MBA Entrepreneurship
Vilnius University Business School

MBA ENTREPRENEURSHIP PROGRAMME

Giedrius Vėgelė

THE REPORT ON THE EXPERIENTIAL ENTREPRENEURSHIP PROJECT

WHATSOEVER-COMMERCIALIZATION OF ENVIRONMENTAL AND BUSINESS OPEN DATA THROUGH A DATA PLATFORM

Advisor Jorge Fernando Pereira Campos

Submitted on 22/05/29

Word count: 3966

Vilnius

Table of Contents

INTRODUCTION:

Data is the main source of decision-making in this age. Many companies are looking into new revenue opportunities to monetize data. "Industry leaders across healthcare, financial services, retail, manufacturing, and many more have been prioritizing data as an organizational asset and have been thinking about the right quality data, at the right level, to the right people, at the right time, for the right decisions." - Forbes, 2021

Open data is data that anyone can access, use, and share. Open data becomes usable when made available in a common, machine-readable format. Open data must be licensed. Its licenses must allow people to use the data in any way they want, including transforming, combining, and sharing it with others, even commercially.

Examples of open data include the following:

- Government financial data.
- The stock information you see scrolling on your newsfeed.
- Market data statistics that you can access via a search engine.
- Published academic research.
- Kindergarten placement data for the city municipality.
- Municipality traffic accident statistics.

A data platform is an architecture in which data processing and data transformations are enabled. According to the Database company MongoDB, " A data platform is an integrated set of technologies that collectively meets an organization's end-to-end data needs. It enables the acquisition, storage, preparation, delivery, and governance of your data, as well as a security layer for users and applications. "

Opendata commercialization can be defined as taking existing open data obtained from the environment, business operations and turning it into a new revenue stream. Briefly, this should enable the data-driven ecosystem to supply economic value not just for data companies and their clients, but also for end-users.

So the whole idea of Whatsoever project is to deliver open data datasets that could turn into

revenue streams. In other words, this should be an analytics platform that combines open data networks with business information where you "insert" the metadata at one end and get the new datasets at the other end and this allows decision-makers or researchers to have insights for decision making and these could be bought on the data marketplace.

PROBLEM/OPPORTUNITY

Technology has allowed us to capture data in ways and volumes that many of us never imagined. Increasingly sophisticated data analytics tools allow us to parse data in new ways to discover trends and findings that have shaped crucial business decisions. These tools allow us to **see connections between data** that have never been explored. Data is nothing compared to answered questions for businesses like in [Figure 1].

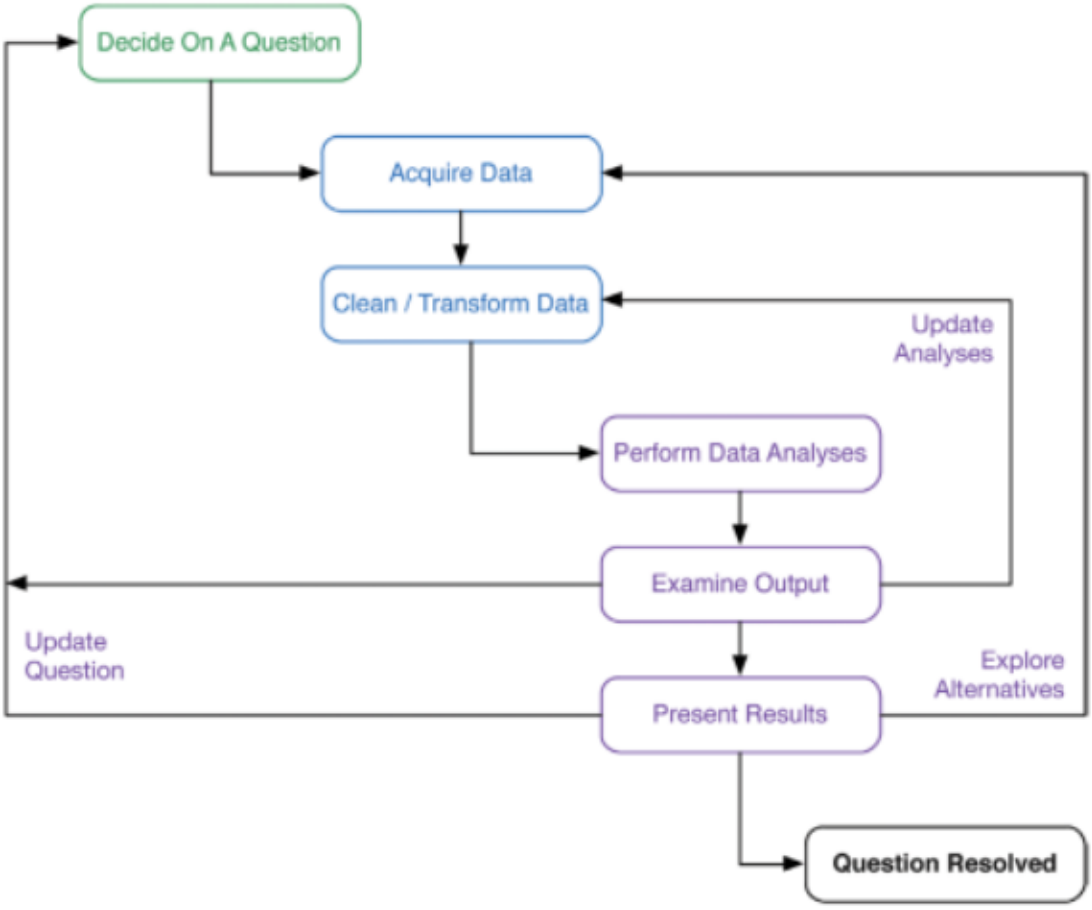

Figure 1. Decision Making Data workflow

We have been looking into the commercialization of open data for some time already as there

