# Home

# **Big Data For Small Organizations**

We are a Netherlands-based start-up company that brings the benefits of big data to small organizations. We want to make sure that small organizations can remain competitive in the era big data, they have access to trustworthy AI, and they can validate with regulators if third-party, corporate sales and marketing algorithms do not have negative biases against them. Our diverse team is particularly experienced in music, the creative industries, and digital humanities.

Our flagship demo projects are the [Listen Local](http://localhost:4321/project/listen-local/) ethical music recommendation system based on our [Demo Music Observatory](http://localhost:4321/project/music-observatory/) data integration and knowledge sharing platform. We have validated our product/market fit in the prestigious [Yes!Delft AI+Blockchain Lab](http://localhost:4321/post/2020-09-25-yesdelft-validation/)

 See our [services](http://localhost:4321/#services): [data curation](http://localhost:4321/service/data-curation/), [open data access](http://localhost:4321/service/open-data/), [survey harmonization](http://localhost:4321/software/retroharmonize/), [reproducible research](http://localhost:4321/service/research/) and validated [trustworthy AI applications](http://localhost:4321/service/trustworthy-ai/).

 Download our [introduction](http://localhost:4321/media/docs/reprex_introduction.pdf).

 Follow [news about us](http://localhost:4321/#news) or the more comprehensive [Data & Lyrics](https://dataandlyrics.com/) blog.

 Contact [us](http://localhost:4321/#contact) .

## Download text

Separate Google Doc in the same folder!

# Services

This is the most important part, and the downloadable introduction. The website puts some excerpts from our blog, which is a problem of its own – but let’s start here. The most visible parts are the ‘summary’ abstracts on the main page.

## Data curation

[**Data Curation**](https://reprex.nl/service/data-curation/)

[We find the relevant data for your business or policy scorecard, market or scientific research project. We are constantly mapping, testing, validating new data sources in the creative and cultural industries.](https://reprex.nl/service/data-curation/)

Last updated on 2021-01-25

[Slides](https://reprex.nl/slides/data-curation/) [Follow](https://twitter.com/dataandlyrics)

[](https://reprex.nl/service/data-curation/)

**If you do not find the right data for your policy evaluation, your consulting project, your PhD thesis, your market research, or scientific research project, it does not mean that the data does not exist. It does not mean that the data is not available for free. In our experience, up to 95% of available open data is never used.**

Everyday, thousands of new datasets are becoming available via the EU open data regime, US and other jurisdictions freedom of information legislation, or open science and scientific reproducibility requirements — but as they are packaged or processed for a different primary, original use, usually only experts of open data find this data and can be bring it to a usable form for re-use in business, scientific or policy research.

The creative and cultural industries are often not participating in government statistical programs, because they almost exclusively consist of microenterprises that are exempted from statistical reporting and file only simplified financial statements and tax returns. This means that finding the appropriate private or public data sources requires particularly good data maps.

Data curation means that we are continuously mapping potential data sources, and we are sending requests to download and put though quality tests newer and newer data sources. Our CEEMID project has produced several thousand indicators, out of which a few dozen is available in our [Demo Music Observatory](http://localhost:4321/project/music-observatory/). If you have special data needs for a scientific research, policy evaluation or business project, we can find you the most suitable, best value for money data for analysis or [ethical AI applications](http://localhost:4321/localhost:4321/service/trustworthy-ai/).

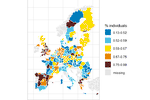
## Open data

[**Open Data**](https://reprex.nl/service/open-data/)

[We provide our clients with simple datasets, databases, harmonized survey data, and various other rich data applications; we provide them continuous access to high-quality, re-processed, re-usable public sector and scientific data.](https://reprex.nl/service/open-data/)

Last updated on 2021-01-25

[Slides](https://reprex.nl/slides/example/) [Follow](https://twitter.com/dataandlyrics)

[](https://reprex.nl/service/open-data/)

**Everyday, thousands of new datasets are becoming available via the EU open data regime, US and other jurisdictions freedom of information legislation, or open science and scientific reproducibility requirements.**

These datasets are usually very reliable, high value datasets, and their cumulated estimated value is close to 100 billion euros in the EU alone. But they offer no free lunch: originally collected to evaluate policies, measure the inflation, control tax returns, predict the weather or traffic jams, they usually require a thorough reprocessing for their re-use in the music industry or another new context. Our software solutions automate this process – for example, whenever price data is collected to measure the inflation in EU countries, we take out the price data of cultural products automatically.

