https://doi.org/10.51408/issi2025_124

Mapping and Quantifying the Boundaries in Research Data Sharing based on Data Policy

Yizhan LI [1], Mingze ZHANG [2], Lu DONG [3], Zexia LI [4]

[1] liyz@mail.las.ac.cn, [2] zhangmingze@mail.las.ac.cn, [3] donglu@mail.las.ac.cn, [4] lizexia@mail.las.ac.cn

National Science Library, Chinese Academy of Sciences, Beijing (China)

Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing (China)

Abstract

Balancing research data openness with security concerns necessitates regulatory constraints, yet the absence of standardized quantitative thresholds complicates cross-institutional and cross-border data sharing. This study examines 72 policy documents from the US, EU, and UK. Using a large language model (LLM)-based prompt engineering approach, we extract and quantify data-sharing constraints through a two-stage framework: (1) Constraint Identification, detecting access limitations, and (2) Quantitative Relation Extraction, identifying key metrics such as data scale, durations, etc. Our findings categorize data-sharing boundaries into three types: mandatory restrictions (red line), conditional constraints (blue line), and ambiguous areas shaped by evolving technologies. A comparative analysis of key quantitative constraints, like embargo periods, reveals inconsistencies across policies, highlighting the need for regulatory alignment. Additionally, we identify subject-specific access restrictions that resemble controlled data list. Future research will refine constraint mapping, analyze policy evolution, and explore interdisciplinary data governance. These efforts aim to enhance policy clarity, enhance operational efficiency, and support international research collaboration.

Introduction

The rapid growth and large-scale accumulation of research data have shifted it from being a mere byproduct of research activities to a foundational resource for scientific investigation. Fields such as earth sciences, life sciences, materials science, and computer science increasingly exemplify the defining features of data-intensive knowledge discovery. This transformation has been propelled by initiatives like *the Global Open Science Movement* and *the Fourth Paradigm of Scientific Research*, which emphasize open access and data-driven discoveries. These efforts have enabled the unprecedented reuse and interconnection of geospatial, ecological, personal sensitive, health, and agricultural data (Xiang & Cai, 2021; George, 2019). However, this openness introduces significant challenges, including risks to security, personal privacy, intellectual property rights, commercial interests, and ethics (Li et al., 2023; Amiri-Zarandi et al., 2022; Majeed, 2021; Zigomitros et al., 2020). These issues, exacerbated by the rapid development of emerging and disruptive technologies, underscore the growing importance of research data security. Nations also have faced fundamental disagreements over principles governing the cross-border flow of data, further complicating efforts to safeguard data (Ducato, 2020).

To address these challenges, national laws establish overarching guidelines, while major funding agencies, research institutions, and international scientific programs implement policies to regulate the sharing and use of research data. These measures

aim to mitigate security risks by creating a multi-tiered framework of regulations and intangible boundaries that define the flow and usage of data. The unique characteristics of research data, such as shareability, non-exclusivity, asymmetry, transferability, long-term accumulation, and its public interest nature—further complicate the balance between openness and security (Li et al., 2024). These characteristics result in diverse priorities and roles for national authorities, funding agencies, researchers, and data contributors within the data sharing and value chain(Li et al., 2022).

In real-world contexts, constraints on research data sharing are often principle-based, with sensitive data classified primarily by the harm or loss they may cause. For researchers, such guidelines often lack practical applicability. While some rules employ quantitative metrics and thresholds, these face challenges such as inconsistent standards and thresholds that evolve with technological advancements and shifting risk factors. This paper focuses on research data sharing policies and seeks to address the following questions:

Q1. What are the current boundaries of research data sharing, and in what forms or manifestations do they appear?

Q2. Can the boundaries of research data sharing be quantitatively defined?

By combining policy text analysis with a quantitative framework, this paper aims to bridge the gap between principle-based and operational rules. This approach enables researchers to navigate data-sharing complexities with greater clarity, consistency, and security. Additionally, it standardizes guidance and fosters a benchmark for dialogue across institutions, organizations, and countries.

Dataset Construction, Processing and Methodology

Research data policy collection and its metadata

This study provides a systematic examination of the legal frameworks and regulatory instruments governing research data across three jurisdictions: the United States (US), the European Union (EU), and the United Kingdom (UK). A total of 72 policy documents were comprehensively collected from publicly accessible sources (Table SI-1), with corresponding metadata and access information recorded (Table 1). These documents encompass a broad range of national legislative acts, directives, regulations, rules, guidance materials, and executive orders related to the data domain in the US, EU, and UK. While not all documents specifically target research data, it is evident that research data—as a critical subset of broader data ecosystems—must adhere to these overarching policies, particularly with respect to data sharing and security. The corpus also includes strategic policy documents that outline anticipated developments and policy trajectories for data sharing in the coming years. In addition, the study reviews data management requirements issued by major funding agencies (e.g., the US National Science Foundation and UK Research and Innovation), prominent research institutions, and international scientific collaboration initiatives. These requirements frequently reflect disciplinary particularities and address diverse data modalities, including text, tables, images, and audio.

