

Grey Literature in Transition: Toward a Paradigm Shift in Research Communication

Gina Pavone, Donatella Castelli, and Stefania Biagioni, CNR-ISTI
Dominic Farace, GreyNet International

Abstract

The scientific communication ecosystem is undergoing deep transformation, driven by technological change, evolving research practices, the growing adoption of Open Science approaches, and a broader shift toward transparency, interoperability, and inclusiveness in research workflows. In this context, Grey Literature (GL) has grown substantially in volume, diversity, and strategic relevance, becoming an essential component of the research lifecycle. Historically, GL has enabled rapid dissemination, supported policy and decision-making, and captured methods, processes, and intermediate outputs not adequately represented by traditional scholarly publishing.

Despite its central role, the definition, management, and infrastructural treatment of grey literature remain anchored to outdated conceptual models, limiting its integration into contemporary scholarly systems. These legacy models - rooted in a print-based, journal-centric conception of “literature” - no longer reflect the heterogeneity of formats and functions that characterise today’s GL landscape.

This position paper argues that a paradigm shift is required to recognise GL as a first-class element of scientific communication. Such a shift entails rethinking its epistemic and functional status and recognising that GL encompasses a wide family of scientific resources extending far beyond text-based documents.

To support this shift, the paper identifies key conceptual, technical, and organisational challenges and proposes a coordinated set of actions articulated along five strategic axes: reliability, citability, discoverability, interoperability, and long-term preservation. Together, these axes provide a practical roadmap for embedding GL systematically within scholarly infrastructures, research assessment frameworks, and Open Science practices. The recommendations extend well-established practices for traditional outputs while addressing the distinctive characteristics of grey literature, including its heterogeneity, contextual production, and close integration with operational and policy-relevant workflows.

By framing GL as a strategic asset rather than an exceptional or marginal category, the paper emphasises its potential to strengthen the transparency and accountability of the research lifecycle as a whole. Improving the management and visibility of GL

benefits its producing organisations, enhances the quality and reusability of scientific outputs, and supports a more inclusive and resilient knowledge ecosystem.

Ultimately, the paper calls on institutions, funders, and information professionals to treat grey literature as an integral part of the scientific record and a key driver in the evolution of contemporary scholarly communication.

The article is inspired by ideas, insights, and reflections discussed at the Pisa2025 joint conference¹.

1. Introduction

Contemporary scientific communication is no longer defined solely by journal articles and monographs, but by a heterogeneous ecosystem of outputs that document research processes, intermediate results, and societal engagement. Within this evolving environment, grey literature (GL) has expanded significantly in volume, relevance, and the diversity of its formats, becoming a core component of the research lifecycle by contributing to the production, documentation, and communication of scientific knowledge at each stage. Long before the emergence of recent open science initiatives, GL supported rapid dissemination, informed policy and decision-making, documented methods and processes, and enabled the circulation of research outputs that were not adequately captured by traditional journal publishing² (Lawrence, 2017). Despite its strategic role, the definition, treatment, and infrastructural support of GL remain largely anchored to outdated conceptual models that no longer reflect contemporary scientific practice. These models originate in historically dated definitions developed in a print-based, journal-centric environment and have persisted well beyond their original context, shaping research assessment practices, information infrastructures, and scholarly culture³ (Schöpfel, 2010). As a result, GL is still frequently characterised as unpublished, not peer-reviewed, difficult to access, ephemeral, non-scientific, or peripheral to libraries. So it remains constrained by a set of inherited assumptions that affect not only its perceived status but also the interpretation of the very term literature, which traditionally evokes written texts and reinforces the assumption that GL mainly consists of textual documents and represents a “non-conventional” or second-tier source of knowledge. While such assumptions may once have reflected genuine limitations in visibility, dissemination, or standardisation, they no longer correspond to current research practices⁴ (Lawrence, 2017).

¹ Pisa2025 Conference - <https://iamslic.wildapricot.org/Pisa2025/home>

² <https://doi.org/10.26069/greynet-2025-000.634-gn>

³ <https://doi.org/10.26069/greynet-2020-000.179-gg>

⁴ <https://doi.org/10.26069/greynet-2025-000.640-gn>

Today, GL encompasses a wide spectrum of formats that extend well beyond written documents and include datasets, software, reports, multimedia materials, workflows, technical standards, and policy briefs⁵ (Adams et al., 2016). These resources are disseminated through persistent, openly accessible, and citable channels, often supported by robust quality assurance processes tailored to specific scientific and societal purposes. They constitute a structured body of purpose-driven outputs produced and used at multiple stages of the scientific process. This diversity of resource types is mirrored by the growing quantitative significance of GL in scientific practice. Although no global, uniform estimate exists of the proportion of GL used in research, numerous empirical studies and large-scale scholarly infrastructures consistently show that these heterogeneous resources form a substantial component of the scientific evidence base (Paez, 2017; Angioni et al., 2026). This breadth of formats and functions also invites a deeper reflection on how these resources fit within the broader scholarly communication system - both in terms of their distinctive characteristics and in the need to reconsider the terminology traditionally used to describe them.

