Cross-Domain Intelligence: The Translation Problem in American R&D

An Evidence-Based Analysis of the Infrastructure Gap in Scientific Knowledge Routing

Victor Joaquin Estrada
DarkSigma, Inc.

March 24, 2026

Abstract

The United States has accumulated the largest body of scientific research in history, yet a significant portion of this knowledge remains inaccessible across disciplinary boundaries due to domain-specific language and institutional silos. This paper presents evidence that the resulting inefficiency constitutes a measurable, multibillion-dollar structural failure in American R&D. Drawing on published data from NIH, The Lancet, the Government Accountability Office, and peer-reviewed studies in cognitive science and information behavior, the analysis quantifies the cost of disconnected knowledge at approximately $28 billion per year in irreproducible preclinical research alone, with pharmaceutical R&D productivity declining 80-fold since 1950. The paper establishes market precedent through case analysis of Bloomberg ($15B annual revenue), Palantir ($330B market capitalization), and the $1.4 trillion economic value generated by GPS following military-to-civilian domain transfer. A strategic assessment examines China's state-managed research intelligence infrastructure under the 14th Five-Year Plan, current U.S. federal initiatives including the Genesis Mission ($613M+ committed) and NAIRR, and allied nation investments. The analysis concludes that while the U.S. is investing significantly in AI-for-science compute and data infrastructure, no existing program addresses the specific problem of cross-domain knowledge routing, the infrastructure layer that translates research findings between scientific fields operating in different vocabularies. This gap represents both a strategic vulnerability and a market opportunity.

Keywords: cross-domain knowledge routing, research infrastructure, R&D productivity, information markets, science of science, research intelligence

1. Introduction

This document lays out the evidence for a single claim: the United States has a translation problem in its research and development enterprise, and that problem is measurable, expensive, and getting worse. The argument is organized across three pillars. The Waste Case (Section 2) quantifies the cost of siloed research: $28 billion per year in irreproducible preclinical work, an 80-fold decline in pharmaceutical R&D productivity since 1950, and late-stage drug trial failures costing up to $1.4 billion per candidate. The Precedent Case (Section 3) shows that information infrastructure markets are real, durable, and enormous: Bloomberg earns $15 billion per year routing financial data; Palantir is valued at $330 billion for routing intelligence data; GPS has generated $1.4 trillion in economic value since its transfer from military to civilian use. The Urgency Case (Section 4) establishes that this is a strategic competition: China is building state-managed research intelligence infrastructure under its 14th Five-Year Plan, the UK is investing £1.6 billion in AI for science, and the EU is building the European Open Science Cloud. The U.S. has significant AI-for-science programs underway, including the Genesis Mission and NAIRR, but none of them address the specific problem of cross-domain knowledge routing.

Every number in this document is sourced and traceable. The argument is conservative by design. It does not require believing in breakthroughs on demand. It requires only one belief: that better information routing produces better outcomes. The evidence strongly supports that belief.

2. The Waste Case

Modern science produces an unprecedented volume of data while simultaneously failing to coordinate the knowledge that data represents. The U.S. research environment is the most sophisticated in history, but the efficiency of this system is in measurable decline. The primary cause: research findings remain isolated within domain-specific languages and institutional silos. When knowledge is not routed across these boundaries, the result is a multibillion-dollar expenditure on problems that have already been solved in other fields.

2.1 The Reproducibility Crisis

The most direct measure of scientific inefficiency is the reproducibility crisis. If a finding cannot be replicated by an independent lab, it cannot serve as a reliable basis for further work. The data on this is severe.

In 2012, researchers at Amgen attempted to reproduce 53 landmark studies in preclinical cancer research. They replicated the findings of only six, an 11% success rate (Begley & Ellis, 2012). Nearly 90% of foundational work in one of the most heavily funded areas of medicine was not robust enough to build on.

Bayer conducted a parallel internal review of 67 drug-target validation projects. Two-thirds showed results inconsistent with published literature (Prinz et al., 2011). Meta-analyses of preclinical research across life sciences indicate irreproducibility rates exceeding 50% (Freedman et al., 2015).

In 2015, Freedman, Cockburn, and Simcoe published a study in PLOS Biology estimating that the United States spends approximately $28 billion every year on preclinical research that cannot be reproduced. This estimate was derived by analyzing total annual preclinical spending and applying the documented failure rates from industry surveys.

Table 1. Preclinical Reproducibility Studies

Study	Institution	Success Rate	Focus Area
Begley & Ellis (2012)	Amgen	11%	Landmark Cancer Research
Prinz et al. (2011)	Bayer	21-25%	Preclinical Drug Targets
Freedman et al. (2015)	GBSI	<50%	General Life Sciences
Nature Survey	Nature	48% (crisis perception)	Multi-disciplinary

This waste is partly methodology, but a significant portion comes from researchers working in isolation. When scientists are unaware of what adjacent fields have already figured out, they replicate flawed experiments or use outdated models. Freedman and colleagues identified four categories of irreproducibility: study design, biological reagents and reference materials, laboratory protocols, and data analysis and reporting. Many of these errors could be caught if the information were routed to experts in adjacent domains who have already developed standards for those specific methodologies.