is a large gap in the usability of this data. There is a growing debate about who should control this data; the corporations and governments who collect it, or the public that benefits from it? Opendata is not always machine-readable, easy to access, does not always come in a good, standardized way, and licensing is not clear.

Using the "Whatsoever" application we would let B2X users get this open data easily by just submitting metadata of the needed dataset and getting the prepared dataset as expected [Figure 2].

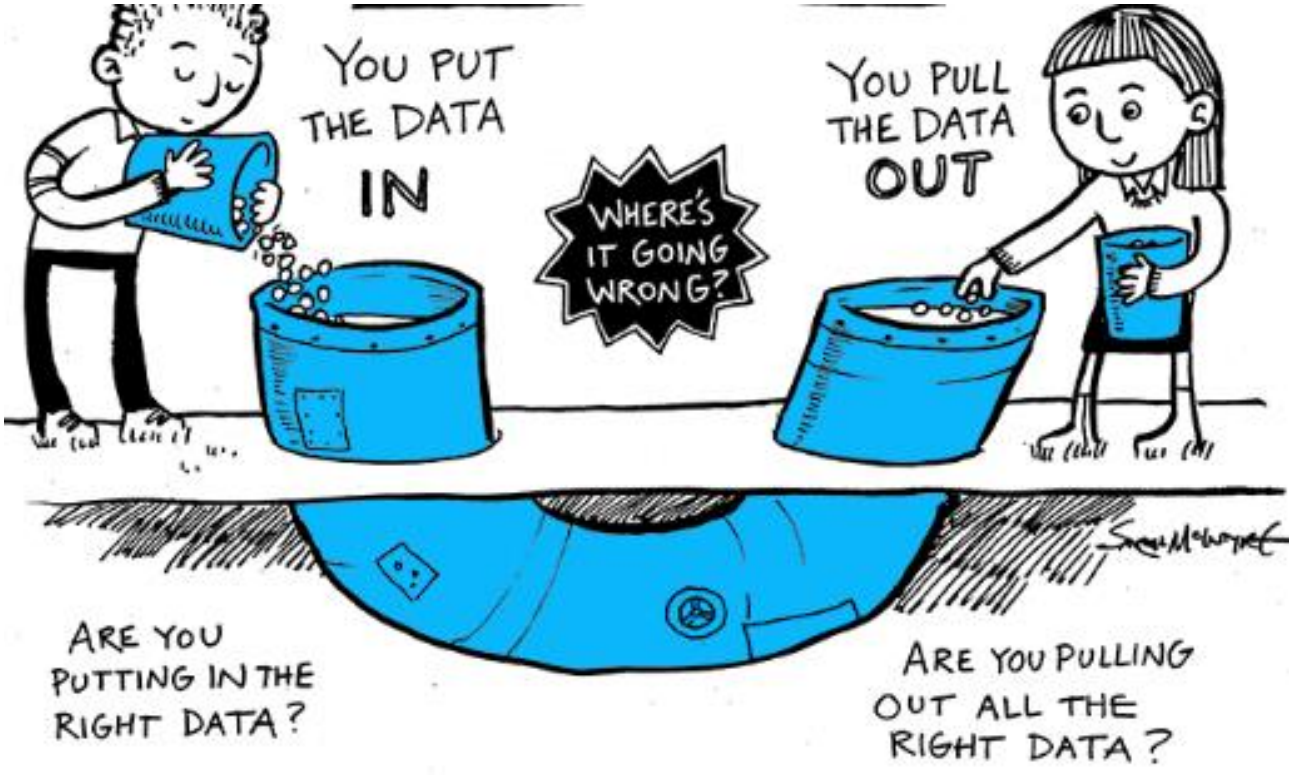

Figure 2. "Data warriors" caricatures

Market opportunities are exceptionally large. According to the "data.Europe.EU" in 2016 direct market size of Open Data was expected to be 55.3 bn EUR for the EU 28. Between 2016 and 2020, the market size increases by 36.9%, to a value of 75.7 bn EUR 2020, including inflation corrections. For the period 2016-2020, the cumulative direct market size is estimated at 325 bn EUR. By the end of 2021, it is expected to have 100,000 Open Data jobs. The accumulated cost savings for the EU27+UK in 2021 are forecasted to equal 1.7 bn EUR.

According to "digital-strategy.ec.europa.eu" the public sector information (PSI) value is expected to increase from a baseline of €52 billion in 2018 for the EU27 + UK to €194 billion in

2030.

