



Immediate and Ongoing Benefits and Monetization Outcomes: Parkinsons Institute and Clinical Center (PICC), "Parkinson's Insight"

Background

For over 30 years PICC has procured a valuable collection of clinical data. There is no other single institution or organization in the world that can match the quality and quantity of PICC's Parkinson's Disease related patient data.

In 2015 Melissa Informatics software and services were deployed at Parkinson's Institute and Clinical Center in Sunnyvale California (PICC), to integrate a substantial portion of the Institute's data.

Summary of Technical Goals and Data Strategy

Technical goals included creating a new, unified "Parkinson's Insight" data resource. Business goals include researching and publishing discoveries from that data, and engaging remunerative partnerships based on the new data resource.

Having achieved the initially defined technical and business goals, Melissa Informatics and PICC continue to expand data integration and partnership opportunities, in large part funded by revenue achieved from the original projects.

Melissa Informatics demonstrated value quickly in a "Phase 1" proof-of-concept, by exploiting disconnected internal and external data resources at PICC that, if harmonized and linked, showed promise for scientific discovery. To do this, Melissa Informatics first integrated "Study Data" from multiple clinical trials and patient data studies reflecting over 70 separate data sources. This included information from over 3,400 biological samples, and approximately 11,000 Parkinson's disease-related assessments, from approximately 3,500 individuals both with and without Parkinson's disease.

Company:

Parkinsons Institute and
Clinical Center (PICC)

Website:

<http://www.thepi.org>

Industry:

Nonprofit

Customer Since:

2015

Team Size:

51-200 employees

Why Melissa Informatics?

- Overcome barriers to meet your data-driven business goals
- Ensure long-term, growing value from your current and future data resources
- Turn your data ore into gold

// Working with Melissa Informatics to realize high quality integrated data from internal and external sources, we've pioneered new applications, widely cited publications and innovative revenue-bearing partnerships. Melissa has made it possible for us to transform our complex and diverse data into a unified, research-ready knowledge resource.

//
– Carrolee Barlow, M.D., Ph.D.,
Chief Executive Officer, PICC

Next, in "Phase 2", we integrated content from PICC's Electronic Medical Records (EMR) systems. This data was housed in two EMR applications that could not be effectively searched independently and could not be searched at all as a unified resource. Melissa Informatics Sentient software was applied to integrate over 70,000 diversely structured, poorly structured and unstructured clinical notes, from over 5,000 patients, collected over more than a 10 year period. The initial integration included Medications, Vital Signs, Race and Ethnicity, Hoehn & Yahr Stage and UPDRS Part III Motor Score Assessments. All of the data was made ready for analysis by statistical packages such as SAS and R, and was de-identified for research purposes - although identified data is also available for permitted PICC clinicians and researchers.

Results - multiple high value Use Cases including publications and revenue

First, the Sentient software system was applied to create an integrated Study Data resource. A key element of this project was to catalogue the various tissues and create a table that defined all of the various resources at the Institute. Subsequent analysis of identified data resources helped uncover and describe poorly understood relationships between different genes and different Parkinson's disease patients, and supported publication of a major Nature Genetics article in late 2018. This publication has become one of the most highly cited articles in the Parkinsons Disease research space (See J William Langston, et al. Multisystem Lewy body disease and the other parkinsonian disorders. Nature Genetics volume 48, (2016)).

Second, Sentient was applied to create an integrated "Real-World-Evidence" resource, drawing from historically and actively acquired EMR data. PICC and Melissa Informatics were able to generate a deep research quality dataset for one Pharma partner, centered on a specific drug of interest to do a "virtual clinical trial". Melissa identified a research cohort (treated with the medication of interest and meeting other requirements) and a Control group for the research. This partnership engaged in a radically innovative method development to utilize EMR data, including content from both structured (XML, tables) and unstructured (notes) content. Traditional extract, transform, load (ETL) methods were combined with advanced machine reasoning through application of semantic artificial intelligence ("AI") to create a dataset suitable for research.

Data acquisition for clinical trials (commonly called electronic data capture or "EDC") can be very expensive, commonly adding \$500, up to \$1,000 or greater cost for each patient visit. Using Sentient, data added to an EMR system as part of a normal visit can be extracted and transformed to deliver clinical trials research quality data, directly from the EMR, without any special additional data capture. Saving \$500 or more per visit, for trials addressing thousands of patients, promises substantial savings in time and cost for clinical trials.

