

# **D4.4 DaaS platform and Experimentation Facilities**

**Coordinator: Luis-Daniel Ibáñez (Soton)**

**With contributions from: Fabrizio Orlandi  
(Fraunhofer)**

Reviewer:

Orsola De Marco (ODI)

Deliverable Nature	R
Dissemination Level	PU
Contractual Delivery Date	31/07/2017
Actual Delivery Date	01/08/2017
Version	03
Total Number of Pages	20
Keywords	SME Services

Executive Summary	3
1. Introduction	4
2. Infrastructure and tools offered	5
3. Technical support report	7
3.1 Technical support procedure and use case	10
3.2 User research for open data tooling	11
Survey	12
Responses	12
Consuming Data	12
Publishing Data	12
Reporting	12
Followup	13
4. Open data repositories and unlocked datasets	13
5. Developers reached	16
6. Conclusions	17
Appendix: Questionnaire for Technical Support	18
<b>Infrastructure</b>	18
<b>Software</b>	18
<b>Open Data Provision</b>	19

# Executive Summary

As part of their offer to selected SMEs, the Open Data Incubator for Europe provided infrastructure, software and data services, to help them overcome technical challenges in their road towards a successful open-data centred business idea.

This document updates the description of the infrastructure and services offered by ODINE with respect to the first version delivered at the first quarter of the project. It also reports statistics about the infrastructure and tools used by the companies and how they profited from ODINE's offer. It also reports on what aspects of Open Data (provenance, licensing, accuracy, timeliness, ease of access, format, documentation, support) could be improved in the opinion of a sample of incubated companies. Finally, it also provides an estimation of the number of developers reached via Open Data powered APIs developed by ODINE-funded companies.

The information contained in this document is of interest for organisations interested in running incubation programs including technical and/or infrastructure support for its participants. It is also of interest for Open Data publishers that want to improve the way their data is consumed by industrial/commercial re-users.

The summary of the findings is:

- Most companies have already chosen a technological infrastructure and have a set of preferred tools to use. Their need from additional help from ODINE was limited.
- Commercial cloud providers are the preferred choice for those in need of Cloud infrastructure. Amazon Web Service is the most prevalent choice.
- From the infrastructural offer, the only one that was taken advantage by several companies (seven) was our agreement with Amazon on free Web Service credits.
- Tool offers were almost always ignored. Technical support requests were more for consultancy to solve a problem, rather than for a specific tool.
- The most important data quality feature for companies is accuracy.
- From a data-access perspective, several companies expressed their wishes for a JSON API to access data.
- We estimate 1177 developers using APIs developed thanks to ODINE funding.

# 1. Introduction

One aspect of the services that ODINE offers to SMEs in its programme is the provision of infrastructure to support the lifecycle of Open Data. This includes their unlocking, their storage and processing, and the generation and publication of further Open Data when applicable. As such, in Work Package 4, ODINE four options to support them:

1. Reaching out to the consortium network to help SMEs unlock datasets. On top of this, the partners pro-actively engaged industrial and commercial actors to unlock datasets that we considered of high impact.
2. Advising SMEs on how to publish and generate open data and standards, thus, contributing to the wider Open Data community.
3. Offering a hosting and computing infrastructure, including technical support, for SMEs that required it.
4. Developing open source tools for data processing developed by members of the consortium or where they had extensive expertise.

Version 1 of this deliverable described the infrastructure provided during the first half of the project. The cloud resources provided through the ODINE consortium during that time were based on FIWARE Lab. Additionally, we offered packs of credits in commercial cloud platforms, through agreements with third parties. However, during the execution of the project we found that SMEs preferred in most cases to have control of their own infrastructure, instead of the Fi-Ware platform; that resulted in a lack of demand in the above, thus the eventual departure of Telefónica from the consortium. Version 1 also described the list of tools we offered, which we maintained available until the end of the project.

