Response to NOT-TR-16-002, Request for Information (RFI): Soliciting Input for the National Center for Advancing Translational Sciences (NCATS) Strategic Planning Process

Elsevier values its multi-faceted and synergistic relationship with the NIH and is appreciative for the opportunity to provide a response to NOT-TR-16-002, Request for Information (RFI): Soliciting Input for the National Center for Advancing Translational Sciences (NCATS) Strategic Planning Process.
Submitted on behalf of Elsevier by:

Holly J Falk-Krzesinski, PhD
Vice President, Strategic Alliances
Global Academic Relations
h.falk-krzesinski@elsevier.com
Elsevier
New York, NY, USA

For NOT-TR-16-002, some examples of particular issues of interest that apply across the translational science spectrum include:

A. Breaking down professional, cultural and scientific silos across the translational science spectrum
B. Focusing on inter-operability of data systems (such as integrating patient data and electronic health records into pre-clinical research)
C. Expanding research efforts at NCATS into new therapeutic modalities
D. Focusing on patient-driven research and patient/community engagement
E. Forming innovative partnerships with a wide variety of stakeholders
F. Identifying skillsets and competencies needed for training the next generation of translational scientists
G. Utilizing modern communication and dissemination tools to expand awareness of translational science to a wide variety of stakeholders

A. Breaking down professional, cultural and scientific silos across the translational science spectrum

Elsevier Clinical Solutions is singularly focused on helping improve clinical outcomes. Elsevier provides over 25% of the world's clinical content and serves over 20 million healthcare professionals. We offer solutions that address the needs of care providers and patients and help healthcare organizations and educational institutions empower clinicians and patients while improving administrative excellence, care, and financial outcomes. We work with our stakeholders to incorporate trusted content into clinical workflows, optimize IT investments, and collaborate with clinical and revenue cycle teams to ensure they have the skills, best practices, and competencies necessary to maintain consistent top performance and reduce variability in care delivery.

Elsevier Clinical Solutions products and services are designed to help caregivers, professionals, and students improve practice, encourage broad and deep adoption of exceptional professional practice guidelines, and promote a culture of quality across the enterprise or academic institution.
We serve a wide range of stakeholder groups, with expertise in facilitating interactions between groups:

- Healthcare institutions, such as hospitals, long-term care facilities, and managed care organizations;

- Individual practitioners and clinicians including physicians, nurses, pharmacists, therapists, and others;
- Retail pharmacies and clinics;
- Academic institutions and students;
- Electronic health record (EHR/EMR) and other health information technology (HIT) companies;
- Government agencies;
- Professional and trade associations

B. FOCUSING ON INTER-OPERABILITY OF DATA SYSTEMS

Elsevier's high-quality clinical information solutions and services are driving the advancement of medical knowledge and improving individual patient health and efficiency of health systems. Elsevier Health Analytics, based in Berlin, Germany, is a leader in predictive analytics and data mining on the German health care data. The team is at the forefront of traditional health analytics: health outcomes research; predicting drug-drug interactions; predictive modeling for population health management; gaps in care; and, pharmacovigilance.

Elsevier Health Analytics has made it its mission to optimize the supply of the German statutory health insurance (GKV). Advanced statistical methods based on GKV-routine data are developed in collaboration with health insurance companies and medical supply solutions. We translate the insights gained into concrete clinical recommendations. Together with our partners, we operate health services research to improve existing treatment offerings, which includes guidelines-based therapy as well as the assessment of innovation in healthcare.

Since HIPAA issues make it near impossible to obtain actual health records, a test/gold set of anonymized Electronic Health Records (EHR) would be a great resource to Elsevier to develop and test point of care applications we are currently developing. Also, a test bed EHR/EMR system would be incredibly valuable, where different content providers could plug in applications to show added value of relevant data at the point of care.

Elsevier's Health Analytics group is especially interested in developments with regards to EHR/EMR. We are supportive of: 1) Central, anonymized linked patient databases (including detailed clinical encounters in primary and secondary care, medication, genetic data, etc.) for research; and, 2) Central patient records, or at least interoperability standards (including federated search or HIEs) as a method of improving care delivery to individual patients. We encourage the NCATS to work in coordination with the Office of the National Coordinator for Health Information Technology to drive both of these initiatives. We also want to make sure that NCATS is aware of our high-performance computing (HPC) capabilities to analyze data for patterns. RELX Group, Elsevier's parent company, is one of the very few companies in the world that has analytical HPC capabilities and is expert and experienced in dealing with highly confidential and very private data.

Regarding linked patient databases, more (diverse) and bigger (simply more) is better. Broad accessibility (under appropriate safeguards) to the anonymized, longitudinally linked for-research data, including by industry, is desirable for Elsevier's Health Analytics. Industry finances applied research and product development that brings universities' basic research to the point of care and to actually

benefit patients. Broad accessibility will also drive innovation from big data, which is currently hindered by selective access. Heath Analytics currently conducts substantial research projects granting us securely anonymized patient data access together with healthcare systems in Europe.