The creative and cultural industries are often not participating in government statistical programs, because they almost exclusively consist of microenterprises that are exempted from statistical reporting and file only simplified financial statements and tax returns. In the creative & cultural sector, industry-organized data collection is particularly important. We are mapping all important global and European data collection, and we often carry out primary data collection for European stakeholders. Access and sharing these data on a continuous, programmatic basis is similarly possible (after the necessary approvals of data owners), just like in the case of open data, open science, or freedom-of-information sources.

## **European open data**

The European data regulation uses the expression open data for anonymized or non-confidential data that has been collected by a public sector organization. Such data is usually open for re-use.

While re-usable public sector information (often covered in other jurisdictions by freedom of information (FOI) legislation is free or requires only a marginal cost reimbursement, is usually not directly usable. The data was collected and processed for the needs of a primary use, such as tax administration, or short-term meteorological forecasting, or understanding current public opinion about a policy proposal. To make use of the information content, the data usually must be reprocessed, normalized, units and currency must be translated, and various other processing steps must be carried out without error.

Check out our [iotables](https://iotables.dataobservatory.eu/" \t "_blank) software that helps the harmonization of various European and African standardized surveys.

## **Open Science**

First class scientific journals require the publication of the empirical data or other data that supported the published scientific claim. This means that a large part of the empirical data input into the world’s high-level scientist output is freely available. Similarly to re-usable public sector data, this data is high-quality and cheap, but it is processed to the needs of the original scientific inquiry. The data usually needs to be reprocessed exactly the same way as re-used public sector information.

Check out our [regions](https://regions.dataobservatory.eu/) software that helps the harmonization of various European and African standardized surveys.

## **Survey harmonization**

In economics and social sciences, market research, politics and policy design, surveys are the primary sources of information. Online, telephone and personal interviews are transcribed into various datasets that are, in the case of public sector or scientific origin, are often available for free re-use.

The cost of surveying has exploded, and because of the increasing number of short, online surveys, survey fatigue is plaguing the population. The cost of high-quality surveys representative for the general (consumer, voter) population of the European Union is around a million euros; even national or regional surveys are alone often out of the reach of smaller research organizations.

Our solution to this is a software supported retrospective survey harmonization. We are looking up survey questionnaires where your question has been asked before, and create accompanying surveys that show either the change, or a more specific detail. This allows timewise, international, inter-regional comparison, and can reduce surveying costs significantly.

We also try to syndicate surveys. Our music professional surveys, for example, has been asking harmonized questions in a growing number of countries music professional population, allowing a comparison of music economics since 2014.

Check out our [retroharmonize](https://retroharmonize.dataobservatory.eu/" \t "_blank) software that helps the harmonization of various European and African standardized surveys.

[**Reproducible Research**](https://reprex.nl/service/research/)

[we create document-, presentation-, website-, and newsletter templates that are auto-refreshing their data contents, visualizations, citations, and other technical documentation items, taking care of error-prone and time-consuming revisions of technical documents, such as consulting reports, scientific publications, litigation documents and regulatory filings.](https://reprex.nl/service/research/)

Last updated on 2021-01-25

[Slides](https://reprex.nl/slides/research-automation/) [Follow](https://twitter.com/dataandlyrics)

**we can provide our clients with document, presentation, website, newsletter templates that are auto-refreshing their data contents, visualizations, citations and other technical documentation items, taking care of error-prone and time-consuming revisions of technical documents, such as consulting reports, scientific publications, litigation documents and regulatory filings.**

Related on our Data&Lyrics blog: [Creating An Automated Data Observatory](https://dataandlyrics.com/post/2020-09-11-creating-automated-observatory/) — [Reproducible research in practice: empirical study on the structural conditions of book piracy in global and European academia](https://dataandlyrics.com/post/2020-12-04-pirate-libraries/)

## **Advantages of reproducible research**

* Improves work habits, enhances the efficiency of analysts.
* Increases teamwork, makes the integration, substitution and replacement of team members far easier.
* Avoids duplication and multiplication of efforts. Dramatically reduces the time spend on data manipulation and debugging, error searching, formatting. This can save up to 80% of working time in analyst and consultant roles.
* Relief senior staff as oversight is far more efficient, as most errors are captured automatically.
* Better suited for cumulative growth of data, information and knowledge. In the medium run it reduces data and information cost significantly, and over the long run it produces a very strong competitive edge.
* Replication, reproducibility and the higher standards of confirmabiliy and auditability are not only scientific standards, but they are often set by market regulators, professional standards and internal working guidelines — these are far easier to achieve with automation.
* Access to the growing body of open data in the EU (such as survey data, raw data used to calculate the inflation, etc.), which is as raw data free or almost free, but has large processing costs, as it is offered by public bodies on an as-is basis. This can replace costlier and less valuable data acquisitions.

