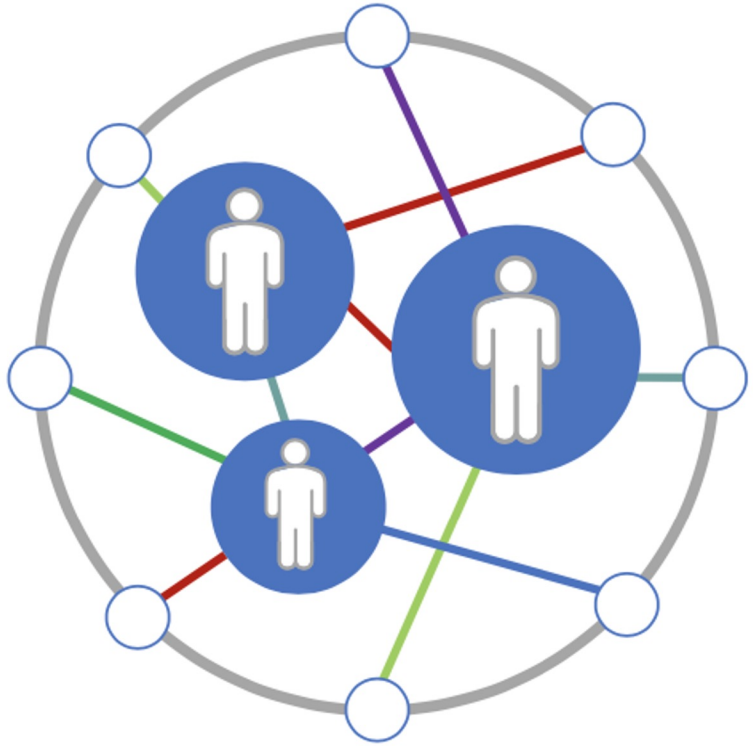
CRDSA

Clinical Research Data Sharing Alliance

PHUSE Data Transparency Summer 2022

Review of Data Sharing Policies and Protection Methodologies

June 29, 2022



What we'll cover

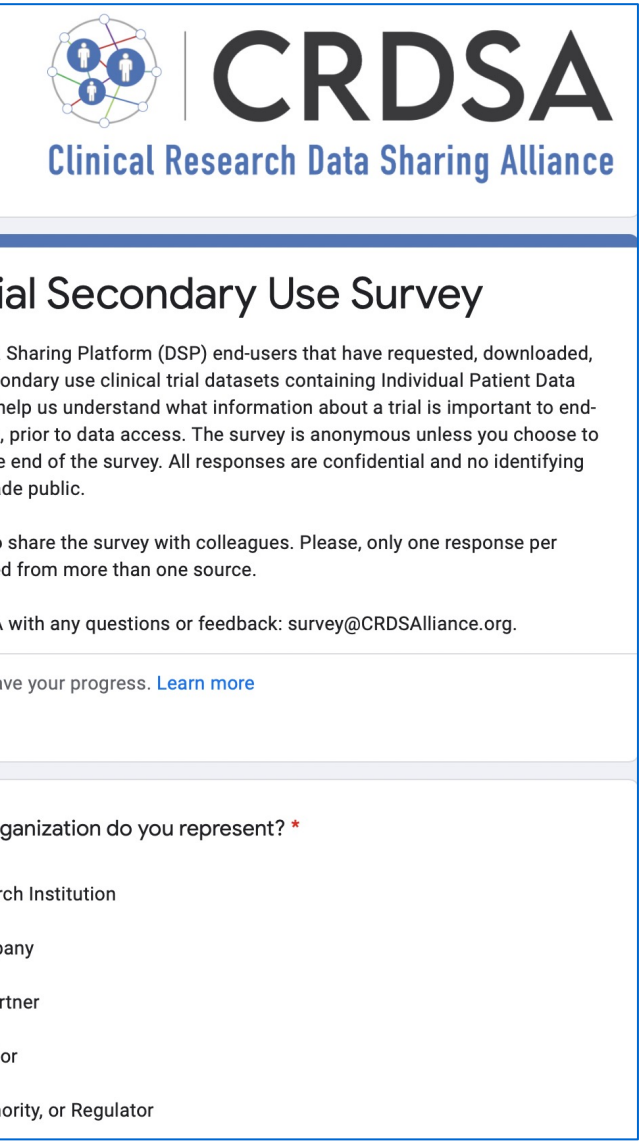
1. Strategy and Objectives
2. Methodology: A Systematic Review of Data Sharing Contributor Policies and Data Protection Approaches
3. Preview: Results - Datasets and Documentation

Before we dive in

If you haven't already, please take our Clinical Trial Secondary Use Survey!