In this updated version, we cover:

1. The infrastructure offered by the consortium in the second phase of the project, with metrics about its adoption rate by the SMEs.
2. The methodology used to carry out a study on the infrastructure and tools used by SMEs via a questionnaire the SMEs completed at the start of their acceleration,
3. A list of datasets unlocked thanks to ODINE's involvement, and the Open Data and standards published by SMEs participating in the ODINE programme.

## 2. Infrastructure and tools offered

During the first half of the project (up to cohort 4), ODINE provided, through Telefónica, access and support to FIWARE Lab. FIWARE Lab is a cloud hosting environment built on top of the Open Stack platform which, in combination with FIWARE Cloud Generic Enablers provides the following major features:

- Infrastructure as a Service, as Virtual Machines deployable and manageable through a web portal and accessible through SSH and other services. A set of VM images with pre-installed open source operating systems (e.g. Ubuntu, CentOS) and software components (e.g. FIWARE GEs and ODINEs supported tools) are provided for easy deployment.
- Security aspects, as integrated authentication based on OAuth 2.0., Security SSH certificates to access VMs and security groups to manage firewall rules to control access to VMs.
- Object Storage service.
- Networking capabilities, and IP access to resources.
- Software deployment tools (known as blueprints), to configure and one- click -deployment of software architecture designs (bundles of tools) in various layers.
- Global (shared) instances of tools that can be used for tools testing and prototyping. A global instance of the CKAN open data portal is also provided, can be used for accessing existing data sources and publishing new output open data.

However, during that time, only one incubated SME used FIWARE Lab. After Telefónica left the consortium, ODINE stopped offering FIWARE Lab support.

Since then, we updated the offer of deals with commercial infrastructure and service providers. We describe them below.



### **Sendgrid Accelerate**

SendGrid has a cloud based email platform that can deliver marketing and transactional emails for your company, moving over 15 billion emails a month, for more than 180,000 companies worldwide.

SendGrid has a startup program called SendGrid Accelerate. The Goal of SendGrid Accelerate is to help SMEs grow your business, by providing mentorship, strategic introductions and our scalable email platform for free, to deliver emails that matter. ODINE incubated SMEs were entitled to the following deal as part of SendGrid Accelerate: 12 months of SendGrid's Pro 300K absolutely free, with a total value of \$2400



### **IBM Softlayer Catalyst Startup Program**

Catalyst is core to IBM's mission to enable next generation technologies and empower the entrepreneurs that will bring them to market.

ODINE incubated SMEs could have access to:

- Free infrastructure: Members get \$120,000 USD in annual credits.
- Mentoring: Pick the brains of the industry veterans and über-geniuses who innovate the SoftLayer platform.



### **Amazon Web Services (AWS) Activate**

The team at Amazon Web Services has a startup program for our early stage companies who have raised up to \$1.5 million in total capital. ODINE incubated companies had access to the AWS Activate program, including the following:

- \$10,000 in AWS credits valid for 2 years
- Free access to the AWS Essentials 1 day web-based or instructor-led training (normally ~\$600/course), AWS Technical Professional Accreditation course, plus 8 tokens for self-paced labs (normally ~\$30/lab)
  - One year of premium AWS Business Support (up to \$5,000)
  - Free access to 1:1 virtual office hours with AWS solutions architects
  - Special offers from other companies that help startups, including Chef, Bitnami, Amazon Payments, SOASTA, Podio, CopperEgg and more.

### **Startup Blueprint (Braintree\_PayPal)**

Startup Blueprint is a global program that partners with top tier startups, incubators, accelerators and VCs. Startups will get significant benefits, including:



Top notch dedicated technical and business customer service provided by specialized Startup Advisors. Ongoing mentorship with thought leaders and our Startup & Developer Relations team.

ODINE SMEs could get: from PayPal: free transaction volume up to \$1.5 million USD (\$50,000 USD of PayPal fees waived for 18 months). From Braintree: free transaction volume up to \$100,000 USD.

7 companies took advantage of the Amazon Activate offer. None of the other offers was redeemed.

Software tools with support from the ODINE team did not change with respect to the first version of this deliverable.

