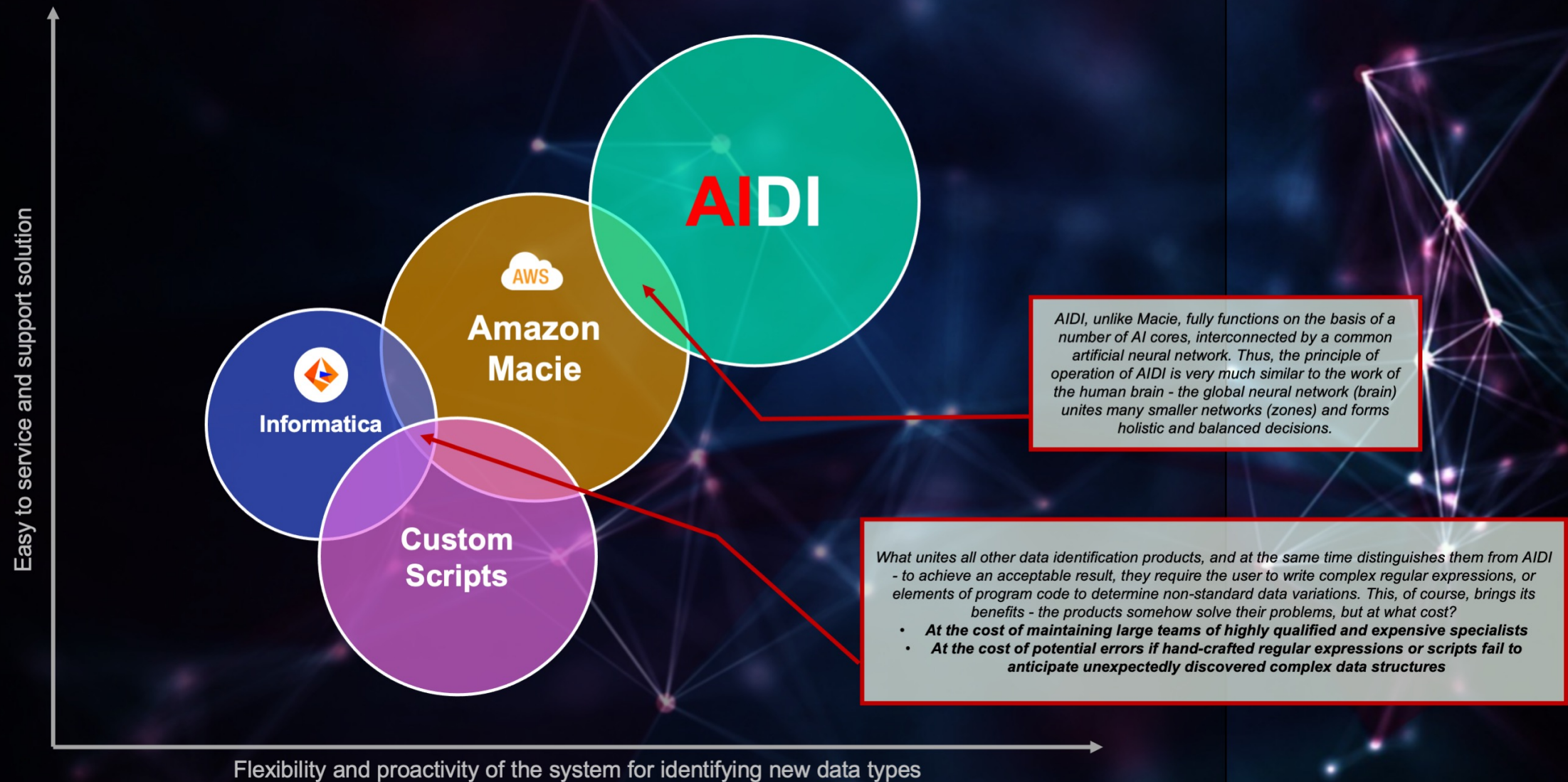




Think Like a human,
see Like a human ...