The downstream cost hits hardest in pharma. When a company tries to validate an academic finding for a new drug target, it typically requires 3 to 24 months and $500,000 to $2 million. When the original finding proves irreproducible, that entire investment is lost (Prinz et al., 2011).

2.2 The Lancet Waste Analysis

A 2014 series of articles in The Lancet argued that biomedical research fails to provide value because of a lack of coordination and transparency. The researchers estimated that up to 85% of all research investment is wasted. This waste occurs at multiple stages: during priority-setting, during study design and conduct, and during reporting of results.

The first category of waste is the failure to set appropriate research priorities. Researchers often do not have access to a systematic review of existing work, especially work done outside their immediate field. This leads to duplicate efforts or investigation of questions already answered in other disciplines. The second category involves flaws in study design, where many researchers lack formal training in advanced statistics. The third category is failure to report results transparently: less than 40% of clinical trial results from leading academic medical centers are shared within two years of completion (Yale News, 2016). When results stay hidden, other researchers repeat the same mistakes.

Table 2. Categories of Biomedical Research Waste (The Lancet, 2014)

Category of Waste	Driver	Consequence
Priority Setting	No synthesis of cross-domain work	Redundant research on solved problems
Study Design	Inadequate statistical power	False positives and non-reproducible data
Conduct/Reporting	Selective reporting, p-hacking	Misleading results in the literature
Accessible Information	Failure to publish negative results	Publication bias and duplicate failures

2.3 Pharmaceutical Attrition

The pharmaceutical industry spends more than $300 billion annually on R&D (DrugPatentWatch). Despite this, only about 10% of drug candidates that enter Phase I clinical trials ever reach the market. The journey from bench to bedside spans over 12 years and costs an average of $2.23 billion to $2.6 billion per approved drug (Insilico Medicine).

The most expensive failure points are Phase II and Phase III trials. Data from Arrowsmith and Miller shows roughly half of all late-stage trials fail. For oncology drugs, failure rates reach 60% (Pharmaceutical Technology). The primary reason for failure in Phase II and III is lack of efficacy, accounting for approximately 50% of terminations (Arrowsmith & Miller, 2013). The drug worked in the lab and in animal models but failed in humans. In many cases, the underlying biological mechanism was already being studied in basic science labs in another field, but that information never reached the drug development team.

Table 3. Clinical Trial Attrition by Phase

Trial Stage	Success Rate	Primary Cause of Failure	Financial Impact
Phase I	~52%	Safety/Toxicity	\$100M - \$300M
Phase II	~28.9%	Lack of Efficacy	\$300M - \$800M
Phase III	~55%	Lack of Efficacy/Commercial Value	\$800M - \$1.4B
FDA Filing	~92%	Strategic/Regulatory Issues	>\$1.4B (sunk costs)

Safety and toxicity failures account for 20-30% of late-stage terminations (Arrowsmith & Miller, 2013). These often involve cardiotoxicity or neurological effects that animal models did not predict (PMC/Annual Reviews). If researchers had better access to cross-domain knowledge, such as materials science data on drug delivery or comparative physiology from aerospace medicine, many of these failures could be identified earlier in the pipeline.

2.4 Eroom's Law: The Productivity Collapse

The declining efficiency of pharmaceutical R&D has a formal name: Eroom's Law, coined by Jack Scannell in 2012. It is Moore's Law in reverse. Since 1950, the number of new drugs approved per billion dollars of inflation-adjusted R&D spending has halved approximately every nine years. That represents an 80-fold decline in productivity over six decades (Scannell et al., 2012). This has occurred despite massive advances in high-throughput screening, genomics, and computational drug design.

Table 4. Drivers of Eroom's Law

Factor	Mechanism	Consequence
Better than Beatles	New drugs must beat high-quality generics	Higher efficacy bar, smaller markets
Cautious Regulator	Lower FDA/EMA risk tolerance post-scandals	Larger trials, more stringent safety data
Throw Money at It	High profits lead to indiscriminate R&D spending	Reduced discipline, project overruns
Brute Force Bias	Reliance on HTS over clinical observation	More candidates with low predictive validity

The "Brute Force" bias is the most relevant information problem. The industry shifted from whole-animal phenotypic screening to target-based high-throughput screening. This is faster and cheaper in the lab, but it often fails in trials because it focuses on a single protein or pathway and ignores biological complexity. Breaking Eroom's Law requires integrating knowledge from multiple disciplines, including systems biology, network medicine, and immunology (Arrowsmith & Miller, 2013).

2.5 Cross-Domain Case Studies

The theoretical argument becomes concrete through historical cases where a solution existed in one field for years or decades before being applied in another. In each case, the connection was eventually made by a person who happened to work across both domains, or by accident. No systematic process existed to make the transfer happen.

mRNA Vaccines: Messenger RNA was discovered in 1961. It was not successfully applied to human vaccines until the COVID-19 pandemic. The gap had nothing to do with molecular biology. The bottleneck was drug delivery. The field of materials science had been developing lipid nanoparticles and polymer coatings for years, but that work was published in polymer chemistry journals that biologists rarely read (PMC, 2025). The convergence of materials science and biology was one of the highest-return events in the history of medicine. It took roughly 40 years to occur.