### 3. Technical support report

To better understand the technological and data needs of the incubated SMEs, we developed a technical questionnaire (cf. Annex 1) divided in three sections:

- Infrastructure: To know what cloud provider they use, if any), and if they require support for obtaining and/or managing cloud resources.
- Software: What Big Data processing tools they used, if any, and if they require any specific support on data processing, with special emphasis on those supported by the consortium.
- Open Data Provision: What datasets would they like to see open, what are the most important features for them in an open dataset and how would they improve the current datasets.

The questionnaire included some questions specific to TID expertise: they were:

In the Infrastructure section

*Do you use or plan to use any of FIWARE Generic Enablers?*

*Do you need specific support or training on any of FIWARE GEs?*

In the Open Data Provision Section

*Do you use or integrate IoT data or any other type of real time data (e.g. from cities) in your solution?*

*If so, with which platform or APIs or providers? do you need any specific support or training on it?*

After TID's departure from the project, that is, for cohorts 5, 6, 7 and 8, these questions were removed from the questionnaire.

The questionnaire was sent by e-mail to each cohort just after starting their incubation period. Answering was voluntary; non-response of the questionnaire was interpreted as lack of interest in any technical support. 26 companies answered the questionnaire, and actions were taken to provide the requested support, when possible.

In the following, we analyze the responses from a quantitatively and qualitatively point of view. The result of the survey should provides useful insights about infrastructure and software needs of European Open Data SMEs, for anyone interested in running a similar project to ODINE in the future.

**Infrastructure:**

Figure 1 provides a summary of the Cloud Providers used by the SMEs that answered the questionnaire. The most popular provider is AWS (35%) followed by Azure (19%) and Digital Ocean (11%). 5 companies declared to be using their own infrastructure for development, testing and deployment. A relatively high number of companies answered that they were not using Big Data technologies, as they were not working yet with it.

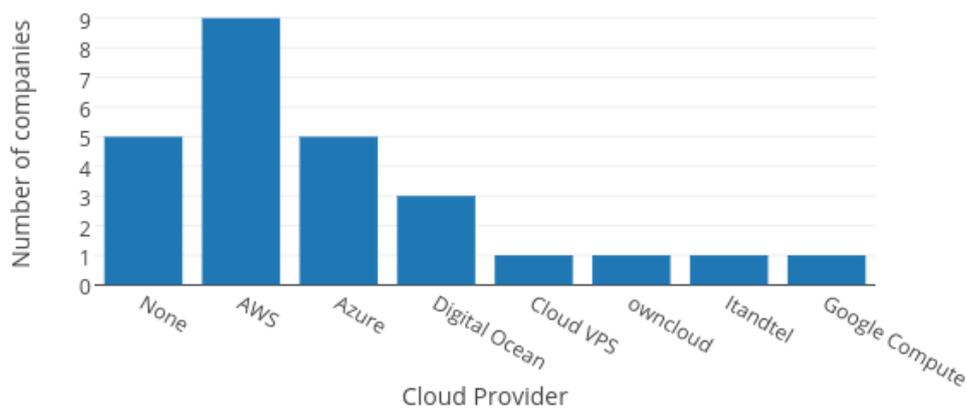


Figure 1: Cloud providers used by ODINE incubated companies

To the question "Do you need support to get new cloud resources?". 23 SMEs answered "No", 2 SMEs requested AWS credits and 1 SME that was not using any cloud provider requested assistance to migrate to AWS. The latter request was fulfilled by the ODINE team.

With respect to the FIWARE related questions, only 1 out of the 12 SMEs that answered the questionnaire in rounds 1-4 wanted to use the ORION Context Broker (it was also using Azure). The ODINE team from TID provided support for this request.

## **Software:**

We asked SMEs about the tools they were using for Big Data and Linked Open Data processing.

10 out of the 26 SMEs generated or processed Linked Open Data. None of them was familiar with the LInDA workbench, but 4 of them were interested in learning more. A LinDA webinar was organised on 27th November of 2015 with the participation of 2 ODINE companies. We noted that regarding Linked Data Technologies, companies could be separated in two groups: those that never heard or had very little idea about it, and companies founded by people with strong Linked Data background (e.g. Ph.D on the subject)

Concerning Big Data, 8 SMEs answered positively to the question of using BigData tools, while 18 said that they were not using any Big Data tool. Six of them specified the tool they were using, 2 of them used Hadoop, 1 Google Tensor Flow, 1 Apache Spark, 1 NodeXL and 1 MongoDB. Table 1 summarizes the results.