Artificial Intelligence Data Identification

Analysis of Data Identification and Data Classification product markets



AIDI is a progressive data type identification system based on machine learning.

By default, AIDI is trained to identify basic data types (cities, phone numbers, email addresses, and so on), but as you use it, you yourself, without noticing it, will teach it to recognize more and more new types of data, and you do not need to be programmer.

Below are some of the uses for AIDI:

- Data Discovery and Data Identification of Databases and Text Files
- Streaming data authentication (REST Full)
- Revealing data of complex structures (blocks of program code, JSON structures, XML, and so on)

What types of data can be identified in AIDI?



For your convenience, as well as the possibility to test the work of AIDI "out of the box", her mind will be trained to recognize the following types of data:

Data Type	Examples of data
First name	<i>Kevin; Mike; Alena; Ivan ...</i>
Second name	<i>Stivens; Sidorov; Watson ...</i>
Web address (URL)	<i>https://google.com/ ...</i>
Email	<i>example@domain.net ...</i>
Geographic address	<i>221B Baker Street ...</i>
Phone number	<i>+375-29-111-12-34; +1(22) 446 45 22 11 ...</i>
Country	<i>Belarus; USA; China; PLN ...</i>
State	<i>Texas; Arkansas; California ...</i>
City	<i>London; Warsaw; Oslo; NY ...</i>
Name of the company (organization)	<i>Google Inc; Novartis Pharma ...</i>
Date / Time	<i>2020.12.18 15:44, 2012.03.15</i>
Profession	<i>Doctor; System Administrator ...</i>
Process statuses	<i>Completed; Planned; Waiting for support ...</i>
Currencies	<i>BYN; Rubl; USD; Dollar USA ...</i>
Time intervals and frequency	<i>Days; weeks; years; monthly ...</i>

With each new update of AIDI, the standard set of recognized data types, as well as the quality of data identification, will grow.

Thus, you can easily train AIDI to recognize almost any type of data.

For example, it can handle identifying the following content in database cells:

- **Json and xml structures**
- **Codes in programming languages**
- **Web Markup and Styling Languages (HTML / CSS)**
- **Files of images and other documents stored in databases**