Regarding central patient records, comprehensive (all individual patient encounters) and timely is better. As an example, Denmark has introduced a shared medication record. Physicians there can see their colleagues' prescriptions. This transparency among providers is dramatically transforming the Danish healthcare system, already one of the best in the world. Physicians now feel responsible for the full array of prescriptions, even those of their colleagues. Also patients can access and review their complete personal health record, which makes them a responsible partner in their health management. The networking of all players improves patient outcomes substantially.

C. EXPANDING RESEARCH EFFORTS AT NCATS INTO NEW THERAPEUTIC MODALITIES

Elsevier R&D Solutions

Better R&D outcomes depend on data, but finding, interpreting and integrating diverse datasets to quickly discover actionable insights is a universal challenge. Elsevier R&D Solutions offers tools that solve these data challenges, enabling research teams to predict outcomes based on past experience — an essential task in lead identification, drug development, pharmacovigilance and more — which helps them make better informed research and business decisions.

- Disease Biology: Elsevier helps organizations understand the biological mechanisms and biomarkers of diseases or abnormal states, enabling the development of new drugs and therapies.
- Lead Optimization: Elsevier provides high-quality external data and supports harmonization with internal data to improve hit identification and lead optimization processes.
- Drug Candidate Selection: Elsevier enables unique access to FDA and EMA drug approval documents and data, and biomedical literature, to enable critical decisions in drug candidate selection, study design and risk mitigation.
- Pharmacovigilance: Elsevier provides optimal literature monitoring and triage strategies to ensure that drug and medical device companies monitor all mentions of adverse events efficiently, and stay compliant on regulatory reporting.

Elsevier R&D Professional Services further provides custom data stewardship and bioinformatics solutions for life sciences research. Properly informed decisions in drug development reduce costs and accelerate the road to market. New insights rely on access to relevant and accurate data — but having access to data is not the same as having answers. Normalizing and organizing data from disparate sources in integrated silos facilitates the comparison, analysis, interpretation and sharing of data, increasing the opportunities to drive pharmaceutical research in innovative directions. Elsevier is a trusted partner in the curation, normalization and integration of life science data.

Composed of experts in life sciences research and informatics solutions, the R&D Solutions Professional Services team enables stakeholders to increase R&D productivity, increase return on information and reduce cost of IT support. Dedicated data stewardship solutions increase R&D productivity and reduce risk across the development spectrum — from discovery through preclinical and clinical trials to post-launch activities:

- Expertise for data retrieval and mining: Creating customized, searchable databases of biological and biomedical information using high-quality text-mining solutions supports early discovery tasks such as target identification and drug repurposing assessments.
- Expertise for data integration: Integration and normalization of pharmaceutical and chemistry data from in-house ELNs with Elsevier data to create a single, searchable database facilitates the complex computational tasks of hit identification and lead optimization.
- Expertise for monitoring adverse events in literature: We offer customizable, automated search strategies that thoroughly search the available literature for adverse events and other crucial data. Integrating these search strategies with automated processes for literature triage speeds up the process of assessing the data and finalizing the required adverse event reportings.

Medical Graph

Elsevier is engaging with research institutions to develop a Medical Graph: the automated extraction and combination of diverse biomedical and health information and data and representation as a weighted, directed graph. The goals are twofold: (1) to link biomedical research with clinical data to aid research, and (2) to better predict patient-individual levels of risk for disease occurrence, progression, or adverse events and to suggest patient-specific treatment options to providers. The key enabling technologies for the Medical Graph are:

- **EMMeT, Emtree:** Elsevier has developed the Emtree biomedical thesaurus and the EMMeT medical ontology. EMMeT links together multiple standard taxonomies (including MeSH, Snomed, WHO ICD, WHO ATC, LOINC, and medical procedures e.g. CPT) around clinical concepts (e.g. type 2 diabetes or Alzheimer's Disease). Emtree is a hierarchically structured, controlled vocabulary and thesaurus for biomedicine and the related life sciences.

- **Routine claims database:** Elsevier Health Analytics' longitudinally linked, six year comprehensive German medical claims database (ICD diagnoses, prescriptions, procedures, demographics from primary and secondary care) of currently 6 million anonymized patients constitutes a large retrospective trial of real-world evidence including multimorbidity. Within the Medical Graph, EMMeT groups the claims data around clinical concepts and enables linking to underlying biomedical mechanisms to allow hypotheses to be posed to real life data. Among other benefits, the claims database shows the incidence and prevalence of indications and comorbidities for all indications, and the probability of disease progression.
- **Elsevier Text Mining** (ETM; see more details in the "Information Extraction" section below): Elsevier is a world leader in natural language processing for knowledge discovery and extraction. We text mine the majority of the world's biomedical publications to extract protein interactions in support of drug discovery and development. ETM is now being expanded to include other biomedical information.
- **Scientific publications**: as the world's leading STM publisher, Elsevier offers electronic access to an unrivalled corpus of biomedical publications (the Science Direct database). Scientific publications summarize the research and underlying data, and represent a vast body of evidence. Elsevier's citation analysis determines the impact of individual publications, a measure of weight in the body of evidence. Clinical trials expose concise medical cause and effect relationships. Elsevier has license agreements in place with other publishers to extract information (e.g. for drug discovery and development). Embase, based on the Emtree biomedical taxonomy, abstracts and indexes the vast majority of biomedical evidence from many publishers.