Open Data results in efficiency gains as real-time data is used that enables easy access to information that improves individual decision-making. For example, Open Data has the potential of saving 7000 thousand lives a year by supplying resuscitation earlier. Furthermore, applying Open Data in traffic can save 629 million hours of unnecessary waiting time on the roads in the EU.

Total 1 306 513 open datasets are available from EU27+, 168 catalogs, and an enormous amount of pre-processed data.

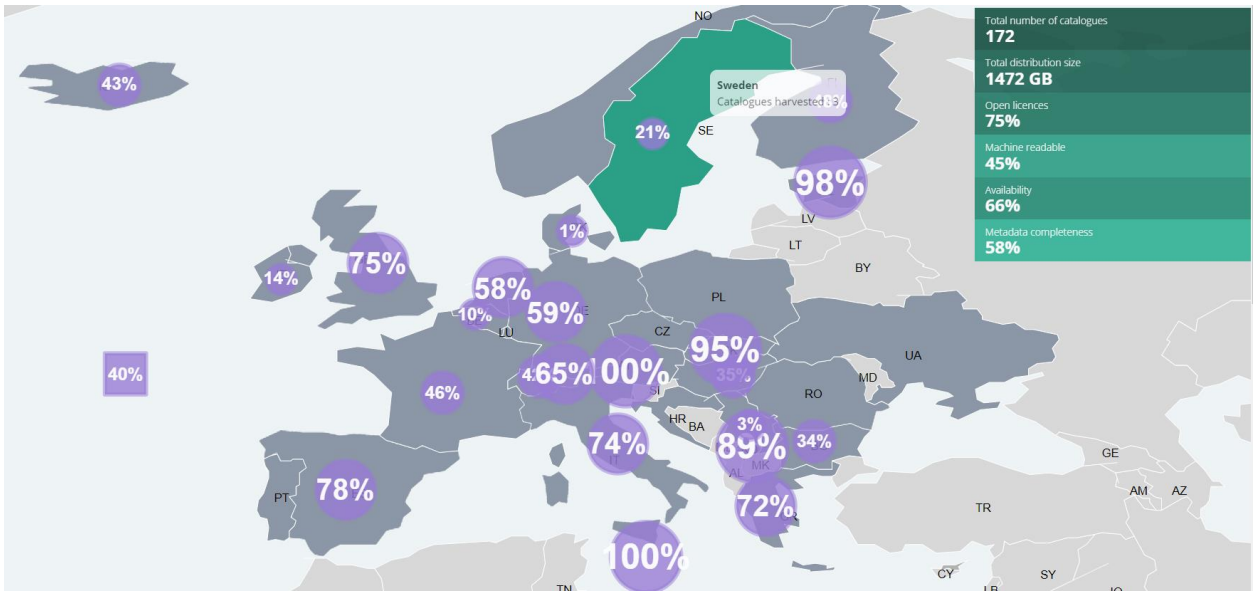

Figure 3. Open data completeness by Country in Europe

Summary: there is a real year-by-year growth of 30% of general data value. There is at least 30% year by year growth of data engineers. 95% of businesses have issues with unstructured data management, so they are not even thinking of including open data in their assets and this is where we can help them to make the process easier. 79 zettabytes of data generated in 2021 and 180 zettabytes of data are expected to be generated in 2025, so this is a 25% yearly growth in quantity and the value would be growing an additional 5% yearly. Due to poor data quality, the US economy loses approximately $3.1 trillion annually.

Said all this there is a huge market opportunity, increased demand, and poor current solutions to manage all the needed open data and handle this with ease.

VALUE PROPOSITION

This research process is carried out in five phases, the overview is shown in Figure 4. We have found the way more personas, use cases, pain points, and revenue streams when we display in this report due to the limitation of words.