Delivering these datasets has resulted in substantial (undisclosed) revenue events for PICC. This new "Parkinson's Insight" resource, built on Sentient software, provides ongoing new sales and partnership opportunities, and recurring revenue opportunities as current Pharma partners come back for more data and next trials.



Phase 1: Proof of feasibility, Searchable Internal and External Integration, Nature Genetics Publication

This initial integration made it possible for PICC to better understand the value of the underlying data, and allowed PICC to ask the following types of questions:

- Find all patients in the “LRRK2 Study” that are under 65 that have a DNA sample taken on a given date
- Find all patients that participated in at least two studies and have family history of Parkinson’s disease
- Report all cryo-preserved fibroblast cell lines for female patients in the PICC’s Induced Pluripotent Stem Cell Inventory

Challenges Solved by Melissa Technology for Study Data Integration

Prior to integration within Parkinson’s Insight, study data for patients were all kept in different “unlinked” data sources in varying formats. Following the successful first feasibility project, the data has been logically integrated for centralized querying. For the proof of feasibility, queries used the technically advanced “Knowledge Explorer” software pictured below.

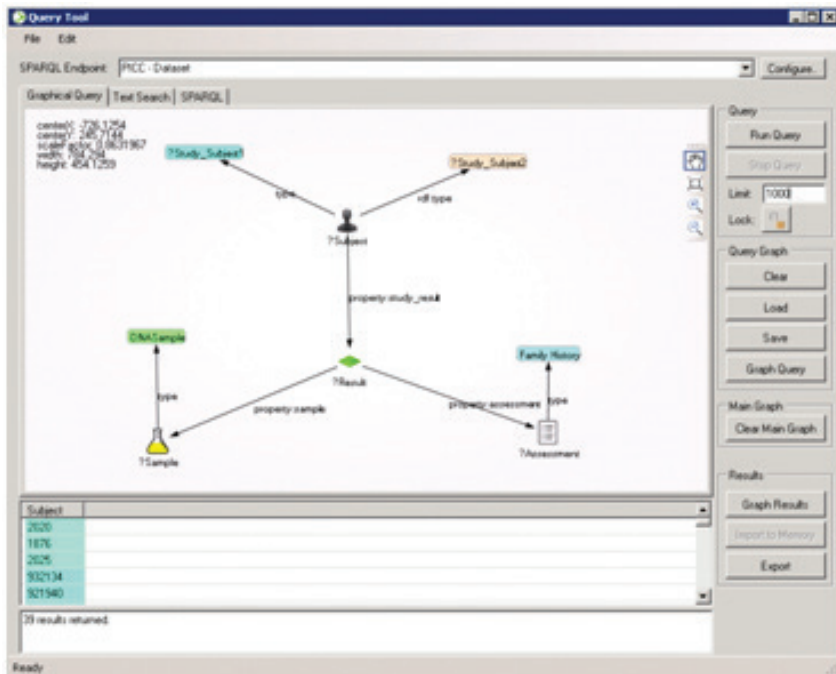
The screenshot shows the Query Tool interface with a SPARQL query graph and a table of results. The query graph is a logical model for the query: `LRRK2Subject` is a `?subject` of type `LRRK2Subject`. The `?subject` has a `property:age` of `?age < 65`. The `?subject` has a `property:study_result` of `?result`. The `?result` has a `property:diagnosis` of `?Diagnosis` and a `property:assessment` of `?UPDRS_Assessment`. The `?Diagnosis` is of type `?Diagnosis`. The `?UPDRS_Assessment` is of type `UPDRS_Assessment`. The `?result` has a `property:DNA_sample` of `?DNASample`. The `?DNASample` is of type `DNASample`.

subject	age	DNASample	UPDRS_Assessment	Diagnosis
202	59.7	bn-5ba8522-650a-439	bn-d3adfe71-ad5c-40e6-b3b4	PD
932	50.7	bn-22455d9-6020-407	bn-344c049b-fbbl-4735-a577	PD
985	49.2	bn-0ee164d4-8574-4f	bn-c75c6cbd-405c-4a6a-97b	PD
202	60.5	bn-ad056ee-9636-4c	bn-58082bd-9593-4d46-a655	PD
203	62.6	bn-10bdb639-c823-4f2	bn-2534a33-73a5-4cec-84dc	PD
946	31.5	bn-252237c6-5bed-43	bn-da3405b6-56c5-46f5-8114	PD

7 results returned.

