Strengthening Trust in News: Implementing Identity Graphs for Authors and Sources

by Dinis Cruz and ChatGPT Deep Research and Claude 3.7, 2025/04/21

Executive Summary

Widespread exposure of fabricated “experts” such as the *Barbara Santini* hoax has revealed a systemic vulnerability in modern newsrooms: speed-driven workflows too often publish quotes from unverified or entirely fictional commentators, eroding public confidence in journalism. This white paper argues that manual vetting cannot keep pace with increasingly sophisticated fakery—and that trust in news will continue to decline unless audiences can see, not just assume, who is behind every claim.

We propose a newsroom-wide adoption of **Identity Graphs**: dynamic, data-rich profiles that verify and continuously update the credentials, affiliations, and past media appearances of both authors and quoted sources. Built on the same semantic-graph principles already used for content metadata, an identity graph links each person to authoritative records (licences, academic rosters, ORCID, etc.), assigns a machine-readable credibility score, and plugs seamlessly into editorial and publishing workflows.

Key benefits include:

- **Editorial assurance** – instant, evidence-based checks of a source’s legitimacy before publication.
- **Audience transparency** – clickable profile cards on published stories that let readers inspect an expert’s verified background.
- **Brand and revenue upside** – differentiation through demonstrably trustworthy content, new subscription tiers, data-licensing opportunities, and AI-readiness for fact-checking tools.

By transforming source vetting from an ad-hoc task into a graph-driven, continuously verified system, news organizations can hard-wire accountability into every article, rebuild audience trust, and create monetizable “trust services” for the wider information ecosystem.

Introduction

In April 2025, major UK news outlets scrambled to remove or amend dozens of articles after the discovery that a widely-quoted "expert" – *Barbara Santini* – might not be who she claimed ('Immediate red flags': questions raised over 'expert' much quoted in UK press | National newspapers | The Guardian). Santini, billed as an Oxford-educated psychologist, had provided commentary on topics ranging from pandemic stress to vitamin supplements across publications like *Cosmopolitan*, *The Telegraph*, and even the BBC. However, investigations raised *"immediate red flags"* about her credentials and identity, suggesting her entire persona could be an elaborate hoax (Virtual reality: The widely-quoted media experts who are not what they seem - Press Gazette). Newsrooms quickly pulled Santini's quotes and articles amid the revelations. This high-profile case is not isolated: an industry probe found multiple *"virtual"* experts infiltrating news features – commentators with impressive resumes who turned out to be fictitious or unverified. Even a White House adviser once invented a fake economist "expert" (the infamous *Ron Vara*, an anagram of his own name) to bolster his arguments in print (White House Adviser Peter Navarro Calls Fictional Alter Ego An 'Inside Joke' : NPR). These incidents underscore a growing threat to journalistic integrity: fake or unverifiable sources are making their way into reputable media content.

Such breaches of authenticity erode public trust in journalism. If readers cannot be confident that quoted experts or even authors are genuine and qualified, how can they trust the information being presented? This white paper makes the case that news organizations need to tackle this challenge head-on. It proposes the adoption of **identity graphs** for authors and sources – data-rich, interconnected profiles that verify and contextualize the people behind the news. By leveraging techniques from semantic knowledge graphs and emerging verification technologies, media companies can restore transparency, credibility, and trust. The paper will examine the fake-expert problem and its impact on trust, explain the identity graph solution and how it builds on semantic graph work, outline implementation steps, explore monetization opportunities, and compare this approach with other trust initiatives (like NewsGuard and The Trust Project).

The Rise of Fake or Unverifiable Sources in Journalism

Modern newsrooms face increasing pressure to produce content quickly, often relying on third-party *"expert comment"* services to provide quotes for lifestyle, health, and finance stories. Unfortunately, this speed-over-substance dynamic has opened the door to *"dubious commentators"* who are either misrepresenting themselves or outright fabrications (Fake experts: Publishers delete articles after Press Gazette report). The **Barbara Santini** saga is a case in point. Over recent years, Santini had been cited in dozens of articles as a psychologist offering advice on relationships, wellness, and more. Only upon closer scrutiny did journalists discover that her only verifiable online presence was as a consultant for a sex toy retailer, and that she had no recognizable academic or professional footprint – not even a social media profile. Major publishers like *Reach* (owner of the *Mirror* and *Express*) and *News UK* reacted by removing Santini's contributions from their archives, treating the case as a wake-up call.