Table 1. Metadata of policy documents related with research data sharing topic.

Field Name	*Description*
OID	Unique ID
File Name	The official name of the document
Type	The type of policy document, including *Act*, *Directive*, *Regulation*, *Rule*, *Strategy*, *Guidance*, etc.
SubType	*Rule* is subdivided into *Rules_Government*, *Rules_Sponsor_Public*, *Rules_Sponsor_Private*, *Rules_Project*, *Rules_Institution*, *Rules_International Organization*, *Rules_International Project*, etc.
Issuing Authority	The name of the organization that issued the document
Country/Region	The geographical scope where the document applies
Issuing Date	The official issuing date of the document
Enforceability	Mandatory or not
Access Address	URL or PDF file download from the official website
Policy Language	English, etc.

Paragraph extraction and analysis with LLM

In the field of policy informatics, several foundational studies have outlined common methods and procedures for the quantitative analysis of policy texts. Automated policy text analysis typically involves three main tasks—classification, clustering, and scaling(Grimmer & Stewart, 2013)—and follows a general workflow that includes preprocessing, stemming, bag-of-words model, category development and coding, reliability and validity checks, and content interpretation(Cao & Zhang, 2022; Bardach & Patashnik, 2019; Lucas et al., 2015). These methods have been applied to various types of policy documents, such as legislative acts and international treaties(Yang et al., 2020), often focusing on entities or clauses as the unit of analysis. In recent years, the emergence and widespread adoption of large language models have made policy text analysis more streamlined and fine-grained. This study employs a structured prompt engineering methodology, integrating template construction and iterative optimization to extract policy constraints on research data sharing (Figure 1). Using LLMs like ChatGPT-4o and DeepSeek-R1, we propose a two-stage framework: (1) Constraint Identification – domain-adapted prompts guide LLMs to detect data-sharing restrictions (e.g., access limits, usage boundaries); (2) Quantitative Relation Extraction – refined templates identify constraint-related metrics (e.g., temporal restrictions, user quotas). Our prompt engineering follows the "Role-Objective-Skill-Workflow-Constraint-Output" framework (Figure SI-1). A test set (20% of 72 policy documents) was iteratively optimized, with representative policies selected from different jurisdictions, policy types (e.g., Act, Directive) (Caufield et al., 2024; Chen et al., 2024; Durmaz et al., 2024; Yang et al., 2024).

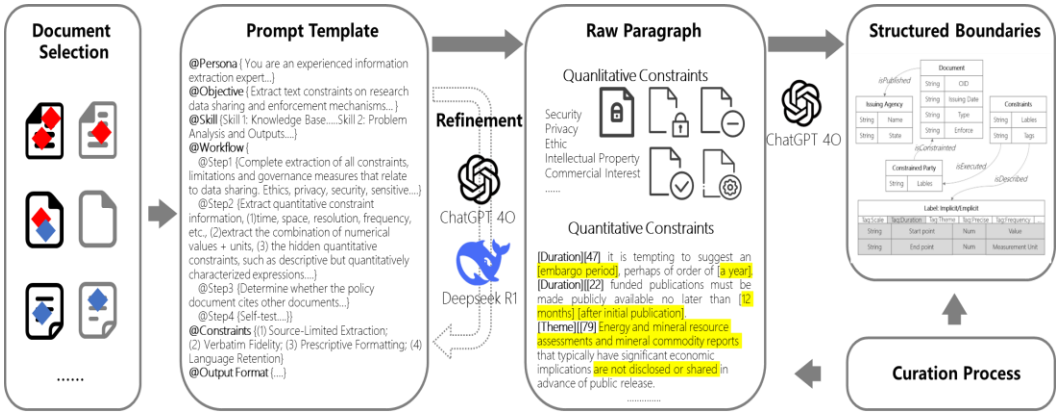

Figure 1. Overview of paragraph extraction process.

All policy documents used in this study are publicly available and contain no personal or sensitive information. Nonetheless, the use of LLMs for data extraction raises concerns about output accuracy and interpretability. To address these issues, a curation process—implemented as a human-in-the-loop review—was used to manually verify and refine all extracted results, ensuring their reliability.

Discussion

Conceptual description and Forms of boundaries

The analysis of policy texts reveals that the openness and sharing of research data are subject to certain boundaries. These boundaries vary depending on the type of research data governance documents and the nature of the institutions that issue them. They can be categorized into three types: mandatory boundaries (Red line), conditionally negotiable boundaries for targeted sharing (Blue line), and areas of uncertainty that remain undefined (Figure 2).

Red line: This category includes confidential data related to national security, data sovereignty, and personal privacy, which are clearly defined by national or regional laws, regulations, and confidentiality agreements.

Blue line: This category refers to data that can be shared under specific conditions, such as restrictions on the use of research data in particular network environments, among defined user groups, or within a controlled scope of access.