- **Graph technology, parallel processing**: a major recent development in information technology has been the emergence of graphs to represent the relationship between entities. Graphs are easily expandable to accommodate new types of information. Stable, large-scalable open-source knowledge graph libraries are now available (e.g. Berkeley Spark graphX). EMMeT and Emtree contain the core semantic building blocks to construct a Medical Graph. Affordable, large-scale parallel computing can now easily process vast amounts of information (Spark MLib, HPCC).

The Medical Graph has the potential to be the next-generation solution for biomedical knowledge discovery and extraction, linking basic research with the clinical level to deliver insights to complex problems.

Information Extraction

- Metadata Analysis Tools and Methodology for Extracting New Information and Knowledge from Studies in Public Data Repositories

Elsevier has a long track record of data and metadata standards, dating back to the 1990s when we led the TULIP project. The Elsevier XML specifications for journal articles and book chapters are widely known and in use for 3000+ propriety and society journals and the metadata for 20,000+ journals. Content, including 12M journal articles, resides in a content repository that is accessible through restful APIs. Its metadata model is described using RDF serialized as JSON-LD. The API payloads and responses in JSON-LD are treated in the same way as our main content standards.

Our content is stored in multiple content-type-specific "warehouses." Through a metadata repository, this is made in to a virtual whole, called our Virtual Total Warehouse. Our content model and metadata standards are especially focused on content versioning. "Generations" of content assets keep various files together that together constitute a version. This Virtual Total Warehouse (VTW) plays a role in acquisition, editing and curating content (in our case, journal articles, book chapters, drug monographs, patents, patient education, and much more) and a Content Enrichment Framework takes this content and can, in principle, run any semantic process on the content, depositing the results back in VTW.

Elsevier also has a linked data repository adhering to the standards of linked data and linked open data. Elsevier's approach to unstructured information: The vast majority of information exists as an unstructured text which makes it unsuitable for efficient analysis by humans. The area of computational assistance to analysis of large volumes of textual information is traditionally split into two (somewhat overlapping) approaches - information retrieval and information extraction.

Information Retrieval (IR) systems concentrate on finding documents containing information deemed relevant to a particular topic of interest. Usually this is done by analyzing the word content of the documents using statistical methods based on keywords or word co-occurrence. IR methods are by their nature generic and to a large degree language-independent; the output of IR systems is intended for human readers.

Unlike IR, Information Extraction (IE) focuses on extracting information contained within the documents in a form suitable for automatic processing. IE systems use an ontology (or knowledge representation schema) as a model of a particular domain, and thus are domain-specific. The simplest form of an ontology is a list (or, even better, a hierarchical tree) of concepts relevant to the domain. More advanced forms of ontology also specify possible semantic types of relationships between the

concepts. Extracting information with high precision involves deep understanding of the actual meaning of the text; as a result, IE systems are language-specific.

In developing solutions for vertical markets, Elsevier takes the IE road. Instead of building one generic, language- and domain-independent system that deals with large number of topics but provides little depth when it comes to the subject matter, we focus on extracting structured information specific for a particular domain from English text.

- Elsevier Text Mining (ETM)

Within its Elsevier Text Mining portfolio, Elsevier has developed a proprietary natural language processing (NLP)-based technology called MedScan for extraction of structured information from unstructured text. It is a good fit for automatic indexing of NIH's content as the MedScan Thesaurus/Taxonomy was built mostly based on NIH thesauri and has all the NIH identifiers integrated (MeSH Headings, NCI Metathesaurus IDs, Entrez Gene IDs, Organism Tax IDs, etc.). The technology works by first recognizing domain-specific named entities (concepts) in the input text, and then uses natural language processing techniques to extract attributed, directional semantic relationships between them. The relationships can be of any complexity from simplest binary (X affects Y) to n-ary (X protects Y from Z) and complex multi-level nested ones (effect of X on Y depends on Z).

Elsevier IE technology has modular architecture. Each module performs its specific function and has well-defined and documented input/output format. Modules with compatible interfaces can be combined into different text processing pipelines, as required by the application. All modules are written from scratch to achieve our flexibility/precision/performance goals. The modules are portable C/C++ applications interacting via files and pipes.

MedScan Technical Description: MedScan is a proprietary natural language processing (NLP)-based technology for extraction of structured information from unstructured text. Structured information is captured and formally represented using a conceptual model (ontology) of the domain. The ontology consists of a set of conceptual named entities (e.g. Proteins, Small molecules, Cellular processes, Diseases, etc) and a set of categorized relationships (Binding, Protein Modification, Expression regulation, Molecular Transport, etc) between them.

