Research Support Structures as Translators in Computational Social Science: A Socio-Technical Framework and Lessons from Practice

Susanne de Vogel [1], Lena Steinmann [1], and Rolf Drechsler [1]

Abstract: This paper conceptualizes Computational Social Science (CSS) as a socio-technical field in which methods, data, standards, and responsibilities move between social science, computer science, and related fields. We propose a framework of five friction points that require translation work and argue that support structures for data-intensive research play a central role in negotiating these frictions. Drawing on practical insights from the interdisciplinary DataNord help-desk, we show how support teams translate between disciplinary assumptions, computational workflows, data governance, and standards of responsible research.

Keywords: Computational Social Science, Socio-Technical-Systems, Interdisciplinary Research, Research Support Units, Data Competence Centers, DataNord

1 Introduction

Computational Social Science (CSS) is a highly interdisciplinary field that combines the social sciences with computer science, statistics, and related fields to investigate human behavior and social systems at scale [JP23; LPW23]. It develops and applies computational methods to analyze complex, often large-scale datasets, including digital trace data from social media platforms, telecommunications, administrative records, sensors, and historical archives [Ed20; La20].

Societal analysis is increasingly shaped by a two-directional movement of methods and expertise: social scientists adopt computational approaches from computer science and data science, while STEM-trained researchers use digital traces, large-scale datasets, and AI tools to address questions about social behavior and societal dynamics. This creates new opportunities for knowledge production, but also generates tensions. Computational methods carry implicit assumptions about valid evidence, uncertainty, data quality, and the kinds of claims that can be made from model outputs. When these assumptions move between sociology, computer science, data science, and adjacent fields, they may not fit the epistemic standards, ethical responsibilities, or methodological traditions of the new context. A model that performs well predictively, for example, may still be difficult to interpret as social-scientific evidence if its variables are weakly connected to theoretical constructs,

[1] University of Bremen, Data Science Center, Bibliothekstr. 5, 28359 Bremen, Germany,
devogel@uni-bremen.de, https://orcid.org/0000-0002-6253-5137;
lena.steinmann@uni-bremen.de, https://orcid.org/0000-0001-5443-0581;
drechsler@uni-bremen.de, https://orcid.org/0000-0002-9872-1740

its data are biased or non-representative, or its outputs are treated as explanations without sufficient conceptual grounding.

We argue that these challenges are socio-technical rather than merely technical. CSS depends on the alignment of researchers, disciplinary norms, forms of expertise, data infrastructures, software tools, legal requirements, ethical responsibilities, and institutional settings. From this perspective, the movement of computational methods into and out of the social sciences requires translation work: the practical effort of making assumptions, concepts, methodological choices, quality standards, and responsibilities understandable and appropriate across disciplinary contexts. This includes clarifying how data-driven features relate to theoretical constructs, what kind of claim a statistical result can support, or how legal and ethical requirements become concrete data practices. Building on Socio-Technical Systems theory and concepts from science and technology studies, this paper identifies five recurring friction points where such translation becomes necessary: epistemic friction, measurement friction, operationalization friction, quality-measures friction, and accountability friction. These friction points mark situations in which different field logics, research practices, standards of evidence, and responsibilities meet around shared objects such as datasets, labels, models, pipelines, platforms, or documentation. Rather than treating them as isolated problems, we conceptualize them as systematic sites of negotiation in interdisciplinary CSS.

We further argue that this translation work is increasingly carried out by support structures for data-intensive research, such as Data or Digital Science Centers and Data Competence Centers, as currently funded by the BMFTR for a period of three years[2]. These units operate at the interface between disciplines through consulting, training, and networking. Empirically, the paper draws on practice-based insights from the interdisciplinary help-desk of the Data Competence Center for the Bremen region "DataNord", hosted at the Data Science Center of the University of Bremen. Using two anonymized examples from consulting practice, we illustrate how friction points arise in concrete research situations and how interdisciplinary support teams translate between social-scientific questions, computational workflows, statistical reasoning, data governance, and standards of responsible research.

The contribution of this paper is twofold. First, we introduce a framework that conceptualizes the transfer of epistemic paradigms, methods, quality standards, and responsibilities in CSS as a socio-technical translation process structured by identifiable friction points. Second, we highlight the role of institutional support structures for data-intensive research as key actors in this process and derive implications both for CSS as an interdisciplinary research field and for Data and Digital Science initiatives [Bo25] that design services, training formats, and governance structures. In doing so, we aim to contribute to a shared understanding of how to support responsible, transparent, and methodologically sound interdisciplinary research.

[2] https://www.bildung-forschung.digital/digitalezukunft/de/bildung/digital-_und_datenkompetenzen/datenkompetenzzentren_fuer_die_wissenschaft_ordner/datenkompetenzzentren_fuer_die_wissenschaft_node.html

2 CSS as a Socio-Technical System

2.1 A Socio-Technical Perspective on Interdisciplinary Research Fields

Before applying this perspective to CSS, we briefly introduce Socio-Technical Systems (STS) theory and its relevance for interdisciplinary research. STS helps to show why the movement of computational methods across disciplinary boundaries is never merely technical. It also involves practices, tools, institutional conditions, shared artefacts, collaboration spaces, and forms of domain-specific and cross-domain expertise that need to be aligned and negotiated.

STS can be understood as work or knowledge systems in which human actors, technical artefacts, and organizational environments are interdependent elements of a larger system [BH77; TB51]. The *social subsystem* includes individuals, relationships, roles, skills, values, organizational structures, and work-related practices, while the *technical subsystem* comprises tools, technologies, methods, devices, and infrastructures. Both operate within an *environmental subsystem*, which refers to the organizational, institutional, cultural, economic, and political conditions under which the system is situated [Tr81]. The central design principle is *joint optimization*: social and technical elements should be aligned so that they mutually support each other and contribute to the functioning of the system as a whole [Tr81]. A technical solution may fail if it does not fit existing practices or forms of expertise; conversely, social arrangements depend on appropriate infrastructures, information flows, and boundary arrangements.