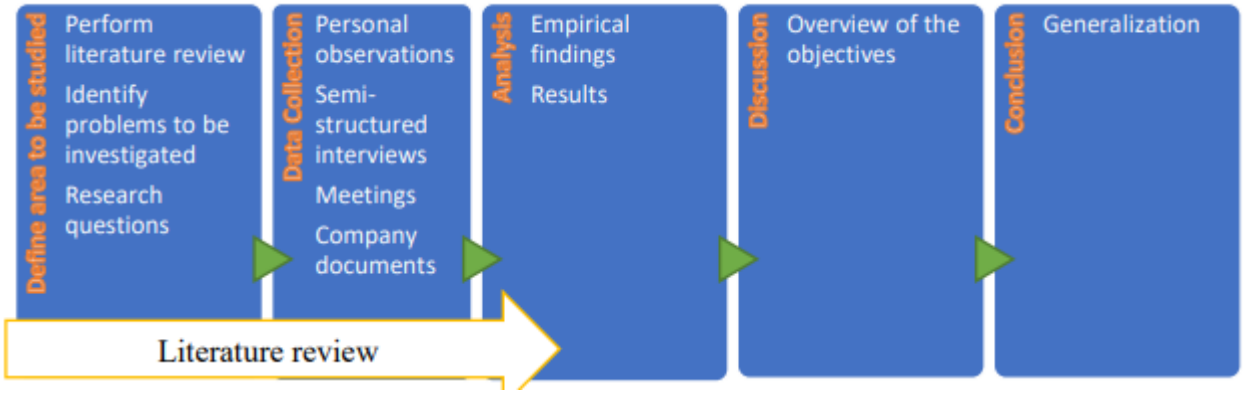

Figure 4. Overview of the discovery process

We there keen to understand different solutions that can support the "Whatsoever" vision and have been looking to create an interactive demo prototype for open data in support of the Value proposition with an assessment of our current thinking with recommendations that will help us understand how this new open data availability approach will significantly improve the customer experience that would be willing to pay the price. This Prototype and recommendations would then be used to support a wider development pitch for investment to implement this strategy.

To achieve this, we have taken a "Persona" driven approach to understand how our customers currently consume our "Whatsoever platform."

Customers of "Whatsoever" fall broadly into the following high-level personas:

- **Non-technical decision-making person**- additional open data sets would be interesting if these would come without any burden, no time would be wasted on the development of data, and the data acquisition process would be easy.
- **Researchers-Advisors**- who cannot use open data easily, there are large "size" requirements as of day "0" for datasets, and high CPU processing power is needed. The datasets they require must be affordable and the quality should be good as well so that they could immediately consume the open data in their research.
- **Data Engineers**- who can do an additional enrichment to their prepared datasets with open

data, the only thing they need is a painless process of getting the open data datasets at their end, so a programmatic interface(API) to fetch the data and straightforward way to pay for the fetched lines.

- **CTOs\CIOs**- add additional data points to the decision making, open data addition to their firm datasets could do a breakthrough in answering the questions that have not been touched or engaged before. Adding additional areas of open data might be a clever way to catalyze the revenue streams for CTOs and CIOs.

GO-TO-MARKET PLAN AND VALIDATION

METHOD- Data Collection

Our analysis is based on qualitative research focusing on value propositions at several companies and individuals who are potential customers of the platform. The participating companies aimed to stand for a balance between business-to-business and business-to-consumer business models. Not all companies and individuals that we initially contacted were prepared to grant the level of details that there required by me. The investigated companies include (a) an insurance company, (b) an insurance broker, (c) a Scandinavian bank (d) freelancers working as data engineers (e) a governmental statistics agency (f) a municipality that has no platform for open data (not Vilnius) (g) journalist-researcher. The companies and governmental firms vary in size between 10-10000 employees.

After the initial introductions and discussions on purpose with our contact person (or firm representative), We have always focused on potential use cases and value that could be brought in trying to check the validity of a) platform functions b) open data types c) commercialization and potential pricing model d) usage patterns e) customer potential f) scaling of the users g) MVP needs(interfaces used, best personas in the firm, technology used, etc.). By doing this we have collected data by interviewing and observing the members of firms, freelancers, and researchers. Focusing on particular Topics and questions enabled me to generate thick descriptions of the value propositions and to collect naturally occurring data using the language and concepts of interviewees. We have conducted several rounds of interviews for each value proposition with different personas.

We have predominately used the Value Proposition Canvas as developed by Alexander Osterwalder and distributed via Strategizer to focus on the core value proposition contained within a Business Model Canvas. The model focuses on the value proposition of customer segment. Fully understanding the products/services, gain creators, and pain relievers in comparison or mapping to

customer gains, customer pains, and customer jobs/tasks. Customer discovery starts with understanding customers' pain points, and these are reflected in an MVP value proposition. Defining and prioritizing personas was an essential part of customer discovery also. Interviewing was a critical aspect of discovery and journey mapping layouts in the entire process and helped turn discovery into action.