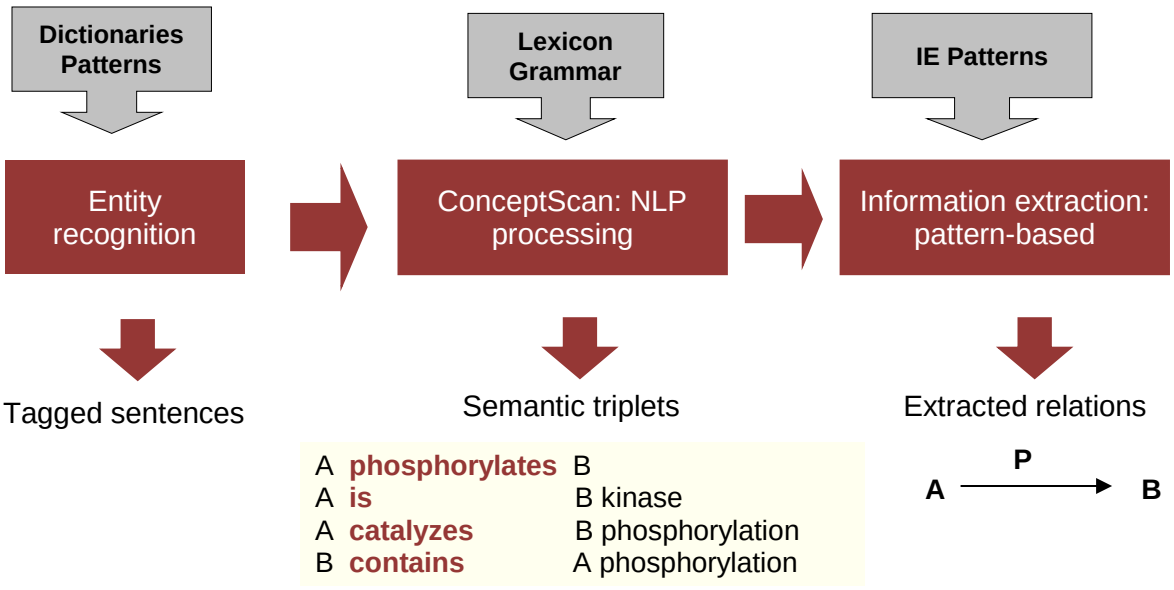


Figure 2. An overview of MedScan Architecture

MedScan first recognizes different domain-specific named entities (gene/protein names, cellular processes, cellular components, diseases, tissues, organs, etc.) in the input text, and then extracts functional relations (binding, regulation, association, molecular transport, etc.) between them. Figure 2 shows an overview of MedScan architecture.

The Entity Recognizer module utilizes hand-crafted dictionaries of domain-specific entities in combination with an advanced matching algorithm to detect them in input text.

To extract entity relationships from the text, MedScan utilizes two modules. The natural language processing module, ConceptScan, analyzes the sentence structure and decomposes each sentence into a deterministic set of Subject-Verb-Object triplets, each representing a single semantic relationship between two singular noun phrases. Next, Pattern Matcher matches carefully designed linguistic patterns over the triplets to extract and encode the entity relationships.

MedScan has been field-tested and is proven to be fast, efficient, and accurate information extraction technology. It is currently used to process the content of the entire Medline database along with more than 40 freely available full-text journals in order to extract more than 3.5 million individual facts (relations) about functions of proteins with an overall accuracy of 90% and recall of 70%. The entire processing cycle can be completed in less than 24 hours on a regular PC.

Dictionaries and Named Entity Recognition:

Entity type	Number	Main sources
Proteins	136,000	Entrez Gene
Prot. Classes	7,500	GO, Enzymes, PubMed
Cell components	740	GO, PubMed
Cell processes	5,200	GO, PubMed
Diseases	6,300	MESH, PubMed
Small Molecules	270,000	MESH, PubChem, PubMed
Tissues	100	MESH, UMLS, NCI, EVoc
Cell types	360	MESH, UMLS, NCI, EVoc
Organs	2,875	MESH, UMLS, NCI, EVoc
Clinical parameters	1,786	Pubmed, ClinicalTrials.gov
Cell lines	2,500	PubMed

Table 1. MedScan Dictionaries

The Entity recognition module of MedScan utilizes hand-curated dictionaries of biomedical entities to detect them in the input text. Dictionaries are manually compiled and curated from the number of various public-domain resources (EntrezGene and SwissProt for protein names, PubChem and MESH for small molecules, GO for cell processes and components, MESH for diseases, NCI thesaurus for organs, tissues and cells, etc). Whenever possible the entities are hyperlinked to those outside resources for reference. Many additional aliases and terms are also added directly from the literature resources, e.g. PubMed. Table 1 shows the content of MedScan dictionaries. MedScan uses number of different algorithms to achieve accurate detection of entities in text. It can also use rule- and regular expression-based approaches to detect specific types of entities (abbreviations, numbers, dates, etc). The dictionaries are in a simple tab-delimited format so they can be easily extended or modified.

The input text can be in various formats (plain text, Microsoft Office, HTML, reasonable forms of PDF, zip/tar/gzip archives of the above, etc.) The output of the entity recognition step consists of individual sentences labeled to preserve their origin with identified named entities marked up with entity IDs, using **ID{***number***=….}** format.