Ambiguity area: This category pertains to areas that are still under debate or evolving alongside technological advancements. For example, the development of gait recognition technology allows surveillance data from public spaces to be used for identifying individuals based on gait features. As a result, this data has been classified as personal information and recognized as a form of biometric data,

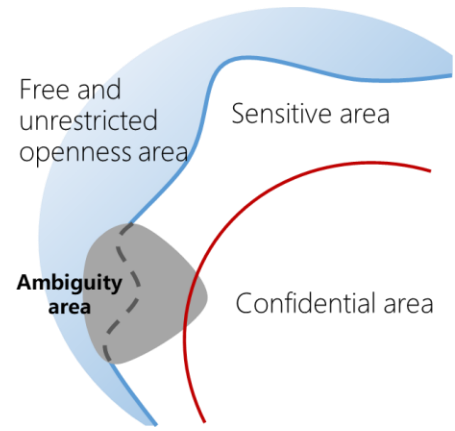

Figure 2. Three forms of research data sharing boundaries.

similar to fingerprints or voiceprints. This is a typical case of how advancements in technology lead to changes in data sensitivity, resulting in a contraction of data sharing boundaries.

Spectrum of boundaries

Figure 3 presents the relationship between the classification of 72 policy documents and the defined boundary constraints, along with the document types and key elements of these constraints. The red-to-blue gradient denotes mandatory regulatory changes, whereas the green-to-yellow gradient represents a shift from qualitative to quantitative constraints.

Mandatory legal regulations typically prioritize qualitative, principle-based constraints. For instance, research data sharing is generally governed by principles such as national security, ethics, privacy protection, and intellectual property rights, etc. Moreover, certain parameters may be subject to principle-based restrictions, meaning that while requirements such as assessments and reviews for large-scale data sharing are imposed, specific quantitative thresholds are not explicitly defined. However, current document analyses indicate that explicit quantitative thresholds are seldom specified, underscoring the need for supplementary regulatory frameworks.

Each NO. represents a document, and the colors of the NO.s represent different countries /region. Legend US, EU, UK, Multi		Red Line			Blue Line					Ambiguity Area	
					Rule						
		Act	Regulation	Directive	Governament rules	Sponsor rules	Project rules	Institution rules	International Organization rules	Guidance	Strategy
1. Qualitative Constraints		58, 61,	33,	62,	45, 65, 72,	01, 13, 14, 15, 17, 44, 52, 53, 56, 57,			26, 29,	42,	27, 43,
2.Quantitative Constraints	2.1 Implicit	24, 25, 69, 70,	23,	32,	08, 68	12, 18, 71,	19, 20,		39,		28,
	2.2 Explicit		30, 31,	51,	04, 05, 09, 21, 22, 40,	02, 03, 06, 10, 11, 16, 38, 46, 47, 48, 50, 59, 60, 63, 66, 67,	64,	36,	34, 35, 54, 55, 65,	41,	
	Scale	24, 25, 69, 70,	23,			38, 66,					
	Precise	24, 25,	23,		08,	66,			55, 65,		
	Frequency	24, 25, 69, 70,		32,	68,	60,			65,		
	Duration	69, 70,		51,	04, 05, 09, 21, 22,	02, 03, 06, 10, 11, 12, 16, 18, 38, 46, 47, 48, 59, 63, 67	19, 20	36,	39, 54, 65		28,
	Scope					46, 47, 60,			35, 55, 65		
	Data Source					46, 47, 66,	64,		65,		
	Theme	24, 25, 69, 70,	23,			16, 49, 50, 59, 63,	64,		34, 54, 55, 65	41,	
	Multi-combine				40,	66,			65,		
	Others		30, 31,		09,						

Figure 3. Spectrum of boundaries across mandatory change.

Regulations issued by research funding agencies, research initiatives, academic institutions, and international scientific organizations further delineate responsibilities. While many documents specify restrictions on research data sharing, disciplines such as astronomy and geoscience tend to emphasize open data policies, whereas life sciences often impose stricter sharing constraints. These restrictions may encompass factors such as data scale, precision, timely, frequency, duration, spatial scope, data sources, themes, and multidimensional conditions.

In regulatory ambiguity areas, countries often issue guidelines and strategic documents to outline potential future measures and directions. A relevant example

is the CODATA publication, *Open Data in a Big Data World*, which states: *"Although it is tempting to suggest an embargo period, perhaps on the order of a year, it would be preferable for individual disciplines to develop procedures attuned to their specific needs, while avoiding undue delays."* This ambiguity is particularly evident in domains such as AI training datasets, cross-border data flows, and emerging technologies like quantum computing, where policy frameworks are still evolving. In February 2025, the OECD released a report titled *Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data*, highlighting that research institutions and universities frequently employ data scraping techniques for academic research and scientific inquiry. Although such activities are typically pursued for legitimate purposes, the use of international datasets may give rise to copyright and data privacy compliance challenges. For instance, scraped content used in studies on academic dissemination, social behavior, or public opinion trends may contain copyrighted materials—such as news articles, scholarly publications, or images—as well as personally identifiable information from sources like social media, user comments, and online forums. Cross-border scraping further raises the risk of triggering foreign data protection laws. Given the divergence in national copyright exceptions and the absence of a unified international framework, there is a growing expectation for the establishment of a coordinated governance mechanism for cross-border data scraping. If the proposed establishment of registration or transparency mechanisms for research-related data scraping were to be implemented, it could potentially reshape compliance requirements for research data in certain disciplinary fields.

