melissa
INFORMATICS

A **MELISSA** WHITEPAPER

7Cs for Data Quality
in Healthcare and Life Sciences

Your data quality scorecard

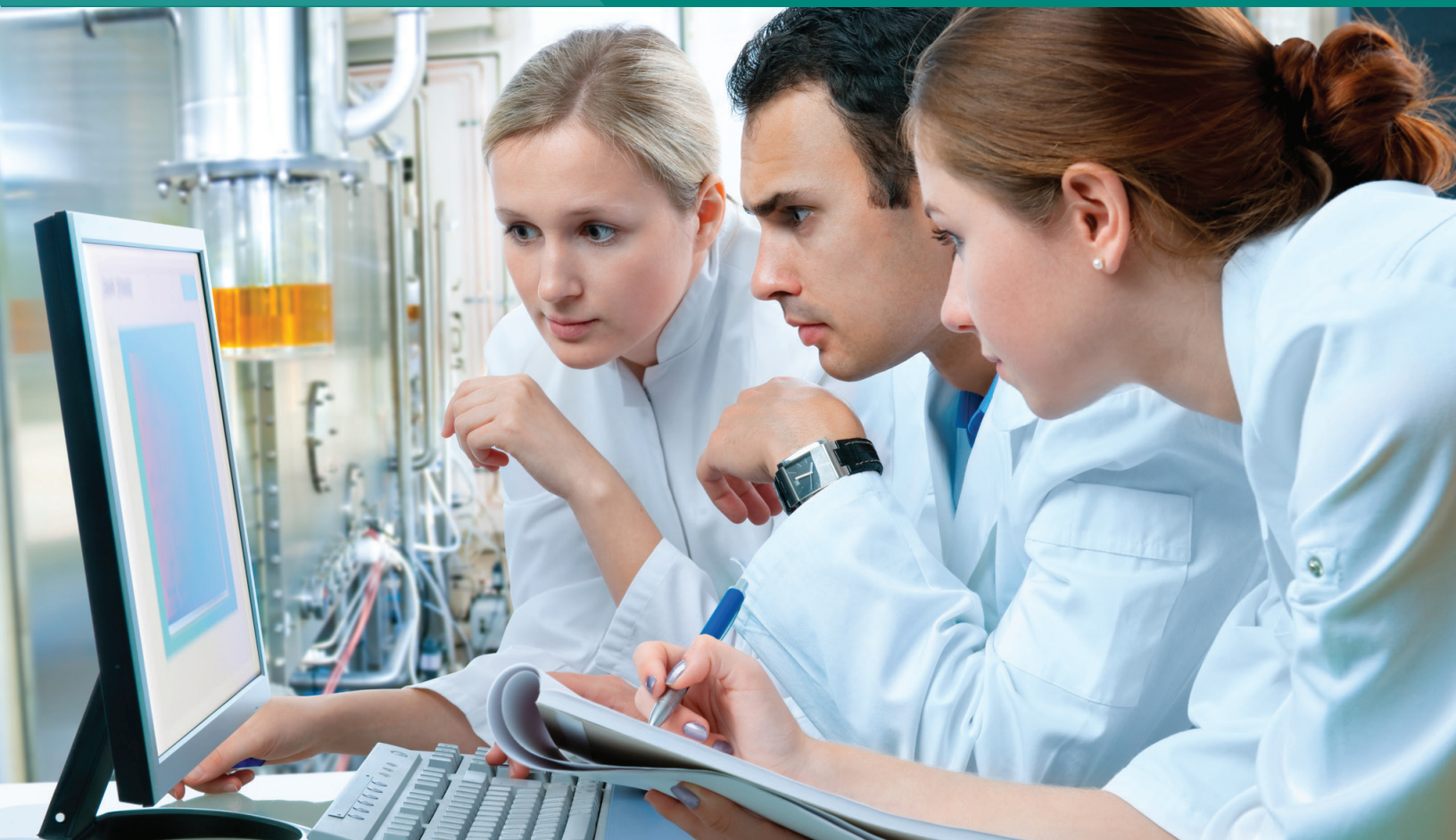


TABLE OF CONTENTS

HOW DO YOU SCORE ON THE 7 CS OF DATA QUALITY?