Press Gazette's investigation in April 2025 revealed that Santini was far from alone. The report identified a pattern of **"widely-quoted media experts" who *"are not what they seem"***. Some were real people exaggerating their qualifications, while others turned out not to exist at all. For example, *"Rebecca Leigh"* was cited in numerous stories on topics from employee benefits to music streaming, yet when contacted, the company behind her admitted *"the name and the photo are not real"*. Another phony persona, *"Charlotte Cremers,"* posed as a London physician to give health advice in exchange for backlinks to a commercial site. In the United States, *Business Insider* reported being approached by a supposed cancer survivor offering commentary – who turned out to be AI-generated. And in one particularly ironic case, former Trump adviser Peter Navarro confessed that the oft-cited "economics expert" in his books, *Ron Vara*, was a fictitious alter ego he used as an "inside joke" and to reinforce his policy views. From niche trade publications to mainstream news, fake experts have infiltrated the media ecosystem.

Several factors drive this phenomenon. **Digital PR agencies** and opportunistic marketers have discovered they can game the system by fabricating experts who submit ready-made quotes to journalists. The payoff is valuable: each quote published brings a mention of their affiliated brand and sometimes a link, boosting SEO rankings for products like CBD oil, diet supplements, or academic essay services. The launch of advanced AI tools like GPT-3/4 has only *"exacerbated"* the issue, making it trivially easy to generate authoritative-sounding commentary and even realistic headshots for fictitious personas. As one media editor noted, *"AI tools make it far easier for bad actors to invent supposed experts for their own purposes"*. The result is a *virtual carousel* of fake pundits slipping into news stories via journalist inquiry platforms (e.g. ResponseSource or Qwoted) before adequate vetting occurs.

The journalistic community is beginning to reckon with this. In the wake of the Santini exposé, publishers like **Yahoo News**, **The Sun**, and others purged over a hundred articles that had relied on suspect sources. Internal memos and industry commentary are warning writers and editors to double-check identities lest they fall victim to a hoax. Yet, manual vetting alone may not be sufficient when time is tight and fake identities are growing more sophisticated. The persistence of this problem indicates a systemic gap in how newsrooms verify *"who"* is contributing information to their stories.

How Lack of Transparency Erodes Trust in Media

The prevalence of unverifiable sources strikes at the heart of journalistic credibility. News is a business of trust – readers trust reporters to present facts and trust that the people cited are who they claim to be. When that trust is broken, the fallout is severe. Public confidence in media is already fragile, with surveys showing record-low trust levels. (For example, only **31% of Americans** in 2024 expressed even a "fair amount" of confidence that news is reported fully and fairly (Americans' Trust in Media Remains at Trend Low).) High-profile embarrassments like fake expert scandals can further erode this confidence. If one quoted "psychologist" turns out to be a mirage, audiences naturally wonder how many other experts or even reporters might be fictitious or misrepresented. In the digital age, such revelations spread quickly on social media, feeding narratives that mainstream media is sloppy or deceptive.

Beyond general audience perception, there is a deeper transparency issue: readers often have no easy way to *verify the identity or expertise of sources* quoted in articles. Traditionally, journalism relies on bylines and occasional author bios, and on the assumption that news organizations have vetted their sources behind the scenes. But as missteps come to light, that implicit trust is undermined. Media scholars emphasize that transparency about sources is critical to maintaining trust. The Trust Project – a consortium for journalism standards – includes *"Journalist Expertise"* and *"Citations/References"* among its eight Trust Indicators, underscoring that audiences should be able to find out **who** the journalist or source is and evaluate their credibility (Trust Indicators - The Trust Project). In practical terms, this means readers appreciate when an article provides a clear description of an expert's credentials or a link to learn more about that person. When such context is absent, or worse when a supposed expert's background is vague, it breeds skepticism. As the Trust Project notes, *"When a journalist shows their sources, we can check their reliability for ourselves."*

Lack of transparency not only impacts readers' trust but also harms internal morale and brand integrity. Genuine journalists do not want their hard work tainted by dubious quotes that later require correction or removal. News organizations stake their reputation on accuracy and honesty; having to issue clarifications like the one Yahoo News appended (noting an earlier version included a source whose expertise "may not be valid" can be embarrassing and damaging. Each incident gives ammunition to those who claim that *"the media lies"* or doesn't do its homework. In short, opaque sourcing and unchecked identities create a vacuum where misinformation and doubt flourish. To shore up trust, media must go beyond reactive corrections and adopt proactive measures ensuring transparency about *who* is behind the news and the information within it.