Embargo period as a case of inconsistency detection

Standardizing and aligning quantitative constraints across legal and regulatory frameworks of varying levels and enforceability enable cross-national, cross-regional, and cross-institutional comparisons. This is crucial for identifying conflicts among these constraints, which pose significant challenges when research data is transferred across institutions, regions, or projects. Addressing such inconsistencies is one of the major operational difficulty researchers face in data-sharing practices.

Through the Prompt-based analysis of relevant policy documents, we identified a set of quantitative constraints. Among them, control over the embargo period is one of the most precisely quantified measures, with the embargo period itself serving as a key indicator. Figure 4 presents a schematic representation of embargo-related quantitative constraints, visually illustrating variations in start time, duration, and enforceability across different regulations. For instance, Document No. 02 mandates a two-year embargo period starting from the completion of data collection, whereas Document No. 28 recommends only one year. These discrepancies necessitate coordination and negotiation, as seen in the case of UK research funding agencies aligning embargo constraints when engaging in CODATA's international collaborations.

Similarly, quantitative constraints on data volume and frequency can be systematically mapped and compared, much like embargo periods. In contrast, subject-specific constraints on research data function more like controlled data catalogs, where sharing is restricted based on predefined classifications.

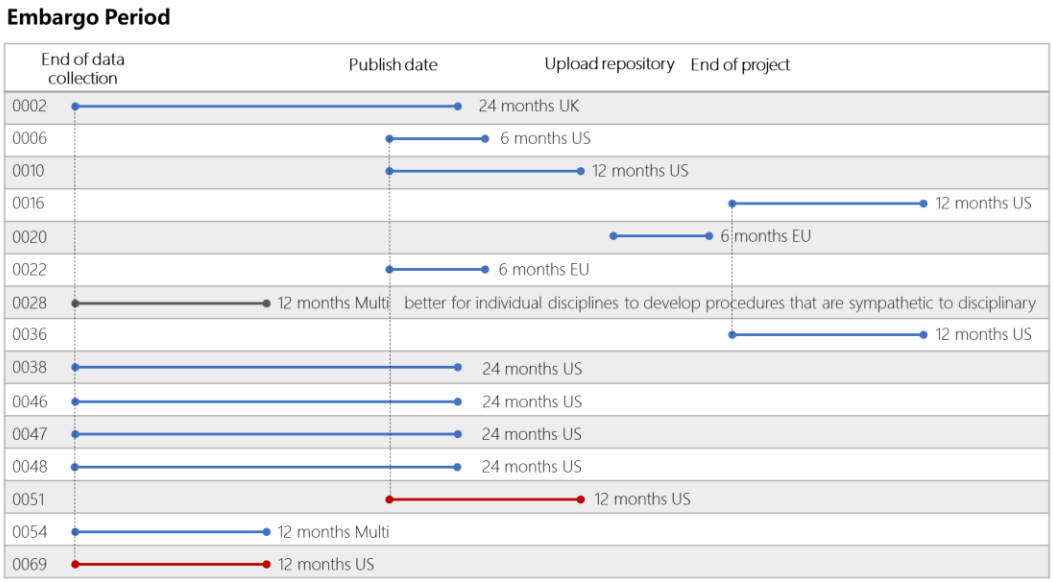

Figure 4. Schematic representation of quantitative boundaries for the embargo period indicator.

Preliminary remarks and limitations

Examining quantitative constraints in research data sharing policies provides a unified reference point for cross-domain collaboration, offering practical value for policy alignment. Our preliminary exploration has demonstrated that: (1) while some manual intervention and content review are still required, the prompt-based extraction method has proven successful and can be further refined into structured data. (2) Structured data effectively supports the visualization and mapping of quantitative constraints, enabling a more intuitive understanding of constraint variations, reducing the complexity of policy interpretation, and improving implementation efficiency.

Although this study highlights the importance of aligning research data policies across jurisdictions, achieving such coordination is fraught with legal and political complexity. From a legal perspective, civil law systems (e.g., the EU, Japan) rely on codified statutory exceptions, whereas common law systems (e.g., the United States) adopt interpretive doctrines such as fair use. Divergent views on data ownership, national sovereignty, and legal entitlements to access further complicate harmonization. Differences in regulatory culture, and institutional trust shape how jurisdictions approach research data governance. Even where overarching goals—such as advancing open science—are nominally shared, substantial asymmetries in enforcement capacity and legal infrastructure remain critical barriers to policy convergence.

Acknowledgments

This research was funded by the National Social Science Fund of China (Grant No. 22CTQ031).