For interdisciplinary research fields, STS can be extended through concepts from science and technology studies. *Trading zones* describe spaces in which communities with different vocabularies, methods, and standards coordinate their practices without fully merging into one discipline [Ga96]. *Boundary objects* are shared artefacts, such as datasets, models, tools, protocols, or infrastructures, that can be used across communities while retaining different meanings for each of them [SG89]. Different forms of expertise specify how actors participate in such settings: *contributory expertise* refers to the ability to carry out recognized work within a specific domain, whereas *interactional expertise* describes the ability to understand and use another expert community's language well enough to engage in meaningful exchange without becoming a full contributor in that field [CE07]. This perspective is particularly useful for analyzing interdisciplinary research fields under conditions of technological change. New data sources, computational tools, and analytical methods do not merely expand technical capacities; they can reorganize workflows, redistribute expertise, introduce new standards of evidence, and shift responsibilities for data quality, documentation, interpretation, and accountability. Because technologies become embedded in specific practices, institutions, and knowledge cultures [BH09], STS provides a lens for analyzing how technical change reshapes epistemic standards, responsibilities, and socio-technical risks in data-driven research [Hä19] and artificial intelligence (AI) [Ka22].

Previous STS-informed work illustrates this analytical strength. Studies of digital humanities show how collaboration between historians, computational linguists, computer scientists, archivists, and other experts depends on shared tools, resources, working languages, and interactional expertise, while epistemic cultures remain distinct [Ke21; LM24]. Edwards et al. [Ed11] describe similar tensions in data-intensive climate research as *science friction*, highlighting the communicative work required when data, tools, metadata, institutions, and disciplinary expectations meet. Ribes [Ri19] further shows that data science itself is organized around the crossing of domains, seeking to make distinct fields interoperable through tools, infrastructures, and hybrid forms of expertise. Finally, Smith et al. [Sm20] add a social science-specific perspective by showing that data management and digital scholarship require continuous negotiation between protection and disclosure, openness and restriction, and disciplinary needs and infrastructural possibilities.

Taken together, these studies show that STS makes visible the translation work required in interdisciplinary research fields under condition of rapid technological change. It directs attention to the alignment of social, technical, and environmental subsystems, the negotiation of standards and responsibilities, and the role of shared artefacts and interactional expertise in enabling cooperation across disciplinary boundaries.

2.2 A Socio-Technical Perspective on CSS

Building on this perspective, we understand CSS as a socio-technical research field in which knowledge about social phenomena is produced through the interaction of people, data, methods, technologies, and broader research contexts [AM23; BH09]. Figure 1 illustrates this conceptualization. In the following, we specify CSS through its social, technical, and environmental subsystems and show how their alignment shapes the conditions under which CSS can produce credible research [AM23; Ap97].

The **social subsystem** of CSS includes researchers from social science, computer science, data science, and related fields; support actors such as technicians, librarians, data stewards, and student assistants; and the people whose actions become part of the data, for example through digital traces. These actors bring different skills, values, and standards into the research process: social scientists may emphasize interpretation, explanation, and theoretical grounding, while computational researchers may focus more strongly on prediction, classification, or technical performance [Ho21]. The social subsystem therefore also covers how actors communicate, develop shared understanding and interactional expertise, and negotiate concepts, methods, and responsibilities across disciplinary boundaries.

The **technical subsystem** comprises large-scale datasets such as digital trace data or administrative records; infrastructures for storing, processing, and accessing data, such as servers, platforms, sensors, and APIs; as well as statistical and computational methods such as machine learning (ML), natural language processing (NLP), network analysis, and agent-based modeling ABM) [BMY16; Ch21; JP23; La20], as well as software and programming languages (e.g. R and Python). The technical subsystem is not limited to final analysis, but also includes data collection (e.g. via APIs or web scraping), curation, cleaning, operationalization, documentation, and visualization [JP23]. At the same time, digital data are not neutral representations of social reality. They are produced through platforms, algorithms, interfaces, access regimes, and curation practices that shape what can be observed and analyzed [AA21; BH09]. Access to data, infrastructures, and computational skills is therefore unevenly distributed across actors and institutions [DHC20].

The **environmental subsystem** refers to the broader conditions under which CSS operates [AM23]. It shapes what can be studied, which data can be accessed, and how results may be produced and shared. This includes ethical norms regarding privacy and consent [KHB24]; legal and regulatory frameworks such as data protection law, intellectual property rights, the Digital Services Act (DSA), and the EU AI Act [La20; LPW23]; as well as expectations of openness, reproducibility, and compliance with FAIR [Wi16], CARE [Ca20], and FACT [VBH17] principles. Since private corporations often control the generation of and access to digital trace data, CSS also depends on platform conditions [Ba24; Br19; Hä19; La20; Pa15]. More broadly, processes of datafication make social life increasingly available as data, while social norms, public values, and institutional priorities shape which research questions appear legitimate or urgent [BH09; DHC20; Ki14].

From an STS perspective, CSS' central challenge is one of joint optimization [Tr81]: Powerful computational methods may still be inadequate if it does not fit the theoretical assumptions, data limitations, or explanatory goals of social research. At the same time, social-scientific inquiry increasingly depends on technical systems that shape data access, operationalization, and possible forms of analysis. Datasets, codebooks, annotation schemes, preprocessing pipelines, models, protocols, or software tools can function as **boundary objects**: they enable coordination across communities while retaining different meanings for different actors [Ke21; LM24]. A dataset, for example, may be treated as empirical material by sociologists, as training input by computer scientists, and as an object requiring documentation, access control, and anonymization by data stewards. These objects support collaboration, but they do not remove the need to negotiate meanings, assumptions, and standards.

