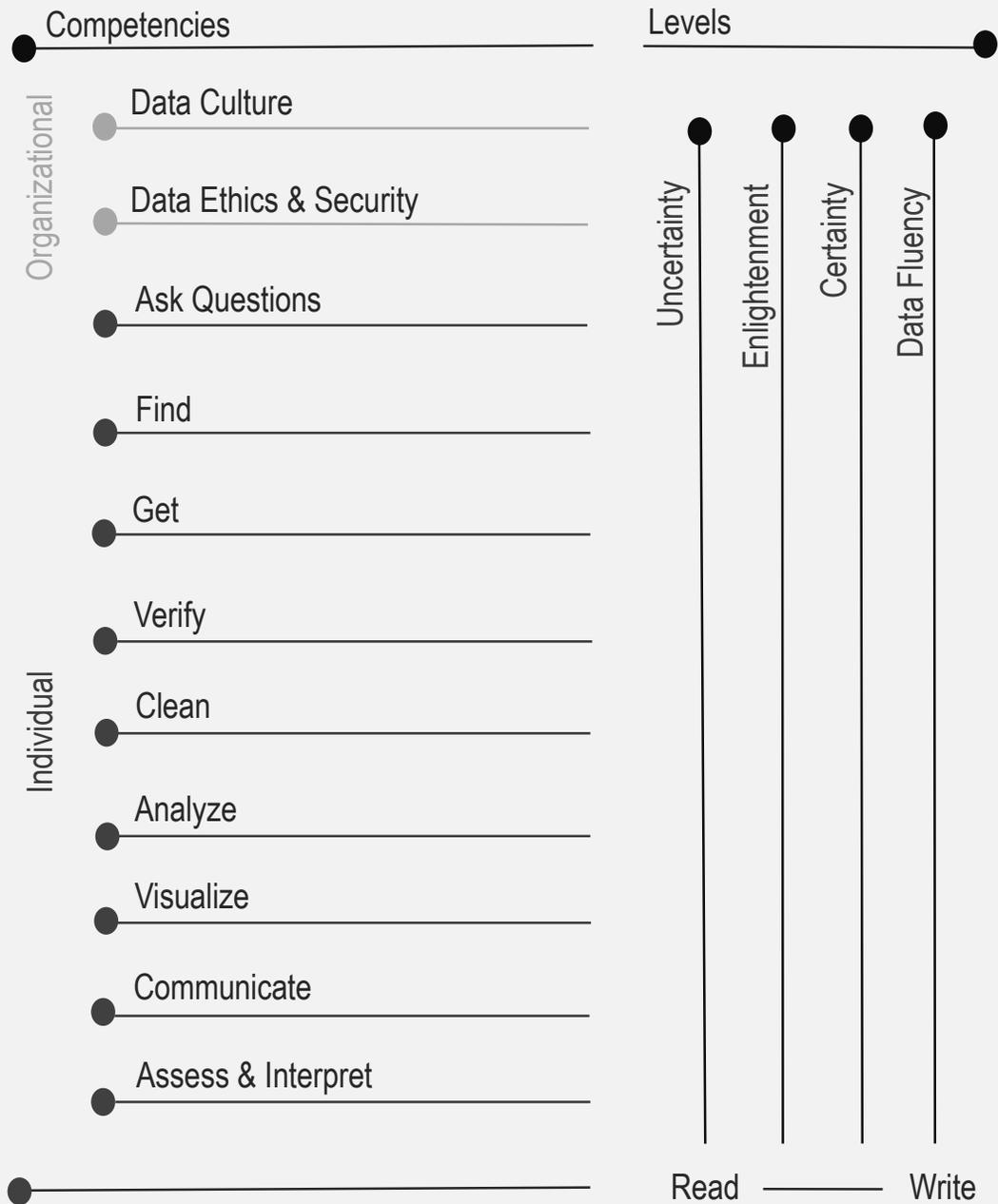
Who should use this model and when might it be used?

The model will be useful for individuals and organizations that work digitally no matter whether you are a newcomer to working with data or are already experienced. Of course, considering the individual and organizational context is key during the evaluation. It will be of most value when it is used as part of an ongoing reflective process building on your personal experiences in an organizational context.