Process for crafting a value proposition used

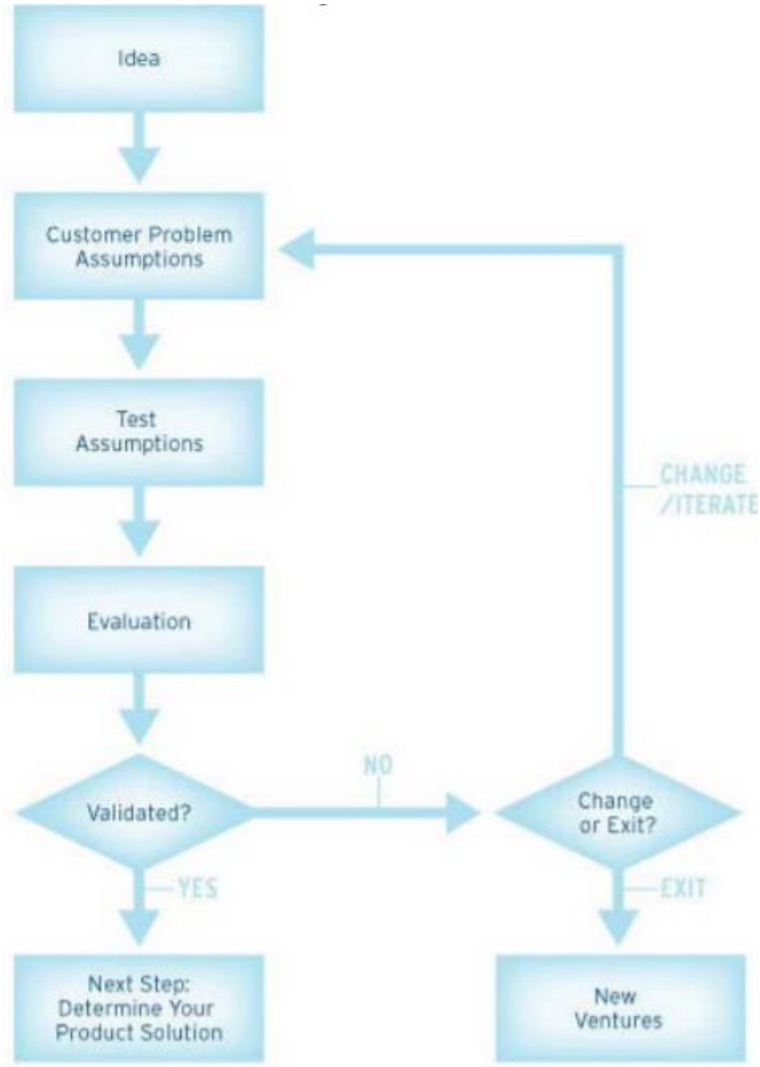

Figure 5 Value verification process

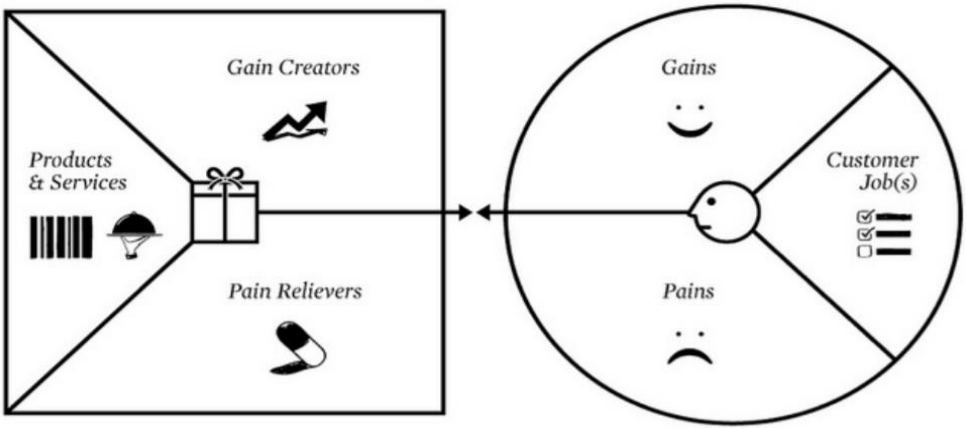

Figure 6 The Value Proposition Canvas (Osterwalder et al., 2014)

Assumptions approach

- Brainstorming on target customers done, mainly the questions on what problems there about to be solved with the Whatsoever there drafted
- Important or urgent problems rethought several times and revised and still being investigated
- How to solve one or another problem, what means are needed, provisional investment sizes thought through
- What are the current competitors and how do they go about the problem-solving?
- How unique my platform is and how it is different in problem-solving

Validation of assumption done by using

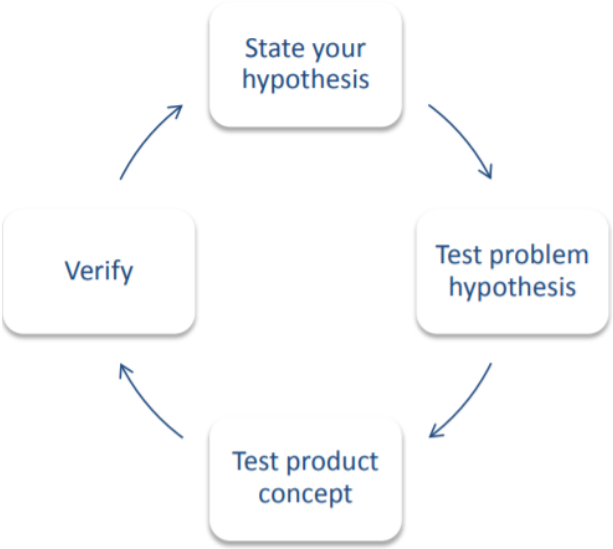

Figure 7 Blank, Steve Four Steps to Epiphany flow for Customer Discovery

Revenue model

"Whatsoever" is a B2X company that sells services to other businesses or individuals (researchers, students, clerks, etc.) on either pay-for-use or freemium models. Subscription per open data lines model would be used to secure initial revenue from B2B early customers.

Markets

First target markets are seen as UK, Sweden, and Estonia because of the availability of open data sets and fulfillment of open data.

BUSINESS MODEL

So first things first we will have a marketplace for selling open data. A marketplace is where two or more types of users meet to transact. The marketplace would derive its revenue through what is called a marketplace rake, typically a percentage of transacted revenue on the marketplace. Instead of a percentage rake, we might want to charge fixed transactional fees for the data. We would try various marketplace fee models to get the targeted paying groups happy as per the above value proposition.