Such negotiations take place in **trading zones** in which social scientists, computer scientists, data scientists, and support actors coordinate practices despite different vocabularies, standards, and epistemic cultures [Ke21; LM24]. In CSS, trading zones may emerge around research questions, construct operationalization, data access, method choice, model interpretation, quality assessment, or ethical and legal constraints. Expertise is central to these negotiations. Researcher from different domains may each possess strong contributory expertise in their own field. In CSS, this may include sociological expertise in theory-building and causal reasoning, computational expertise in modeling, classification, and prediction, or data-scientific expertise in data collection, engineering, and architecture [Hä19; Ri19]. However, science friction [Ed11] arise when this expertise remains domain-specific and individuals lack the **interactional expertise** [CE07] needed to understand the other community's language and reasoning well enough to ask informed questions, recognize limitations, and negotiate methodological decisions. In the following, we conceptualize these tensions as friction points in the translation of computational methods into and out of sociology.

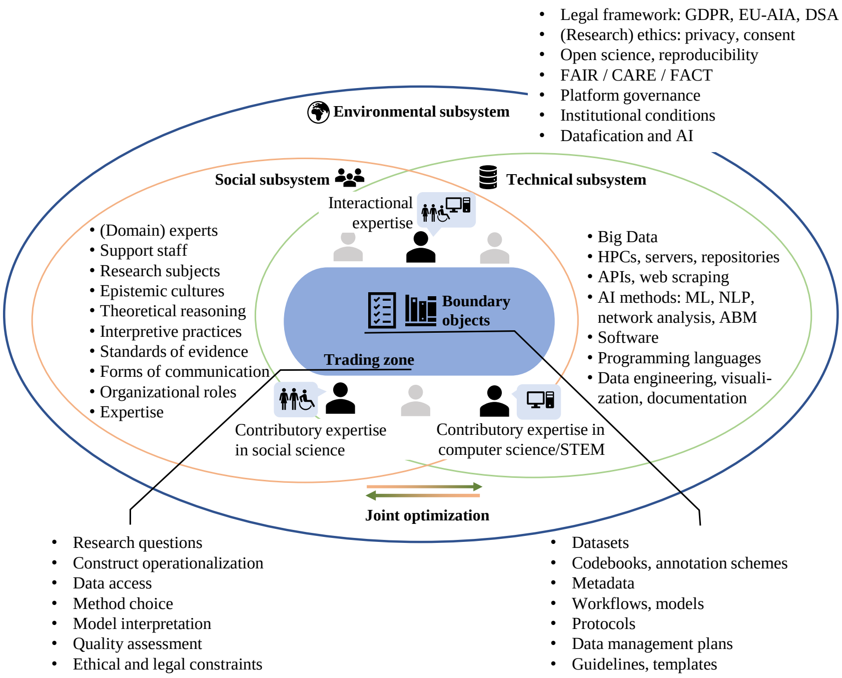

Fig. 1: Computational Social Science as a Socio-Technical System

Note. Adapted and further developed from Kaminsky [Ka22] and informed by concepts of socio-technical systems, boundary objects, trading zones, and expertise studies [BH77; CE07; Ga96; SG89; TB51].

3 Friction Points as Sites of Translation in CSS

As argued above, friction points in CSS arise from two closely connected sources. First, they emerge when different epistemic cultures, research practices, and forms of contributory expertise meet around shared objects such as datasets, labels, models, pipelines, or platforms, while actors lack the interactional expertise needed to understand how other fields attach different meanings, standards, and responsibilities to these objects. Second, these negotiations are shaped by partly conflicting requirements of the social, technical, and environmental subsystems: CSS should be theoretically grounded and technically feasible, open and reproducible yet privacy-preserving and legally compliant, and methodologically rigorous while remaining responsive to limits of data access, infrastructure, and expertise. Building on the existing literature, we identify five recurring friction points in CSS that mark where translation work becomes necessary (Table 1).

Tab. 1: Friction points in CSS

Friction point	Core tension	Translation needed
Epistemic friction	Explanation, causality, and interpretation vs. prediction and performance	What kind of claim can this result support?
Measurement friction	Theoretical constructs vs. labels, variables or proxies.	What exactly is being measured, and does the proxy fit the construct?
Operationalization friction	Theoretical assumptions and causal mechanisms vs. model inputs and computational features.	How can theoretical assumptions be represented in data, features, or models?
Quality measures friction	Validity, reliability, representativeness, and bias vs. accuracy, robustness, scalability, and efficiency.	Which quality standards apply, and how can they be combined?
Accountability friction	Openness and reproducibility vs. legal compliance, ethics, and opaque infrastructures.	Which responsibilities follow regarding data practices, documentation, and sharing decisions?

Epistemic friction concerns the different knowledge claims that meet in CSS. Social-scientific research, both quantitative and qualitative, is often oriented toward theory-building, interpretation, causal reasoning, and context-sensitive explanation [GRS21; Ho21]. Computer-scientific and machine-learning approaches, by contrast, have been shaped by Big Data and increasing computational power and therefore often prioritize prediction, scalability, benchmark performance, and evaluation on unseen test data [GRS21; Ho21]. This opposition is not fixed: current CSS increasingly acknowledges sequential, iterative, and partly abductive forms of inquiry, in which theory, measures, hypotheses, and models are refined through engagement with data [GRS21; JP23]. Friction arises when these epistemic paradigms remain implicit. Algorithmically detected patterns are not meaningful by themselves; without theoretical grounding, data mining can produce spurious correlations that are statistically striking but lack social reality [Hä19; Ki14]. Likewise, predictive accuracy may be mistaken for explanation, model performance for theoretical validity, or causal claims for mere pattern recognition [Ho21]. What needs to be translated is therefore the epistemic status of a computational result: whether it describes a pattern, measures a construct, predicts an outcome, explains a mechanism, or supports a causal claim [Ho21].