Aerospace to Cardiology: NASA’s biotelemetry devices, developed to monitor astronaut health in flight, were eventually adapted for hospital use to monitor patients with cardiac abnormalities (NASA Technical Reports). Digital Twin technology, originating in aerospace to track spacecraft behavior in real-time, is now being explored for "human-on-a-chip" models and personalized medicine. These transfers required specialized "Biomedical Applications Teams" acting as human routers for technology (Scholarly Commons, ERAU).

Self-Healing Concrete: Civil engineers developed self-healing concrete by embedding dormant bacteria that produce calcium carbonate to repair cracks (ResearchGate). This required understanding both structural engineering and microbiology, two fields that typically operate in complete isolation.

Table 5. Documented Cross-Domain Transfer Delays

Source Domain	Target Domain	Solution Applied	Transfer Delay
Polymer Chemistry	Molecular Biology	Lipid nanoparticles for mRNA delivery	~40 years
Aerospace Engineering	Cardiology	Biotelemetry for heart monitoring	~15 years
Aerospace Engineering	Personalized Medicine	Digital Twin simulations	~20 years
Civil Engineering	Biotechnology	Microbial calcite precipitation	~10 years

2.6 Federal R&D Inefficiency

The federal government is one of the largest R&D investors, but efficiency is hampered by agency silos and coordination failures. The Government Accountability Office has repeatedly flagged redundant R&D as a concern. A GAO review of DOE’s advanced reactor R&D found the agency funding multiple competing technologies without a clear strategy for commercialization (GAO-14-545). A 1992 GAO report on DoD estimated costs for certain research verification protocols at $1.4 billion (GAO NSIAD-92-149). A NASA OIG investigation identified duplicate SBIR contract deliverables worth approximately $28.6 million (NASA OIG IG-11-010-R).

The National Academy of Sciences has identified "intellectual turf" and "disciplinary disdain" as major barriers to federal research efficiency. Scientists are trained in a single discipline and adopt its specific analytical constructs. This professional socialization makes it hard to appreciate work done in other fields (National Academies, 2005). Different agencies continually rediscover each other's findings because they have different names for them (NCBI Bookshelf).

Stanford economists found that maintaining performance growth rates in industries like semiconductors now requires 18 times more inflation-adjusted R&D spending than in the 1970s (McKinsey). The "easy discoveries" have been made. Future progress requires better routing of existing knowledge to find new combinations.

2.7 The Researcher Perspective

A 2024 survey of federally-funded investigators found that nearly 90% chose to collaborate across disciplines. However, 80% worked with colleagues from fields very different from their own, creating significant communication and data integration challenges (Frontiers in Education, 2024). The primary barrier: jargon. Scientists speak the language of their discipline, which acts as a wall to understanding and being understood by others (NCBI Bookshelf). A researcher in immunology might work on a pathway already characterized by a cell biologist, but because they use different terms for the same phenomenon, they never find each other's papers.

The NSF Research Traineeship Program was designed to train scholars who can move beyond disciplinary boundaries (PMC/PLOS One). These programs face resistance from traditional departments that prioritize professional identity over collaboration (National Academies, 2005). The current tenure and promotion system often fails to account for interdisciplinary contributions, discouraging young scientists from taking the risks required for cross-domain breakthroughs.

2.8 The Cognitive Barrier

The root cause of cross-domain failure is cognitive. Dedre Gentner's structure-mapping theory explains it: analogy is a mapping of knowledge from a base domain to a target domain that conveys shared relational structure (Gentner & Smith, 2012). The movement of a crowd through a narrow gate is analogically similar to electricity flowing through a resistor. The objects (people and electrons) share no surface resemblance, but the relational structure of "impeded flow" is identical.

The problem: human memory retrieval is heavily biased toward surface features (Gentner, 2025). People are much more likely to recall a prior situation if it looks like the current one, even if the underlying mechanics differ. This "surface similarity bias" makes it extremely difficult for researchers to spontaneously retrieve structural solutions from other domains.

Kevin Dunbar's in vivo studies of microbiology labs confirmed this. Scientists frequently use analogies to interpret unexpected findings, but 98% of these analogies are "local" or "regional" (Schunn & Dunbar). Spontaneous long-distance analogies were nearly non-existent in the daily practice of top-tier scientists. The cross-domain connection depends on rare individuals who happen to hold two domains in working memory at once. That is the gap a systematic routing tool fills.

Table 6. Analogical Retrieval by Similarity Type (after Gentner)

Similarity Type	Cognitive Mechanism	Ease of Retrieval	Strategic Value
Surface Features	Concrete objects/attributes	Very High	Low (Incremental)
Relational Structure	Causal/mathematical patterns	Very Low (Spontaneous)	Very High (Breakthrough)
Structural Alignment	Mapping objects to parallel roles	Requires effort/expertise	High (Insight)
Higher-Order Relations	"Cause," "Impedes," "Equivalence"	Extremely High (Expertise)	Highest (Paradigm Shift)

3. The Precedent Case

Information infrastructure markets are real, large, and defensible. The pattern is consistent: someone builds a routing layer for a data category that professionals depend on, it becomes embedded in workflows, and it becomes nearly impossible to displace. This has happened in financial data (Bloomberg), intelligence synthesis (Palantir), and research indexing (Clarivate/Elsevier). The research version of this infrastructure does not yet exist.