The extracted triplets capture the main facts expressed in a sentence:

Triplets:

Axin2	associate	beta-catenin abundance
Axin2	inhibit	beta-catenin function
Axin2	associate	beta-catenin abundance
Axin2	inhibit	beta-catenin function
Axin2	affect	MEF cell line proliferation
Axin2	work	negative feedback pathway
Axin2	regulate	Wnt signaling
Axin2	control	apoptotic process

The ConceptScan is used in conjunction with named entity detection algorithm to index relationships between biomedical entities and to extract entity relationships. ConceptScan parses sentences in several sequential algorithmic steps (see figure below).

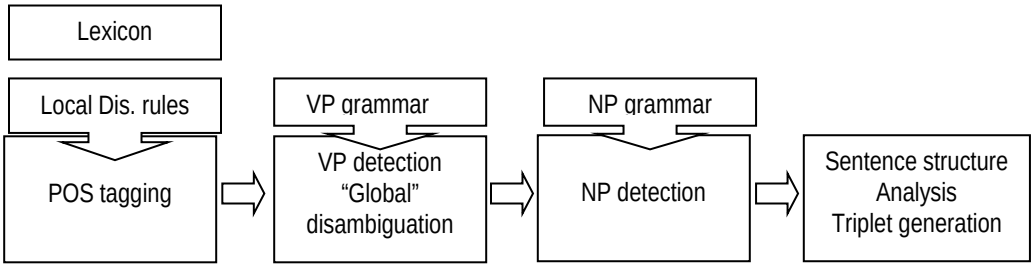


Figure 3. ConceptScan algorithm

The first step of NLP is part-of-speech tagging and local disambiguation. During this step, the words in a sentence are reduced to all possible uninflected forms, looked up in the lexicon and annotated with the respective syntactic categories. After initial POS tagging, the local disambiguation algorithm, encoded by a set of contextual regular expression-like rules, is applied. Notably, not all ambiguities can be resolved locally. The unresolved ambiguities are preserved for subsequent processing steps. The next step is identification of verbal phrases. Verbal phrase (VP) grammar is encoded in a single but complex deterministic finite-state automaton (DFA), with more than 25,000 states. It is matched over the sequence of syntactic categories assigned to sentence words at the POS-tagging step. NP grammar is matched after detection of verbal phrases is complete. Similarly to VP grammar, it is encoded by a DFA. The structure of NP grammar covers prepositional attachment, conjunctions, relational constructs, appositions and exemplifications. Once VPs and NPs have been identified, ConceptScan analyzes the structure of the entire sentence.

Information extraction: The specific relationships between entities are extracted using separate module - Pattern Matcher. It utilizes a formalism closely resembling regular expressions to detect specific linguistic constructs expressing entity relations and to capture the expressed relations. It is specifically tailored to deal with linguistic input; it operates on the level of individual words rather than symbols and supports advanced linguistic features like matching all word forms and multi-word lexemes. Pattern matching also supports all regular expression features: wildcards, sets, negation, etc. The figure below shows a sample information extraction pattern.

MedScan output: The output of MedScan is in an XML-based format describing entities and relation between them.

MedScan Ontology of Relationships: Elsevier has developed ontology of different types of relations between biological entities. Each type of relation has a very specific semantic definition and is typically attributed with additional information, e.g. sign of relations (e.g. positive, negative or unknown) or mechanism (e.g. phosphorylation, methylation, etc). There are three set of patters currently used by MedScan to extract biological relations – patterns focused on extraction of different aspects of protein functions, small molecule functions and disease biomarkers. Table 2 below shows the scope of biological relationships currently extracted by MedScan.

- Protein -> Protein
  - Binding
  - Protein modification
  - Expression (positive/negative/unknown)
  - Promoter regulation/Binding
  - Regulation (positive/negative/unknown)
- Protein -> Small Molecules
  - Synthesis/Degradation
  - Mol. Transport
- Protein -> Cell processes
- Protein -> Disease
  - Positive/negative regulation
- Disease -> Protein/Small molecules
  - Changed concentration/expression (positive/negative/unknown)
  - Mutations
  - Activity (positive/negative/unknown)
- Small molecules -> Protein
  - Binding
  - Direct regulation
  - Expression
  - Indirect regulation (positive/negative)
- Small molecules -> Disease/Cell processes (positive/negative/unknown)

Table 2. Relationships currently extracted by MedScan

The current scope of the information extracted by MedScan can be extended by developing new dictionaries covering other aspects of biomedical domain (e.g. focused more on medical or clinical entities) and/or by developing novel information extraction patterns to capture other types of entity relationships. The Pattern Matcher is extremely fast: it runs through more than 16,000,000 entity-tagged sentences from the entirety of Medline in less than 20 minutes.