One way or another, almost everyone in health care and life sciences industries must address the issue of data quality. Why? Because dirty, disconnected data represents lost intellectual property, while clean, well-connected data is prepared to deliver knowledge, reduce costs and boost the bottom line. Data quality makes innovative analytics possible and reduces cost and time to market, while improving efficacy for new drugs and precision medicine products. Organizations with healthy data and effective long-term data quality and integration strategies save money on research, safety, trials and delivery. They formulate effective scientific, clinical and business strategies and marketing campaigns based on accurate data. Their customers experience higher levels of satisfaction. But how many business understand how to evaluate and correct issues related to data quality?

In today's data-driven marketplace, data quality issues can no longer be downplayed or farmed out to internal development. The health of your company's data impacts initiatives and departments ranging from early stage discovery, development, safety and efficacy to regulatory submissions, trials, marketing, sales, billing, accounting, clinical care and compliance.

Melissa Informatics' aim is to help researchers, business managers and executives, marketing professionals and other personnel understand what data quality is, why it is important, and how they can quickly clean up and connect their company's mission-critical data.

That's why the 7 Cs of data quality in healthcare life sciences are essential. We'll look at each one of the 7 Cs in detail so you can absorb and apply the fundamental principles of data quality. The 7 Cs are:

1. **Clean –** curated by deep data identification and correction methods for certified accuracy
2. **Connected –** meaningfully linked and searchable across previously disconnected data sources
3. **Coordinated –** improved interoperability with other existing and new data sources
4. **Compliant –** aligned with FAIR, HIPAA, EMEA, GDPR guidelines as required
5. **Cost Effective –** prepared for revenue-bearing value at lower near and long-term cost
6. **Consumer Centric –** drug-centered, patient-centered, consumer-centered data as required, to more effectively profile, communicate with and bring life saving value to your market
7. **Confident –** certified accuracy through automated QA, with provenance and QC functionality, so you can be confident your data is correct

The 7 Cs are building blocks for data quality. They provide a quick reference 'scorecard' to help businesses - whether large or small - assess the health of their organization's mission critical data. No matter whether your business is located in Europe, Asia or North America, data quality is a must for the 21st century.

HOW CLEAN IS YOUR DATA?

Do you know what percentage of your historical clinical trials data was correctly entered?*

Companies are easily spending from 20 to 80 million dollars or more for Phase 1-3 clinical trials. Electronic Data Capture (EDC) for phase 1 trial alone, with 200 patients tracked over 10 visits, costing $500 / visit for data acquisition, can cost over a million dollars – just for a small, preliminary trial.

How can you be confident that the information in your new database is accurate? Business can now employ a solution that will validate, correct, and standardize clinical trials data – even without using expensive specialized EDC applications and methods.

Melissa Informatics can take data directly from EMR systems to create high quality research data, aligned as needed with FDA guidelines for clinical trial submissions. This is true of data from any source, including from discovery and safety studies, historical trials data, ongoing EMR data collection.

Save money by using the data you are already acquiring, rather than by creating expensive new EDC data acquisition efforts.

**"Optimizing the Use of Electronic Data Sources in Clinical Trials: The Landscape, Part 1". Ed Kellar, Susan M. Bornstein, Aleny Caban. October 13, 2016 Sage Publications, Therapeutic Innovation & Regulatory Science 2016, Volume: 50 issue: 6, page(s): 682-696. Research Article http://journals.sagepub.com/doi/full/10.1177/2168479016670689*

IT'S TIME TO GET YOUR DATA CONNECTED

Dirty, disconnected data costs the Pharma industry billions of dollars a year to address. Even with these costs, more projects fail (and more money is spent) due to challenges with data harmonization and integration.

Extensive data harmonization may be required to connect two datasets. Spelling errors and data entry typos aside, one database may use the term "acetaminophen", while the other calls the drug "Tylenol". Additionally, duplicate data need to be handled and preferred terms need to be defined. Traditional methods for harmonizing and connecting datasets are brittle. You may choose to harmonize to one standard (for example, FDA drug terminology). What if you need a different set of preferred terms later? Harmonizing and re-harmonizing data to different standards takes too much time and money.

Now you can align your data to different standards for harmonization "out of the box". Melissa provides rich resources for standards alignment and easy re-alignment to different preferred terms and standards as needed, without refactoring the entire database.