A concrete paradigm shift is now required to realign the conceptualisation and management of GL with contemporary research practices and to enable its effective integration into the broader scholarly communication ecosystem. GL should be reconceptualised as a first-class component of scientific communication, fully aligned with open science principles and current research workflows. Its management should be harmonised with that adopted for other scholarly outputs to support a coherent and inclusive model of scientific communication capable of exploiting the full spectrum of research products.

To support this shift, this paper identifies *the main technical, organisational, and conceptual challenges* that currently hinder the systematic integration of GL into scholarly infrastructures. At the core of this analysis lies a fundamental conceptual clarification: GL should not be understood as belonging solely to the domain of “literature” in a purely humanistic sense, but to a distinct and complex area of information science (Farace, 2025). While it shares certain characteristics with other digital scientific resources, such as journal articles, datasets, and multimedia objects, it also presents specific requirements and challenges that cannot be fully addressed through generic solutions. Its heterogeneity in format, origin, purpose, and role makes its management inherently more complex than that of more uniform scholarly resources.

On this basis, the paper proposes a set of actionable measures aimed at improving the visibility, reliability, interoperability, citability, and long-term preservation of GL resources. By framing these measures as both a direct benefit for producing organisations and a contribution to the evolution of scientific communication as a

⁵ <https://guides.library.illinois.edu/c.php?g=1310347&p=9630544>

whole, the paper positions GL as a key driver of a more transparent, inclusive, and resilient research ecosystem.

Much of the analysis presented here builds on the dialogue developed during the Pisa2025 Joint conference⁶, which brought together researchers, practitioners, and institutional actors from multiple disciplines.

The remainder of the paper is structured as follows: Section 2 further characterises the required paradigm shift, Section 3 highlights a coordinated set of technical and policy-oriented actions required to implement this change, and finally, Section 4 proposes a forward-looking vision for the role of GL in the coming decade and concludes the paper.

2. The evolving role of GL

The evolving role of GL within today's scientific landscape requires more than incremental adjustments to existing scholarly communication systems. It calls for *targeted technological solutions and explicit policy choices* capable of supporting a genuine paradigm shift, one that enables GL resources to become fully communicable, discoverable, citable, interoperable, and reusable for scientific purposes.

Integrating GL resources is inherently more complex than managing traditional scholarly outputs. Conventional publishing infrastructures are optimised for relatively homogeneous products, journal articles and monographs, supported by standardised workflows and commercial actors operating within consolidated evaluation and dissemination models (Adams et al., 2016, Kamei et al., 2021). GL resources, by contrast, challenges these norms in fundamental ways.

First, GL resources are *intrinsically heterogeneous* in terms of formats, origins, and functions, and no universally accepted standards exist for many of these resource types. Importantly, this challenge is no longer confined to GL alone. The broader push toward open science, transparency, and reproducibility is exposing similar limitations within mainstream scholarly publishing, where datasets, software, protocols, and workflows increasingly lack stable and widely adopted standards.

Second, GL resources serve a *distinct functional role* within the knowledge ecosystem. They convey situated, operational, and often time-sensitive knowledge produced in response to concrete needs within specific projects, institutions, or communities of practice. Rather than addressing a universal scholarly audience, they frequently support policy development, operational decision-making, or research workflows (Lawrence et al., 2014). As a result, their organisation, assessment, and

⁶ <https://iamslic.wildapricot.org/Pisa2025/home>

dissemination require *context-sensitive and flexible approaches*, rather than rigid bibliographic models inherited from journal publishing.

Finally, GL plays a *dual role*: it acts both as a *vehicle* for rapid knowledge transfer to relevant stakeholders and as a *driver* of further research, innovation, and policy development. Many GL resources underpin or precede formal publications; others remain authoritative outputs in their own right.

These characteristics give rise to persistent challenges related to *reliability, citability, discoverability, interoperability, and long-term preservation*. This paper takes the position that addressing these challenges should not be seen as an exception or a compensatory effort, but as an opportunity: *solutions developed for GL resources can anticipate and inform the future evolution of scholarly communication as a whole*.