Tool	None / Not needed	Yes but unspecified	Hadoop	Tensor Flow	Apache Spark	Node XL	Mongo DB
#SMEs	8	12	2	1	1	1	1

Table 1: Use of Big Data Tools among SMEs that answered the technical questionnaire

## **Open Data Provision:**

This section of the questionnaire aimed at identifying (i) The most important quality features for the companies. This knowledge may be useful for Open Data publishers to focus resources in improving the features considered as most important for re-users. (ii) Additional datasets that they would like to see open. The intention was to identify datasets transversal so several companies, where efforts for unlocking could be focused.

We asked SMEs to rate the importance of 8 quality features of Open Data, using the following scale: 1 (not important at all); 2 (slightly important) ; 3 (reasonably important); 4 (very important) ; 5 (critically important). Figure 2 shows the average rating, truncated at the first decimal digit. Accuracy is the feature perceived as more important, followed by timeliness. Format and support were considered the less important.

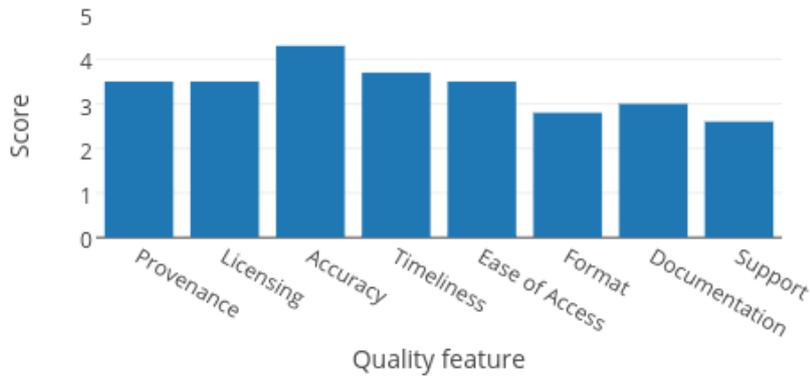


Figure 2: Average importance rating of data quality features as perceived by ODINE-incubated companies

We also asked *Which types of data and specific datasets would you like to see open?* The intention was to gather insight to open data publishers about the data needs of SMEs actually working on Open Data. Answers were disparate, each company described specific data relevant for their domain or business. From the available sample, we did not identify any dataset or dataset type common to more than one company.

We also asked *Which types of open data and specific open datasets would you like to see improve, and how?*, and as with the previous question, answers were very specific to the domain of the SMEs and correlated with the features they considered as critical on the question about quality. Nevertheless, eight companies considered an important improvement to be able to access datasets via an API. Six of those precised that they would like a JSON API.

### 3.1 Technical support procedure and use case

For providing technical support, ODINE followed the following procedure:

1. Sending the questionnaire described above. Companies needing technical support were invited to contact the technical team with their requests.
2. The technical team analysed the request and assessed if the relevant expertise was available within the consortium
  - a. If this wasn't the case, we reached out to our networks and provided introductions to providers with the right expertise. The companies then had the choice to contact them directly and decide if they were interested in the

services offered. SMEs were able to use the ODINE grant to pay for the services (amending their originally agreed budget ,if necessary).

- b. In case the expertise existed in-house, we facilitated the connection of the SME and the expert through videoconference. In this case, there was no additional cost for the company.

We detail below one use case of support where technical expertise existing in the ODINE team helped one of the incubated companies.