In addition to a marketplace we will have the Data as a service business model which would entail delivering prepared data to the customers, typically data to SMB and/or enterprise clients, and charging a subscription (e.g. pay per data lines or pay peruse in our case) for the data. We might think of including service contracts also as there is potential in that for deriving revenue from custom support or integration. We thought of having a freemium model for personal use (e.g. researchers, scientists, data engineers in our case), wherein entry-level plans are free(e.g. 200 lines of each dataset are free, or metadata is partly free), and more robust plans are paid (e.g. full data access pay per lines of data).

To get all mechanics out of the way we would have an API business model also to serve the data and monetize datasets programmatically without any human intervention. This is similar to Data as a service and specifically pay per data lines or pay per use model.

Summary: We are a data startup where we deliver value from collecting, extracting, cleaning, reformatting, transforming, analyzing, and selling open data. There would be at least two new business models or revenue streams when we would establish well with the current business lines and processes which would include public data scraping and analytical dashboards on top of any dataset

we produce.

COMPETITOR ANALYSIS

We have been trying to identify direct and indirect competitors when looking at the commercial market and NGOs. It came out that there are three pilar groups in the field we should be caring about:

NGOs

- https://www.dataportal.se/en

  In mid-2012, The Ministry of Enterprise and Innovation within the government of Sweden orders the Swedish Innovation Systems Agency, Vinnova, to develop a technical platform called “oppnadata.se” or the dissemination of data made available for reuse. The purpose of “oppnadata.se” is to promote the reuse of available data in such a way to support the development of e-services made by citizens, companies, and other users in focus. Particular attention should be paid to how public information can be facilitated for citizens, businesses, and other actors with efforts considering privacy and other security aspects.

  Umeå School of Business in cooperation with the Swedish Innovation Systems Agency established Project Ladds - Lab for the Data-Driven Society- The lab will act as a platform and an incubator for innovations where companies, different organizations, and individuals can use or upload data hence encouraging new ways of co-developing products or services, by experimenting with data, capturing new data, analyzing relationships between texts and images to help find patterns and deviations and commercialize innovations. The lab is, however, financed through public funds provided by Vinnova to enhance sustainability and create public value as value capturing. (Biedenbach & Bostrom, 2018). This is an indirect competitor where we could use them as our partners in the future.

- https://www.trafiklab.se/

  Trafiklab was founded as a collaboration between different Swedish entities including Samtrafiken, SL and RISE. The aim is to link and create interconnections between Sweden´s public transportation companies. Trafiklab is a community for open traffic data where developers and stakeholders easily access data and APIs for public transport, to develop new services. Through the APIs, Trafiklab offers access to all of Sweden's public data in real-time and the necessary tools to create and share new products and services. Trafiklab also

offers to help to connect designers and developers to co-develop new products and services. It is a platform that collects everything from experimental test projects to apps and other activities to improve the platform further. (Trafiklab, n.d.) This is one more good example of tailored open data usage, we will consider partnering up with Trafiklab later in the process to get already prepared data sets and APIs for reselling.

Commercial institutions:

- https://dev.socrata.com/ and https://www.tylertech.com/

  Socrata platform- Tyler's Data & Insights enables governments to use data as a strategic asset in the design, management, and delivery of programs. Data flows easily between staff and departments leading to more efficient programs and better decision making. This is a direct competitor of ours that has an Open data catalog, data APIs, Metadata, Pipelines management, and metadata management.

- https://openinsurance.io/

  Open insurance framework and tech connecting different data exchanges between different insurers including usage of open data in the insurance processing.

- Any customer720 platform

  TATA, Cognizant, and Accenture do have such platforms or sales where 360 come internally from processing the data on your customer and another +360 come from either internet (LinkedIn, Facebook, Twitter,) or open data providers. Two business models are possible with such platforms where data would be acquired from us, or we provide such a platform ourselves.

People

- https://www1.nyc.gov/site/opportunity/poverty-in-nyc/poverty-in-nyc.page

  NYC Opportunity is committed to the use of data and evidence in formulating poverty reduction policies. Within NYC Opportunity, the Poverty Research Team generates the alternative poverty measure for New York City, a keystone in innovative, rigorous data analysis. This is an indirect competitor through a use case that could be used elsewhere for commercial purposes. We would like to absorb such use cases going mature with our product.

  There are a bunch of other niche businesses built around specific public data sets (e.g., SEC Edgar, patent filings, etc.).

Summary: As we are looking into ENVIRONMENTAL AND BUSINESS OPEN DATA

commercialization as milestone 1, we do not have so many direct and indirect competitors and most probably we would use some of the indirect competitors as our partners for data provisioning, data fulfillment, metadata acquisition, API capabilities, etc.

PROJECT TEAM

Operations

What we want to achieve initially is to be a) AGILE AND FLEXIBLE TO MEET DEMANDS b) DRIVE EFFICIENCIES AND CLOSE COLLABORATION IN OUR SMALL TEAM c) GET COST BENEFITS IN OUR PRODUCT/PLATFORM

To support this we will have a FLAT structure at the start of the company with some governance and steering boards where decisions on different subjects would need to be taken.