- 10-15 minutes
- Results will be published and made publicly available later this year
- Open for PHUSE through **Thursday, June 30th**

Survey Link: https://forms.gle/b5xoeX8Dns25uuBq7



CRDSA
Clinical Research Data Sharing Alliance

Clinical Trial Secondary Use Survey

This survey is for Data Sharing Platform (DSP) end-users that have requested, downloaded, accessed, or used secondary use clinical trial datasets containing Individual Patient Data (IPD). The survey will help us understand what information about a trial is important to end-users and researchers, prior to data access. The survey is anonymous unless you choose to share your email at the end of the survey. All responses are confidential and no identifying information will be made public.

You are encouraged to share the survey with colleagues. Please, only one response per person, even if received from more than one source.

Please contact CRDSA with any questions or feedback: survey@CRDSAlliance.org.

Sign in to Google to save your progress. Learn more

* Required

Q1. What type of organization do you represent? *

- Academic Research Institution
- BioPharma Company
- CRO / Service Partner
- Technology Vendor
- NGO, Health Authority, or Regulator

CRDSA's High-Level Objectives

Current State:

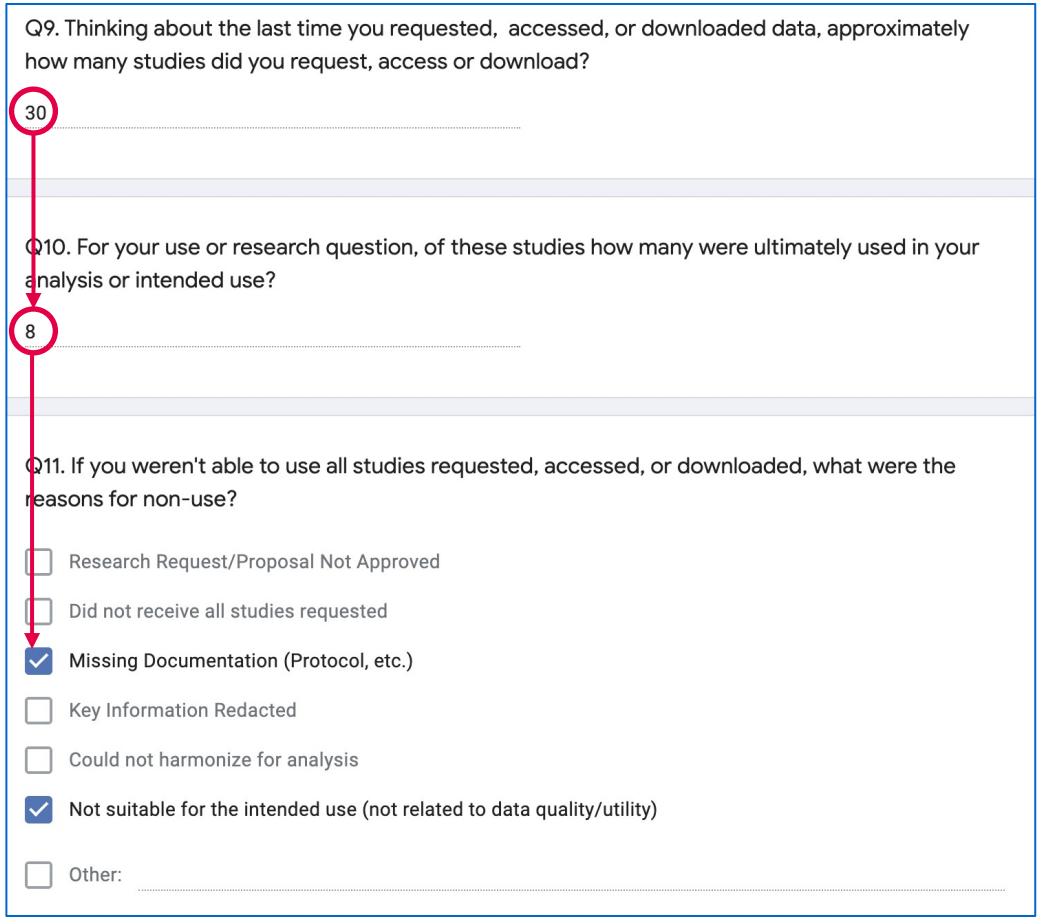
- A limited number of data contributions
- Lack of consistently high data utility
- Often high barriers to researcher access
- Data contributors face significant logistical and resourcing challenges