3.1 Bloomberg Terminal

Before Bloomberg, financial data existed but was fragmented across paper records, scattered electronic services, and manual processes. Trading volumes in the 1960s-70s caused such severe operational problems that the NYSE was forced to close on Wednesdays just to let firms catch up with paperwork. Several banks and brokers collapsed because they could not manage the information volume (LSE Working Paper).

Michael Bloomberg was fired from Salomon Brothers in 1981 with a $10 million partnership settlement (Wikipedia). He bet that Wall Street would pay a premium for high-quality information delivered instantly on computer terminals. The first Bloomberg Terminal was installed at Merrill Lynch in 1982. It combined bond and stock prices, news, and analytics that had previously existed only in scattered formats (FundingUniverse).

In 2024, Bloomberg achieved an estimated $15 billion in annual revenue (Wikipedia/Bloomberg L.P.). The terminal accounts for more than 85% of this, with some estimates as high as 92% (PortersFiveForce.com). Bloomberg holds 33.4% of the $30 billion financial data market. A single terminal costs $31,980 per year. Customer retention exceeds 90%.

Table 7. Bloomberg Terminal Financial Metrics

Metric	Value	Source
Annual Revenue (2024)	$15 Billion	Bloomberg L.P., Wikipedia
Market Share (Financial Data)	33.4%	Sparkco
Global Subscribers	325,000+	PortersFiveForce.com
Single Terminal Price (2025)	$31,980/year	PortersFiveForce.com
Customer Retention Rate	>90%	PortersFiveForce.com
Financial Analytics Market (2024)	$10.7B - $12.5B	Fortune Business Insights
Projected Market (2034)	>$27 Billion	Fortune Business Insights

The stickiness comes from workflow integration and network effects. The proprietary Instant Bloomberg (IB) chat has become the de facto communication layer for Wall Street. If every counterparty uses Bloomberg, you must use Bloomberg (The Good Investors, 2026). This is the proof point: build the infrastructure layer that professionals depend on, and you build a durable, high-value business. A research equivalent does not yet exist.

3.2 Palantir Technologies

Palantir's origin story is tied to the intelligence-sharing failures that led to 9/11. The 9/11 Commission Report concluded that the government's principal failure was a failure to connect the dots (Domestic Preparedness). Information was stored in varied corners of government, but siloed thinking and bureaucratic policies blocked sharing. Before Palantir, most CIA and FBI databases were siloed, forcing analysts to search each one individually. Piecing together fragments of an attack campaign took weeks or months (Google Patents, US11601442B2).

Palantir built a platform to fuse unstructured data points into a single analyzable whole. The core technology is the ontology system in Palantir Foundry (Palantir Documentation). This layer represents decisions, not just data. It unifies disparate sources from fragmented systems into coherent objects, properties, and links, creating a digital twin of the organization.

Table 8. Palantir Technologies Financial Performance (FY 2025)

Financial Metric	FY 2025 Result	Source
Total Revenue	$4.48 Billion	Wikipedia
Revenue Growth (Y/Y)	56%	Palantir Investor Relations
U.S. Commercial Revenue	$1.47 Billion	Palantir Investor Relations
U.S. Government Revenue	$1.86 Billion	Palantir Investor Relations
GAAP Net Income	$1.63 Billion	Wikipedia
Market Cap (March 2026)	~$330-342 Billion	Public.com
U.S. Commercial Growth (Q4 2025)	137% Y/Y	Palantir Investor Relations

3.3 Historical Cross-Domain Breakthroughs

The most valuable technology in everyday life came from research in a completely different domain. The pattern is consistent.

GPS was developed as a U.S. military navigation system starting in 1973 (Economy Insights). GPS has generated roughly $1.4 trillion in U.S. economic benefits since the 1980s. Telecommunications accounts for 51% of these benefits, telematics 24%, and location-based services 16%. Precision agriculture alone has generated over $5.8 billion in U.S. benefits since 1998.

The transistor was invented at Bell Labs in 1947 as a replacement for vacuum tubes in the telephone network (Computer History Museum). By 1960, most new computers were transistorized. This domain transfer from telecommunications to data processing created the foundation for the entire modern electronics industry.

The internet began as ARPANET, a DoD project in the late 1960s. The first node was at UCLA in 1969 (Wikipedia/ARPANET). NSF launched NSFNET in 1986 as the internet backbone. In 1995, backbone functions transferred to the commercial sector (NSF Impacts). The digital economy is now valued at approximately $16 trillion, representing 15% of global GDP (IDCA, 2025).

3.4 Clarivate and the Research Indexing Market

Table 9. Major Research Indexing Companies

Company	Annual Revenue (Approx.)	Market Cap (March 2026)
Clarivate	$2.5 Billion	$1.6 Billion
Elsevier (RELX, journal segment)	$1.76 Billion	N/A
Springer Nature Group (journal segment)	$1.33 Billion	N/A

The global STM journal market is estimated at $10 billion (SPARC). Academic research databases as a tool category were valued at approximately $280 million in 2024 and growing (Proficient Market Insights). University library collection budgets for journal and database access at a medium-sized institution run around $3.6 million per year (UNCG Libraries), with about a third of total library expenditure going to database and journal subscriptions (ACRL, 2024).