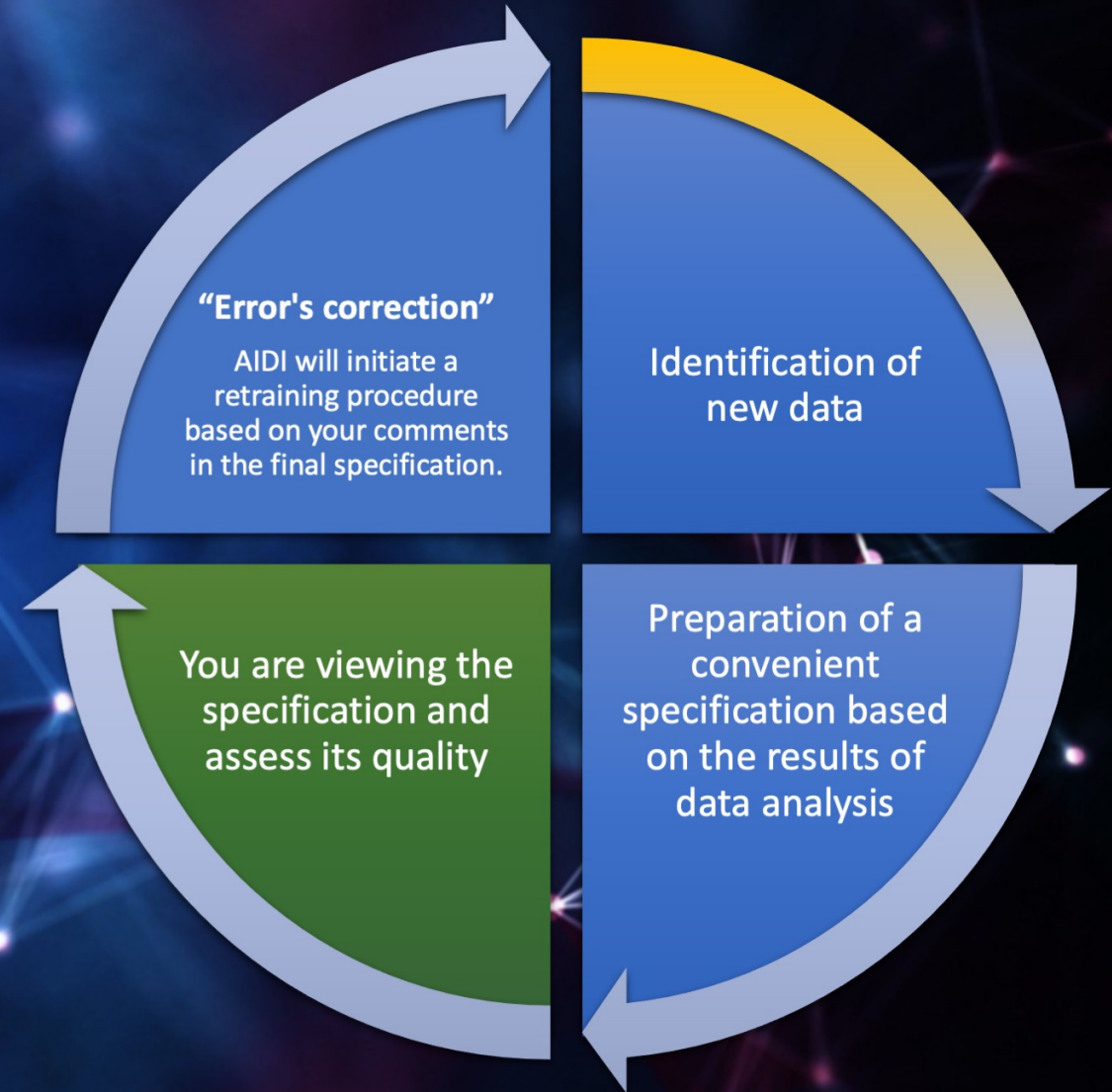


Data identification and **AIDI** training

Data identification processes as well as additional training for your AIDI are closely related. In the process of solving your work tasks, AIDI will adjust its work, fix errors and improve its qualifications based on your assessments of its work.

Thus, the process of improving the quality of AIDI's work is ongoing. AIDI independently tracks its mistakes, seeks to understand them and improve the quality of its work. In this case, your participation is reduced only to a convenient verification of the quality of its work in a few mouse clicks.

The following slides describe each of these steps in more detail.

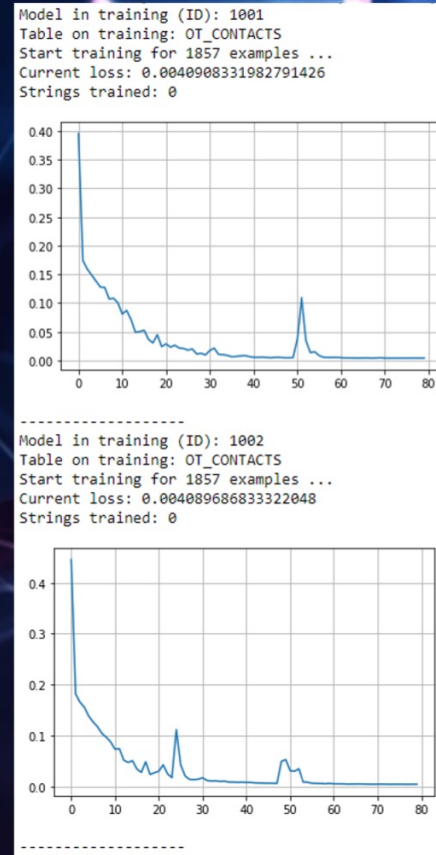


Data identification and AI training

A few mandatory quality requirements are imposed on the learning process, as well as its results. Compliance with these requirements is monitored by the system automatically. For example, these include:

- **The requirement for a variety of training data** - training data should be as unique as possible, and also reflect the variability of their presentation as much as possible
- **The requirement to reduce probabilistic errors** - training continues until the error in identifying new data by the system becomes less than 0.001%

The image shows the process of assimilation by neural connections of new interpretations of already known data types. As can be seen from the graphs, the system actively assimilates new data during the first 30 repetitions of their study, and by 60 repetitions the probability of errors in data identification reaches acceptable values.