References

Caufield, J. H., Hegde, H., Emonet, V., Harris, N. L., Joachimiak, M. P., Matentzoglu, N., Kim, H., Moxon, S., Reese, J. T., Haendel, M. A., Robinson, P. N., & Mungall, C. J. (2024). Structured Prompt Interrogation and Recursive Extraction of Semantics (SPIRES): a method for populating knowledge bases using zero-shot learning. *Bioinformatics*, *40*(3), btae104, Article btae104. https://doi.org/10.1093/bioinformatics/btae104

Chen, X., Li, Y., Fan, S. H., & Hou, M. S. (2024, Apr 26-28). *Judicial Text Relation Extraction Based on Prompt Tuning* 2nd Asia Conference on Computer Vision, Image Processing and Pattern Recognition (CVIPPR), Xiamen, China. https://doi.org/10.1145/3663976.3664029

Durmaz, A. R., Thomas, A., Mishra, L., Murthy, R. N., & Straub, T. (2024). An ontology-based text mining dataset for extraction of process-structure-property entities. *Scientific Data*, *11*(1), 1112, Article 1112. https://doi.org/10.1038/s41597-024-03926-5

Li, Y. Z., Dong, L., Fan, X. X., Wei, R., Guo, S. J., Ma, W. Z., & Li, Z. X. (2024). New roles of research data infrastructure in research paradigm evolution. *Journal of Data and Information Science*, *9*(2), 104-119. https://doi.org/10.2478/jdis-2024-0011

Yang, Y. R., Chen, S. S., Zhu, Y. P., Liu, X. M., Ma, W., & Feng, L. (2024). Intelligent extraction of reservoir dispatching information integrating large language model and structured prompts. *Scientific Reports*, *14*(1), 14140, Article 14140. https://doi.org/10.1038/s41598-024-64954-0

Li, W. W., Leung, A. C. M., & Yue, W. T. (2023). Where is IT in information security? The interrelationship among IT investment, security awareness, and data breaches. *Mis Quarterly*, *47*(1), 317-342. https://doi.org/10.25300/misq/2022/15713

Amiri-Zarandi, M., Dara, R. A., Duncan, E., & Fraser, E. D. G. (2022). Big data privacy in smart farming: A review. *Sustainability*, *14*(15), 9120, Article 9120. https://doi.org/10.3390/su14159120

Cao, L. J., & Zhang, Z. Q. (2022). Research Progress on Quantitative Methods of Policy Texts from the Perspective of Policy Informatics. *Library and Information*(6), 70-82.

Li, Y., Liu, X., Li, Z., Yin, X., & Wu, M. (2022). Study on conceptual analysis model of scientific data security boundary: from the perspective of stakeholders. *Bulletin of National Natural Science Foundation of China*, *36*(2), 339-347.

Majeed, A. (2021). Towards privacy paradigm shift due to the pandemic: a brief perspective. *Inventions*, *6*(2), 24, Article 24. https://doi.org/10.3390/inventions6020024

Xiang, D., & Cai, W. (2021). Privacy protection and secondary use of health data: strategies and methods. *Biomed Research International*, *2021*, Article ID 6967166, Article 6967166. https://doi.org/10.1155/2021/6967166

Ducato, R. (2020). Data protection, scientific research, and the role of information *Computer Law and Security Review*, *37*, 105412.

Yang, C., Huang, C., & Su, J. (2020). A bibliometrics-based research framework for exploring policy evolution: A case study of China's information technology policies. *Technological Forecasting and Social Change*, *157*, 120116. https://doi.org/https://doi.org/10.1016/j.techfore.2020.120116

Zigomitros, A., Casino, F., Solanas, A., & Patsakis, C. (2020). A Survey on Privacy Properties for Data Publishing of Relational Data. *Ieee Access*, *8*, 51071-51099. https://doi.org/10.1109/Access.2020.2980235

Bardach, E., & Patashnik, E. M. (2019). *A Practical Guide for Policy Analysis: The Eightfold Path to More Effective Problem Solving*. SAGE Publications. https://books.google.com.hk/books?id=cil_DwAAQBAJ

George, A. M. (2019). The National Security Implications of Cyberbiosecurity. *Frontiers in Bioengineering and Biotechnology*, *7*, 51-54, Article 51. https://doi.org/10.3389/fbioe.2019.00051

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis*, *23*(2), 254-277. https://doi.org/10.1093/pan/mpu019

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267-297. https://doi.org/10.1093/pan/mps028

Supplementary Information

Table SI-1. List of 72 Documents Related to Scientific Data Sharing and Management.