Some ideas about how and when to use the framework include:

o Using it to value existing competencies and to make decisions on training and development needs.

o Supporting individuals and organizations at the start of a data assignment or project to understand the competencies required, to identify their personal starting points and as a reflection and development tool.

o Taking individual competencies and using them as a starting point for a data mentoring partnership.

o Using it as a planning and reflection tool for individual or collaborative projects with other data practitioners as part of feedback sessions with partners and practitioners when the reflection of others can be used alongside personal reflections.

MODEL OVERVIEW

Competencies

Organizational

- Data Culture
- Data Ethics & Security

Individual

- Ask Questions
- Find
- Get
- Verify
- Clean
- Analyze
- Visualize
- Communicate
- Assess & Interpret

Levels

- Uncertainty
- Enlightenment
- Certainty
- Data Fluency

Read ——— Write

MATURITY MODEL LEVEL DESCRIPTIONS

Organizations are unaware of the need for data literacy skills and have no or very vague understanding on what is required. Individuals might have a certain interest in data and work digitally, but are unsure about the different steps that exist when working with data.

UNCERTAINTY

Organizations are experimenting with the application of data-related topics. Describes a state where a lot about data has already been understood theoretically, but cannot be applied in many cases and has to be further trained.

ENLIGHTENMENT

Organizations perform data handling steps with confidence and have built data-driven activities into their routine processes wherever it makes sense. Generic procedures and standards on how to handle data are formalized and widespread. Benefits are understood at all levels of the organization.

CERTAINTY

Organizations have established a data-informed culture throughout all levels. Data is actively used to improve processes and create workflows.

DATA FLUENCY

DATA CULTURE

Data is perceived as an ambiguous term which causes insecurities.

Data is perceived as an interesting concept and benefits are appreciated. Insecurities exist regarding use cases and what exactly to expect.

Data is not perceived as a source of insecurities, but rather understood as an enabler for progress and support for existing and planned activities. Higher management and leaders support data initiatives.

Psychological barriers of data have been brought down (e.g. insecurities, fear, resignation) and comfort around data is promoted. Higher management and project managers understand and support importance of dedicated resources (time, budget, human resources) for data handling and conversion.

UNCERTAINTY

ENLIGHTENMENT

CERTAINTY

DATA FLUENCY

DATA ETHICS & SECURITY

No awareness for guidelines that ensure confidentiality, integrity and availability of data.

Rising awareness and uncoordinated attempts to promote the importance of the responsible use of data. No defined guidelines.

Awareness of the impacts of data use. Guidelines for responsible data handling are defined and incorporated internally to activities.

Processes are in place to ensure confidentiality, integrity, and availability of data. Only data that is necessary is collected/used. Consistent, companywide policies for secure and ethically sound data handling are constantly redefined and updated.

UNCERTAINTY

ENLIGHTENMENT

CERTAINTY

DATA FLUENCY

ASK QUESTIONS TO DATA

Lacking ability to formulate questions to find meaningful answers in data. No feeling about which questions can be answered by data.

UNCERTAINTY

Questions can be asked to data in limited number of situations and answers are provided through simple queries.

ENLIGHTENMENT

Questions to data are formulated precisely and target-oriented to find meaningful answers in most of the cases.

CERTAINTY

Entire projects are based on multidimensional questions. Answers to informational needs can be found consistently in data, because of the high awareness of what questions can be answered by data (no overinterpretation).

DATA FLUENCY

FIND DATA

Limited understanding of possible data sources. Use of basic search engines to find data. No experience for identifying and selecting most relevant data sources.

UNCERTAINTY

Knowledge only limited to a few data sources. Advanced use of search engines, use of internal data sources and data requests at public institutions are common practice.

ENLIGHTENMENT

Broad understanding of different data sources, most relevant ones can be chosen from a selection of data sources. Awareness and use of data portals for specific topics.