Accumulated experience

- This experience did not allow us to solve the last task at 100%.
- The number of identification errors was more than 0.01%



New instructions

- New data types
- New representations of already known data types

The updated experience is used in further iterations of data identification and retraining.

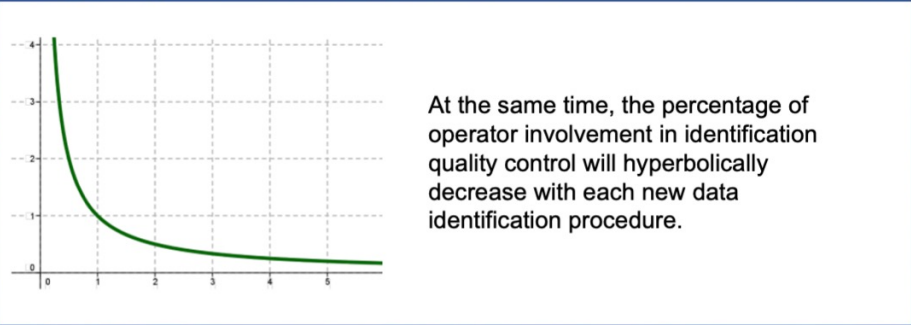
Updated experience

- Successfully solves all previous data identification tasks
- Reduces the likelihood of potential errors for future identification tasks

Data identification and AIDI training

Column name	Data Examples	Identification Results
PO_CHANGED_ON	07/21/2020 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
DATE_ID	07/30/2013 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
REQUISITIONER	DERIDMA1	Наименование юридического лица 1 <input type="checkbox"/> Login 521 (Novartis Pharma) 1 <input checked="" type="checkbox"/>
INVOICE_ENDORSER	DRUTIER1	Наименование юридического лица 1 <input type="checkbox"/> Login 521 (Novartis Pharma) 1 <input checked="" type="checkbox"/>
PO_CREATED_BY	STROESO3	Наименование юридического лица 1 <input type="checkbox"/> Login 521 (Novartis Pharma) 1 <input checked="" type="checkbox"/>
PO_VALID_FROM	06/28/2019 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
PO_VALID_TO	02/27/2021 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
WBS_ELEMENT_STATUS	OPEN	Наименование юридического лица 1 <input type="checkbox"/> Activity Status 1 <input checked="" type="checkbox"/> First Name 0.51 <input type="checkbox"/>
VENDOR_NAME	EUROFINS LANCASTER LABORATORIES INC	Наименование юридического лица 1 <input checked="" type="checkbox"/> Profession 0.82 <input type="checkbox"/>
COUNTRY	FR	State 1 <input type="checkbox"/> Country 1 <input checked="" type="checkbox"/> Наименование юридического лица 1 <input type="checkbox"/>
COUNTRY_NAME	France	Country 1 <input checked="" type="checkbox"/> Наименование юридического лица 1 <input type="checkbox"/> Activity Status 0.69 <input type="checkbox"/> City 0.59 <input type="checkbox"/>
FINAL_INVOICE	Closed	First Name 1 <input type="checkbox"/> Наименование юридического лица 1 <input type="checkbox"/> Activity Status 1 <input checked="" type="checkbox"/> City 0.99 <input type="checkbox"/>
PO_DOCUMENT_DATE	11/27/2018 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
FIRST_INVOICE_DATE	07/09/2015 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
FIRST_GR_DATE	01/29/2020 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
LAST_INVOICE_DATE	08/28/2014 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
LAST_GR_DATE	01/26/2021 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
DOCUMENT_CURRENCY	GBP	Наименование юридического лица 1 <input type="checkbox"/> Currency 1 <input checked="" type="checkbox"/> Activity Status 0.57 <input type="checkbox"/>
LOCAL_CURRENCY	CHF	Наименование юридического лица 1 <input type="checkbox"/> Currency 1 <input checked="" type="checkbox"/> Activity Status 0.67 <input type="checkbox"/>
INVOICE_DATE	11/30/2009 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
BASE_DATE	01/11/2010 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
CLEARING_DATE	09/28/2018 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
NET_DUE_DATE	07/25/2020 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>