Identity Graphs: A Data-Driven Solution for Verification

To address this credibility gap, we propose implementing **Identity Graphs** for authors and sources as a core newsroom technology. An *identity graph* is essentially a centralized, dynamic profile that links all relevant data about a person's identity and qualifications, much like a knowledge graph but focused on individuals. In the marketing world, identity graphs are used to connect disparate consumer data points into a unified profile (each person is a node, and their devices, emails, and behaviors are linked as edges) (7 Reasons Publishers Should Build Their Own Identity Graph). In a news context, an identity graph would consolidate and verify information about the people *behind the content* – the journalists who write stories and the experts or sources they quote.

What would such an identity graph contain? For a journalist (author), it could include their real name, a verified headshot, a biography with credentials (education, specialties, awards), links to their past articles, and social media handles – all interlinked. For an expert source, the graph profile might store their full name and aliases, professional titles and affiliations (e.g. *Dr. Jane Doe – Professor of Economics at XYZ University*), verified credentials (degrees, licenses, membership in professional bodies), previous appearances in media, and references to any published research or books. Crucially, the identity graph isn't just a static bio – it's a network of data points that can be updated and *cross-checked automatically*. For example, an identity graph entry for *"Barbara Santini"* would have immediately

revealed the lack of any university alumni records or professional memberships to back her “Oxford-educated psychologist” claim. It might have also flagged the unusual detail that her online footprint was limited to a retail website profile – a red flag for editors.

Verification is at the heart of the identity graph approach. Each profile in the graph can be linked to external authoritative sources: a node for an academic expert might link to their **ORCID** or Google Scholar profile (to verify publication history), a node for a medical expert could link to a government registry of licensed practitioners, a node for a journalist could link to their employer’s staff directory or a journalism accreditation database. These connections would be periodically validated, either by automated scripts or during editorial workflows, ensuring that if a person claims to have a Ph.D. or a certain title, the system confirms it. The identity graph could also incorporate **credibility scores** or trust signals. For instance, a source who has been quoted by five reputable outlets and vetted each time might accumulate a high trust score in the system, whereas a new source with no external references would start at a lower trust level. This concept is similar to what media inquiry platforms are now exploring: ResponseSource has indicated it will introduce *“credibility scores”* and peer reviews for the expert profiles in its network (allowing reporters to “thumbs down” dubious sources and highlight trustworthy ones). An identity graph in a newsroom would extend that idea – verifying education claims, tracking if a source consistently delivers reliable information, and even recording if any story required a correction due to that source.

By turning identity verification into a data-driven graph, news organizations gain a powerful tool: **a single source of truth about people associated with their content**. Before an editor publishes a quote from an expert, they could quickly pull up that person’s identity graph entry and see a vetted summary: e.g., *“Dr. X has a valid medical license (verified via government database on DATE), has been quoted in 3 articles on health since 2023, and has a credibility score of 8/10.”* If any aspect is unverified or there are warnings (say, another outlet flagged this person as questionable), the editor is alerted to do further checks or reconsider using the source. In essence, identity graphs bring to people the same rigor of cross-referencing and provenance that we strive for with facts. Just as a well-linked knowledge graph can show the sources behind a piece of information, an identity graph shows the background behind a person’s claims.

Building on Semantic Knowledge Graphs and Provenance

The concept of identity graphs for journalism builds naturally on advancements in **Semantic Knowledge Graphs** that some media organizations are already experimenting with. Semantic graphs organize information in networks of entities (people, places, topics) and relationships, enabling better content personalization and fact provenance. Dinis Cruz’s own work on *Semantic Knowledge Graphs for personalized news* is a prime example: in an MVP project, news articles were ingested and transformed into a graph of entities to provide **provenance** – i.e., showing *why* each article was recommended to each persona (How it works - myfeeds). In that system, Large Language Models (LLMs) helped extract entities and link content, creating a transparent chain from source data (RSS feeds) to delivered news, with the goal of deterministic, explainable recommendations. This demonstrates how graph-based approaches, combined with AI, can bring clarity and traceability to news delivery.