Measurement friction arises around the translation of theoretical constructs into computational proxies. In social-scientific research, concepts such as social capital, identity, democracy, or hate speech are not directly observable, but require theoretical definition and careful operationalization [Sa18]. In machine-learning workflows, the same problem often appears as a question of labels, variables, proxies, or ground truth: an observable marker is selected or annotated so that an algorithm can learn from it [LBJ22; RJ20]. Friction arises when such labels are treated as if they were already valid measurements of the underlying social concept. For example, annotator disagreement may be treated as noise to be reduced technically, although the deeper problem may be an ambiguous or contested construct [RJ20]. Technical solutions, for example, cannot decide what hate speech means in a specific research context. What needs to be translated is the relation between construct, proxy, and procedure: abstract concepts have to be turned into explicit coding rules, annotation guidelines, and validation strategies, while computational labels and outputs have to be translated back into theoretically meaningful claims.

Operationalization friction concerns how theoretical assumptions about social processes are represented in computational models. While measurement asks whether an indicator captures a theoretical construct, operationalization asks how assumptions about mechanisms, contexts, and causal relations are translated into model inputs. In many social-scientific approaches, explanation is built around a theoretically selected set of variables whose estimated parameters are interpreted in relation to a mechanism or causal argument [GRS21; Ho21]. Feature selection is therefore not only a technical preprocessing step, but part of theoretical model building [RJ20]. Machine-learning approaches, by contrast, often focus on accurate predictions of an outcome and therefore work with high-dimensional predictor sets, flexible models, automated feature extraction, and validation on held-out data [GRS21; LBJ22]. Friction arises when features improve prediction but do not correspond to theoretically meaningful processes, or when black-box models make it difficult to connect inputs and outputs to social mechanisms [JKH21; LPW23; RJ20]. This is amplified in platform-based data, where observed patterns may reflect platform design or recommendation systems rather than the social process under study [Sa18]. What needs to be translated is the relation between causal reasoning and computational representation.

Quality measures friction addresses the criteria by which CSS research is evaluated as "good" or "valid". Social-scientific quality criteria typically include validity, reliability, representativeness, bias, uncertainty, and theoretical plausibility: research is considered strong when data and measures fit the concept under study, findings can be generalized to the defined target population, and interpretations remain sensitive to the data-generating process [LPW23; Sa18]. Computer-scientific and machine-learning approaches, by contrast, often evaluate quality through predictive performance, robustness, scalability, and efficiency, using metrics such as accuracy, precision, sensitivity, specificity, or AUC, especially on test data [Ch21; GRS21; LPW23]. Friction arises because these standards do not automatically map onto one another. A model may perform well according to predictive benchmarks but still be weak from a social-scientific perspective if it is opaque, biased, non-representative, or lacks explanatory power [Ho21; LPW23]. What needs to be translated is the quality claim attached to a result: whether “good” means accurate, robust, representative, interpretable, theoretically plausible, reproducible, or useful for a specific research purpose.

Accountability friction captures the tension between scientific standards, ethical and legal requirements, and CSS’s dependence on private and opaque infrastructures. Social-scientific perspectives emphasize that digital data are not neutral raw material, but products of socio-technical processes [AA21]. Their quality depends on how they are generated, classified, transformed, accessed, and interpreted; in this sense, data quality is inseparable from control over the statistical chain and the infrastructures that produce it [DHC20]. Legal and ethical issues such as privacy, consent, confidentiality, copyright, and data protection are therefore not external constraints, but part of responsible knowledge production [KHB24; Le23]. Computational approaches, however, often treat platforms, APIs, infrastructures, and pipelines primarily as enabling conditions for analysis. Friction arises when this technical orientation underestimates that data are socially produced, legally constrained, and politically consequential. CSS is expected to be open and reproducible while remaining legally compliant and privacy-preserving under conditions of platform dependency, unclear data access, proprietary infrastructures, and algorithmic opacity [La20; Pa15]. What needs to be translated are responsibilities across infrastructures: legal and ethical requirements must become concrete data practices, access rules, documentation routines, and sharing decisions, while technical choices about APIs, scraping, preprocessing, model use, and data sharing have to be translated back into their social, legal, and political implications.

4 From Friction Points to Practice

4.1 Institutional Drivers of Friction in CSS

Having outlined the five friction points, we now return to the two sources from which they arise: the insufficient alignment of CSS's social, technical, and environmental subsystems, and the uneven distribution of interactional expertise needed to negotiate assumptions, standards, responsibilities, and constraints across disciplinary boundaries [CE07]. Science-sociological work on CSS links both problems to the organization of universities and to persistent gaps in social science methods training [La20; Ru13].

A first problem is institutional. CSS depends on cooperation between social science, computer science, data infrastructures, ethics and legal support, libraries, and IT services, yet these competencies are often distributed across disciplinary and administrative silos [La09; La20]. Within universities, researchers are typically housed in separate units with few formal mechanisms to facilitate exchange between domains. Budgeting, infrastructures, and evaluation systems still tend to privilege disciplinary work over multi- or interdisciplinary collaboration. As a result, many universities lack shared environments that combine computational power, secure access controls, data protection routines, and usable workflows for research with large-scale and sensitive data [La20].