Based on the results of the identification carried out, the system generates a consolidated report that informs the user (operator) about which classifications it suggests attributing certain data. The user's task is to validate this data and tick the correctly identified data.



At the same time, already at the first iterations of identification (as shown in the example below), the system manages to identify such a complex semantic data object as the name of the company.

PO_VALID_FROM	06/28/2019 00:00:00	Telephone, Fax 1 <input checked="" type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
PO_VALID_TO	02/27/2021 00:00:00	Telephone, Fax 1 <input type="checkbox"/> Date / Time 1 <input checked="" type="checkbox"/>
WBS_ELEMENT_STATUS	OPEN	Наименование юридического лица 1 <input type="checkbox"/> Activity Status 1 <input checked="" type="checkbox"/> First Name 0.51 <input type="checkbox"/>
VENDOR_NAME	EUROFINS LANCASTER LABORATORIES INC	Наименование юридического лица 1 <input checked="" type="checkbox"/> Profession 0.82 <input type="checkbox"/>
COUNTRY	FR	State 1 <input type="checkbox"/> Country 1 <input checked="" type="checkbox"/> Наименование юридического лица 1 <input type="checkbox"/> Country 1 <input checked="" type="checkbox"/>

So why AIDI is the best choice

1. Unlimited data identification possibilities
2. Clear and convenient process of working with the system for everyone
3. The process of training and improving the quality of work is integrated into the data identification process, which significantly speeds up the AIDI training process
4. Intelligent training AIDI allows it to independently adapt to your data identification needs
5. No need to hire highly qualified staff to support infrastructure
6. The cost of the AIDI is lower than that of analogues
7. Successfully conducted POC tests on real-world tasks have shown that *the untrained AIDI core is capable of classifying data with an accuracy of more than 85%. In turn, the trained kernel classifies data with a marginal error of less than 0.001% (that is, the data classification accuracy is more than 99.999%).* This significantly outperforms many other competitors.
8. Large consumer market - a number of large companies are already interested in this solution, paying special attention to the safety of their own data. All that is needed is to translate the conceptual solution into the product to be sold.

CERTIFICATE

This certificate confirms that the software based on Artificial Intelligence and Machine Learning “**AIDI**” (Artificial Intelligence Data Identification) has been functionally tested in conditions as close as possible to Production.

The ability of this software to identify various types of data (including complex data types of non-standard structure) was demonstrated and confirmed during this functionality testing process.

Date of PoC: **29 April 2021**

Date of presentation for Novartis
ISRM Data Masking Team: **3 May 2021**

The author, developer and copyright
owner of this software :

Yauhen A. Panimatchanka

Panimatchanka Yauhen
Digitally signed by Panimatchanka Yauhen
DN: dc=com, dc=novartis, ou=external-gold, serialNumber=2416165, cn=Panimatchanka Yauhen
Reason: I am approving this document
Date: 2021.05.04 16:14:57 +03'00'

SME Expert, NBS IT SRM

The validity of this document is confirmed by Novartis with digital signature certificates of signers.