CREATING THE GOLDEN RECORD

The process of producing a golden record (also known as survivorship) is a key goal for connected datasets. This is a central step in the record matching process involving partially duplicated data. It requires identifying the record with the best data quality. There are three common techniques in determining the surviving record:

Most Recent – the most recent record can be considered elegible as the survivor.

Most Frequent – matching records containing the same information are also an indication for correctness

Most Complete – records with more values populated for each available field are also viable candidates for survivorship.

Melissa Informatics can go beyond "generic survivorship" techniques to leverage context and reference data. This makes more sophisticated understanding of data contents possible. With Melissa, you can pull from multiple records, applying machine reasoning and reference information in order to create golden records – even when original data quality is poor or inconsistent.

Now you can employ Melissa Informatics' solution to connect data in innovative ways. Data models or "ontologies" link data in ways that make connections easier to create, easier to modify for different purposes, more useful for creating "golden records", and ready for deep machine reasoning.

Save time and money getting and keeping data connected – and open up new opportunities for discovery – by connecting your data under flexible and meaningful NoSQL "ontologies".

OH NO, WE DON'T WANT ANOTHER BIG DATA WAREHOUSE!

Along with the problem inconsistent, disconnected data, another "dirty secret" plagues the healthcare life sciences industry. This is the lack of coordination across existing and new data resources. Traditional databases aren't readily extensible, or prepared for integration with new data.

How many times has your company completed a big new "integrated data warehouse" project only to find that the expensive data warehouse isn't prepared to connect to that other data resource you just realized you need?

With Melissa, any new integrated data resource you create is prepared for connection to other resources without refactoring the original database schema and without rebuilding an entirely new data warehouse. Adding new data is as easy as adding a new friend to your social network. Just connect the data to any common classes and elements within the existing data network, and add new links for the new data. It's that easy.

If you have two or more datasources that have been cleaned and connected by Melissa, integrating them is as easy as "drag and drop". Pull the two datasets together within Melissa's semantic database environment and they will automagically connect on all common classes and terms.

Melissa has customer examples where two completely different groups, without explicit cooperation, each applied the same open semantic standards for their integration process. Later it was desired to connect these two databases. Following the semantic methods employed by Melissa, these databases connected without additional integration effort. The customer calls this "coordination without cooperation". This represents a huge step forward for data interoperability.

Ensure long-term value for your integrated data resources by making sure your methods are coordinated to ensure extensibility and interoperability at lower time, effort and cost.

ENSURE COMPLIANCE WITH REGULATORY GUIDELINES AND STANDARDS

If you've never been a member of a standards definition body, or if you've never been a part of the discussion about what standard(s) you need to align your data with, consider yourself lucky. There has been much confusion and disagreement over the past decades about standards. Billions of dollars have been spent defining standards, determining what standards should be used, why they should be used, and how they should be applied.

Standards are useful. They provide critical common concepts and terms and useful disciplining structure. With Melissa you can align your data with the standard you need, "out of the box". Melissa brings a curated set of published standards to you. We can help you meet FAIR, HIPAA and other guidelines and standards at lower time and cost. We can also combine, create, apply and even modify standards to meet your specific needs

THE HIPAA EXAMPLE

HIPAA provides an excellent real-world example of the relationship between data quality and compliance. In response to HIPAA, health care providers must comply with specific government regulations regarding data handling, data sharing, data access, submission claims, even billing procedures.

One provision of HIPAA 5010 deals specifically with, for example, address contact data. To be in compliance, providers must include a full 9-digit ZIP Code for billing provider and service locations on all claim submissions. Another HIPAA rules prohibits the use of a PO Box for the billing provider address. If a health care provider's data is in chaos and they cannot comply, they will have to pay the price.

While the specifics of compliance are different depending on the industry and regulation in question, the bottom line is that compliance issues intersect with almost every facet of most health care and life science data, forcing businesses to look long and hard at improving and maintaining their data quality and compliance.