An example of the problem:

Overarching Objectives:

- Sponsors **share openly**, **timely**, with contributions maintaining **high data utility.**
- Researchers know **where to go**, **what to expect**, and **how to** make the best **use** of secondary data.

From our Secondary Use Survey:



Data Protection and Secondary Use Standards WG's

Data Protection

Policies and Protection Methods Review Whitepaper

- Establishes data sharing **benchmarks** for sponsors

Data Protection Playbook

- Targeted at **executive-level** pharma decision makers
- How to put it into practice (for **internal and external** data sharing)

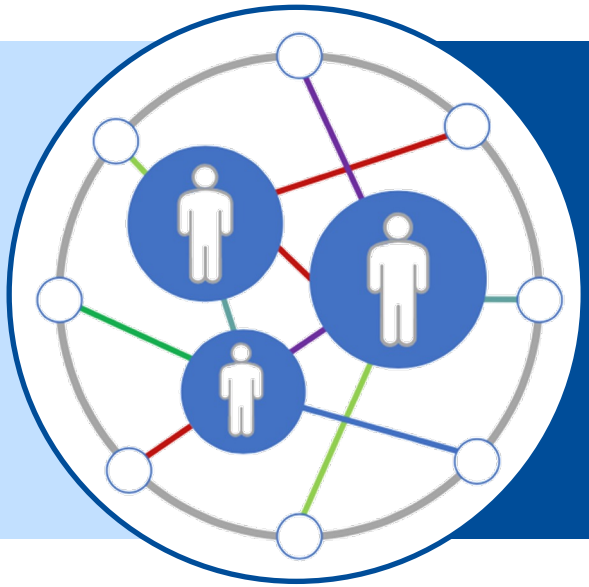
Secondary Use Standards

Secondary Use Survey and Manuscript

- Establishes **what researchers need** in order to make best use of contributed data
- Gives sponsors **meaningful anchor points**
- **Path to** reports/scorecards and **Standards**

What are we trying to change?

1. **Mindset**: Sponsors actively engaging with the data sharing ecosystem, as both contributor and data consumer
2. **Governance**: Sponsor policies that responsibly balance patient privacy while maintaining end-user data utility
3. **Resourcing**: Sponsors allocate sufficient internal resources to data contributions, moving towards universal prospective data sharing



Systematic Review Methodology

Building on the Good Pharma Scorecard*

The 2021 Good Pharma Scorecard scorecard examines 5 data-sharing measures (data transparency is evaluated separately):

1. whether they have a public policy committing to sharing analysis-ready datasets and clinical study reports (CSRs) for applicable studies;
2. whether their policy explains how such data can be requested;
3. whether the policy commits to making data available by 6 months after approval by the FDA or European Medicines Agency or 18 months after a trial's completion date, whichever was later;
4. whether the company reports the number of data requests received and how each was handled (granted or denied);
5. the proportion of 'data sharing applicable' trials registered in a public registry.

* Bioethics International

Our focus: End-User Research Utility

- Of the 42 sponsors in the 2021 GPS survey, **55%** (23) met criterion #1 by contributing both analysis-ready datasets and CSRs
- Of the 29 sponsors in the CRDSA review:
  - 15 overlap with the GPS scorecard
  - **69%** (20) meet the GPS criterion

The Data Protection Work Group review builds on the GPS work by:

1. **Expanding** the range of datasets and documentation
2. **Adding** analysis of applied data protection methodologies

Next, let's look at the 4-step process applied in our systematic review…

Step 1: Determine Sources

Data Sharing Platforms:

- Clinical Study Data Request (CSDR)
- Vivli
- YODA

Sponsors

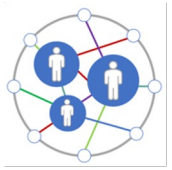
This section of the site provides information on Sponsor's criteria for listing studies and other relevant Sponsor specific information.
Select the Sponsor's logo to view this information.