Researchers report significant dissatisfaction. Existing recommender systems suggest items similar to what users have already searched for, limiting novel discovery across fields (SciTePress, 2016). Citation networks are inherently intra-domain: papers in different fields rarely cite each other even when they solve the same problem. 50% of the journals institutions subscribe to account for only 5-10% of total article readership (SPARC).

3.5 AI-Native Vertical Intelligence

Venture firms are actively investing in vertical AI knowledge tools. a16z argues that founders must translate model capabilities into industry-specific value. Sequoia frames AI as capturing portions of labor spend by performing work. Lux Capital uses outcome-focused frameworks to identify high-impact use cases in science and tech (Contrary Research; a16z; Lux Research).

Current players are building real companies in this space, but they remain single-domain tools: Hebbia raised $130 million Series B from a16z, focuses on finance and legal (Hebbia Blog). Consensus and Elicit conduct literature reviews but do not proactively route knowledge across fields (ResearchGate). Primer focuses on intelligence synthesis for defense, now partnered with Palantir through FedStart (PrimerAI). Capital is flowing into AI-native knowledge infrastructure. None of the current players solve the cross-domain routing problem. They optimize search within a field. Cross-domain routing requires a structurally different approach.

4. The Urgency Case

This is a strategic competition, and the United States faces a specific capability gap in research intelligence infrastructure. China is building it through state policy. The UK, EU, and Israel are investing through targeted programs. The U.S. has significant AI-for-science investments underway, but none of them address the specific problem of routing knowledge between scientific fields. Meanwhile, American R&D productivity is declining at a measurable rate, and the cost of finding new ideas is compounding.

4.1 China's Science-of-Science Investments

The Chinese government has formalized research intelligence as a national security priority. The 14th Five-Year Plan and the Long-Range Objectives through 2035 describe a transition toward quality-driven development (PMC, 2021). The leadership is implementing the Strong Authoritarian Mobilization and Original Innovation model (SAMI 2.0), built on the assumption that China must generate original breakthroughs through a centralized, party-led system (UC IGCC/Cheung). The New Whole-of-Nation System combines state resources, market activity, and academic institutions. China has formed two dozen central-level innovation consortia. The Central Science and Technology Commission, established in 2023, provides oversight for the entire apparatus.

China is building specific tools to manage its national research output. The National Science and Technology Information System uses data warehouse technology to integrate information from distributed sources (IEEE/ICSS, 2021). The Institute of Scientific and Technical Information of China is developing a National Science and Technology Information Big Data Center (EKT, Greece). Chinese institutions lead in bibliometrics and citation network analysis, using tools like CiteSpace (PMC).

4.2 U.S. R&D Spending and Declining Returns

The FY2025 federal budget proposal includes $202 billion for R&D, a 4% increase over the previous year (CRS R48307). Six agencies receive nearly 95% of all federal R&D funding, with DoD and HHS accounting for more than 70% of the budget.

Table 10. Federal R&D Budget by Agency (FY2025 Proposed)

Agency	FY2025 Proposed R&D ($B)	% Change from FY2024
Department of Defense	92.8	+2%
Health and Human Services	51.3	+8%
Department of Energy	23.4	+5%
NASA	11.7	-1%
National Science Foundation	8.1	+4%
Department of Agriculture	3.3	-3%

A major study by Bloom, Jones, Van Reenen, and Webb found that research productivity in the United States falls in half every 13 years (MIT IDE; Stanford). In semiconductors, the number of researchers needed to double chip density per Moore's Law is now 18 times larger than in the early 1970s, with sector productivity falling 7% annually. In medicine, the productivity of creating new molecular entities has fallen by a factor of 11. In medical research, it takes an average of 17 years for original research to reach patient care (PMC). Only about 14% of research eventually reaches clinical practice (WashU Public Health). About 35% of published studies are effectively lost because of inconsistent indexing (Oxford Academic).

4.3 Defense Sector

The number of prime defense contractors has dropped from 51 in the 1990s to fewer than 10 today (EurAsian Times). Firms specializing in defense with little commercial business now account for 61% of major DoD programs by value, compared to 6% in 1989. DARPA is the exception: it funds cross-domain work because it produces outsized results. Programs like Polyplexus (CHIPS/DONCIO), CASCADE (Issues in Science and Technology), and the Adapting Cross-domain Kill-webs program (Dr. Tompkins, Senate Testimony) demonstrate the value of synthesis. But these are specialized efforts, not common infrastructure.

4.4 Current U.S. Federal Initiatives

It would be inaccurate to say the U.S. government is doing nothing. Multiple federal agencies are making large, real investments in AI for science. The question is whether any of these initiatives solve the specific problem of cross-domain knowledge routing. They do not. They solve adjacent problems, and they validate the market thesis, but the translation layer between scientific fields remains unbuilt.