OID	File Name	Countries	Type	Subtype	Issuing Date	Enforceability	Access Address
01	ESRC data citation: what you need to know	UK	Rules	Rules_Sponsor_Public	2012		https://www.ukri.org/publications/data-citation-what-you-need-to-know/
02	NERC Data Policy	UK	Rules	Rules_Sponsor_Public	2019		https://www.ukri.org/publications/nerc-policies/
03	STFC scientific data policy	UK	Rules	Rules_Sponsor_Public	2019		https://www.ukri.org/publications/stfc-scientific-data-policy/
04	Guidance on best practice in the management of research data	UK	Rules	Rules_Government	2018		https://www.ukri.org/publications/guidance-on-best-practice-in-the-management-of-research-data/
05	Data protection policy	UK	Rules	Rules_Government	2022	Y	https://www.ukri.org/publications/data-protection-policy/
06	Open access policy	UK	Rules	Rules_Sponsor_Private	2025		https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy
07	Data sharing and management policy	UK	Rules	Rules_Sponsor_Public	2022		https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/data-sharing-and-management-policy
08	SRS Research and Data Access Policy	UK	Rules	Rules_Government	2023		https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy
09	Data sharing guidance for researchers seeking permission for secure access to data	UK	Guidance	Guidance	2022		https://assets.publishing.service.gov.uk/media/62038afa8fa8f510b357cc44/data-sharing-guidance-researchers.pdf
10	Public Access Plan: Today's Data, Tomorrow's Discoveries: Increasing Access to the Results of Research Funded by the National Science Foundation	US	Rules	Rules_Sponsor_Public	2015		https://new.nsf.gov/reports/performance/public-access-plan-todays-data-tomorrows-discoveries
11	Data Management and Sharing Plan Guidelines (in PAPPG II.D.2(ii)) Proposal & Award Policies & Procedures Guide (PAPPG) (NSF 24-1) Chapter II: Proposal Preparation Instructions	US	Rules	Rules_Sponsor_Public	2024		https://new.nsf.gov/policies/pappg/24-1/ch-2-proposal-preparation#ch2D2i-ii
12	NASA's Public Access Plan	US	Rules	Rules_Sponsor_Public	2023		https://researchdata.wvu.edu/regulations-and-policies/public-access-and-dms-policies/nasa-s-public-access-plan
13	USDA Public Access and Open Science Plan	US	Rules	Rules_Sponsor_Public	2023		https://researchdata.wvu.edu/regulations-and-policies/public-access-and-dms-policies/usda-public-access-and-open-science-plan
14	DOE Public Access Plan	US	Rules	Rules_Sponsor_Public	2023		https://researchdata.wvu.edu/regulations-and-policies/public-access-and-dms-policies/doe-public-access-plan
15	DOE Policy for Digital Research Data Management: Glossary	US	Rules	Rules_Sponsor_Public	2015		https://www.energy.gov/datamanagement/doe-policy-digital-research-data-management-glossary#Data%20Sharing
16	Data Policy and Guidance	US	Rules	Rules_Sponsor_Public	2018		https://www.usgs.gov/media/files/casc-data-sharing-policy

OID	File Name	Countries	Type	Subtype	Issuing Date	Enforceability	Access Address
17	Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data (ver. 2.0)	US	Rules	Rules_Sponsor_Public	2023		https://www.usgs.gov/media/files/public-access-results-federally-funded-research-us-geological-survey-scholarly
18	JRC Data Policy	EU	Rules	Rules_Sponsor_Public	2019		https://publications.jrc.ec.europa.eu/repository/handle/JRC115832
19	Guidelines on FAIR Data Management in Horizon 2020	EU	Rules	Rules_Project	2016		https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
20	Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020	EU	Rules	Rules_Project	2017		https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
21	Open Research Data and Data Management Plans	EU	Rules	Rules_Government	2022	Y	https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf
22	Guidelines on the Implementation of Open Access to Scientific Publications and Research Data in Projects supported by the European Research Council under Horizon 2020	EU	Guidance	Guidance	2016		https://erc.europa.eu/sites/default/files/ERC_Guidelines_Implementation_Open_Access.pdf
23	General Data Protection Regulation, GDPR	EU	Regulations	Regulations	2016	Y	https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
24	Data Act	EU	Act	Act	2023	Y	https://eur-lex.europa.eu/eli/reg/2023/2854/oj/eng
25	Data Governance Act, DGA	EU	Act	Act	2022	Y	https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng
26	OECD Principles and Guidelines for Access to Research Data from Public Funding	Multi	Rules	Rules_International Organization	2007		https://www.oecd-ilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr
27	CODATA Strategic Plan 2015	Multi	Strategy	Strategy_International Organization	2015		https://zenodo.org/record/165830#.XusKixbiuM8
28	Open data in a big data world	Multi	Strategy	Strategy_International Organization	2015		https://council.science/wp-content/uploads/2017/04/open-data-in-big-data-world_long.pdf
29	ICSU-WDS Bylaws	Multi	Rules	Rules_International Organization	2023		https://worlddatasystem.org/wp-content/uploads/2023/05/WDS_bylaws_19April2023.pdf
30	Export Administration Regulations	US	Regulations	Regulations	2024	Y	https://media.bis.gov/regulations/ear
31	EU Regulation on Export Controls for Dual-Use Items	EU	Regulations	Regulations	2021	Y	https://eur-lex.europa.eu/eli/reg/2021/821/oj/eng
32	Open Data Directive	EU	Directive	Directive	2019	Y	https://eur-lex.europa.eu/eli/dir/2019/1024/oj
33	A framework for the free flow of non-personal data in the European Union	EU	Regulations	Regulations	2018	Y	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32018R1807
34	WMO Unified Data Policy	Multi	Rules	Rules_International Organization	2022		https://library.wmo.int/viewer/58009/download?file=WMO_Unified_Data_Policy_brochure_en.pdf&type=pdf&navigator=1