"Sponsor" means a person who takes responsibility for and initiates a clinical investigation. The Sponsor may be an individual or pharmaceutical company, governmental agency, academic institution, private organisation, or other organization.

astellas, BAYER, CHUGAI (A member of the Roche group), Eisai, GRÜNENTHAL, gsk, NOVARTIS, ono, sanofi, SHIONOGI, Sumitomo Pharma, sunovion, ViiV Healthcare, wellcome

Sponsor Specific Details: Bayer

Studies listed	Clinical trials in patients sponsored by Bayer that have been submitted to the US and EU regulatory agencies as part of the request for marketing authorization for new medicines or indications for products that have been approved by both of the agencies on or after Jan 01, 2014. For products that have not received market approval in either US or EU before Jan 01, 2014, data will be shared if they have been approved on or after Jan 01, 2014 by only one of the agencies, provided clinical trials are not part of an ongoing review or are going to be part of a planned submission to a highly regulated regulatory authority such as EMA or FDA.
Exceptions	Bayer will not share data when we believe that there is a reasonable likelihood that the individual could be re-identified, for example, clinical studies of rare diseases, single-center clinical studies, or clinical studies with a very small number of subjects. Bayer may not be able to share data in case of substantial practical constraints to provide technical data access. For example, there may be issues related to the size of databases from genetic studies. Where clinical data were collected subject to legal, contractual or consent provisions that prohibit transfer to third parties, or in case of substantial conflicts of interest, data access under this policy may be precluded.
When studies are listed	Studies are considered for inclusion in the "Study list" (see section "Studies listed") after regulatory approval. In case of planned scientific publications, studies will be listed once the primary manuscript describing the results has been accepted for publication.
Additional conditions for data access	When patients agreed to take part in Bayer clinical studies they gave permission (through informed consent) to use their data to study the medicine or disease Bayer was researching. Further research must therefore study the medicine or disease that was researched in the original studies. For studies Bayer plans from May 01 2014, patients will be asked to give permission to anonymization of their data so further research can be conducted with data from these studies. For partnered compounds data access will be provided if the development partner agrees to data access and to related data sharing agreements with the requesting third party. Data Access can be provided when it has been technically implemented (planned by Dec 2014). In exceptional circumstances, access to data may be declined by the sponsor, for example, where there is a potential conflict of interest or an actual or potential competitive risk.
Datasets and documents provided	Where available, the following anonymised patient level data and information is provided for each clinical study. **Raw dataset.** This is the dataset collected for each patient in the clinical study. **Analysis-ready dataset.** This is the dataset used for Bayer's analysis. **Protocols with any amendments.** This describes the objectives, design, methodology, statistical considerations, and organization of a clinical study. **Annotated case report form.** This is a blank case report form with descriptions of the data collected and how they are described in the dataset. **Reporting and analysis plan.** This describes methods of analysis, procedures for data handling Bayer used for the study. **Dataset specifications.** This is the meta-data which describes the datasets e.g., variable labels, variable descriptions, code lists, formats. **Clinical study report.** This is the report of efficacy and safety data from the study. It forms the basis of submissions to regulatory authorities such as the Food and Drug Administration (FDA) and the European Medicines Agency (EMA). Appendices that contain no patient relevant data (e.g. audit certificates or manuals from providers of various diagnostic services) as well as individual patient level data (e.g. narratives) are not included. The data from narratives will of course be contained - in aggregated fashion - in the datasets provided by Bayer. Documents will be redacted to protect personal data of study participants, study personnel, and Bayer employees, and to protect Bayer's commercially confidential information, including intellectual property rights.
Enquiries	Responses to enquiries will be provided for studies listed. For studies not listed please refer to Bayer's TrialFinder application
Access to study documents without patient-level data	Access to study documents without patient-level data will be provided for studies listed. For studies not listed please refer to Bayer's TrialFinder application.
Review criteria for enquiries	Please see section above and refer to the sections "Studies listed" and "Exceptions".
Anonymisation standards	



Step 2: Collect

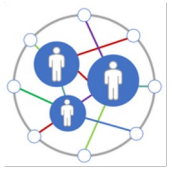
Collected and organized information supplied by a total of **29 pharma** sponsors

- Our analysis was limited to the information supplied on the platform site
- Detailed academic sponsor information was not available, therefore not evaluated
- Some information may be out-of-date but…

This is what researchers will see (and rely on)!