Report all patients in the LRRK2 Study that are under 65, diagnosed with PD, that have both a DNA sample and a UPDRS assessment.

The result is that PICC researchers have access to more comprehensive information than ever before. Below is an example of another search created during the proof of principle.



Find ALL patients with a known Family History of Parkinson's disease who have participated in at least two clinical studies.

For one more example of value emerging from the early days of the Parkinson's Insight Project, Melissa Informatics converted a previously un-searchable version of PICC's iPSC (cell line) inventory into one that allows researchers to run complex searches.

Show/Hide Query Summary

Your search returned 34 record(s)

Export: Excel Text (TSV) HTML XML Open in Knowledge Explorer Chart

Sample	commit_by	Passage	rack	removed	lowest_passage	visal_bld	remain	commit_date	initial	bank	box	notes
ipsc_sample_1285_FBC	Clinic - redacted		5	6	1	9 GREEN		9/09/23/2010	10	4	7	
ipsc_sample_1600_FBC	Clinic - redacted		5	4	0	1 RED		1/04/13/2010	1	4	7	
ipsc_sample_1679_FBC	Clinic - redacted		5	1	0	4 RED		1/01/27/2010	1	4	8	
ipsc_sample_1679_FBC	Clinic - redacted		5	1	0	4 GREEN		2/01/31/2010	2	4	8	
ipsc_sample_1679_FBC	Clinic - redacted		5	4	0	4 WHITE		1/01/27/2010	1	4	2	
ipsc_sample_1754_FBC	Clinic - redacted		5	4	2	4 BLUE		0/02/17/2010	2	4	6	
ipsc_sample_1754_FBC	Clinic - redacted		5	3	0	4 YELLOW		1/02/09/2010	1	4	9	
ipsc_sample_1754_FBC	Clinic - redacted		5	11	0	4 N/A		2/02/04/2013	2	4		11 2009 BIOPSY
ipsc_sample_1754_FBC	Clinic - redacted		5	11	1	4 N/A		1/02/04/2013	2	4		11 2012 BIOPSY
ipsc_sample_1775_FBC	Clinic - redacted		5	6	5	4 WHITE		4/06/13/2009	9	4	7	
ipsc_sample_1789_FBC	Clinic - redacted		5	4	0	1 N/A		1/08/27/2010	1	4	5	
ipsc_sample_1811_FBC	Clinic - redacted		5	1	1	7 BLUE		1/01/28/2010	2	4	5	
ipsc_sample_1811_FBC	Clinic - redacted		5	1	0	7 YELLOW		2/01/31/2010	2	4	5	
ipsc_sample_1811_FBC	Clinic - redacted		5	11	0	7 N/A		1/10/04/2012	1	4	6	
ipsc_sample_1811_FBC	Clinic - redacted		5	11	0	7 WHITE		3/03/22/2013	3	4	11	
ipsc_sample_1815_FBC	Clinic - redacted		5	1	1	16 YELLOW		1/03/29/2010	2	4	9	
ipsc_sample_1815_FBC	Clinic - redacted		5	1	0	16 BLUE		5/05/20/2010	5	4	7	
ipsc_sample_1815_FBC	Clinic - redacted		5	11	1	16 RED		7/02/05/2012	8	4	6	
ipsc_sample_1815_FBC	Clinic - redacted		5	11	0	16 WHITE		3/03/22/2013	3	4	11	
ipsc_sample_1816_FBC	Clinic - redacted		5	2	0	3 BLUE		2/02/09/2010	2	4	9	

1 2

Show/Hide SQL

Report all cryo-preserved fibroblast cell lines for female patients in the PICC's Induced Pluripotent Stem Cell Inventory.

Integration of this iPSC cell line data makes it possible to see how the cell line / tissue sample information is related to other information. Importantly, this saves researchers' time when it comes to identifying biological samples required for a specific research plan.