A second problem is pedagogical. Social science research must adapt to new data types and computational methods without abandoning its theoretical and methodological standards [LPW23]. However, methods training in the social sciences has often not kept pace with Big Data, AI-driven methods, programming, data engineering, and platform-based research [La20; LPW23; Ru13]. Standard curricula remain focused on qualitative methods, statistics, and survey methodology, while essential data science skills such as programming in R or Python are still less systematically integrated [Ed20]. This gap is reinforced by a lack of computationally literate faculty [Ke21]. At the same time, social-scientific questions are often addressed by computer scientists, data scientists, or other STEM-trained researchers who bring strong computational expertise, but are not necessarily trained in social-scientific theory, research standards, or the specific responsibilities involved in working with data about people [KHB24; La09; La20; Sa18].

Thus, the negotiation of friction points cannot simply be left to individual researchers. Following Smith et al. [Sm20], CSS requires institutional support structures that provide translation, shared standards, and accessible interactional expertise across fields.

4.2 Institutionalizing Translation Work: Support Structures for Data-Intensive Research

New institutional arrangements in the data- and digital-science landscape aim to close this gap. These support structures for data-intensive research include, most prominently, **Data Science Centers** (DSCs) [Pf24; St23b]. DSCs are predominantly established as interdisciplinary, cross-faculty research structures within higher education institutions. They therefore fulfill a critical "hinge" function, mediating between methodological development—specifically in the fields of computer sciences, mathematics, and statistics—and their diverse application disciplines. Besides technical infrastructure such as GPU servers, DSCs provide research support services including consulting and educational programs to enhance data literacy in working with Big Data and AI methods across the research data lifecycle. DSCs are typically staffed by interdisciplinary teams of data scientists and/or data stewards, designed to bridge the gap between methodological expertise, research data management and domain-specific applications. To facilitate knowledge transfer and research collaborations, DSCs often additionally maintain an extensive network of members and associated partners from various faculties and research institutes, facilitating knowledge transfer collaborations.

Complementing DSCs, **Data Competence Centers** (Datenkompetenzzentren, DKZs) represent a more recent nationwide initiative to strengthen data-related competencies across the research landscape [Pf24; St23b]. Importantly, DSCs often act as key stakeholders within these centers. In Bremen, for example, the DSC is one of the central pillars of the DKZ DataNord[St25] alongside the doctoral training programme "Data Train" [Hö21]. BMFTR-funded as regional, inter-institutional support structures for data-intensive research, DKZs are designed to combine learning, consulting, and networking activities to support researchers across disciplines, career stages, and levels of prior expertise. The landscape of DKZs is heterogeneous: some centers are explicitly interdisciplinary, such as DataNord for the Bremen region [St25]. Other DKZs are more domain- or method-specific [GBK25]. For CSS, KODAQS is particularly relevant as a data competence center focusing on social and behavioral sciences. It supports researchers in assessing and improving the quality of research data, including digital trace data, and administrative or process-produced data [De24]. DKZs bring together universities, non-university research institutes, infrastructure providers, and existing data-related initiatives. They therefore function as interfaces between local research practices and broader national infrastructures, such as the National Research Data Infrastructure (NFDI), and extend the hinge function of involved DSCs.

By bundling expertise across domains and institutions, DSCs and DKZs address the two problems identified above: they break down disciplinary silos and provide training opportunities that compensate for gaps in traditional disciplinary education. In this sense, they create institutional spaces in which friction points around research questions, data practices, methods, infrastructures, ethical and legal requirements, and standards of good research can be negotiated across fields. In line with Smith et al. [Sm20], they also create boundary objects, such as guidelines, checklists, templates, workflows, metadata standards, and training materials [St25], which provide common reference points for collaboration while remaining adaptable to different disciplinary and project-specific needs. Their interdisciplinary staff provide the interactional expertise needed to translate between computer science, data science and domain of application.

5 Translation in Action at the DataNord Help-desk

5.1 DataNord - Interdisciplinary Data Competence Center for the Bremen Region

The following section shifts from the conceptual framework to support practice. If the five friction points mark where translation into and out of social sciences becomes necessary, the everyday work of the interdisciplinary DataNord help-desk shows how such translation is carried out in practice. Drawing on our consulting experience, we illustrate how support structures for data-intensive research can function as practical negotiation spaces in which CSS-related frictions are identified, translated, and worked through.

Institutionally, the help-desk is embedded in DataNord, an interdisciplinary data competence center for the Bremen region [St25]. It is one of eleven national centers and the only one of its kind in Northern Germany. DataNord is strategically anchored in the U Bremen Research Alliance (UBRA), which connects the University of Bremen with thirteen non-university research institutes. Beyond this core network, DataNord also integrates further regional partner institutions, such as the Bremen City University of Applied Sciences, research data centers and repositories such as PANGAEA and Qualiservice, as well as NFDI consortia with Bremen participation. This regional embedding is further strengthened through DataNord's connection to the broader *Northwest Data Space* [Sc26] of data-related initiatives and infrastructures in North-West Germany.

The help-desk itself is centrally located at the *University of Bremen's Data Science Center* [St23a] and acts as the first point of contact for researchers within the DataNord network seeking support in data science and research data management related questions. It handles most inquiries directly and, where specialized expertise is needed, connects researchers with other relevant services of e.g. Qualiservice, the NFDI consortia, or other data competence centers [St25]. It is staffed by an interdisciplinary team of five data scientists representing the region's established scientific profile areas: (1) Environmental and Marine Sciences, (2) Social Sciences, (3) Material and Engineering Sciences, (4) Health Sciences, and

(5) Humanities. Beyond their respective domain expertise, team members are trained in statistics, computational and AI methods, programming languages, research software, data engineering and research data management. Their expertise is based on prior academic and professional experience as well as on continuous mutual learning from each other. In this way, the team combines complementary forms of domain-specific and methodological knowledge and develops the interactional expertise required to translate between disciplinary contexts. Drawing on Lazer's [La20] terminology, this positions the team members as *computationally literate social scientists* and *socially literate computational and STEM scientists*. Support for data-driven research is structured along three core activity areas: (1) training, (2) consultation services, and (3) networking activities:

- **Regular and on-demand training formats.** In the context of CSS, these include methodological workshops and short input talks on machine learning, text mining, automated audio transcription, and geospatial data analysis, as well as courses on programming languages such as R and Python for qualitative and quantitative research. Further training topics include tools and infrastructures such as Git, SQL, and high-performance computing using the DSC computing cluster, as well as research data management, reproducible research, and the responsible handling of personal and sensitive data. For researchers without prior training in the social sciences, the help-desk also offers introductory courses on basic statistics, questionnaire design, and empirical data collection methods.
- **Individual consultations**. These range from short-term *exploration* consultations, such as data management plans or technical troubleshooting, to more mid term, in-depth *guidance* on more complex methodological and analytical challenges, and long-term *realization* support in project-based collaborations. Consultations regarding CSS-related issues are usually handled as a team effort, combining one data scientist with in-depth domain knowledge in quantitative or qualitative social sciences and, where needed, additional colleagues with complementary methodological expertise.
- **Networking activities**. In events like the Data Community Club researchers and data practitioners (e.g. librarians, data stewards, and data scientists) can discuss current challenges, share experiences, and identify opportunities for collaboration across fields.

These formats provide complementary entry points for method translation: training builds shared methodological foundations, consultation enables context-specific translation in concrete research projects, and networking activities facilitate the exchange of experiences and practices among researchers of different fields.

5.2 Translation in and out of Social Science: Insights from Practice

To make this translation work visible, we draw on two short, anonymized examples from our consulting practice at the DataNord help-desk. They illustrate both directions of translation:

social scientists who seek to use computational methods for social-scientific questions, and researchers from computational or STEM-oriented fields whose data and methods raise social-scientific questions.

Case 1: Turning a qualitative social media corpus into a structured data workflow. The first case concerns a qualitative research project working with a large corpus of manually collected social media posts and comments. The researcher initially planned to analyze the material in qualitative analysis software, but the corpus became difficult to manage: the data were heterogeneous, inconsistently structured, and contained personal information. The consultation therefore focused on developing a data organization and wrangling workflow that linked posts, comments, subcomments, metadata, and screenshots while preserving the interactional context needed for qualitative interpretation.

Accountability friction arose around the handling of personal data: names, pseudonyms, document identifiers, and metadata had to be separated while preserving the link between posts, comments, and interactional context. *Quality measures friction* emerged around the reliability and consistency of the corpus: the manually assembled material had to be transformed into a reproducible and less error-prone data structure. Finally, this example involved *operationalization friction*: interactional structures, such as comments, replies, and participation patterns, had to be translated into analyzable metadata while still remaining meaningful for qualitative interpretation.

The translation work was carried out by a consulting team consisting of three data scientists. A data scientist with qualitative social research expertise preserved the epistemic logic of the project in ensuring that the workflow remained compatible with interpretive analysis. A data scientist with expertise in database structures and Python helped to time-efficiently restructure the data into machine-readable text tables and metadata and supported in the automatization of repetitive tasks. A data scientist with quantitative-oriented expertise in social sciences contributed to the pseudonymization workflow, ensuring that analytical usability and data protection requirements were considered together. This workflow functioned as a boundary object: it translated qualitative research needs into structured data practices and translated technical decisions back into questions of interpretation, accountability, and analytic usability.

Case 2: Aligning experimental data analysis with social-scientific measurement standards.
The second case concerns a team of computer scientists working with experimental data from a participant-based perception study. The initial request focused on statistical analysis: which tests were appropriate, how experimental conditions could be compared, and how ranking and rating information should be interpreted. As the data involved variables related to perception, evaluation, and subjective experience, during the consultation, however, the issue shifted from statistics to measurement. The consultation added a social-scientific perspective on theoretical constructs, operationalization, reliability, and validity.

Measurement friction arose because participant ratings and ranking could not simply be treated as self-evident variables. They had to be understood as operationalizations of underlying constructs that require theoretical justification and empirical validation. *Quality-measures friction* became visible in the discussion of appropriate statistical procedures, reliability checks, effect sizes, non-parametric tests, and uncertainty estimates. *Epistemic friction* appeared because the computer-scientific research question was initially oriented toward testing whether the experimental material produced the expected participant responses, while the social-scientific perspective shifted to focus what kind of evidence the participant responses could actually provide: a descriptive pattern, a valid measurement, a reliable comparison between conditions, or a broader interpretive claim.

The consultation team consisted of two data scientists combined social-scientific and mathematical expertise. The mathematician was primarily involved for statistical support, but their disciplinary proximity to the research team also helped bridge communication gaps, while the social-scientific perspective clarified questions of measurement and interpretation. The consultation thus functioned as a trading zone in which technical terminology, statistical reasoning, and social-scientific measurement logic could be aligned.

6 Implications for CSS and the Data and Digital Science Community

The framework developed in this paper suggests that the central challenge of CSS is not only to combine social-scientific questions with computational methods, but to make the conditions of this combination explicit. Computational methods enter social research together with epistemic assumptions, quality standards, data practices, infrastructures, and responsibilities. The five friction points show where these assumptions become practically relevant: when predictive performance is mistaken for explanation, when labels are treated as constructs, when features are interpreted as mechanisms, when technical metrics are equated with research quality, or when openness and reproducibility meet legal, ethical, and infrastructural constraints.

For **CSS as a research field**, this firstly implies that **interactional expertise should be treated as a core methodological competence**. Researchers need enough cross-domain understanding to ask informed questions, recognize limitations, and negotiate methodological decisions. Second, **friction points should be addressed early in the research process**. Questions of construct validity, operationalization, data quality, documentation, accountability, and interpretation should be built into project design, data workflows, and collaboration agreements, rather than postponed until analysis or publication. Third, **shared materials such as codebooks, annotation guidelines, metadata schemes, data management plans and documentation should be used deliberately** as they help stabilize communication across fields and make methodological decisions transparent.