Bradzetskaya Iryna
Digitally signed by Bradzetskaya Iryna
DN: dc=com, dc=novartis, ou=external-gold, serialNumber=2443309, cn=Bradzetskaya Iryna
Date: 2021.05.04 16:09:07 +03'00'

Solution Developer, NBS IT SRM

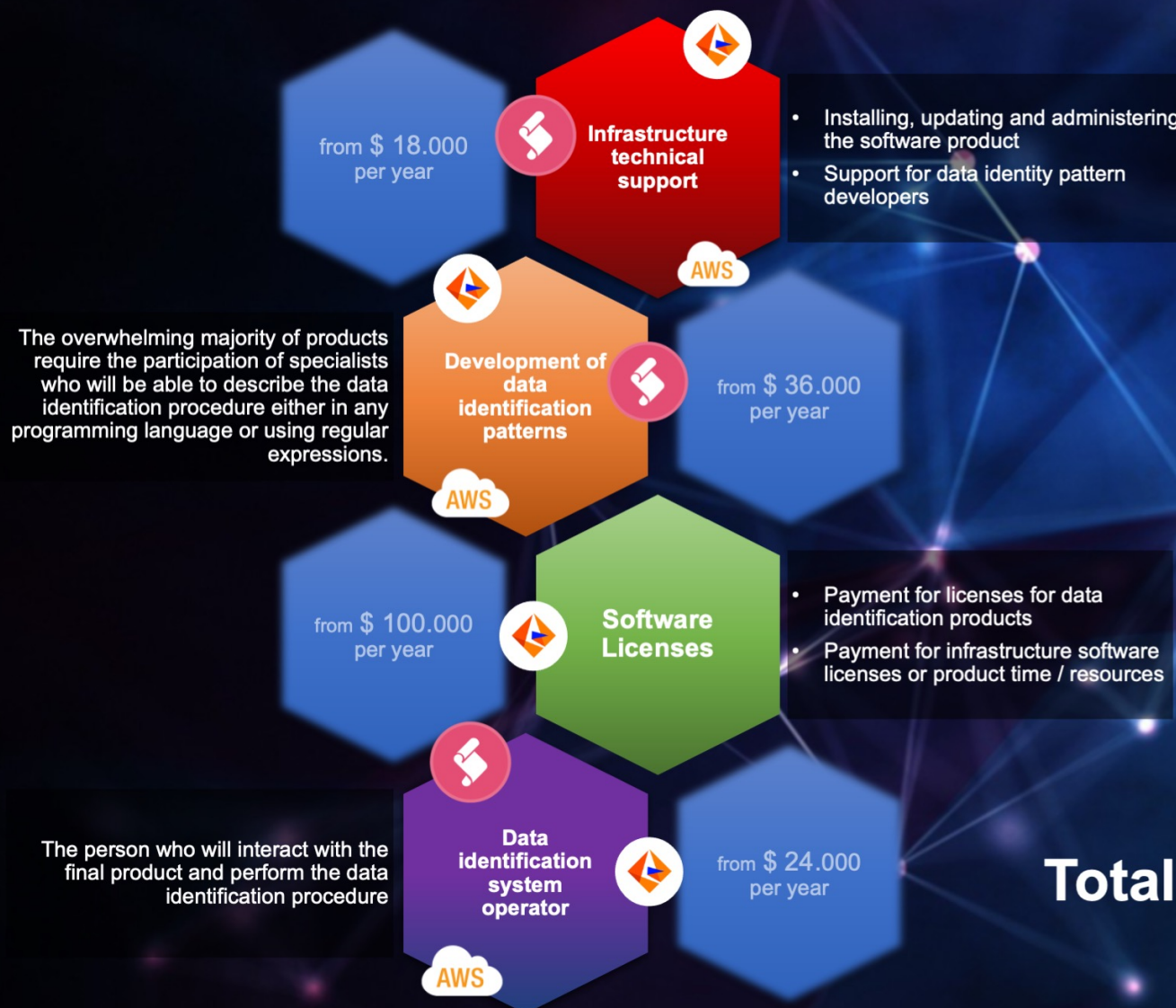
Software development was carried out in accordance with the current legislation of the Republic of Belarus, as well as international norms for the recognition of copyright:
- solely by the author of the software
- using the author's personal computing resources and in his spare time.