The Genesis Mission (DOE, November 2025). Launched by Executive Order 14363 on November 24, 2025, led by the Department of Energy with the stated goal of doubling R&D productivity within a decade (White House, EO 14363). Under Secretary for Science Darío Gil leads the initiative (DOE, Nov 2025). As of March 2026, 51 organizations have signed collaboration agreements, including NVIDIA, OpenAI, IBM, Microsoft, Google, and Oracle (Inside Government Contracts, Mar 2026). DOE has announced $293 million in funding across 21 challenge areas (DOE RFA, Mar 17, 2026). A separate $320 million has been invested in the American Science Cloud and the Transformational AI Models Consortium (Inside Government Contracts, Mar 2026). The Genesis Mission Consortium launched February 6, 2026 (DOE Consortium Announcement). CSIS described Genesis as a "high-conviction bet" but flagged that "past attempts at coordination have struggled or failed outright" (CSIS, Feb 2026). What Genesis does: it connects DOE's 17 national laboratories into an integrated AI platform. What it does not do: route knowledge between scientific fields across agencies.

NAIRR: National AI Research Resource (NSF-led, 2024). A multi-agency pilot with 15 federal agencies and 28 private-sector partners. It has supported over 600 research projects and 6,000 students across all 50 states (NSF NAIRR). NSF is establishing a NAIRR Operations Center to transition the pilot into a sustained program (NSF Solicitation 25-546, Feb 2026). NAIRR solves access to compute and data. It does not solve cross-domain knowledge routing.

NIH Bridge2AI ($130M, Advancing to Stage 2). Invested $130 million over four years to generate AI-ready biomedical datasets (NIH Common Fund). In January 2026, the NIH Council of Councils approved Bridge2AI to advance to Stage 2, deploying datasets through "Innovation Funnels" (NIH, Jan 29, 2026). Bridge2AI operates within the NIH's biomedical scope. It does not bridge between biomedicine and other scientific domains.

NSF AI Research Institutes (29 Institutes, $100M New in 2025). NSF expanded its network with a $100 million investment in partnership with Capital One and Intel (NSF AI Institutes, Nov 2025). NSF invests over $700 million per year in AI research overall (NSF AI Focus Area). Each institute operates within its target domain. There is no institute or program tasked with routing knowledge between them.

Taken together, these programs validate that R&D productivity is a recognized national priority. They invest in compute, data, and domain-specific AI. The translation layer that routes knowledge across field boundaries is the piece that remains missing.

4.5 Allied Nations

United Kingdom: UKRI is investing £1.6 billion in the AI sector for 2026-2030 (UKRI). A major component is the AI for Metascience mission (Centre for British Progress). The Alan Turing Institute is working on data-efficient AI to accelerate discovery across scientific domains (Alan Turing Institute). European Union: The European Open Science Cloud (EOSC) is a federated system for data across borders and fields (EOSC). Israel: Elite military units like Unit 8200 act as incubators for technology companies, with a national AI program strengthening research infrastructure (Israel National News).

5. Technical Timing

Cross-domain knowledge routing requires the ability to identify structural similarity between concepts described in completely different vocabularies. The technical milestones that made this tractable are recent: Word2Vec in 2013 demonstrated semantic relationships in vector space. BERT in 2018 enabled contextual language understanding. GPT-3 in 2020 showed that large language models could reason across domains. Domain-specific embeddings followed. The academic literature on cross-domain semantic matching is largely post-2018, confirming the timing.

Previous attempts relied on keyword matching, ontology bridging, or manual curation. These could not scale. The jargon problem defeated them. Two fields can describe the same phenomenon with entirely different vocabulary, and without the ability to decompose concepts at a structural level, no automated system could find the connection. That capability is new.

6. Buyer Landscape and Market Sizing

The buyers have large R&D budgets. The total addressable market spans sectors where budgets already exist: Pharmaceutical and Biotech (U.S. pharma R&D exceeding $100 billion annually); Defense Contractors and National Labs (top five primes holding over $155 billion in DoD contracts); Semiconductors (requiring 18x more researchers than in the 1970s); and Government Agencies (federal R&D budget of $202 billion with no unified cross-domain system). Pricing precedents are strong: Bloomberg charges $32,000/seat/year, Palantir's contracts are public record, and pharma companies already pay for tools like PatSnap, CAS SciFinder, and Cortellis.

7. Conclusion

The data across all three pillars converges on a single conclusion. The United States has a translation problem, and it is measurable, expensive, and getting worse. The waste is documented at $28 billion per year in irreproducible research, with an 80-fold decline in pharmaceutical productivity and a 17-year average lag from discovery to clinical practice. The market for information infrastructure is proven at scale by Bloomberg ($15B), Palantir ($330B), and GPS ($1.4T). The strategic competition is real: China has formalized research intelligence as state policy, allied nations are investing, and the U.S. federal government is investing billions in AI-for-science but not in the specific layer that routes knowledge between fields. The technology to build this layer became tractable only recently. The window is open.

> *The country that builds the most effective system for navigating the global body of scientific knowledge sees the next frontier of science before its competitors. It knows which research is likely to fail and which is likely to produce a breakthrough. In a world where ideas are getting harder to find, the advantage goes to the one who knows where to look.*

References

Organized by source document. Bracketed numbers correspond to the numbered works cited list within each research file. Full URLs are available in the original documents for direct verification.