"Whatsoever" is a B2X company that sells services to other businesses or individuals (researchers, students, clerks, etc.) on either pay-for-use or freemium models.

As this is a digital company all the marketing and sales would need to be automated in an "effective way," to save money and effort and get the most benefits out of this. So in a nutshell marketing and sales automation tools would be used for this. In parallel, we would do many specific showcases with different firms to support the management and to have really good showcases on what you can do on the platform and open data.

Accounting and billing would be automated from the early beginning as there is not much of the budget assigned to this, so only automation can cope with this without human intervention.

People culture and management function would be with CEO while the company trues up. Such things as KPIs/WIGs/MbO/OKR would be left for the people concerning their performance and to the CEO concerning the product.

First target markets are seen as UK, Sweden, and Estonia because of the availability of open data sets. There might be some deviations due to the governance of the data and ES policies vs UK policies, but this is a job to be done.

Summary: In a nutshell, this is a continuous job where we will be just starting and looking out how to true this up on the go.

Core Team

- CEO/COO in one person for 3 years while the platform is built, and the business model stood up
- VP Sales and Marketing in one person for the first 3 years while the business model is stood up. Potentially product features would be driven by the VP of sales and Marketing as the first milestone.
- CTO/Tech lead would be the main person to drive the technological landscape and agenda.
- 3xDevelopers where some of them would need to have either DevOps or data engineer capabilities while we build out the whole solution. (There is an option to hire a development service company with the needed expertise and working in this field and make the whole stand up a bit more rapid. Cost wise this should be similar.)

All of these would be the essential key people for starting up the platform.

FINANCIAL PROJECTION AND KEY METRICS

Sales: it is expected to onboard up to 100 users during the first year, an additional 500 in the 2nd year of operations, and another 2400 in the third year. Each user's monthly cost is 25EUR or a yearly cost of 300EUR.

Financial analysis shows that we will be having -457738EUR loss for the first year, -420200 loss for the second year, and during the 3rd year we will be equal to zero with a potential reinvestment of 140kEUR into further marketing activities at the end of year 3.

Below tables reflect the financial figures explained.

Income Statement Summary

	Year 1	Year 2	Year 3
Revenue	20000	150000	700000
Cost of Goods Sold	3000	5000	10000
Gross profit	17000	145000	690000
Total expenses	495200	590800	818150
Income before Tax	-458363	-420200	47985
Income Tax	0	0	0
Net Income	-458363	-420200	47985

Balance Sheet Summary

	Year 1	Year 2	Year 3
Assets	113200	48800	64400
Liabilities	-283900	-701963	-876 363
Equity	-458363	-916727	-1008741

Start-up costs

Table describing cost incurring during the first 6 months, so this includes all the costs associated with starting the business.

Startup	Actual	Budgeted
Accounting Services		**7200**
Advertising and Promotion for opening		**0**
Architectural Design		**0**
Cash		**0**
Decorating		**0**
Deposits for Utilities		**0**
Equipment		**7000**
Expected Taxes		**3000**
Headhunting or other Hiring Costs		**0**
Installation of Equipment		**0**
Insurance		**0**
Legal Costs		**0**
Licenses and permits		**0**
Moving		**0**
Office Supplies		**0**
Print Design		**0**
Printing		**0**
Remodeling, Buildout		**0**
Rent Deposits		**3000**
Salaries		**160000**
Signs		**0**
Software		**2000**
Starting Inventory		**10000**
Unanticipated Expenses		**6800**
Vehicles		**0**
Website		**2000**
Other		**600**
Total Start-up Costs		**201600**

Break-Even Analysis

Knowing how much we need to charge to make a profit is making all the difference in our business's potential growth and scalability. A detailed break-even analysis prevents us from accidentally underestimating costs and ending up in a financial black hole.

There are some caveats around break event analysis such as the market as a whole which can be unpredictable due to some circumstances or there might be some discounts that have been issued by the sales and this was not taken into the trajectory. Seasonal aspect e.g. students using the platform more when it is time for their exams.

Break-even Analysis Summary	
	Amount
Variable Cost per Unit	**3**
Fixed Cost (Total)	**654150**
Expected Sales (in Units)	**29734 (monthly subscription unit)**
Price (per Unit)	**25**
Total Revenue	**743352**
Total Variable Costs	**89202**
Profit	**89 202**
Profit by Unit Sold	**3**

The Break-even quantity is 2477 monthly subscriptions assuming that these all are signed for a calendar year, so 12 months

It is expected that we will mitigate any risks for selling this number of subscriptions. This number does not analyze the demand.

Exit Strategy

The plan is to true up the platform and exit through a merger or acquisition with a data company.