MAINTAIN COMPLIANCE OVER TIME… WITH AGILITY

Standards can also be constraints! What if you aligned your data with one standard or set of standards and lexicons, then learn that you would like to use the data for other purposes, requiring other standards? This often happens. For example, you may have defined your data according to a narrowly defined "SDTM" standard for study data. Later your boss determines that this data is ready for submission to FDA, perhaps for a New Drug Application (NDA). This requires the CDISC standard.*

With traditional methods, realigning your data to a new standard can take more time and cost than you have to spend. By the time you've redefined your database to handle the changes classes and relationships required by the new standard, the window of opportunity for that submission has passed. You need technology that can make it possible to align your dataset with new standards without refactoring the database.

With Melissa, you can benefit right away by the application of formal standards for data harmonization, integration and compliance. However, you aren't trapped by your initial decisions. Now you can more efficiently align and re-align your data with different standards as you need them.

**https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM511237.pdf*

What does FAIR stand for?

To meet FAIR guidelines*, data must be findable, accessible, interoperable and reusable while maintaining regulatory compliance for protected health information, contractual and business compliance, security, and confidentiality. To ensure Findable data, Melissa Informatics can add structure and metadata to your data, including searchable globally unique persistent identifiers. We apply the World-wide-web Consortium's (W3C) Resource Description Framework (RDF) standard for unique resource identifiers (URIs) to be provided for all data and metadata elements in a secure, standards-based data source. To be Accessible, (meta)data can be made be retrievable by their identifier using a standardized communications protocol that is universally implementable with authentication and authorization; and metadata will be managed to ensure accessibility. To be Interoperable, all data can be set to follow accessible, shared, and broadly applicable language and methods for knowledge representation, using standards and lexicons that follow FAIR principles. To be Reusable, data can be described with a plurality of relevant attributes, released with a clear and accessible data usage license, associated with provenance, meeting domain-relevant community standards. Ask us about helping you meet FAIR guideline with your data!

**https://www.nature.com/articles/sdata201618*

HOW MUCH IS BAD DATA COSTING YOUR ORGANIZATION?

U.S. businesses lose over $600 billion a year because of bad data, with more than 25% of that due to customer data entry errors. This problem has grown exponentially for clinical data acquisition. Melissa Informatics has examples from electronic medical records system where a single drug was recorded in over 190 different ways. Dealing with this sort of complexity in order to create unified, accurate, research or submission-ready datasets can be very expensive

How much money would your company save every year if your data could be prepared for revenue-bearing value at lower near and long-term cost? Are you sure you're finding all of the internal data (and sufficient external published data) relevant to your research project – before you start? Can you be sure every patient in this cohort are really not taking any contraindicated drugs for the clinical trial treatment? To answer these sorts of questions, you need to understand how much bad data is costing your organization right now.

KEEP BAD DATA OUT

Consider the "1-10-100 Rule" which posits that it takes $1 to verify the accuracy of a record at point-of-entry, $10 to clean it in batch mode, and $100 per record if nothing is done (this includes costs associated with repeated experiments, undetected safety issues, clinical trials rejections and over-prescribed drugs, etc.). Therefore the best ROI can be attained by employing a "data quality firewall" at point of entry, to immediately verify the accuracy of information. For historical data, best ROI can be attained by automating the data validation process.

With Melissa, real-time data entry solutions are available via services, and enterprise integration platforms can be provided to automate historical data cleansing and integration. Save money by keeping bad data out and by turning your old 'dirty' data into "golden' data, ready for accurate research and business analysis.

BRING YOUR DATA TO LIFE OR LEAVE IT IN A "DATA TOMB"

Has your organization completed major data acquisition or data cleaning and integration projects only to realize the data wasn't effectively targeted to the actual requirement or consumer? It can be difficult if not impossible in many cases to understand every question your internal or external customer will want to ask of the data… before they've seen it.

Billions of dollars are spent annually on new data acquisitions and new integrated data warehouses, in projects that end up languishing and under-utilized because the 'final' dataset isn't meeting the actual, ongoing needs of the customer.

RIGHT DATASETS, RIGHT STANDARDS, RIGHT TIME

In order to be sure you have the data your customer needs, you need to be able to more efficiently add new data, take data out (for example, quickly remove all required identifying information when needed for patient research data), align the data with new terminologies and linkages as needed to meet your research, partnering or business goals.