Pillar 1: The Waste Case

Source file: Research_Plan__Waste_Case_Analysis.docx

[1] "Reproducibility2020: Progress and Priorities." PMC/NIH. pmc.ncbi.nlm.nih.gov/articles/PMC5461896/

[2] "Reproducibility in Pre-Clinical Life Science Research." Culture Collections.

[3] "The Reproducibility Crisis in Preclinical Research: Lessons from Clinical Research." Trilogy Writing & Consulting.

[4] Freedman, L.P., Cockburn, I.M., & Simcoe, T.S. (2015). "The Economics of Reproducibility in Preclinical Research." PLOS Biology.

[5] Freedman, L.P., Cockburn, I.M., & Simcoe, T.S. (2015). "The Economics of Reproducibility in Preclinical Research." PLOS Biology. journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165

[7] "Improving Transparency and Scientific Rigor in Academic Publishing." ResearchGate.

[8] "Many Clinical Trial Results Not Shared, Creating 'Blind Spot.'" Yale News, 2016.

[11] "Reviving a Pharmaceutical R&D Pipeline." DrugPatentWatch.

[13] "Case Study: Insilico's Transformation." Insilico Medicine.

[15] Arrowsmith, J. & Miller, P. (2013). "Phase II and Phase III Failures: 2013-2015." ResearchGate.

[17] "Preventing Phase III Failures." Pharmaceutical Technology.

[19] "Next-Gen Therapeutics: iPSCs, Genomics, AI." PMC/Annual Reviews.

[21] Arrowsmith, J. & Miller, P. (2013). "Trial Watch: Phase II and Phase III Attrition Rates 2011-2012." Nature Reviews Drug Discovery.

[22] "Medical Applications of Aerospace Technology." Scholarly Commons, ERAU.

[24] Scannell, J. et al. (2012). "Eroom's Law." Wikipedia.

[29] "Revolutionizing Drug Delivery: Advanced Materials Science." PMC.

[31] "Aerospace Medicine and Biology." NASA Technical Reports Server.

[35] "Microbial Biotechnology in Concrete Engineering." ResearchGate.

[38] "Advanced Reactor Research: DOE Technologies." GAO-14-545.

[39] "Improved Coordination of Arms Control Research Needed." GAO, NSIAD-92-149.

[42] "Review of NASA's Management of Its SBIR Program." NASA OIG, IG-11-010-R.

[43] "Barriers to Interdisciplinary Research and Training." National Academies, 2005.

[44] "Barriers to Interdisciplinary Research." NCBI Bookshelf.

[45] "How AI Is Driving R&D Productivity." McKinsey.

[46] "Post-Award Grant Collaboration: Cross-Disciplinary Authorship." Frontiers in Education, 2024.

[47] "Lessons from Interdisciplinary NSF Research Traineeship Programs." PMC/PLOS One.

[56] Foster, A. & Ford, N. (2003). "Serendipity and Information Seeking." Aberystwyth University.

[59] "Neuroscientific Insights into Analogical Reasoning." PMC.

[60] Gentner, D. & Smith, L. (2012). "Analogical Reasoning." Northwestern Psychology.

[66] Gentner, D. (2025). "Analogy." OECS. Northwestern Psychology.

[68] Schunn, C. & Dunbar, K. "The Relationship of Analogical Distance to Analogical Function." University of Pittsburgh.

Pillar 2: The Precedent Case

Source file: Researching_Information_Infrastructure_Precedents.docx

[1] "The Terminal Revolution: Reuters and Bloomberg 1960-2020." LSE Working Paper.

[2] "Bloomberg Terminal." Wikipedia.

[3] "Bloomberg L.P." Wikipedia.

[6] "History of Bloomberg L.P." FundingUniverse.

[8] "How Does Bloomberg Company Work?" PortersFiveForce.com.

[9] "Bloomberg Industry Disruption Forecasts 2025." Sparkco.

[12] "Financial Analytics Market Size [2034]." Fortune Business Insights.

[15] "IB Chat / Network Effects." The Good Investors, 2026.

[17] "Palantir: From Secretive Startup to Data Superpower." Medium/Evergreen LLC.

[19] "Palantir Technologies (PLTR) Market Cap." Public.com.

[21] "Palantir." Wikipedia.

[22] "Palantir Q4 2025 Revenue Report." Palantir Investor Relations.

[24] "The Ontology System." Palantir Documentation.

[26] "Overview: Ontology." Palantir Documentation.

[28] "Repeated Intelligence Failures: Not Connecting the Dots." Domestic Preparedness.

[31] "US Patent US11601442B2." Google Patents.

[32] "The Economic Power of GPS Technology." Economy Insights.

[40] "Inventing the Transistor." Computer History Museum.

[45] "History of the Transistor." Wikipedia.

[47] "Birth of the Commercial Internet." NSF Impacts.

[49] "ARPANET." Wikipedia.

[52] "Global Digital Economy Report (2025)." IDCA.

[53] "Progress Deferred: Lessons from mRNA Vaccine Development." IFP.

[54] "Messenger RNA-Based Vaccines: Past, Present, Future." PMC.