We propose to extend the same principles of **graph-driven provenance** to the domain of identity verification. Instead of (or in addition to) mapping how facts relate, we map **who is behind the facts**. A semantic graph of authors and sources would interconnect with the content graph – for example, an article node links to its author node and to the nodes of each person quoted. This interlinking means one can traverse the graph in powerful ways: click on a source's node to see all other pieces of content they contributed to, or query the graph to find all climate science stories written by authors with a certain expertise. The identity graph becomes a layer of metadata enriching each piece of news with context about its human origins.

Importantly, leveraging AI can make maintaining these identity graphs feasible at scale. Just as LLMs were used to extract topics and entities from text, they can assist in extracting and updating biographical data. For instance, an AI system could routinely scan the web for new information on a known expert (did they receive a new credential? change jobs? get mentioned elsewhere?) and update the graph. However, unlike the freeform use of AI to generate content (which could introduce errors), here AI's role would be constrained to *augmenting verification* – the graph structure provides a deterministic framework where every data point for a person can be traced back to a source or validation method. In technical terms, the identity graph functions as a **"trust network"** within the organization's knowledge system. It implements a *"continuous verification chain"* that maintains clear provenance for identity data (Monetising Trust and Knowledge: How News Providers can leverage Personalised Semantic Graphs - Dinis Cruz - Documents and Research). Each credential or claim in the graph is linked to evidence (e.g., a link to a university's degree confirmation or a professional society membership list), ensuring that the profile's credibility evolves only as new verified evidence comes in. Just as a semantic content graph can track how information flows and evolves, the identity graph tracks a person's credibility over time – perhaps through *temporal trust analysis* that notes if a source's predictions often pan out or if their info has been disputed.

By building on existing semantic graph infrastructures, news organizations can integrate identity graphs without reinventing the wheel. The same graph databases or platforms used for content metadata can house the identity data. And the concept of *provenance*, which is gaining traction to fight misinformation, naturally extends to identities: not only *"What is the source of this fact?"* but also *"What is the source of this quote (and who is this source)?"* Both need to be transparent. In summary, identity graphs complement semantic knowledge graphs by adding the *who* dimension to the *what* and *how*, thereby closing the loop in establishing trust.

Implementation: From Data Collection to User Experience

Implementing identity graphs in a newsroom workflow involves several components: **data collection**, **verification processes**, **graph integration**, and **user interface design**. Below, we break down each aspect, outlining how news organizations can practically build and use identity graphs for authors and sources.

- **Data Collection and Profile Creation:** The first step is to gather authoritative data about each person of interest. For staff journalists and regular contributors, much of this data is internal (HR records, resumes, beats/areas of expertise, links to their work). For external sources such as experts or interviewees, data

collection may involve querying external databases. For example, if a reporter wants to quote a medical expert, they (or a research assistant tool) would enter the expert's name into the system. The identity graph backend could pull from APIs or open datasets: university alumni databases, professional license registries (e.g., a state medical board for doctors), LinkedIn or ORCID for work history and publications, etc. The result is an initial profile node with raw attributes like *education: Ph.D. from X University (year)*, *position: Lecturer at Y Institute*, *works_for: Y Institute*, *credentials: Member of ABC Association*, etc. In addition, any past media appearances could be logged – many experts list press quotes on their personal or company pages, which the system can scrape to link those article URLs as part of the profile. **Importantly, the identity graph should assign a unique identifier** to each person (to avoid confusion of similar names) – this could be facilitated by existing identifiers like ORCID IDs for researchers or creating an internal UID for others.