Successful Results: Immediate and Growing Benefits from Phase 1

Transformative Insights from Data – Nature Genetics Use Cases Example

The initial proof-of-concept integration led quickly to scientific publication in one of the leading healthcare / life science journals (See J William Langston, et al. Multisystem Lewy body disease and the other parkinsonian disorders. Nature Genetics. Volume 48 (2016)). Rapid publication in one of the most prestigious life science journals demonstrates an unusually high level of success from a proof-of-concept. Access your Nature Genetics subscription or ask Nature Genetics for this review paper to learn more!

All costs in this case were applied to creating an unusually comprehensive, longitudinal reusable and easily extended dataset. The immediate benefits that emerged from the successful first Phase completed by Melissa Informatics leaves PICC with growing long-term benefits as additional research is applied to this dataset, and as new data is added.

Phase 2: Electronic Medical Records harmonization => Monetized Pharma Partnerships

In Phase 2 of the relationship, beginning in 2016, Melissa Informatics successfully accessed, curated, harmonized and integrated Electronic Medical Records (EMR), including medications (current and past), family history, and related medical conditions. All of the data has been de-identified for research and partnership purposes, although identified data will be available for clinicians and permitted PICC staff.

Scientific and Business Goals

The primary technical goal for Phase 2 was to curate and integrate structured EMR data (e.g., in tables, csv and decrypted XML formats) with over 70,000 unstructured clinical notes from over 5,000 patients, collected over a 10 year period.

This was intended to support a long-term data management strategy to support state-of-the-art clinical informatics and health outcomes research, and to support a Pharma collaboration – to understand performance of a specific drug, and to a second effort to provide clinical trials quality research data for an ongoing natural history study.

Challenges

This data was originally acquired by two separate EMR applications over almost 14 years. EMR 1 was applied from September 2004 to February 2011, and contained diverse unstructured notes and wildly diverse use of structured fields, including medical evidence mixed together with application information (for example, width of a border or color of a button in the software application). EMR 2 was applied from February 9, 2011 to the present. This EMR provided substantial additional structure, and data from 2016 forward is more highly structured due to improved clinical practice, following feedback from data quality review and related clinical leadership.

This data could not be deeply and comprehensively searched, or searched together. Data from provided and recurring EMR “data dumps” was extremely diverse, with, for example, over 190 different spellings identified to report a single drug (e.g., Stalevo, or carbidopa-levodopa-entacapone). Medical notes were not always consistent with structured content. Relationships for linking across content were not well defined and were difficult to identify.

Previously, research quality data was not available from these systems. A pharma partner reported working previously, with a different vendor, to acquire EMR content for research. The partner was not able to work with that data, as it had too many missing or conflicting values. Melissa Informatics and PICC overcame these challenges.

Solutions

Melissa Informatics Sentient software transforms data dumps containing unstructured text, XML, tables, tsv, image content and other data formats, into research quality, well-managed data resources. To make this possible, we have developed advanced, standards-based semantic software technologies. This applies a global open “W3C” standard that can be thought of as HTML for data integration. (See <http://www.w3.org/TR/rdf11-concepts/>)

Melissa Informatics’ semantic technologies apply “ontology-enabled machine reasoning” in combination with disease, drug and treatment lexicons. This makes it possible correctly identify and “normalize” (e.g., to transform to meet an expected terminology or data format standard) healthcare data, such as conditions, treatments and procedures. For ambiguous data and even for missing data, medically relevant ontologies can be applied with machine reasoning to disambiguate and “infer” content that meets context clues. This means the software can look at many different areas in a dataset to make decisions about what a specific data item really should be, to transform ‘tnol’ to “Tylenol”, or even to calculate a disease assessment score by looking at multiple factors that define that score. (See <https://www.w3.org/standards/semanticweb/ontology>)

Sentient finds the needles (important data points) in haystacks, cleans and sorts those needles, and creates connections or “links” across them as required create a beautiful tapestry, that is, to create a well-formed, searchable dataset.