MedScan Customizations: MedScan is flexible platform open for two types of end-user modifications. First, MedScan taxonomy and dictionaries can be extended to include new concepts and even new concept classes. Dictionaries are provided in a simple text-based tabular format and new concepts and concept aliases can be added to the files. Second, the scope of extracted information can be extended to include new relationships by modifying information extraction rules. The rules are recorded in a well-documented textual format and new rules can be created and added to MedScan.

MedScan Features and competitive advantages: Elsevier's IE engine has been designed and implemented from scratch to address flexibility, precision/recall and performance problems of the off-the-shelf NLP tools. Our design efforts focused on issues specific for texts in vertical application domains characterized by complex sentence and relationship structure, highly specialized entity notation, proliferation of abbreviations and synonyms. As a result of this focus, we have surpassed the 90% precision / 60% coverage mark on technical texts in our current application domains (biology and medicine). Our engine has an unmatched performance – it can process up to 1000 sentences per second on a regular PC, which is 2-3 orders of magnitude faster than prevailing NLP technologies. High

performance allowed us to achieve clean separation between modules where traditional approaches intertwine distinct functions like parsing and ontology-based information extraction to cut down on the amount of information exchanged between modules. Also, much attention has been paid to keep domain-specific information in dictionaries and rule files, to simplify maintenance and extending the coverage to other domains.

- Elsevier's Information Extraction (IE) technology: Fingerprint Engine

A back-end software system, the Elsevier Fingerprint Engine mines the text of scientific documents – publication abstracts, funding announcements and awards, project summaries, patents, proposals/applications, and other sources – to create an index of weighted terms which defines the text, known as a Fingerprint™ visualization.

By aggregating and comparing Fingerprints, the Elsevier Fingerprint Engine (FPE) enables institutions to look even beyond metadata and expose valuable connections among people, publications, funding opportunities and ideas.

The Elsevier Fingerprint Engine powers many solutions including Pure, comprehensive information management system, and Reviewer Finder, Elsevier's tool for finding reviewers.

The Elsevier Fingerprint Engine uses a variety of thesauri to support applications pertaining to different subject areas. By applying a wide range of thesauri, Elsevier can develop solutions in but not limited to: the life sciences, engineering, earth and environmental sciences, arts and humanities, social sciences, mathematics and agriculture. Thesauri provided by an institution or specific research domain can also be incorporated.

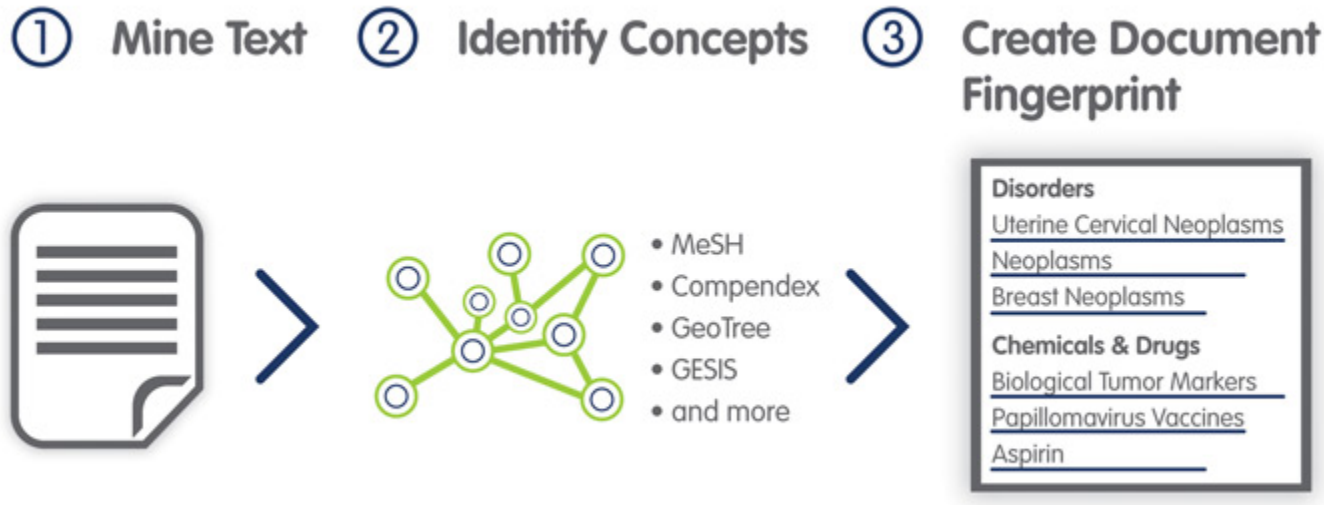


Figure 6: The Elsevier Fingerprint Engine creates Fingerprints via a three-step process

1. The Elsevier Fingerprint Engine applies a variety of Natural Language Processing (NLP) techniques to mine the text of scientific documents including publication abstracts, funding announcements and awards, project summaries, patents, proposals, applications and other sources

2. Key concepts that define the text are identified in thesauri spanning all the major disciplines

3. The Elsevier Fingerprint Engine creates an index of weighted terms that defines the text, known as a Fingerprint.

Applying Fingerprints to inform decision making: By aggregating and comparing Fingerprints of people, publications, funding opportunities and ideas, the Elsevier Fingerprint Engine can reveal insightful connections with practical applications. Here are some examples of how Fingerprints are currently used to bring scholarly business intelligence to institutional data:

- The NIH Research, Condition, and Disease Categorization (RCDC) system uses the FPE to categorize and report the amount the NIH has funded in each of 233 reported categories of disease, condition, or research area. The FPE enables the RCDC to provide consistent and transparent information to the public about NIH-funded research, providing a complete list of all NIH-funded projects related to each category.