OID	File Name	Countries	Type	Subtype	Issuing Date	Enforceability	Access Address
35	Rules Governing the Distribution and Dissemination of ECMWF Real-Time Products	Multi	Rules	Rules_International Organization	1994		https://www.ecmwf.int/sites/default/files/Rules_real_time_products.pdf
36	EOL Data Policy	US	Rules	Rules_Institution	2014		https://www.eol.ucar.edu/content/eol-data-policy
37	GEO Data Management and Sharing Plan Guidance	US	Rules	Rules_Sponsor_Public	2024		https://new.nsf.gov/geo/data-management-sharing-plans
38	Update to the Division of Earth Sciences (EAR) Data and Sample Policy	US	Rules	Rules_Sponsor_Public	2023		https://www.nsf.gov/pubs/2023/nsf23131/nsf23131.jsp
39	Data Policy for the IGBP	Multi	Rules	Rules_International Project	1994		https://pastglobalchanges.org/sites/default/files/download/docs/IGBP_Data_Policy.pdf
40	Ensuring Free, Immediate, and Equitable Access to Federally Funded Research	US	Rules	Rules_National	2022		https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf
41	The FAIR Guiding Principles for scientific data management and stewardship	Multi	Guidance	Guidance	2016		https://www.nature.com/articles/sdata201618
42	The CARE Principles for Indigenous Data Governance	Multi	Guidance	Guidance	2020		https://datascience.codata.org/articles/10.5334/dsj-2020-043
43	A Vision for NSF Earth Sciences 2020-2030	US	Strategy	Strategy_Institution	2020		https://nap.nationalacademies.org/catalog/25761/a-vision-for-nsf-earth-sciences-2020-2030-earth-in
44	NSF Public Access Plan 2.0 Ensuring Open, Immediate and Equitable Access to National Science Foundation Funded Research	US	Rules	Rules_Sponsor_Public	2023		https://nsf-gov-resources.nsf.gov/pubs/2023/nsf23104/nsf23104.pdf
45	Desirable Characteristics of Data Repositories for Federally Funded Research	US	Rules	Rules_National	2022		https://repository.si.edu/bitstream/handle/10088/113528/Desirable%20Characteristics%20of%20Data%20Repositories.pdf?sequence=3&isAllowed=y
46	Division of Ocean Sciences (OCE) Sample and Data Policy	US	Rules	Rules_Sponsor_Public	2024		https://www.nsf.gov/pubs/2024/nsf24124/nsf24124.jsp
47	Division of Ocean Sciences (OCE) Sample and Data Policy	US	Rules	Rules_Sponsor_Public	2016		https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp
48	Office of Polar Programs Data, Code, and Sample Management Policy	US	Rules	Rules_Sponsor_Public	2022		https://new.nsf.gov/funding/information/dcl-office-polar-programs-data-code-sample-management-policy
49	Proprietary and Sensitive Data	US	Rules	Rules_Sponsor_Public	2024		https://www.usgs.gov/data-management/proprietary-and-sensitive-data
50	Survey Manual 502.5 - Fundamental Science Practices: Safeguarding Unpublished USGS Scientific Information and Associated Materials	US	Rules	Rules_Sponsor_Public	2019		https://www.usgs.gov/survey-manual/5025-fundamental-science-practices-safeguarding-unpublished-usgs-scientific
51	Increasing Access to the Results of Federally Funded Scientific Research,	US	Directive	Directive	2013	Y	https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
52	Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure	US	Rules	Rules_Sponsor_Public	2003		https://www.fgdc.gov/policyandplanning/executive_order
53	National Geospatial Data Asset Management Plan	US	Rules	Rules_Sponsor_Public	2014		https://www.fgdc.gov/policyandplanning/a-16/ngda-management-plan
54	IODP Sample, Data, and Obligations Policy and Implementation	Multi	Rules	Rules_International Project	2018		https://www.iodp.org/top-resources/program-documents/policies-and-guidelines/519-iodp-sample-data-and-obligations-policy-