Melissa's semantically enabled "noSQL" data is purpose-built to support rapid extension to add new data sources, "no-re-factoring" modifications to change or remove existing data structure and content as required to meet new goals, and to align and re-align with different terminologies and standards as needed, without breaking the existing database

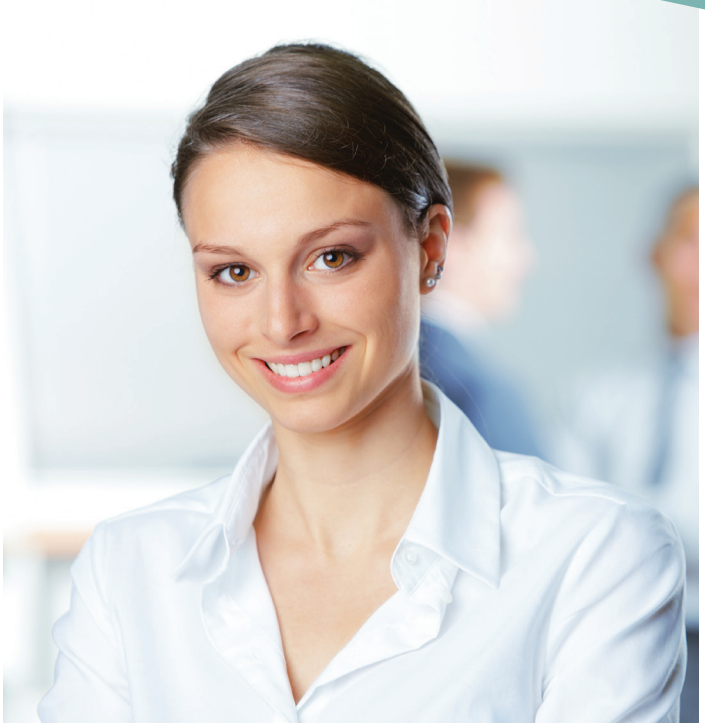
With Melissa, you can be confident that you'll have the right data when you need it, aligned with the right standard for the customer, when you need it.

IT'S TIME TO BECOME E-CONFIDENT

Repeated or failed research due to missing, disconnected data or cost and time required to clean and integrate data costs organizations billions annually. Do you have confidence that you are using the data that you have already created, at great cost and expense? Are you confident that the data you are using is accurate? Does it take too much time to clean and integrate the data you need to meet your goals?

In today's competitive global markets, you can't afford to continue operating under low confidence data environments. Verifying research and medical data at the point-of-entry eliminates inevitable costs of bad data. Automating data integration and curation makes it possible to avoid an exponential increase in downstream costs.

With Melissa Informatics, you can be confident you have the data quality you need to reduce costs and maximize revenue for your business.



Take the Next Step

So now that we've run through the 7 Cs of healthcare life science data quality, how would you rate your business? Are you a master data steward of is there room for improvement?

Regardless of whether data quality is already a top priority at your company or if you are just beginning to discover its key concepts, we're certain you can utilize the 7 Cs of data quality to quickly make a significant contribution to your organization's bottom line. Remember: the data quality investment you make now will pay dividends down the road in terms of improved outcomes, saved time and money.

Now that you've had your introduction to the 7 Cs, it's time to take the next step. Visit Melissa Informatics at **www.melissainformatics.io** to learn more about where your business stands today in terms of data quality, and to explore how we can help you achieve the 7 Cs at the lowest time and cost with the highest ROI.

melissa
INFORMATICS

www.melissainformatics.io

About Melissa Informatics

Melissa Informatics extends the capabilities of Melissa's global intelligence software and services to support world leaders in life sciences, biotechnology, pharmaceutical, and medical industries by harnessing the entire data lifecycle for business, pharmaceutical and clinical data. Our software and services bring data quality and machine reasoning together for insight and discovery by intelligently cleaning, connecting and harmonizing multiple sources to offer interoperable data. Melissa Informatics reduces time and cost to benefit from clean, richly connected data, and reveals deeper data relationships from complex, dynamic data through machine reasoning operations for reliable information in mission critical healthcare and life science informatics.

US

22382 Avenida Empresa
Rancho Santa Margarita, CA 92688-2112

800.MELISSA (635.4772)

UK

+442030510140

GERMANY

+493079788829

INDIA

+918065603282

AUSTRALIA

+61280917500