For the **Data and Digital Science community**, the argument highlights the strategic role of support structures for data-intensive research: Data Science Centers, Data Competence

Centers, and related initiatives create institutional spaces where epistemic paradigms, research standards, methods, and data governance can be aligned. This also has consequences for how support structures need to be designed and evaluated. If their role is translation, their **success cannot be measured only by the number of workshops, consultations, or tools provided** [St25]. It rather depends on whether they help researchers make assumptions explicit, choose appropriate methods, document decisions and handle data responsibly. Support teams therefore need **interdisciplinary staffing, stable career paths, sustainable institutional anchoring, and time for mutual learning** [Bo25]. Short-term or purely project-based funding risks undermining the accumulated interactional expertise that makes such units effective. At the same time, **support structures cannot replace disciplinary training or broader institutional reform**. CSS still requires stronger computational training in the social sciences, stronger social-scientific literacy in computational and STEM fields, and university structures that reward interdisciplinary work rather than treating it as additional labor [La20].

This paper also has limitations. The proposed framework is conceptual and practice-informed rather than empirically validated across a broad range of institutions. The two examples are drawn from one interdisciplinary help-desk and are necessarily selective and anonymized. They illustrate how translation work can take place in everyday consultation, but they do not allow general claims about the effectiveness of support structures across the Data and Digital Science landscape. Future research should therefore examine how different support structures organize translation work across disciplines and institutional contexts, and how the impact of such work can be assessed beyond simple service metrics.

Acknowledgments

The work was prepared as part of the DataNord project, funded under grant number 16DKZ2026A by the BMFTR and financed by the European Union – NextGenerationEU. The authors would like to thank all partners, collaborators, and supporters of DataNord.

References

[AA21] Amaturo, E.; Aragona, B.: Digital Methods and the Evolution of the Epistemology of Social Sciences. In (Mariani, P.; Zenga, M., eds.): Data Science and Social Research II. Series Title: Studies in Classification, Data Analysis, and Knowledge Organization, Springer International Publishing, Cham, pp. 1–8, 2021, http://link.springer.com/10.1007/978-3-030-51222-4_1, accessed: 04/22/2026.

[AM23] Abbas, R.; Michael, K.: Socio-Technical Theory: A review. In (Papagiannidis, S., ed.): TheoryHub Book. 2023.

[Ap97] Appelbaum, S. H.: Socio-technical systems theory: an intervention strategy for organizational development. Management Decision 35 (6), pp. 452–463, 1997, http://www.emerald.com/md/article/35/6/452-463/285914, accessed: 04/21/2026.

[Ba24] Bail, C. A.: Can Generative AI improve social science? Proceedings of the National Academy of Sciences 121 (21), e2314021121, 2024, https://pnas.org/doi/10.1073/pnas.2314021121, accessed: 04/10/2026.

[BH09] Bauer, J. M.; Herder, P. M.: Designing Socio-Technical Systems. In: Philosophy of Technology and Engineering Sciences. Elsevier, pp. 601–630, 2009, https://linkinghub.elsevier.com/retrieve/pii/B9780444516671500264, accessed: 04/21/2026.

[BH77] Bostrom, R. P.; Heinen, J. S.: MIS Problems and Failures: A Socio- Technical Perspective: Part I: The Causes. Management Information Systems Quarterly 1 (3), _eprint: https://misq.umn.edu/misq/article-pdf/1/3/17/730/4_bostrom.pdf, pp. 17–32, 1977, https://doi.org/10.2307/248710.

[BMY16] Borge-Holthoefer, J.; Moreno, Y.; Yasseri, T.: Editorial: At the Crossroads: Lessons and Challenges in Computational Social Science. Frontiers in Physics 4, 2016, http://journal.frontiersin.org/Article/10.3389/fphy.2016.00037/abstract, accessed: 04/22/2026.

[Bo25] Bockermann, C. et al.: Positionspapier zur Zukunft der Data- und Digital-Science-Community, Version Number: V1, Zenodo, 2025, https://zenodo.org/doi/10.5281/zenodo.17475528, accessed: 04/30/2026.

[Br19] Brady, H. E.: The Challenge of Big Data and Data Science. Annual Review of Political Science 22 (1), pp. 297–323, 2019, https://www.annualreviews.org/doi/10.1146/annurev-polisci-090216-023229, accessed: 04/22/2026.

[Ca20] Carroll, S. R. et al.: The CARE Principles for Indigenous Data Governance. Data Science Journal 19, p. 43, 2020, http://datascience.codata.org/articles/10.5334/dsj-2020-043/, accessed: 04/29/2026.

[CE07] Collins, H.; Evans, R.: Rethinking Expertise. University of Chicago Press, 2007.

[Ch21] Chen, Y. et al.: Social prediction: a new research paradigm based on machine learning. The Journal of Chinese Sociology 8 (1), p. 15, 2021, https://journalofchinesesociology.springeropen.com/articles/10.1186/s40711-021-00152-z, accessed: 04/22/2026.

[De24] Dehne, J. et al.: Die Vermittlung von Best Practices zur Messung von Datenqualität in den quantitativen Sozialwissenschaften. ISBN: 9783885797463, 2024, https://dl.gi.de/handle/20.500.12116/45173, accessed: 04/30/2026.