[56] "Who Invented the Microwave." Bajaj Finance.

[59] "Evolution of Materials: Fiberglass and Carbon Fiber." CustomFRP.

[62] "Clarivate (CLVT) Revenue." Companies Market Cap.

[66] "Research Companies: Elsevier." SPARC.

[67] "Global Academic Research Databases Market Report 2025." Proficient Market Insights.

[68] "Collections Budget Central." UNCG Libraries.

[69] "2024 Academic Library Trends Survey." ACRL/CRLN.

[70] "Challenges of Serendipity in Recommender Systems." SciTePress, 2016.

[75] "The Vertical AI Playbook." Contrary Research.

[77] "Context Is King." a16z.

[82] "Hebbia Series B." Hebbia Blog.

[85] "Effectiveness of AI Tools (Elicit, SciSpace, Consensus)." ResearchGate.

[86] "Primer + Palantir." PrimerAI.

Pillar 3: The Urgency Case

Source file: Researching_China_s_Research_Intelligence.docx

[1] "Innovation and Reform: China's 14th Five-Year Plan." PMC.

[3] Cheung, T. "China's Competitive Model for Strategic STI (SAMI 2.0)." UC IGCC.

[5] "Data Warehouse for Science and Technology Management System." IEEE/ICSS, 2021.

[6] "Visit of Delegation from ISTIC." EKT, Greece.

[8] "CiteSpace Bibliometric Analysis." PMC.

[14] "Federal R&D Funding: FY2025." Congressional Research Service, R48307.

[15] Bloom, N., Jones, C., Van Reenen, J., & Webb, M. (2020). "Are Ideas Getting Harder to Find?" MIT IDE.

[16] Bloom, N., Jones, C., Van Reenen, J., & Webb, M. (2020). "Are Ideas Getting Harder to Find?" Stanford University.

[17] "Evidenced-Based Claims About Evidence." PMC.

[20] "Research Priority: Implementing What Works." WashU Public Health Sciences.

[21] "Practice-Based Evidence." Oxford Academic/Family Practice.

[22] Reinhardt, B. "Why Does DARPA Work?" benjaminreinhardt.com.

[23] "Lockheed, Boeing, Northrop: Why U.S. Could Lose the Next War." EurAsian Times.

[25] "U.S. Awards Lockheed Martin Record $68.5B." 24/7 Wall St.

[26] "DARPA Polyplexus Collaboration Site." CHIPS/DONCIO.

[27] "CASCADE/Liminal Design AI." Issues in Science and Technology.

[28] Tompkins, S. "Adapting Cross-Domain Kill-Webs (ACK)." Senate Testimony.

[31] "PubMed Knowledge Graph 2.0." PMC.

[35] "SEAB AI/ML Working Group Report." Department of Energy.

[36] "Public Input on Cyber-Physical Systems Resilience Plan." NITRD/PCAST.

[38] "UKRI AI Strategy." UK Research and Innovation.

[39] "Britain's AI for Science Strategy." Centre for British Progress.

[40] "Turing Researchers: Data-Efficient AI for Discovery." Alan Turing Institute.

[41] "EOSC EU Node." European Open Science Cloud.

[44] "AI Development in Israel." Israel National News.

[46] "Knowledge Transfer, Gaps, and Silos in Citation Networks." PMC.

[47] "Controlling the Innovation Chain." MERICS.

[48] "14th Five-Year Plan for National Informatization." DigiChina/Stanford.

U.S. Federal Initiative Sources

White House. Executive Order 14363: Launching the Genesis Mission. November 24, 2025. whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/

Department of Energy. "Energy Department Launches Genesis Mission." November 2025. energy.gov

Department of Energy. "26 Genesis Mission Science and Technology Challenges." February 12, 2026. energy.gov

Department of Energy. "$293 Million in Funding for Genesis Mission." March 17, 2026. energy.gov

Department of Energy. "Genesis Mission Consortium Launch." February 6, 2026. energy.gov

Department of Energy. "24 Collaboration Agreements." December 2025. energy.gov

CSIS. "The Genesis Mission: Can the U.S. Bet on AI Revitalize Science?" February 9, 2026. csis.org

HPCwire. "DOE Announces $293M as Genesis Moves Toward Operational Phase." March 18, 2026. hpcwire.com

Inside Government Contracts. "DOE Announces Funding Opportunity Under Genesis Mission." March 2026.

National Science Foundation. National Artificial Intelligence Research Resource (NAIRR). nsf.gov/focus-areas/ai/nairr

NAIRR Pilot Portal. nairrpilot.org

NSF Solicitation 25-546: NAIRR Operations Center. February 2026. nsf.gov

GovTech. "2 Years Into NAIRR Pilot." December 2025. govtech.com

NIH Common Fund. Bridge to Artificial Intelligence (Bridge2AI). commonfund.nih.gov/bridge2ai

NIH. "Bridge2AI Advances to Next Stage." January 29, 2026. commonfund.nih.gov/bridge2ai/news

NSF AI Institutes. $100M Investment in 5 New Institutes. November 2025. aiinstitutes.org

NSF. Artificial Intelligence Focus Area: $700M+ Annual. nsf.gov/focus-areas/ai