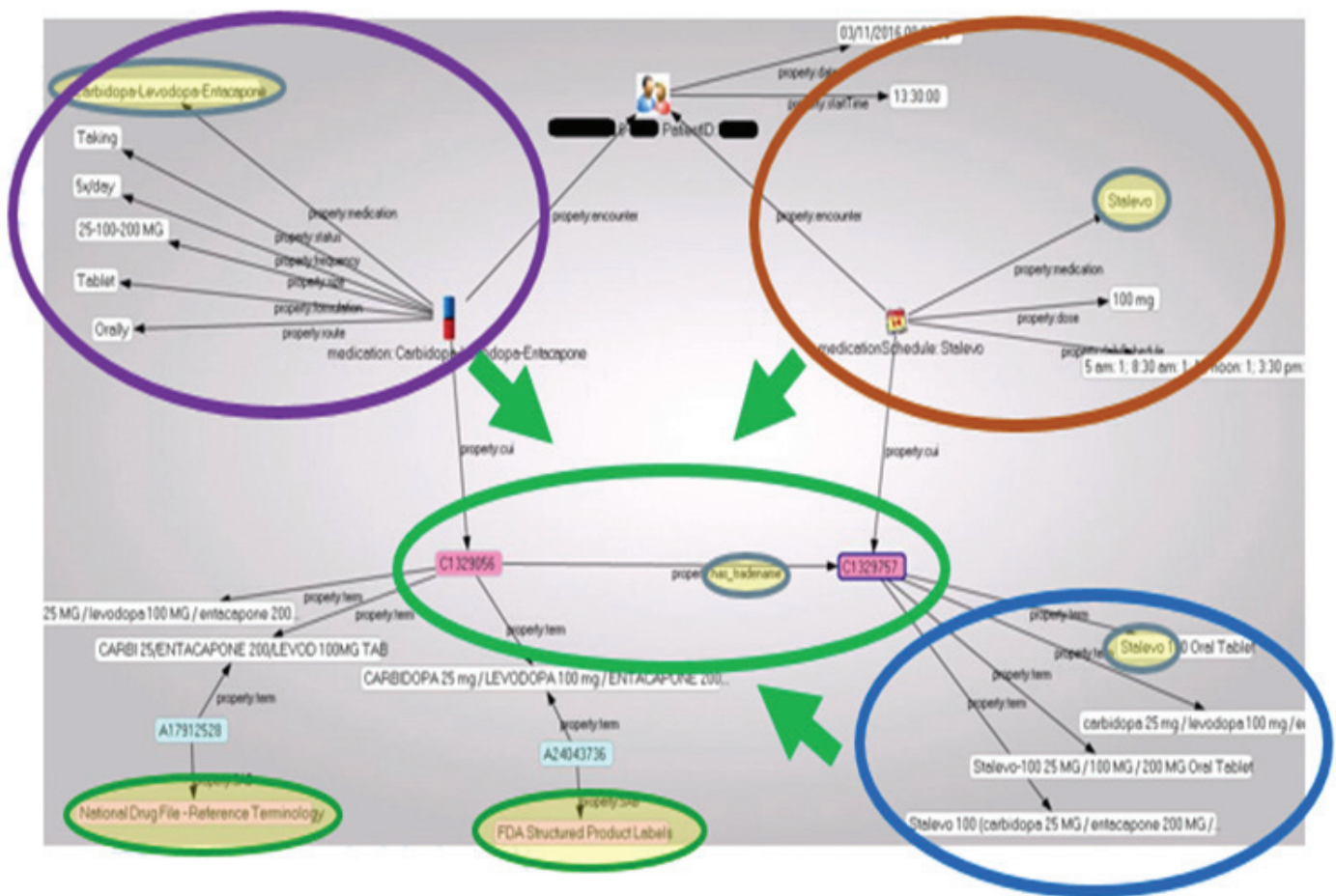
Below are examples of disconnected source data, transformed into effectively integrated data. This data was pulled from PICC content that we have been permitted to use. First, extensive use of unstructured medical notes is important. The system must be able to identify, extract and normalize this data.

~~patient~~ reports that he has resumed taking levodopa now that the reported grogginess issue have been addressed via Stalevo (5x daily). Tremors decrease for about an hour after taking. Patient also reports use of ~~as~~ as complementary, is counseled that possible contraindication with carb/lev/ entacapone is unknown. Exercise is recomendd as patient is unsteady, attributed in part to lack of exercise. Patient resists gym and asstd training. Is recommended to walk 2x daily and agrees.