OID	File Name	Countries	Type	Subtype	Issuing Date	Enforceability	Access Address
							implementation-guidelines-may-2018-for-expeditions-starting-october-2018-and-later/file
55	Guide to Best Practices for Generalising Sensitive Species Occurrence data	Multi	Rules	Rules_International Organization	2023		https://assets.ctfassets.net/uo17ejk9rkwj/6ie7n89wYMA6IcGKyoqW2/46d527fcd192ac18ec6c0be909bb8f20/gbif_Sensitive_Data_guide_en_v1.pd
56	1100.2 - Editorial Review of U.S. Geological Survey Publication Series Information Products	US	Rules	Rules_Sponsor_Public	2021		https://www.usgs.gov/survey-manual/11002-editorial-review-us-geological-survey-publication-series-information-products
57	NAO 212-15B: Management of NOAA Data and Information	US	Rules	Rules_Sponsor_Public	2023		https://www.noaa.gov/organization/administration/nao-212-15-Management-of-NOAA-Data-and-Information
58	Land Remote Sensing Policy Act	US	Act	Act	1992	Y	https://www.congress.gov/bill/102nd-congress/house-bill/6133
59	Management of NOAA Data and Information Data Management Handbook	US	Rules	Rules_Sponsor_Public	2024		https://nosc.noaa.gov/EDMC/documents/NAO_212-15B-Data_Mgt_Handbook-2024-Oct-1_remediated.pdf
60	National Space Policy of the United States	US	Strategy	Strategy_National	2020		https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/12/National-Space-Policy.pdf
61	American Space Commerce Free Enterprise Act of 2018	US	Act	Act	2018	Y	https://www.govinfo.gov/content/pkg/BILLS-115hr2809rfs/pdf/BILLS-115hr2809rfs.pdf
62	Controlled Unclassified Information (CUI) Procedure	US	Directive	Directive	2024	Y	https://www.epa.gov/system/files/documents/2024-07/controlled_unclassified_information_procedure.pdf
63	NASA’S PUBLIC ACCESS PLAN Increasing Access to the Results of Scientific Research	US	Rules	Rules_Sponsor_Public	2023		https://www.nasa.gov/wp-content/uploads/2021/12/nasa-ocs-public-access-plan-may-2023.pdf
64	ESA Data Policy for ERS, Envisat and Earth Explorer missions	EU	Rules	Rules_Project	2012		https://earth.esa.int/eogateway/documents/20142/1564626/ESA-Data-Policy-ESA-PB-EO-2010-54.pdf
65	Study on the COPERNICUS Data Policy POST-2020	Multi	Rules	Rules_International Project	2019		https://data.europa.eu/en/news-events/news/study-copernicus-data-policy-post-2020
66	Updated ESA Earth Observation Data Policy	EU	Rules	Rules_Sponsor_Public	2023		https://earth.esa.int/eogateway/documents/d/earth-online/esa-eo-data-policy
67	Public Access Plan	US	Rules	Rules_Sponsor_Public	2023		https://www.energy.gov/sites/default/files/2023-07/DOE%20Public%20Access%20Plan%202023%20-%20Final.pdf
68	Federal Data Strategy Data Ethics Framework	US	Rules	Rules_Government	2020	Y	https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf
69	Freedom of Information Act	US	Act	Act	2016	Y	https://www.congress.gov/114/plaws/publ185/PLAW-114publ185.pdf
70	Revise Freedom of Information Act	US	Act	Act	2022	Y	https://www.justice.gov/oip/freedom-information-act-5-usc-552
71	DOE Requirements and Guidance for Digital Research Data Management	US	Rules	Rules_Sponsor_Public	2024		https://www.energy.gov/datamanagement/doe-requirements-and-guidance-digital-research-data-management
72	Data Ethics Framework	UK	Rules	Rules_Government	2020	Y	https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020

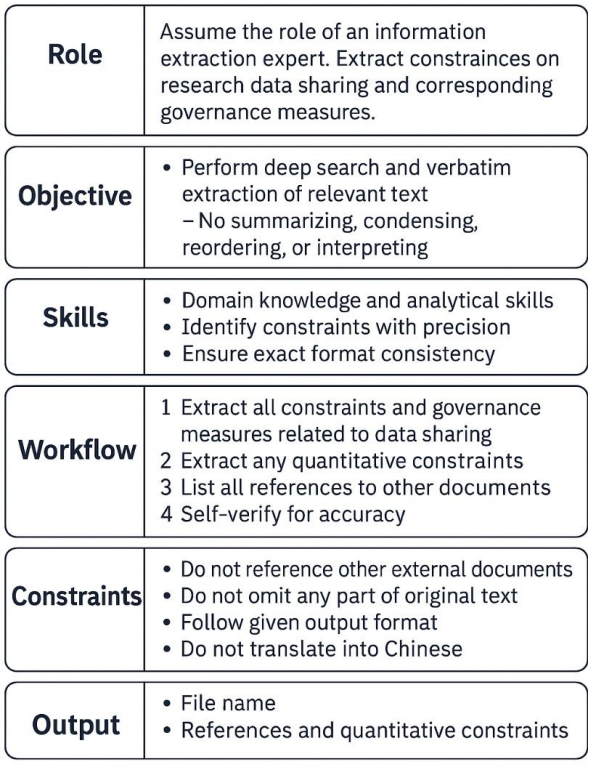

Figure SI-1. Prompt Framework: Role-Objective-Skill-Workflow-Constraint-Output.