CONCLUSIONS AND RECOMMENDATIONS

I would finish this paper with Risks and opportunities rather when recommendations as being an entrepreneur are risk-taking and management of risk rather than recommending something

- Wide functionality platform needs to be built, so this would be a resource-hungry project/business
- This means a "fat" investment needs to be done to "quick-start" immediately

- Complicated use cases which start deviating based on people skillset (seniority should be driving this in anger)
- GDPR and other compliance figures would need to be taken care of to de-risk the data usage and commercialization
- Ability to get funding to expand the business as desired initially might be an issue
- Genuine feedback loop might be an issue because of the size of the platform
- No regulations for open data yet
- No direct competitors, competitors covering only specific areas and some markets only
- There is an opportunity to bring the platform further by adding closed data analysis going forward and doing a mix of open and closed data sources for enterprises or some specific areas where needed, so further commercialization models are possible
- Good developers are becoming like unicorns, so maybe some other markets for remote development capability might be considered and, in this case, the costs could go down for the developer services
- Enterprises are struggling with their data model so there is a good opportunity to expand this platform for closed data usage
- Size of the market is enormous and keeps growing by 30% in size and people skills yearly and the opportunity is huge here!

LITERATURE

Webpage on Fundersclub (Section Business model**)**

No Author. No date. Startup business models overview. https://fundersclub.com/learn/startup-business-models/st/startup-business-models-overview/

Journal of Cleaner Production. (Section Business model, Value proposition**)**

N.M.P.Bocken. (2013) Volume 65.

https://doi.org/10.1016/j.jclepro.2013.11.039

Journal of Production Economics (Section Bussiness model)

C. Gimenez, V. Sierra, and J. Rodon. "Sustainable operations: Their impact on the triple bottom line,." Int. J. Prod. Econ., vol. 140, no. 1, pp. 149–159, 2012.

https://doi.org/10.1016/j.ijpe.2012.01.035

Webpage on www.medium.com (Section Validation)

Hypothesis Driven Design https://medium.com/@furthermore_ux/hypothesis-driven-design-9f78db4dbe91

Webpage on www.slideshare.net **(**Goto market and Value proposition**)**

Design Thinking: Prototyping & Testing. Sankarshan Durgaprasa.

https://www.slideshare.net/SankarshanDurgaprasa/design-thinking-prototyping-testing

Blog Post on www.creately.com **(**Goto market and Value proposition**)**

No Author. (2022). The Easy Guide to the Business Model Canvas

https://creately.com/blog/diagrams/business-model-canvas-explained/

Journal of the Academy of Marketing Science (Introduction, Value proposition**)**

Alam, I. (2002) An Exploratory Investigation of User Involvement in New Service Development. Journal of the Academy of Marketing Science, 30(3), pp. 250–261.

https://link.springer.com/article/10.1177/0092070302303006

Webpage on https://data.gov.lt/ (Introduction, Problem)

Open data in Lithuania and use cases
https://data.gov.lt/usecases/examples

Webpage on https://data.europa.eu/en (Introduction, Problem, Market)
Opendata European Union data availability, use cases, policies
https://data.europa.eu/sites/default/files/report/data.europa.eu_Report_Citizen-generateddataondata_europa_eu.pdf

Webpage on https://www.dataportal.se/en (Introduction, Problem, Market)
Opendata Sweden data availability, use cases, policies, data provisioning, APIs, metadata
https://www.dataportal.se/en/specifications?p=1&q=&s=2&t=20&f=&rt=spec_standard%24spec_profile&c=false

Webpage on https://www.rik.ee/en/open-data (Introduction, Problem, Market)
Opendata Estonia data availability, policies, and licensing
https://creativecommons.org/licenses/by-sa/3.0/legalcode

Webpage on https://www.strategyzer.com/
Osterwalder, A. and Pigneur, Y. (2012) Business Model Generation, Self-Published.

Webpage on https://www.strategyzer.com/
Osterwalder, A. and Pigneur, Y. (2014) Value Proposition canvas.

ANNEXURES

1. Description of the various customers situation addressed in literature

Customers situation	Description	Example	Author(s)
Getting the job done	Customers who needs help to get their job done	Rolls Royce airlines	(Osterwalder and Pigneur, 2012); (Johnson, Christensen and Kagermann, 2008); (Rintamäki, Kuusela and Mitronen, 2007)
Design	Customers who are more into design	Fashion and consumer electronics	(Osterwalder and Pigneur, 2012)
Brand/ Status	Customers value is defined through the brand they use	Rolex	(Osterwalder and Pigneur, 2012); (Rintamäki, Kuusela and Mitronen, 2007); (Skålén et al., 2014)
Price	Customers value is defined through low prices	Southwest, easy Jet, and Ryanair, Nano	(Osterwalder and Pigneur, 2012); (Rintamäki, Kuusela and Mitronen, 2007)
Cost reduction	Customers value is defined through helping them in reducing the costs they spent	Salesforce.com	(Osterwalder and Pigneur, 2012)
Risk reduction	Customers value is defined through the assessing level of risk a product or service has.	Car buyer	(Osterwalder and Pigneur, 2012)
Accessibility	Customers value is defined through making the usage of products or services to many.	NetJets, mutual Funds	(Osterwalder and Pigneur, 2012)
Convenience/ usability	Customers value is defined through making the things convenient to use.	iPod and iTunes	(Osterwalder and Pigneur, 2012)

2. The classic view of the value proposition process (Source: Barnes, Blake and Pinder, 2009)