## **Our Software Solutions**

Our reproducible research solutions go through rigorous unit-testing and anonymous peer-review before release. They are fully open source, their use causes no unforeseen costs or vendor lock-in. You are even free to modify the software if you do not sell our product as commercial software.

[iotables](http://localhost:4321/software/iotables/) is a reproducible research tool that is able to work with national accounts and create some satellite accounts for all EU member states. It was originally developed to calculate the economic impacts of the Hungarian tax shelter before renewal (state aid notification at DG Competition) and for the Slovak Music Industry Report, which used similar methodology to prove that CCS sectors are overtaxed in the country. The iotables open-source statistical software library is used by about 800 practitioners in the world.

[regions](http://localhost:4321/software/regions/) solves the problem that Europe’s regional boundaries have changes in several thousand places over the last 20 years, and therefore member states and Eurostat’s regional statistics are not comparable for more than 2-3 years. This software validates, and where possible, changes the regional coding from NUTS1999 till the not yet used NUTS2021, opening up vast, valuable, untapped data sources for longitudinal analysis, or far more precise panel analysis that only national data would allow. It was originally designed in a research project at IVIR in the University of Amsterdam to understand the geographical dynamics of book piracy. Because of the needs this software fills, it had 700 users in the first month after publication.

[retorhamonize](http://localhost:4321/software/iotables/) is a software that allows the programmatic retrospective harmonization of surveys, such as the last 35 years of all Eurobarometer microdata, or all Afrobarometer microdata. Eurobarometer grew out of the need of certain CEE member states’ need to get comparable data about their music and audiovisual sectors. We commissioned surveys following ESSNet-Culture guidelines and joined their data with open access European microdata-level surveys.

[**Turstworhty AI**](https://reprex.nl/service/trustworthy-ai/)

**We follow the**[**Ethics Guidelines for Trustworthy AI**](https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai)**set up by the European Union for trustworthy AI that is transparent, validated, and free of discrimination. We create ethical recommendation systems or help the validation and assessment of third-party systems to avoid undetected biases. We believe that AI-driven sales bots should comply with competition and consumer protection rules just like the human salesforce must do.**

1. Data integration: Our experience that an investment into AI 80% about data processing and validation, and often in large extent about external data acquisition, and only in very small part about actually training algorithms. We help small organizations to pool large enough datasets and low costs to be able to deploy their own AI systems or detect biases that work against them in large corporate systems. See our [Music Observatory](http://localhost:4321/project/music-observatory/) demo project and our other services to find the relevant, high quality and affordable data to start training algorithms, or test their effect on your business or policies.  
   Related on our Data&Lyrics blog: [Why Did We Start The Demo Music Observatory](https://dataandlyrics.com/post/2020-10-27-why-start-music-observatory/) — [Feasibility Study For The Establishment Of A European Music Observatory & The Demo Observatory](https://dataandlyrics.com/post/2020-11-16-european-music-observatory-feasibility/).
2. Ethical recommendation systems: We help our clients to power educational, sales, evaluation applications based on ethical AI that is free of hidden gender, racial, ethnic, and other prohibited biases. See our [Listen Local](http://localhost:4321/project/listen-local/) demo project.  
   Related on our Data&Lyrics blog:: [Listen Local: Why We Need Alternative Recommendation Systems](https://dataandlyrics.com/post/2020-12-15-alternative-recommendations/)— [The Racist Music Algorithm](https://dataandlyrics.com/post/2020-10-30-racist-algorithm/).
3. Validation: We help our clients to understand what illegal biases may affect their sales and promotion processes by using black-box, third party algorithms on various social media, streaming and sales platforms.  
   Related on our Data&Lyrics blog: [Who Is Not Recommended On Spotify](https://dataandlyrics.com/post/2020-11-17-recommendation-analysis/) — [Forgetify: Popular Music That Nobody Listens To](https://dataandlyrics.com/post/2020-10-24-forgetify_pop_october/" \t "_blank).