Following this position, the next section presents the specificity of challenges in the GL landscape and suggests actions to address them.

3. How to implement the paradigm shift

Addressing the GL paradigm shift described above encompasses both technical solutions and policy innovations that support the systematic inclusion of GL within research assessment practices and scholarly infrastructures. Below, a coordinated set of technical and policy-oriented actions organised by challenges is identified. All of them aim to systematically integrate GL resources into scholarly infrastructures and research assessment frameworks. Some of these actions extend practices already well established for traditional research outputs; others respond specifically to the distinctive nature of GL.

Reliability

Reliability in GL should not be assessed through a single, journal-centric peer review model (Kamei et. al. 2021). Instead, it should be grounded in *transparency, provenance, and fitness for purpose*, reflecting the diversity of GL production contexts.

Key actions include:

- *Documentation of producer credibility* through persistent identifiers (e.g. ROR, ORCID), clear authorship, provenance information, and, where applicable, funder identifiers
- *Use of explicit validation processes*, whether they involve internal review, expert assessment, or stakeholder consultation
- *Clarification of the purpose and intended use* of each GL resource

- *Provision of robust documentation*, including metadata, versioning, and links to related outputs

At the policy level, funders, repositories, and evaluators should explicitly recognise *institutional and contextual validation mechanisms* as legitimate forms of quality assurance. Moreover, the adoption of field-specific policies, such as guidelines for the inclusion of GL in systematic reviews in fields like public health, environmental science, and management studies, further supports its credibility and relevance⁷. These frameworks could establish clear criteria for formats, sources, and evaluation, ensuring that GL resources are integrated responsibly and transparently into evidence syntheses and decision-making processes.

Citability

Grey literature is often perceived as uncitable, primarily due to unstable access and incomplete metadata. This barrier is technical and procedural rather than intrinsic (Farace et. al. 2010).

Priority actions include:

- *Systematic assignment of persistent identifiers* (e.g. DOI, Handle, ISBN/ISSN where appropriate)
- *Provision of standardised citation records*, directly embedded in repository interfaces (“Cite this item”)
- *Maintenance of stable landing pages* with version control and long-term URL persistence

Citation guidelines and style manuals should explicitly incorporate GL resources, reinforcing their legitimacy as citable scientific outputs⁸.

Discoverability

Despite being openly accessible, many GL resources remain difficult to find due to fragmented repositories and inconsistent metadata practices (De Castro et al, 2018). AI-assisted metadata generation and emerging discoverability techniques can largely facilitate this challenge, provided that certain conditions are met⁹.

⁷ <https://icer.org/policy-on-inclusion-of-grey-literature-in-evidence-reviews/>

⁸ <https://guides.library.illinois.edu/c.php?g=1310347&p=9630671>

⁹ https://www.greynet.org/images/GL25_Savic_p._119-129_.pdf

Improving discoverability requires:

- *Use of metadata profiles* that balance standardisation with contextual richness.
- *Exposure through interoperable repositories* using established harvesting protocols (e.g. OAI-PMH)
- *Integration with institutional CRIS systems*, ensuring visibility alongside other research outputs
- *Contextual discovery mechanisms*, linking GL resources to projects, policies, datasets, and related entities
- *Inclusion in global discovery services* (e.g. Google Scholar, WorldCat, Crossref, DataCite).

Repositories, libraries, and information services should align their practices by adopting common metadata schemas, persistent identifier frameworks, and harvesting protocols. Equally important is fostering awareness and capacity among data managers and repository staff, particularly in underrepresented regions, to close the discoverability gap and ensure equitable access.

Interoperability

Interoperability for GL resources must be understood as a *multi-level property*, encompassing technical, semantic, and organizational dimensions.

Core actions include:

- Adoption of *open formats and machine-readable metadata* aligned with widely used standards
- Use of *controlled vocabularies and shared taxonomies* for GL resource types
- *Persistent linking* between GL resources and related entities (projects, datasets, people, institutions)
- Development of *coordinated institutional policies, federated repository networks, and clear licensing frameworks*

Achieving interoperability ultimately requires not only technical alignment but also sustained collaboration and cultural change.

Preservation

The long-term preservation of GL resources is essential to ensure continuity, reproducibility, and reuse of scientific knowledge.