 **NOVARTIS**
PHARMACEUTICALS

 **NOVARTIS**
PHARMACEUTICALS

What do you pay for using other products

Minimum infrastructure requirements for analysing 100 gigabytes of data per month	
Number of infrastructure support specialists	1
Number of Data Identification Technicians Developers	2
Number of data identification operators	2



According to statistics, large companies (with more than 2000 employees) spend from 1.5 to 2 million US dollars per year on solving these problems.

Total: from \$ 178,000 per year

AIDI Costs

Costs and Liabilities

We are currently actively seeking investors and buyers for the rights to further distribute AIDI. If you are interested in making an effective investment in this product with the possibility of further profitability, please read our requirements, as well as an example of strategies for profit from the distribution of AIDI.

From \$ 25.000
(before taxes)

Acquisition of rights to develop and distribute AIDI

Reimbursement of costs incurred for requirements research, design and development of the concept, and POC.

Your benefits:

- By purchasing the rights to develop and distribute the final product, you get a unique chance to get attention from large clients
- the opportunity for further cooperation and development of this project in accordance with your business needs
- you get everything you need to develop the final product that will be ready for sale



Minimum for 1 year

Conclusion of an agreement on further development and product development with iTerra

For at least one year, the process of further product development is carried out by iTerra employees. Depending on your goals, we form the necessary staff and develop a solution. From your side, accordingly, the development process is financed.

What is the benefit:

- Its founders continue to work on already your product - those people who understand the processes of its functioning in detail, and, accordingly, can lead its further development with maximum productivity and quality.
- We, as the founders of the product, and accordingly, also being in the future together with you the benefit of the acquirer, are maximally interested in the product being of high quality and extremely attractive to customers. In this regard, we are ready to apply the maximum conditions for our greatest joint benefit from this transaction.



10% of all sales

Conclusion of an agreement on the distribution of profits

This condition provides for the payment of a monetary remuneration in the amount of 10% of all sales of product licenses as well as the provision of product support services to third parties.

Rationale:

- The payment of interest on the amount transferred by customers determines the low cost of acquiring unique rights to develop and distribute a software package.
- This interest rate, considering the potential income, does not create significant economic burdens on the acquirer of the rights to develop and distribute the AIDI software package and is easily covered by all income received from the sale of licenses for this software and the provision of services for its support.
- The existence of this payout program is a huge incentive for iTerra and the original founder of AIDI to contribute in every possible way to its development and increase its consumer appeal.



AIDI Costs

Recommended Capitalization Strategies

The slide below shows the recommended prices for distributing AIDI to end customers. These prices are much lower than those of competing products in order to form a high attractiveness of AIDI to potential customers.



Server licenses

The basis of the profit is the sale of annual licenses for the use of AIDI computing units.

Example: One license activates AIDI on one physical / virtual server and uses a maximum of 4 CPU cores and 16 GB of RAM.



Recommended cost for new clients:

\$ 25.000 per year
per one license

Basic Offering



Technical support

Sale of AIDI support services on the customer's side, as well as advisory and practical assistance in solving a number of related tasks.

Example: As part of the purchased license, the customer receives free 8 hours of technical support. He can spend these hours, for example, on the installation of AIDI with the help of our specialists, or any updates were carried out by our forces. If it is necessary to involve our employees in performing work in the customer's infrastructure, payment for the time of providing these services is charged additionally.



Recommended cost for new clients:

From \$ 25 per
hour



Advanced packages

Subscriptions for advanced functionality, by purchasing which the customer receives additional functionality. In addition to the fact that this brings additional profit, this approach also opens up flexible opportunities for further development of the product without any significant impact on its architecture and business model.

Example: Development and subsequent sale of solutions for the full cycle of SPI and PII data masking in the customer's environment.



Recommended cost for new clients:

\$ 10.000x per
year per license

Additional Offering

Expected receipts (per year) from one client:

\$ 50.000 - \$ 500.000

Still have questions?
Contact us and we will answer them!



info@ilterra.com



ilterra.com

Artificial
Intelligence
Data
Identification