The company Viomedo is a marketplace for clinical trials. They simplify and speed up the matchmaking process between patients searching for clinical trials and health-service companies or researchers looking for patients willing to participate in clinical trials. They approached the ODINE consortium requesting some additional insights about datasets on clinical trials, such as the challenge of data integration and curation needed when clinical data are integrated in the platform

Responding to the request, the ODINE team at Fraunhofer presented the main RDF datasets in the Biomedical domain and well-known domain ontologies. In particular, they discussed LinkedCT<sup>1</sup> and the steps followed during the creation and integration of this dataset into the Linked Open Data cloud. Additionally, the main properties of LinkedCT vocabularies were analyzed, as well as Biomedical ontologies like NCIThesaurus, Gene Ontology, Disease Ontology, and SNOMED.

A further challenge they faced was the ability to provide a higher granularity of the search feature for patients. Currently, their trials database is only annotated at the disease level, e.g. lung cancer, etc. Their primary research highlighted that patients would like to be able to refine the search by stage and sub-type of disease, instead. With the help of the ODINE team, the problem was broken down into two actions:

1. Extracting relevant concepts from the narrative of the clinical trial.
2. Simulate an 'include/exclude' binary criteria and matching with clients, once the clinical trial narrative is annotated

However, the problem described above was not part of their ODINE milestones, but a further thing that they wanted to explore given that Fraunhofer had the expertise. After the meeting, they decided to focus on ODINE milestones instead, as it was more critical in their roadmap.

## 3.2 User research for open data tooling

In August 2016 the ODI carried out a user research to assess the need for improved open data tooling that could be developed by ODI Labs as part of ODINE task 4.5 (Data providers liaison)

---

<sup>1</sup> <https://arxiv.org/pdf/0908.0567.pdf>

The results show unmet user needs around data cleaning and discovery, and use or creation of data standards, and we recommend that efforts should be focused in building tools to meet these needs.

## Survey

A [survey](#) was sent to all 39 ODINE startups in mid-August. As of 18th August, 9 [responses](#) had been received, a response rate of **23%**. **44%** of respondents were ODI members.

This is a good response rate to an unsolicited survey, so we consider the results below to be indicative of the ODINE cohort as a whole.

## Responses

This section does not include responses to all questions, but instead pulls out strong signals that are of interest.

## Consuming Data

- **89%** of respondents are using open data as part of their business model. **75%** are using data from governments, and **75%** are using data from private sector organisations.
- **75%** of respondents state that the data they are using is poor-quality, and **75%** are having difficulty finding data in the first place is difficult.
- **50%** of respondents have to spend more than 1 day cleaning data before they can use it.

## Publishing Data

- **78%** of respondents are publishing open data as part of their business model.
- Only **29%** stated that they are using a machine-readable format. **57%** selected the “other” option when asked if they publish in CSV, JSON or other standard electronic formats.
- **44%** of respondents would consider publishing industry performance stats, standard identifiers, or product lists as open data. **22%** would consider publishing financial performance data openly.
- **0%** of respondents are publishing data using specific standards.

## Reporting

- **67%** of respondents have to report their performance somehow, normally by email updates.
- Only **11%** were interested in open publishing of their performance data.

- **22%** were interested in free tools for creating dashboards .

## Followup

- **56%** of respondents were interested in taking part in further user research around tools, implying that they do have unmet needs.
- **44%** of respondents were interested in having a case study written about their use of open data.

The numbers showed that:

- There are unmet needs around **data cleaning and discovery**, which would help ODINE startups to more effectively discover and use data.
  - Encouraging further use of Open Data Certificates and CSVlint may help with quality issues, as well as investigating and promoting other data cleaning tools such as DataGraft.
  - There may be utility in working on improving data discovery platforms to help startups identify data that is useful to them.
- While most startups are publishing data (and are interested in publishing more), they are not doing this using simple machine-readable formats and standards such as CSV and JSON. **This severely limits the utility of the published data.**
  - Tools to helping startups discover existing standards for their sectors may be useful.
  - If standards are not available, creation or promotion of simple tools to help startups document their own standards will enhance utility.
  - Data publishing tools that make standards compliance and machine readability simple will help startups to improve the reusability of their data.
- There is no clear demand for dashboarding tools.

## 4. Open data repositories and unlocked datasets

Besides the commented list of data repositories described in version 1 of this deliverable, ODINE engaged in unlocking datasets

ODINE's strategy for unlocking datasets follows four pathways to impact.