Key measures include:

- Adoption of *trusted digital preservation frameworks*
- Use of *persistent identifiers and version control*
- Storage in *open, preservation-friendly formats*
- Preservation of *rich descriptive and provenance metadata*
- Participation in *federated and redundant archiving networks*

Preservation responsibilities should be explicitly embedded in institutional and funder policies as part of responsible research management.

4. Way forward and conclusions

This paper argues that GL is not a marginal or transitional by-product of science, but a *structural component of contemporary scientific communication*. Its diversity of formats, immediacy, and close integration with research practice make GL resources essential for documenting scientific processes, supporting policy and decision-making, and enabling interdisciplinary and socially engaged research.

The central challenge is no longer conceptual recognition, but *systematic integration*. Persisting in outdated distinctions between “published” and “unpublished” outputs continues to marginalise valuable scientific resources and to weaken the coherence of the research ecosystem. GL resources must instead be assessed according to methodological rigor, transparency, and fitness for purpose, irrespective of the commercial or institutional nature of their dissemination channels. Recognising GL resources as first-class scientific outputs is therefore a necessary condition for aligning scholarly communication with open science principles and contemporary research workflows.

This paper has outlined a set of concrete actions, addressing reliability, citability, discoverability, interoperability, and preservation, that provide a practical pathway for this integration. These actions are not designed solely to accommodate GL within existing systems, but to *strengthen scientific communication as a whole*. By improving the management, visibility, and reuse of GL resources, producing organisations

enhance the quality and accountability of their outputs while contributing to a more transparent, inclusive, and resilient knowledge ecosystem.

Advancing GL management and advancing scholarly communication are mutually reinforcing objectives. Acting on GL today means anticipating the future needs of science itself. Institutions, funders, and information professionals are therefore called upon to treat GL not as an exception to be managed, but as a strategic asset to be fully integrated into the scientific record.

Acknowledgements

We gratefully acknowledge the participants in the *Pisa2025* conference Panel Session – “From a Panel Discussion to a Position Paper” for their valuable ideas, insights, and reflections, which significantly informed the development of this position paper.

References

- Adams, J., Hillier-Brown, F., Moore, H. J., Lake, A. A., Araujo-Soares, V., White, M., & Summerbell, C. (2016). Searching and synthesizing “grey literature” and “grey information” in public health: Critical reflections on three case studies. *Public Health*, 138, 21–28.
- Angioni, S., Baglioni, M., Bardi, A., Manghi, P., Mannocci, A., & Pavone, G. (2026). Charting the landscape of Italian Diamond Open Access publishing. *Quantitative Science Studies*. <https://doi.org/10.1162/QSS.a.466>
- Auger, C. P. (2017). *Information sources in grey literature*. De Gruyter.
- De Castro, P., et al. (2018). Grey literature and evidence-based policy making: The role of discoverability and access. *Grey Journal*, 14(3), 150–158.
- Farace, D. J., & Schöpfel, J. (Eds.). (2010). *Grey literature in library and information studies*. De Gruyter Saur
- Farace, D. (2025). *A new approach to the field of grey literature: Let's reduce the noise and raise awareness*. GreyForum Series, 10(1), 12 p. TextRelease. <https://doi.org/10.26069/greynet-2025-000.640-gn>
- Kamei, F., Wiese, I., Lima, C., Polato, I., Nepomuceno, V., Ferreira, W., Ribeiro, M., Pena, C., Cartaxo, B., Pinto, G., & Soares, S. (2021). *Grey literature in software engineering: A critical review*. arXiv. <https://arxiv.org/abs/2104.13435>
- Lawrence, A., Houghton, J., Thomas, J., & Weldon, P. (2014). Where is the evidence? Realising the value of grey literature for public policy and practice. Swinburne Institute

for Social Research. <http://apo.org.au/research/where-evidence-realising-value-grey-literature-public-policy-and-practice>

Lawrence, A. (2017). Grey literature publishing in public policy: Production and management, costs and benefits. In L. Chan & F. Loizides (Eds.), *Expanding perspectives on open science: Communities, cultures and diversity in concepts and practices* (pp. 85–99). IOS Press.

Paez, A. (2017). Grey literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, 10(3), 233–240.

Schöpfel, J. (2010). Towards a Prague Definition of Grey Literature. In *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues, Prague, 6-7 December 2010* (pp. 11-26). TextRelease. hdl.handle.net

Savic, G. (2025). Harnessing the power of AI for managing grey literature. In *Proceedings of the 25th International Conference on Grey Literature* (pp. 119–129). GreyNet International. https://www.greynet.org/images/GL25_Savic_p_119-129.pdf