- Pure aggregates the Fingerprints of individual documents to create unique Fingerprints that reveal your researchers' distinctive expertise. Pure also matches the Fingerprints of funding opportunities in SciVal® Funding to researchers' Fingerprints, recommending appropriate funding opportunities and suggested collaborators.

- Reviewer Finder compares document Fingerprints with researcher Fingerprints, making it easier to identify reviewers and raise awareness about potential conflicts of interest.

- Elsevier Journal Finder helps researchers find journals that could be best suited for publishing their articles. Journal Finder matches abstracts to Elsevier journals, scanning Elsevier's 2,200+ titles in the Health Sciences, Life Sciences, Physical Sciences and Social Sciences.

F. IDENTIFYING SKILLSETS AND COMPETENCIES NEEDED FOR TRAINING THE NEXT GENERATION OF TRANSLATIONAL SCIENTISTS

Elsevier's Science and Technology Books

Elsevier's Science and Technology Books (S&T Books) Division is committed to publishing high-quality content in translational research to support academic, government and industry researchers involved in the discovery of reproducible methods and innovative technologies that will lead to the development, testing and implementation of safe and effective drugs, biologics, diagnostics and therapies. Our goal is to provide researchers with the information, tools and resources they need to ask and answer important research questions that will help them succeed in the translation of basic research to tackle unmet medical needs. Our content is cross-cutting and spans a wide range of disciplines, including genetics and genomics, molecular biology, stem cells, chemistry and biochemistry, immunology, bioengineering, material science, bioinformatics, pharmacology, drug discovery, toxicology, clinical research, public health and much more. Our book editors, authors and contributors

are leading experts who are actively committed to the dissemination of knowledge in order to advance science and technology to improve individual and overall population health. We are currently working with Dr. John Gallin and Dr. Frederick Ognibene from the NIH Clinical Center to thoroughly update and publish the 4th edition of their timely and significant work, *Principles and Practice of Clinical Research*. Within our own division and across divisions at Elsevier, we aim to foster meaningful collaboration and communication while keeping the needs of the overall scientific, technical and medical communities at the center of what we do. We welcome the opportunity to investigate ways we can partner with NCATS on accelerating translational sciences to continue to drive advances in research and forge innovative solutions that address the challenges facing researchers and enables the delivery of safer, faster and more effective treatments and interventions for human health.

Elsevier's S&T Books Academic Press imprint offers professional development books and resources as tools to foster professional development and career advancement. Academic Press books cover a range of topics: clinical research protocols; research proposals; scientific writing and presentation; communication skills; strategies for leadership and mentoring; exploring career opportunities; and, innovation and entrepreneurship. Elsevier often partners with organizations such as the National Postdoctoral Association, American Association of Immunologists, and Association for Women in Science to develop new books and offerings and would welcome the opportunity to work with NCATS and the CTSA Hubs community to develop new resources related to skills and competencies

Elsevier Publishing Campus
The Elsevier Publishing Campus is an online platform offering interactive training, free lectures, and professional advice. Learning modules and resources on the platform currently cover topics such as: writing a journal article and submitting a book proposal; learning how to conduct peer review; understanding research and publishing ethics; writing a successful grant application; and transferable skills such as communication and mentoring. Just as we do for Academic Press books, Elsevier works with a variety of partners worldwide to develop new training and career resources for the platform, and link out to valuable resources housed at partner sites. We would be delighted to work with NCATS to develop new resources related to clinical and translational professional development and career pathways.

Elsevier Clinical Solutions
Within this portfolio (described in more detail above), are the Academic Education offerings, which include interactive student and faculty resources, review and test prep, digital books and journals. Clinical Solutions Academic Education advances world-class learning, test preparation, and reference tools from the most trusted sources, empowering students and faculty to achieve superior results in clinical education.