1. A demand-driven process, where we lobby for the unlocking of datasets explicitly required by SMEs in our incubation program.
2. Some of our incubated companies release datasets that they have collected and curated as part of their milestones.
3. We identified a set of critical companies and institutions holding data whose opening would help the ecosystem, and offered our help to open them.

The list below details the datasets that ODINE helped to unlock, involving both policy and technical support:

- Data on schedule, working elevators, transportation net, stations, buildings and many more from DeutscheBahn<sup>2</sup>
- List of bike stations<sup>3</sup> for the Berlin Bike Hackathon (also ahead of public release)
- Through our incubated company contagt, the release as Open data of public building's indoor navigation guides.
- Implisense requested for an open company register in Germany. We forwarded that request to the relevant government department and provided the information to policy makers and members of the Parliament and political groups. OKFDE will continue working towards this goal
- Fitness/wellbeing sector<sup>4</sup> via open active coordinated by imin. Includes datasets by:
  - British Cycling<sup>5</sup>
  - Good Gym<sup>6</sup>
  - Table Tennis England<sup>7</sup>
  - Flex<sup>8</sup>
- List of extractive industry companies by Openoil<sup>9</sup>
- Some of the data combinations in the data portal on equipment of UK higher education institutions<sup>10</sup>.
- Dataset on UK academic open data services<sup>11</sup>, with a focus on services provided for an entire organisation.
- The ODI worked together with Experian Data Quality in producing a quality assessment of three UK public open datasets:
  - GP Practices and Surgeries<sup>12</sup>
  - Land Registry<sup>13</sup>
  - Companies house corporate dataset<sup>14</sup>
- We assisted the University of Alicante on the provision of more effective data services on top of the datasets of their open data portal<sup>15</sup>.
- Syngenta: the ODI helped them publish their first open data<sup>16</sup> on agriculture efficiency indicators as part of Syngenta's Good Growth Plan.

---

<sup>2</sup> <http://data.deutschebahn.com/dataset?groups=datasets>

<sup>3</sup> <http://data.deutschebahn.com/dataset/data-call-a-bike>

<sup>4</sup> <https://www.openactive.io/use-data.html>

<sup>5</sup> <http://data.goskyride.com/>

<sup>6</sup> <https://goodgym-oa.github.io/>

<sup>7</sup> <http://data.pingengland.co.uk/>

<sup>8</sup> <http://data.joinflex.tv/>

<sup>9</sup> <http://aleph.openoil.net/>

<sup>10</sup> <http://equipment.data.ac.uk/>

<sup>11</sup> <http://hub.data.ac.uk/>

<sup>12</sup> <https://paulmalyon.github.io/gp-practices-and-surgeries-data-quality-overview/>

<sup>13</sup> <https://paulmalyon.github.io/land-registry-price-paid-data-quality/>

<sup>14</sup> <https://paulmalyon.github.io/companies-house-corporate-dataset-column-stats/>

<sup>15</sup> <http://datos.ua.es/>

- Department for Environmental, Food and Rural Affairs of the UK (DEFRA): For the period June 2015 - June 2016, the policy team of the ODI directly participated in:
  - A bulk download of Defra's historical air quality data archive<sup>17</sup> as open data, at the request of SMEs looking to build services monitoring AQ in England.
  - Releasing Environment Agency's LIDAR point cloud data as open data<sup>18</sup>. Both this and the previous datasets are aligned with the goals of our incubated company PlumeLabs.
  - Publishing nearly 30 years of individual food diary data (30,000 diaries per year, from 1972 - 2000) as open data<sup>19</sup>
- Pollution level data in the city of Paris, through our incubated company PlumeLabs<sup>20</sup>
- Amsterdam's IOT Living Lab API<sup>21</sup>, through our incubated company Glimworm.
- Our incubated company Tilde is part of the Meta-Share network<sup>22</sup> that makes Language Resources openly available. As part of their ODINE project "MODEL", they developed a MultiLingual Open Data Corpus for European Languages<sup>23</sup>
- Through our incubated company Zazuko, metadata about archives of several cantons in Switzerland<sup>24</sup>

Besides datasets, ODINE has also engaged in the development of open standards.