- **Verification and Validation:** Once data is collected, it must be verified. This is where the identity graph approach stands apart from ad-hoc methods – every key data point should have a verification status attached. Some verifications can be automated: for instance, an algorithm can check if the "University of X – Class of 2010" entry for a person matches a publicly available graduation list, or it can query the membership roster of a professional society via their API. Other verifications might require manual steps, such as contacting the supposed employer to confirm the person's role, or using services like **Verifiable Credentials** (emerging standards where institutions issue cryptographic proofs of degrees or employment). The system might implement a **tiered verification**: basic identity (does the person exist, is their name tied to a real entity) is one level; credential verification (are their qualifications real) is another; reputation check (have they provided reliable info before) is yet another. Each verified element can be timestamped in the graph. For ongoing reliability, verifications should be periodically refreshed – e.g., an annual re-check of a source's professional license status. The end goal is that before someone is cited in an article, their identity graph entry is marked as "Verified" for the claims being made about them. If a claim cannot be verified, that too is noted (e.g., *claimed Oxford degree – not verified by any records*), alerting journalists to proceed with caution or disclose uncertainty.
- **Graph Integration and Linking:** With profiles verified, the identity graph becomes a living database accessible throughout the editorial process. Journalists writing a story can search the identity graph for an expert by name or topic. Instead of relying on a blind email from "Dr. John Doe, expert," they can see if John Doe has a profile in the system. If yes, they can view his credentials and past interactions; if not, they can initiate creating one (as above) *before* quoting him. Each article in the CMS (Content Management System) would be equipped to link to identity profiles for every person referenced. This could be as simple as tagging the name with an internal ID (similar to how tagging a company or topic might link to a taxonomy). Once linked, the article's metadata would include references to all associated identity graph entries. On the back-end, this means an editor or fact-checker can pull a report of all sources in a piece along with their verification statuses. On the front-end (the published site), this linking enables transparency features (discussed below). The identity graph should also ingest the published article link back into each person's node – effectively closing the loop: *Person X – quoted in [Article title, Outlet, Date]*. Over time, the graph accumulates a network: one can see that *"Person X has been cited 5 times in our publication and 12 times industry-wide"*, etc. This historical context is invaluable – if a person suddenly shows up offering commentary on everything (as Santini did) it becomes evident in the graph, allowing pattern detection of possible inauthentic behavior.

- **User Interface and Presentation:** Implementing identity graphs is not just an internal tool; it's also an opportunity to **show transparency to the public** in a user-friendly way. News organizations should surface this rich profile information in their content. For example, author bylines on articles could be interactive – hovering or clicking the journalist's name might bring up a verified profile card: *"Jane Smith – Investigative Reporter with 10 years experience covering healthcare. Education: Columbia Journalism School. Awards: XYZ. Past articles: [link]."* For sources quoted in the piece, a small icon or hyperlink on first reference could indicate more info is available. Readers clicking a quoted expert's name might see a pop-up or sidebar: *"Dr. John Doe – Board-certified cardiologist, UCLA Medical Center. (Verified) – This source has been vetted by [NewsOrg]'s system. Relevant expertise: 20 years in cardiology, 50+ research publications. Quoted in [list of 3 stories]."* Such a feature turns what is often a blind quote into a mini-profile, instilling confidence that this voice is real and relevant. It's important that the interface clearly denotes verification status – e.g., a checkmark or note that certain credentials were confirmed. If any details are unverified or based on the source's own claims, that can be indicated too (transparency about uncertainty). Visualization-wise, an identity graph could also power an interactive "About our sources" section at the end of an article or a central repository on the news site listing all expert contributors and authors with profiles. This level of openness would be a strong signal that the organization has nothing to hide about *who* is informing its journalism.
- **Governance and Privacy:** Implementing identity graphs will require governance policies – especially regarding privacy and consent. Journalists on staff can be expected to participate as part of their role (much like having an author page). For external sources, especially private individuals, newsrooms should be transparent about the data they collect. It's advisable to limit the identity graph to professional information and avoid sensitive personal data not relevant to credibility. In many cases, sources will be flattered to be featured with a profile (experts often welcome visibility). Nonetheless, clear opt-in procedures and data handling compliant with privacy laws (GDPR, etc.) should be established. The identity graph should also have an *audit trail* – if an entry is edited (say a credential removed or corrected), it's logged, maintaining the history. This is akin to Wikipedia's edit history but for identity data, again supporting transparency and accountability.

In sum, the implementation of identity graphs involves weaving a verification mindset into every stage of journalism production. It equips news staff with an accessible database of vetted information about people, much like a modern encyclopedia of sources at their fingertips. By integrating with publishing tools and presenting information to readers, it transforms what could be an internal database into a public trust asset.

Monetization Opportunities for Verified Identity Systems

Beyond the clear editorial benefits, investing in identity graphs and verified source networks can unlock new **monetization and business opportunities** for media organizations. In an era where trust itself has value, news companies can leverage their credibility-enhancing infrastructure in several ways:

- **Premium "Trust Layer" for Subscribers:** A verified content experience could be a selling point for paid subscribers or members. For example, a news outlet might offer a *premium tier* where readers get access to enhanced transparency features powered by the identity graph. This could include detailed source cards, the ability to drill down into the background of any expert in an article, or even a newsletter that introduces "new voices we've vetted this week." Essentially, subscribers pay not just for content, but for insight into the reporting process and confidence in its integrity. Such offerings align with a growing segment of news consumers who are willing to support quality journalism and appreciate transparency as part of the value proposition. Over time, a reputation for thoroughly vetted content can also reduce subscriber churn – readers stick with outlets they deem trustworthy.
- **Licensing and Data Services:** A robust identity graph of authors and sources could become a valuable dataset in its own right. Media organizations could license access to their "credibility graph" to third parties. Consider *NewsGuard*, which rates the trustworthiness of news websites and licenses these ratings to tech platforms and advertisers as a credibility layer. Similarly, a newsroom's verified identity database could be offered (with appropriate API access) to platforms like search engines, social media, or fact-checking coalitions. For instance, a social media company might integrate a feed of trusted expert profiles to help label or elevate credible voices in their feeds. Research firms or libraries might subscribe to query the database when they need to quickly find a verified expert on a subject. Even other news organizations could become clients – smaller outlets might tap into a larger outlet's identity graph service instead of building their own, analogous to how they subscribe to wire services. This creates a B2B revenue stream where the painstaking verification work done by one newsroom benefits others (for a fee), fostering a *"trust network"* across the industry.
- **AI Training and Enterprise Solutions:** In the AI and tech sector, high-quality structured data is gold. A curated graph of verified journalistic sources and authors could be used to **train AI models** for better information discernment. For example, AI fact-checking systems or large language models could be trained to consult the identity graph to evaluate the credibility of a statement (e.g., if a piece of content cites Dr. X, the AI can check the graph to see if Dr. X is recognized as credible). Media organizations could partner with AI developers or sell access to this data for model training, thereby contributing to the wider fight against misinformation. Additionally, enterprise clients – such as financial firms, law firms, or academia – might license the identity graph as a reference tool. Imagine a hedge fund that wants to quickly assess the credibility of an op-ed's author before making a decision based on it, or a university that wants to teach journalism students about verifying sources using real-world data – they could all be customers of an identity-verified knowledge base.
- **Sponsored Content and Advertising Alignment:** While maintaining strict church-state separation between editorial integrity and advertising, there is a scenario where a reputation for verified content can attract *premium advertisers*. Brands increasingly care about the context in which their ads appear (to avoid brand safety issues). An outlet that can demonstrably show that "100% of our content's sources are verified and credible" might have a marketing edge, allowing it to command higher ad rates or secure sponsorships from institutions that align with promoting truth and expertise (such as universities, professional associations, etc.). Furthermore, the identity graph infrastructure could support *native advertising* or events in a way that's ethical – for instance, hosting a directory of industry experts (from the graph) that businesses can access or sponsor, without compromising the editorial use of that graph for news verification.

- **Value-Added Products:** With a comprehensive graph of experts, news organizations can spin off new products. One idea is an **"Expert Finder"** service – a platform (perhaps consumer-facing or for professionals) where one can search for experts by topic and see verified profiles (essentially a by-product of the internal graph). This could be akin to a *high-end Q&A platform or a consulting marketplace* where, say, a documentary filmmaker or a conference organizer can find and reach out to experts who have been vetted by the news outlet. Monetization can come through subscription to this service or referral fees for connecting experts with opportunities. Another idea is educational content: offering transparency reports or an online dashboard (like a *trust index*) as part of a media literacy initiative, which could attract sponsorship or institutional funding.

In evaluating monetization, it's key to remember that **trust itself has become monetizable**. As one industry analysis put it, content providers can transform their verification and fact-checking strengths into structured services that users and partners will pay for. An identity graph essentially turns a rigorous editorial practice into a scalable asset. By investing in this capability early, news organizations not only differentiate their editorial product (safer, more trustworthy content) but also open up ancillary revenue streams in a news landscape where traditional advertising is under pressure.

Comparisons with Existing Transparency and Trust Initiatives

The idea of bolstering trust in news is not new – across the industry there are several initiatives and tools aimed at increasing transparency and credibility. Implementing identity graphs for authors and sources aligns with the goals of these efforts and in some ways, goes a step further. Here we compare and illustrate how identity graphs complement current approaches:

- **NewsGuard:** NewsGuard is a service that employs journalists to rate entire news websites on a set of credibility and transparency criteria (such as whether the site repeatedly publishes false content, discloses its ownership, provides names of writers, etc.). It then issues "nutrition label" scores for each site, often used by advertisers or readers as a quick trust gauge. While NewsGuard is useful at the *site level*, it doesn't drill down to the granularity of individual article sources. A site might generally be reliable yet still occasionally quote a dubious expert. Identity graphs operate at the *content and contributor level*, catching issues that broad site ratings might miss. In fact, identity graphs could enhance services like NewsGuard – a publication that actively verifies and discloses detailed source profiles would likely earn higher marks on NewsGuard's transparency criteria. (Notably, NewsGuard itself has applauded transparency standards like those of The Trust Project (20 Leading News Groups Join Trust Project Transparency Initiative), which dovetail with what identity graphs provide.) In short, NewsGuard and similar rating systems set the external expectation ("news sites should be transparent about who writes and provides information"); identity graphs are an internal mechanism to fulfill and exceed that expectation, with the added benefit of potentially sharing that data out.
- **The Trust Project:** The Trust Project is a consortium establishing standardized indicators to signal trustworthy journalism. Participating newsrooms commit to displaying these indicators, which include labels such as *Best Practices*, *Author/Reporter Info*, *Type of Work*, *Citations*, and more. An identity graph is

essentially a back-end way to automate and enrich several of these indicators. For example, *Journalist Expertise* (Indicator #2) calls for providing the author's credentials and background – an identity graph ensures that information is thorough and verified, and can feed into the author bio box on each story. The *Citations/References* indicator calls for citing sources in a way readers can evaluate – again, an identity graph can power a "references" section where each source's profile is one click away. By implementing an identity graph, a newsroom would inherently be implementing The Trust Project's principles, but with a robust database behind the scenes to keep those disclosures accurate and up-to-date. This goes beyond simply tagging a name and hoping a reader will take it on faith; it provides an infrastructure to back up those transparency gestures with evidence. In essence, identity graphs operationalize the Trust Project's vision at the newsroom level.

- **Internal Fact-Checking and Standards:** Many large news organizations have internal standards desks or fact-checkers who ensure accuracy of contentious facts. However, verifying the *identity* of sources has traditionally been a less formal process, often left to the reporter's judgment. By instituting an identity graph, a newsroom is effectively encoding source verification into its content management. This is akin to how some outlets now require a *"fact-check tag"* before publishing, to confirm certain steps were taken. Identity graphs would prompt reporters/editors to confirm a source's profile (or create one) as part of the publishing workflow. This method complements traditional fact-checking by extending it to the realm of *"people verification."* It also provides a record for the standards team: if a controversy arises about a source later, the graph's history shows what was checked and when, aiding accountability.
- **Other Efforts (Blockchain and Content Credentials):** There have been experimental efforts to use blockchain or cryptographic signatures to verify news content (e.g., Project Origin, Adobe's Content Authenticity Initiative for photos). Those projects focus on assuring that content (text, images) hasn't been tampered with and comes from a legitimate source. Identity graphs address a different but related point – verifying that the *content of the content*, i.e., the statements and expertise, come from legitimate people. In a way, they are complementary: one ensures the item you read is really published by CNN (for instance) and not a fake copy; the other ensures that the quote in that CNN article is from a real doctor and not a made-up persona. Both aim to bolster trust by fighting different angles of misinformation. A holistic trust-tech strategy for media could eventually integrate these – e.g., a digitally signed article that also carries metadata from the identity graph for each source, all visible to the reader.
- **Community and Crowd-Sourced Verification:** Platforms like Wikipedia or even Twitter Community Notes rely on crowds to flag misinformation or provide context. In journalism, there is less scope for crowdsourcing identities due to time constraints and the risk of doxing or privacy breaches. However, an identity graph could interface with external fact-checkers or communities by providing an authoritative source of truth that those communities can reference. If an observant reader questions an expert's legitimacy in a comment, moderators or staff can quickly check the identity graph and respond with confidence. In effect, the identity graph shields against the scenario of crowds *rightfully* calling out a fake that the newsroom missed – a situation that would be both reputationally damaging and harmful to audience trust. By preemptively vetting, the newsroom stays ahead of the public in validation, rather than playing catch-up.

In all these comparisons, a common theme emerges: **identity graphs reinforce and enhance existing trust-building measures**. They do not replace ethical journalism or the need for good judgment; rather, they provide tools and data that make transparency and verification a seamless part of news production. The approach stands to set a new benchmark: instead of just asking audiences to trust the brand, show them the verified people and expertise behind every story. This level of granularity in trust is the natural next step as journalism adapts to the information age's challenges.