Example 1 - text mining and machine reasoning methods described above help to identify and clean useful data. Spelling for levodopa is extracted, corrected and reported as data, and drug name and treatment protocols are extracted and reported as data.

All of the identified variables from examples 1-3 above are suited for linking. However, different terms may reflect the same concept, and errors in spelling, punctuation and format are also common. With data from EMR systems, substantial normalization is required. When data are transformed in order to normalize (move to a standard) and harmonize (conform to other data) them, “provenance”, or history is kept for QA/QC purposes. Provenance may include data source information, originally reported values or text strings, closeness of fit scores between source and assigned terminology, etc.

The next image below shows the Melissa Informatics’ Sentient “Knowledge Explorer” software product. The identified data can be visualized and queried within Sentient, machine reasoning and other business and integration rules are applied, and data are harmonized and linked within Sentient software. Visual review by a data scientist or “knowledge engineer” is supported by ontology-enabled machine reasoning and automated workflows designed to properly identify and curate data. Lexicons (lower left) are applied for relevant text identification and cross-source identification.



Data ingested into Sentient for integration. Information about a drug (Stalevo, or carbidopa-Levodopa-Entacapone) is combined in this data network or “graph”. Content from data source locations 1, 2, and 3 are linked by common concepts and identifiers provided by standards-compliant FDA and National Drug File resources.

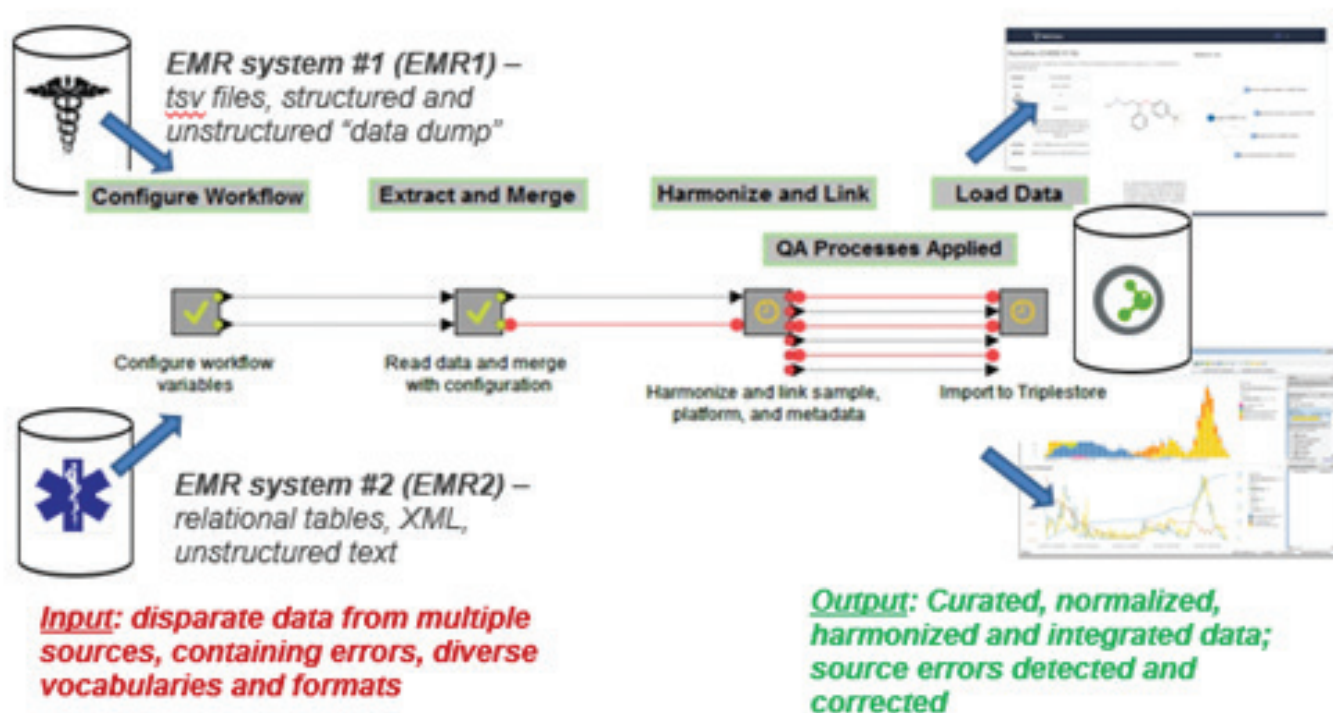
Sentient connects across sources via multiple lexical and contextual comparisons, identifies Common Unique Identifiers (CUIs) and reports common concepts as equivalent – reporting diverse source content as normalized preferred labels. Finally, a clean, well-managed database is created, and queries and reports become possible.