The supplied information is a mix of:

- Informational statements:
  - "Bayer will not share data when we believe that there is a reasonable likelihood that the individual could be re-identified, for example, clinical studies of rare diseases, single-center clinical studies, or clinical studies with a very small number of subjects."
- Data Points:
  - Datasets and documents provided (raw dataset, protocol, etc.)
- Linked in-depth documents:
  - "Bayer Clinical Trial Data Transparency Anonymization and Data Protection Procedures" PDF



Step 3: Collate and Organize

29 Sponsors with interpretable information:

The information was then collated and organized into tiers based on sponsor size (determined by employee count):

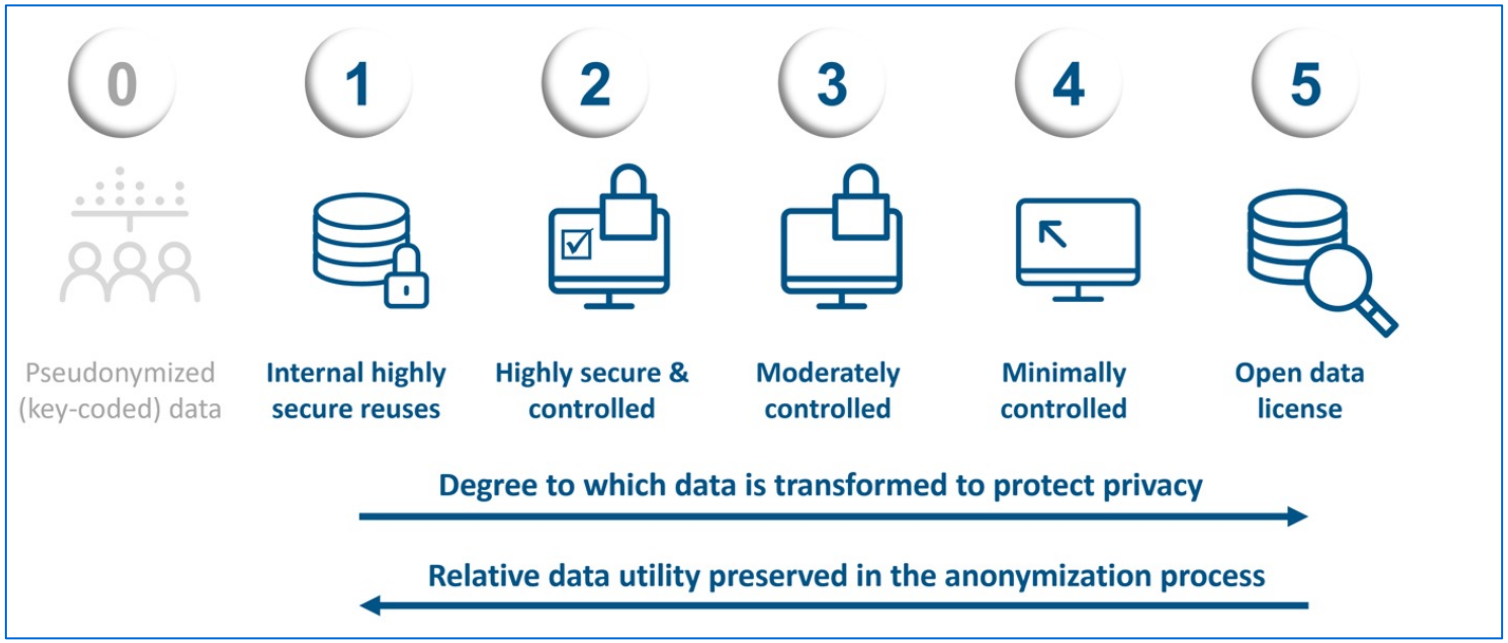
- **12 Tier 1**: Over 25k
  - The smallest 1 is approximately 45k employees
- **11 Tier 2**: 5 to 24,999k
- **6 Tier 3**: Under 5k
  - Range 1,200 to 3,500

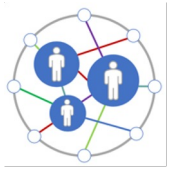
Step 4: Analyze and Interpret

In addition to synthesizing the information supplied, the team developed a consistent methodology to categorize data protection approaches:

Objective: Determine if a sponsor uses a Risk-Based data protection approach

Why: A risk-based approach to data protection is generally thought to enhance the end-user utility of data contributions, eg, SAFE data rating by Bamford et al. (Applied Clinical Trials, 2022)





Step 4: Analyze and Interpret

Criteria: used to classify an approach as ”risk-based:”

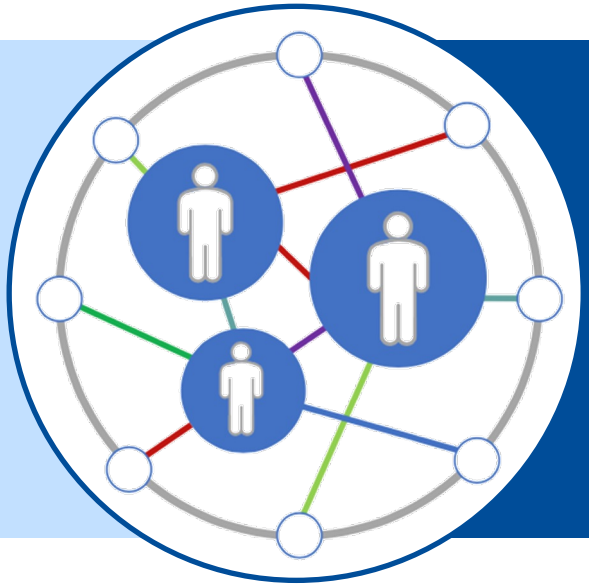
I. Explicit claim by the sponsor that their anonymization approach is “risk-based”

II. A mention of a “risk assessment” in addition to a specified rule-set

III. The extent an approach to anonymization is modified based on specific factors: study population, disease prevalence, data sensitivity, system controls, context risk, various re-id attack scenarios, and adversary profiles

Assessment:

- Criterion (i) is classified as using a risk-based anonymization approach
- Criterion (iii) is classified as using a risk-based anonymization approach and we can also infer the use of a quantitative approach
- Criterion (ii) is possibly a risk-based anonymization approach
- Else, if a sponsor uses a rule set minus the risk assessment component, then we infer a rules-based approach used



Preview: Datasets and Documentation

From the Review: Datasets and Documents

Data Protection	Tier 1	Tier 2	Tier 3	Total
Systematic Review	12	11	6	All Tiers
Datasets and Documents				
Raw	100%	73%	83%	**86%**
Analysis	92%	82%	67%	**83%**
Protocol	100%	73%	83%	**86%**
aCRF	100%	64%	67%	**79%**
Reporting and Analysis Plan / SAP	100%	73%	67%	**83%**
CSR	92%	82%	33%	**76%**
Dataset Specifications	75%	64%	50%	**66%**

- Largely consistent large sponsor commitment provides valuable benchmarks for mid and small sponsors

The full review results and analysis will be available in a forthcoming white paper

Working Together: Review and Survey

Data Protection	Tier 1	Tier 2	Tier 3	Total		Secondary Use Survey
Systematic Review	12	11	6	All Tiers		INTERIM RESULTS
Datasets and Documents						
Raw	100%	73%	83%	**86%**		**76%**
Analysis	92%	82%	67%	**83%**		**86%**
Protocol	100%	73%	83%	**86%**		**80%**
aCRF	100%	64%	67%	**79%**		**70%**
Reporting and Analysis Plan / SAP	100%	73%	67%	**83%**		**52%**
CSR	92%	82%	33%	**76%**		**58%**
Dataset Specifications	75%	64%	50%	**66%**		**80%**

Secondary Use Survey:

Open for PHUSE through June 30

- The Secondary Use Survey analysis will look at multiple response dimensions including:
  - Results by user type (e.g., academic vs. industry)
  - Predominant data use (e.g., Publication vs. Regulatory Use)
- What are the education gaps and opportunities?

Thank You!

Aaron Mann
Chief Executive Officer
Clinical Research Data Sharing Alliance
+1 917-520-2346
aaron.mann@CRDSAlliance.org

Luk Arbuckle
Chief Methodologist
Privacy Analytics
+1 613-816-1453
larbuckle@privacy-analytics.com