G. UTILIZING MODERN COMMUNICATION AND DISSEMINATION TOOLS TO EXPAND AWARENESS OF TRANSLATIONAL SCIENCE TO A WIDE VARIETY OF STAKEHOLDERS

Elsevier's Science Technology and Medical Sciences Journals
Elsevier's Science Technology and Medical Sciences Journals (STM Journals) division is dedicated to publishing high-quality content in translational research to support academic, government, and industry researchers involved in the discovery of reproducible methods and innovative technologies that will lead to the development, testing, and implementation of safe and effective drugs, biologics, diagnostics, and

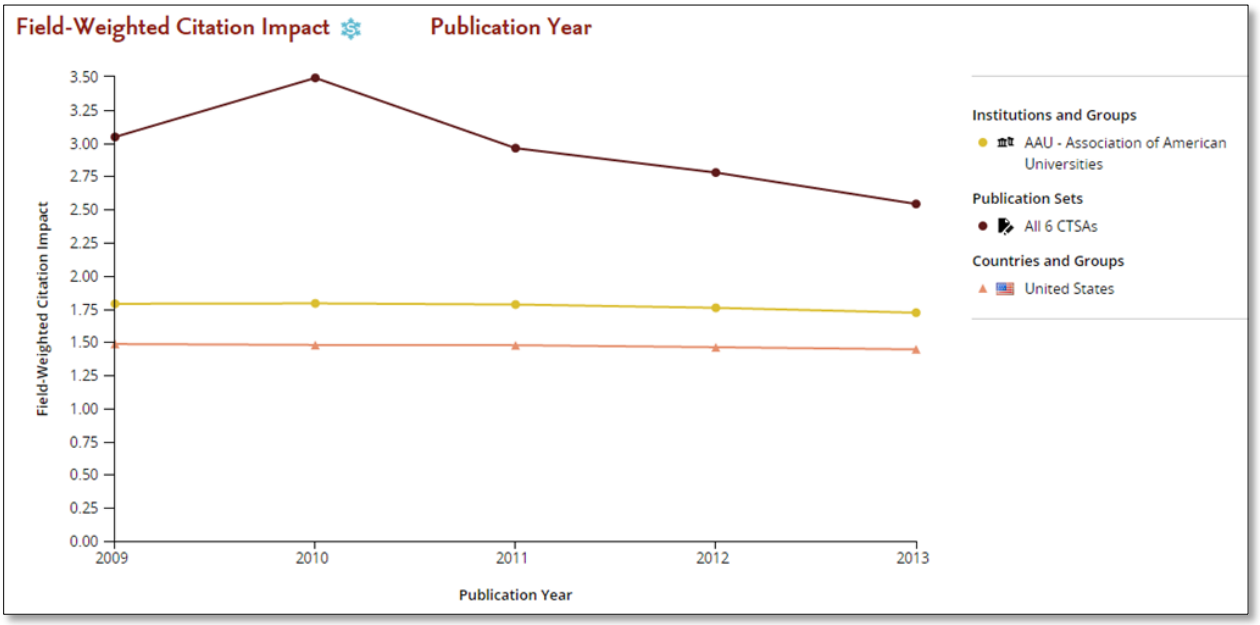
therapies. It has long been recognized that the effective translation of insights gained from biomedical research into improved human health is a global priority. To this end, Elsevier has looked to the leadership of its two leading journal portfolios, *Cell* and *The Lancet*, to guide the launch of a new comprehensive, online-only, open access Elsevier journal, *EBioMedicine*. This new journal is focused on forming a community that spans this interface and creates a valuable opportunity for dialogue and collaboration between the journals' respective audiences.

The scope of *EBioMedicine* covers the entire breadth of translational and clinical research within all disciplines of life and health sciences, ranging from basic science to clinical and public/global health science. In addition, to publishing a range of research study types, *EBioMedicine* publishes commentaries, reviews, and viewpoints that enhance the accessibility and applicability of basic research findings for health professionals, and promote a better understanding of clinical challenges for biomedical researchers. As a fully open access journal, which can be accessed and read for free in its entirety, *EBioMedicine* is committed to serving the clinical and basic research communities and the community partners by offering a multimedia platform to facilitate dialogue where scientific ideas and clinical needs can be defined, experimentally explored, and ultimately solved. *EBioMedicine* is currently indexed in PubMed Central and has been accepted for indexing in MEDLINE.

EVALUATING TRANSLATION RESEARCH AND ITS IMPACT

Powered by data from Scopus®, SciVal from the Elsevier Research Intelligence portfolio offers quick, easy access to the research performance of more than 220 nations and regions and over 6,000 research institutions worldwide. For NCATS-supported CTSA Hubs, SciVal can facilitate meaningful and consistent evaluation of the impact of research outputs from individual CTSAs or the entire consortium. Evaluators can examine CTSA performance across numerous research metrics using either publication set(s) or groups of researchers (e.g., members of a CTSA), and track CTSA progress over time. The tool's rich visualization capabilities and exporting functionality enable further, and even more flexible, data analysis.

For benchmarking, use of a common set of metrics (refer to the SciVal Metrics Guidebook) and an approach to standardize disparate publication sets enables a CTSA Hub to evaluate its output against the entire institution, across groups of CTSAs, the full CTSA consortium, or to comparator organizations. Beyond Hub-level analysis, SciVal allows for clear and compelling showcasing of the overall impact of the CTSA consortium or regional groups of CTSA institutes. See figure below.



CTSAs Outperform in Research Impact: A consortium of six CTSA institutes displays a higher field weighted citation impact (FWCI) over five years compared with both the output of research in the United States and that from the American Association of Universities (AAU) consortium. FWCI is a metric that normalizes for differences in citation activity by subject field, article type, and publication year thus enabling comparisons such as the one shown above.