S MED	S DIRECTION	S STRENGTH	S FORM	S FREQUENCY	S TIME	S ROUTE	S STATUS
ATORVASTATIN (atorvastatin)		40mg		Once a day			Taking
CARBIDOPA AND LEVODOPA (carbidopa-levodopa, carbidopa/levodopa)	take up to 2.5 tab	25/100MG	Tablet	1day as directed	5 am	Orally	Taking
CARBIDOPA, LEVODOPA AND ENTACAPONE (cabidopa-levodopa-entacapone; stalevo)	1 tablet	25-100-200 MG	Tablet	5x/day	5 am	Orally	Taking
ERGOCALCIFEROL (vitamin d3)	1 tablet			Once a day			Taking
IBUPROFEN (ibuprofen)	1 tablet	800 MG	Tablet	As needed/rare		Orally	Taking
POLYTHYLENE GLCOL 3350 (miralax)			Poweder	once daily		Orally	Taking
PRAMIPEXOLE DIHYDROCHLORIDE	TAKE 1 TABLET	0.5MG	Tablet	5x/day	8:30 am	Orally	Taking
RASAGILINE MESYLATE (azilect)	1 tablet	1MG	Tablet	Once a day	8:30 am		Taking
SLO-NIACIN (slo-niacin)	1 tablet	500 MG	Tablet Ext.	Once a day		Orally	Taking
VITAMIN A, VITAMIN C, VITAMIN D, VITAMIN E, THIAMINE, RIBOFLAVIN, NIACIN, PYRI...	1 tablet			Once a day			Taking
probiotic				once daily			Taking
AMANTADINE HYDROCHLORIDE (amantadine hd; amantadine)	1 capsule	100 MG	Capsule	up to twice a day	12 noon	Orally	Taking
ASA (asa)	1 tablet	81 mg		Once a day			Taking

Integrated report from Sentient. This figure shows a report that combines content from multiple data sources.

Substantial data normalization and linking was required to generate the report above. For example, different terms and spellings for the same drug were transformed into common preferred term, with original terms noted in parentheses. Depending on customer needs, additional normalization can be applied.

Finally, following data harmonization and linking, extensive automated QA and QC procedures are applied. An automated workflow is applied for QA, to ensure data integrity and quality and for regular data updating. Data outliers and unexpected values are reported for manual QC review.



Integrated text mining applies ontologies and terminologies with combined machine reasoning, machine learning and traditional business rules, to extract and curate data content and relationships.

Summary: Immediate and Growing Benefits

The relationship between Melissa Informatics and PICC opens up many transformative use cases, including:

- Improved understanding of how genes, proteins and treatments impact Parkinson's disease
- Identification of candidates for clinical trials or alternative treatments
- Earlier identification of treatment opportunities and risks
- Improved care for and knowledge of each patient, including what course of treatment is most likely to be successful based on how similar patients have responded
- Secure - and highly remunerative - data sharing with collaborators
- And many others...

Realize Knowledge and Revenue from your Data

Before Parkinson's Insight, PICC staff needed to piece together data for any research question or collaboration. This would take days, weeks, even months to complete for a narrow set of questions. Now, PICC's data is accessible and useful, and has reached a completely new and unexpected level for value.

Melissa Informatics made it possible to realize knowledge and revenue from clinical data improving medical care and bringing substantial revenue back into the clinic. Now, ground breaking discoveries are published and data-driven partnerships with Pharma companies are successfully realized.

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.

Melissa Informatics
22382 Avenida Empresa
RSM, CA, 92688

1-800-476-7635

info@melissainformatics.io

www.melissainformatics.io