- FIRE standard (Suade)<sup>25</sup> a common specification for granular regulatory data. Regulatory data refers to the data that underlies regulatory submissions and is used for policy, monitoring and supervision purposes.
- Open standard for the fitness industry by openactive<sup>26</sup>. This community represents a group of organisations who share a vision for tackling the global problem of physical inactivity. They use technology to help more people to become more active, by providing open source tools, open standards and open data
- Open Banking Standard report<sup>27</sup>. The Open Banking Standard is aimed at guiding how open banking data should be created, shared and used by its owners and those who access it
- open contracting standard<sup>28</sup> The ODI helped develop the technical components of and helped it to be adopted by multiple governments.

---

<sup>16</sup><http://www4.syngenta.com/what-we-do/the-good-growth-plan/progress/progress-open-data/progress-data-downloads>

<sup>17</sup> <https://uk-air.defra.gov.uk/data/>

<sup>18</sup> <https://defradigital.blog.gov.uk/2016/03/17/exploring-the-laser-cloud/>

<sup>19</sup> <https://www.gov.uk/government/statistics/family-food-open-data>

<sup>20</sup> <http://opendata.aircare.tech/docs/#!/default/measurements>

<sup>21</sup> <http://iotlivinglab.com/amsterdam-iot-living-lab-wiki/>

<sup>22</sup> <http://metashare.tilde.com/>

<sup>23</sup> <http://metashare.ilsp.gr:8080/repository/browse/tilde-model-multilingual-open-data-for-eu-languages/a70af9701f6811e7aa3b001dd8b71c6645b57ed010b84385868c128cdf5807a/>

<sup>24</sup> <http://data.alod.ch/>

<sup>25</sup> <https://suade.org/fire/>

<sup>26</sup> <https://www.openactive.io/technology.html#the-steps>

<sup>27</sup> <http://theodi.org/open-banking-standard>

## 5. Developers reached

One of ODINE's goals was to increase the outreach of Open Data and Open Data based tools to application developers beyond those working for the SMEs incubated by the program. To count the outreach we followed the steps above:

1. From the incubated companies, we identified those whose business model was to sell data as a product or data analysis services. As for our Business Model survey (cf. Deliverable 6.3), 16 companies matched this profile.
2. From those 16 companies, we further identified those that offered an API, meaning that their target client was developers or companies developing applications. We also included companies that developed open standards on which they proposed to reformat a certain type of data. We discarded 4 companies: 3 of them were using ODINE funding to develop a side product/service that did not target developers. 1 was at early stage and used the incubation period to collect the critical mass of data required before launching their API.
3. From the 12 remaining companies, we extracted from their final milestone report the following statistics (i) number of requests for a key for their APIs, (ii) number of external developers that had access to the API (Beta testers, participants in a hackathon, as part of a training session paid by a client), (iii) number of small, medium and large company clients that bought premium access to the APIs.
4. We aggregated the numbers collected in the previous step as follows: For the API key requests, we assumed that half of them ended actually using the API<sup>29</sup>; For each external developer and small client we counted 1 developer using the API; For each medium client we considered a team of 3 developers working with it; and for each large client we considered a team of 6 developers.

Table 3 shows the results. We estimate a total of 1177 developers using Open Data powered APIs and services developed with ODINE funding. This figure largely exceeds the 300+ developers mark initially set as a goal in the initial agreement. Note that the bulk of developers come from subscriptions to publicly released APIs.

---

<sup>28</sup> <http://www.open-contracting.org/>

<sup>29</sup> Unfortunately, we don't have access to data about API calls with enough granularity to provide a better estimate.