CERTAINTY

Profound understanding of the various possible types of data sources. Assessment criteria for selecting the ones most relevant to an informational need are formulated. Ability to detect when a given problem or need cannot be solved with the existing data, and knowledge about research techniques to obtain new data (e.g. complex queries).

DATA FLUENCY

GET DATA

Data is derived from full text and used as base for further processing.

UNCERTAINTY

Use of downloads and data formats such as .csv . Often use of internal programs to access data (e.g. CRM).

ENLIGHTENMENT

Data can be accessed using more complex data formats (e.g. JSON, XML). Use of APIs to get data.

CERTAINTY

Access to data through sophisticated methods (e.g. automated data scrapers / scripts). Ability to convert input format into a form that can be used for further processing and analysis.

DATA FLUENCY

VERIFY DATA

Ability to do data quality assessment independently. Data evaluation criteria regarding authorship, method of obtaining and analyzing data, comparability and quality are precisely defined.

Multiple layers of data checking are implemented in standard procedures.

Critical check of simple data quality measures.

Critical evaluation of data does not exist, data is taken at face value. Data evaluation criteria cannot be described.

UNCERTAINTY

ENLIGHTENMENT

CERTAINTY

DATA FLUENCY

CLEAN DATA

No awareness that given data might have to be checked, cleaned or normalized. Data is further processed as is.

UNCERTAINTY

Awareness that given data most often is not perfect. Awareness of some data quality criteria (e.g. empty fields, duplicates) and manual fixing of errors.

ENLIGHTENMENT

Invalid records can be detected and are removed using programs that support data cleaning (e.g. OpenRefine). High awareness of data quality criteria (e.g. machine processable, empty fields, duplicate detection).

CERTAINTY

Independent ability to remove invalid records and translating all the columns to use a sane set of values through an automated script. Ability to combine different datasets into a single table, remove duplicate entries or apply any number of other normalizations.

DATA FLUENCY

ANALYZE DATA

Full suite of machine learning tools (e.g. clustering, forecasting, boosting, ensemble learning).

Ability to work with advanced statistics (e.g. inferential view of data, linear regression, decision trees).

Ability to work with basic descriptive statistics. Pivot tables for aggregating information, histograms and boxplots.

Bar and pie charts, simple use of data tables and basic summaries of data.

UNCERTAINTY

ENLIGHTENMENT

CERTAINTY

DATA FLUENCY

VISUALIZE DATA

No awareness of the multiplicity of how data can be presented. No understanding of when standard visualizations are chosen, decision based on what looks best (trial and error).

UNCERTAINTY

Ability to find specific outputs in accordance with information that want to be represented (e.g. in Excel).

ENLIGHTENMENT

Creation of interactive charts / dashboards, uncertainties are always visualized along with the data.

CERTAINTY

High awareness of the various forms in which data can be presented (written, numerical or graphic). Sophisticated visualizations are programed, linked, dynamic dashboards that anticipate user requests are designed.

DATA FLUENCY

COMMUNICATE DATA

Insights from data are not communicated or put into a broader context.

UNCERTAINTY

Limited ability to find specific outputs. Simple narrative support static visualizations / key numbers (e.g. reporting to funding partners, newsletters).

ENLIGHTENMENT

Own projects are supported by interactive visualizations and more sophisticated narrative in a broader context. (e.g. data storytelling, conferences, talks, monthly updates, blog posts).

CERTAINTY

Ability to synthesize and communicate in ways suited to the nature of the data, their purpose and the audience (e.g data storytelling, data-driven campaigning, workshops, conferences, monthly updates, blog posts, reproducible research).

DATA FLUENCY

ASSESS & INTERPRET DATA

Data outputs are used at face value without questioning their correctness and message.

UNCERTAINTY

Growing awareness for critically assessing data outputs and interpreting the results. Insecurities regarding what exactly to pay attention to.

ENLIGHTENMENT

Data outputs and results are interpreted confidently and critically. Evaluation criteria are internalized.

CERTAINTY

Data outputs and results are consistently questioned and challenged, interpretation extents the obvious and information are successfully translated into actionable knowledge.

DATA FLUENCY

CONTACT

Helena Sternkopf
sternkopf.h@gmail.com