Conclusion: A Roadmap to Trust and Accountability

The battle for trust in news is won or lost not only on what stories are told, but on how they are told and who is telling them. As the cases of fabricated experts and shadowy sources illustrate, credibility can be severely undermined when basic questions of identity are left unanswered. News organizations owe it to their readers – and to their own reputations – to ensure that every voice they amplify is authentic and accountable. Implementing identity graphs for authors and sources provides a forward-looking solution to meet this obligation.

By creating interconnected, verified profiles of the people behind the news, media companies can dramatically increase transparency. Readers can immediately see the qualifications of an expert source and trust that the newsroom has done its homework. Journalists, in turn, gain a robust support system for vetting sources, helping them avoid mistakes born of haste or misinformation. The organization as a whole cultivates a culture where **verification is data-driven and continuous**, rather than ad hoc. This paper has outlined how such a system can be built – drawing from semantic graph techniques, integrating with editorial workflows, and ultimately surfacing as user-friendly trust indicators in the published content.

The benefits extend beyond trust for trust's sake. In a competitive media environment, being a pioneer in source transparency can be a market differentiator. It aligns with the direction of industry initiatives (like The Trust Project's standards and NewsGuard's criteria) and indeed pushes the envelope further. Moreover, it opens new avenues for revenue and partnerships grounded in the newsroom's core competency: verification. From premium subscriber features to licensing deals and AI integrations, a verified identity graph transforms due diligence into a service. It embodies the maxim that **quality journalism is worth paying for** – because quality is explicitly shown and proven, not just promised.

Of course, adopting identity graphs will come with challenges: technical investment, change management in the newsroom, and ensuring the accuracy of the graph itself. But the tools and knowledge to start are readily available. Many building blocks – from knowledge graph databases to open credentials – exist and continue to mature. What's needed is the vision and will to implement them in service of journalistic integrity. As Charlie Beckett of LSE's journalism AI project aptly noted in response to the Santini affair, *"This is a wake-up call to all of us"* in the media. It's a call to not only react to past mistakes but to innovate against future ones.

In conclusion, news organizations should view identity graphs not as a tech gimmick, but as a strategic asset for the coming decade. In a time of deepfakes and AI-generated texts, doubling down on verifying real human sources is a powerful countermeasure. It reassures the public that journalism is adapting and that credible media are doubling down on truth and transparency. Implementing identity graphs for authors and sources will require effort and collaboration, but it promises a substantial payoff: **restored trust, stronger accountability, and a reinforced contract with the audience that what they are reading is grounded in reality.** The message to media executives and editors is clear – by investing in knowing *who* is behind your content, you invest in the credibility and future of your journalism.

Sources:

1. Michael Savage, *The Guardian* – Investigation into the Barbara Santini fake expert case (‘Immediate red flags’: questions raised over ‘expert’ much quoted in UK press | National newspapers | The Guardian)
2. Rob Waugh, *Press Gazette* – *“Virtual reality: The widely-quoted media experts who are not what they seem”* (exposé on fake media commentators) (Virtual reality: The widely-quoted media experts who are not what they seem - Press Gazette)
3. Rob Waugh, *Press Gazette* – Publishers respond by deleting/amending stories with dubious experts (Fake experts: Publishers delete articles after Press Gazette report).
4. NPR – *“White House Adviser Peter Navarro Calls Fictional Alter Ego An 'Inside Joke'”* (Ron Vara case) (White House Adviser Peter Navarro Calls Fictional Alter Ego An 'Inside Joke' : NPR).
5. The Trust Project – Trust Indicators highlighting author expertise and source transparency (Trust Indicators - The Trust Project) (Trust Indicators - The Trust Project).
6. Dinis Cruz – *“Monetising Trust and Knowledge: Personalised Semantic Graphs”* (trust networks and expert validation) (Monetising Trust and Knowledge: How News Providers can leverage Personalised Semantic Graphs - Dinis Cruz - Documents and Research).
7. ResponseSource (via Press Gazette) – Plans for credibility scores and source ratings in journalist request networks (Fake experts: Publishers delete articles after Press Gazette report).
8. Gallup – Poll on public trust in media (2024): trust at record lows (Americans' Trust in Media Remains at Trend Low).
9. Setupad – Explanation of identity graphs (unified profiles connecting data points) (7 Reasons Publishers Should Build Their Own Identity Graph).
10. MyFeeds.ai (Dinis Cruz) – Use of semantic knowledge graphs to provide provenance in personalized news feeds (How it works - myfeeds).