Criterion	½ * API Subscribers	External Developers	1 * #Small corporate client	3 * #Medium corporate clients	6* #Large corporate clients	Total
Developers	815	105	2	33	26	<b>1177</b>

Table 3: Estimated number of developers using Open Data backed APIs and services provided by ODINE-incubated companies

## 6. Conclusions

One important aspect of ODINE services to incubated companies was to provide them with (1) support with technical tools and infrastructure to develop and maintain their products and services on top of Open Data and (2) support with the unlocking of further datasets that could help them unlock further value, with special emphasis on those that could be transversal.

Regarding (1), ODINE provided support on the FI-WARE platform and in several Open Source data analysis and processing tools, plus several decades of experience in Data Science. However, very few incubated companies used these services, as most of them already had their technological infrastructure in place and required more help in business skills (marketing, business development, etc). The consortium ran a survey on Open Data tooling to find out if there are other tools that open data startups might require, suggesting that there are unmet needs on certain data cleaning tools and

Regarding (2) the ODINE consortium participated in the unlocking of more than 30 datasets and of 4 Open Standards in different sectors, either in joint effort with incubated companies or as an effort considered as key for the Open Data Community. We note that in some cases, the time required to unlock a dataset needed by a company could take more time than the incubation period.

The consortium also measured the outreach of open data through the APIs developed by incubated companies, estimating the number of developers in reached more than one thousand.

# Appendix: Questionnaire for Technical Support

## Infrastructure

ODINE provides different cloud services to SMEs: FIWARE Lab infrastructure (in the terms and conditions of the FIWARE acceleration programme), and agreements with third parties as Amazon AWS credits (subject to request and approval).

*Are you using a cloud provider? Which one? \**

- None, just using on premise infrastructure
- FIWARE Lab
- Amazon Web Services
- Rackspace
- Azure
- Others (please detail) :

*Do you need support to get new cloud resources?*

Yes (explain) / No

*Do you use or plan to use any of FIWARE Generic Enablers? (Note: Question deleted after TID's departure from round 5)*

Yes (explain) / No

*Do you need specific support or training on any of FIWARE GEs? (Note: Question deleted after TID's departure from round 5)*

Yes (explain) / No

## Software

ODINE provides support for a number of tools where the consortium have an special involvement . If you use these tools, it will be easy to get in touch and get direct support from experts. If you use other tools, tell us and we will find a way to help you. Do you use Linked Data tools

Yes (Which ones?) / No

*Do you reuse or generate Linked Open Data? \**

- Yes
- No

If so, with which tools? do you need any specific support or training on it?  
Please explain

*Do you use or plan to use LinDA tools? \**

- LinDA Visualization
- LinDA Transformations
- LinDA Analytics
- LinDA Query Builder
- Not really
- Others:

*Do you need specific support or training on any of LinDA tools?  
(Please name them)*

*Do you use big data tools (e.g. Hadoop)? \**  
Yes (Which ones?)/ No

*Do you need any specific support or training on Big Data tools?  
Please explain*

*Do you use any other tool or process to manage (open) data (curation, transformation, analysis, visualization, etc.)? \**  
Please specify the tools (e.g., CubeViz, Open Refine,...)

*If so, do you need any specific support or training on it?  
Please name the tool(s)*

*Would you need any other type of technical support or help regarding software and tools?  
Please explain.*

## **Open Data Provision**

Through the ODINE partners contacts, it would be possible to engage with open data providers to fulfill specific data needs.

*Which types of data and specific datasets would you like to see open? \**

E.g. More data about agriculture, more real time air quality data, more government tender data, more satellite data, an address file, a full register of UK public sector bodies, BBC archive metadata, etc.

*Which types of open data and specific open datasets would you like to see improve, and how? E.g. Increased granularity of air quality data, all OS data published under Creative Commons open licenses, contract award notice summaries released as csv and through a JSON API, etc.*

*Please indicate the importance of each of the following issues in your company's decision to use open datasets: \**

1 (not important at all)

2 (slightly important)

3 (reasonably important)

4 (very important)

5 (critically important)

Provenance of data

Licensing of datasets

Accuracy of data

Timeliness of data

Ease of access to datasets

Format of data

Accompanying documentation

Help and support from publisher

*Anything else you would like to receive help from ODINE from a technological point of view?*