[DHC20] Diaz-Bone, R.; Horvath, K.; Cappel, V.: Sozialforschung in den Zeiten von Big Data. Die Herausforderungen der neuen Datenwelten und die Notwendigkeit einer Soziologie der SozialforschungSocial Research in Times of Big Data. The Challenges of New Data Worlds and the Need for a Sociology of Social Research. Historical Social Research 45, Version Number: 1.0, p. 314341, 2020, https://www.ssoar.info/ssoar/handle/document/68198, accessed: 04/24/2026.

[Ed11] Edwards, P. N. et al.: Science friction: Data, metadata, and collaboration. Social Studies of Science 41 (5), pp. 667–690, 2011, https://journals.sagepub.com/doi/10.1177/0306312711413314, accessed: 04/24/2026.

[Ed20] Edelmann, A. et al.: Computational Social Science and Sociology. Annual Review of Sociology 46 (1), pp. 61–81, 2020, https://www.annualreviews.org/doi/10.1146/annurev-soc-121919-054621, accessed: 04/21/2026.

[Ga96] Galison, P.: Computer Simulations and the Trading Zone. In (Galison, P. L.; Stump, D. J., eds.): The Disunity of science: boundaries, contexts, and power. Stanford University Press, pp. 118–157, 1996.

[GBK25] Götze, L.; Berger, M.; Korb-King, F.: Übersicht der BMBF-geförderten Projekte zum Aufbau von Datenkompetenzzentren, 2025, https://doi.org/10.5281/zenodo.14883999.

[GRS21] Grimmer, J.; Roberts, M. E.; Stewart, B. M.: Machine Learning for Social Science: An Agnostic Approach. Annual Review of Political Science 24 (1), pp. 395–419, 2021, https://www.annualreviews.org/doi/10.1146/annurev-polisci-053119-015921, accessed: 04/10/2026.

[Hä19] Häußling, R.: Zur Erklärungsarmut von Big Social Data. Von den Schwierigkeiten, auf Basis von Big Social Data eine Erklärende Soziologie betreiben zu wollen. In (Baron, D.; Arránz Becker, O.; Lois, D., eds.): Erklärende Soziologie und soziale Praxis. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 73–100, 2019, http://link.springer.com/10.1007/978-3-658-23759-2_5, accessed: 04/22/2026.

[Ho21] Hofman, J. M. et al.: Integrating explanation and prediction in computational social science. Nature 595 (7866), pp. 181–188, 2021, https://www.nature.com/articles/s41586-021-03659-0, accessed: 04/24/2026.

[Hö21] Hörner, T. et al.: Disziplinübergreifendes Modell zur Ausbildung von Forschungsdatenmanagement und Data Science Kompetenzen: „Data Train – Training in Research Data Management and Data Science“. Bausteine Forschungsdatenmanagement (3), pp. 56–69, 2021, https://bausteine-fdm.de/article/view/8343.

[JKH21] Jarvis, B. F.; Keuschnigg, M.; Hedström, P.: Analytical sociology amidst a computational social science revolution. In: Handbook of Computational Social Science, Volume 1. 1st ed., Routledge, London, pp. 33–52, 2021, https://www.taylorfrancis.com/books/9781003024583/chapters/10.4324/9781003024583-4, accessed: 04/24/2026.

[JP23] Jungherr, A.; Posegga, O.: Computational Social Science. In (Kersting, N.; Radtke, J.; Baringhorst, S., eds.): Handbuch Digitalisierung und politische Beteiligung. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 1–17, 2023, https://link.springer.com/10.1007/978-3-658-31480-4_54-1, accessed: 04/21/2026.

[Ka22] Kaminsky, J.: Editorial. Theory applied to informatics: Socio-Technical Theory. Canadial Journal of Nursing Informatics 3-4 (17), 2022.

[Ke21] Kemman, M.: Trading Zones of Digital History. De Gruyter, 2021.

[KHB24] Knöpfle, P.; Haim, M.; Breuer, J.: Key topic or bare necessity? How Research Ethics are Addressed and Discussed in Computational Communication Science. Publizistik 69 (3), pp. 333–356, 2024, https://link.springer.com/10.1007/s11616-024-00846-7, accessed: 04/22/2026.

[Ki14] Kitchin, R.: Big Data, new epistemologies and paradigm shifts. Big Data & Society 1 (1), p. 2053951714528481, 2014, https://journals.sagepub.com/doi/10.1177/2053951714528481, accessed: 04/22/2026.

[La09] Lazer, D. et al.: Computational Social Science. Science 323 (5915), pp. 721–723, 2009, https://www.science.org/doi/10.1126/science.1167742, accessed: 04/22/2026.

[La20] Lazer, D. M. J. et al.: Computational social science: Obstacles and opportunities. Science 369 (6507), pp. 1060–1062, 2020, https://www.science.org/doi/10.1126/science.aaz8170, accessed: 04/21/2026.

[LBJ22] Lundberg, I.; Brand, J. E.; Jeon, N.: Researcher reasoning meets computational capacity: Machine learning for social science. Social Science Research 108, p. 102807, 2022, https://linkinghub.elsevier.com/retrieve/pii/S0049089X22001181, accessed: 04/10/2026.

[Le23] Leslie, D.: The Ethics of Computational Social Science. In (Bertoni, E. et al., eds.): Handbook of Computational Social Science for Policy. Springer International Publishing, Cham, pp. 57–104, 2023, https://link.springer.com/10.1007/978-3-031-16624-2_4, accessed: 04/28/2026.

[LM24] Láng, B.; Megyesi, B.: An STS analysis of a digital humanities collaboration: trading zones, boundary objects, and interactional expertise in the DECRYPT project. Humanities and Social Sciences Communications 11 (1), p. 618, 2024, https://www.nature.com/articles/s41599-024-03135-w, accessed: 04/21/2026.

[LPW23] Leitgöb, H.; Prandner, D.; Wolbring, T.: Editorial: Big data and machine learning in sociology. Frontiers in Sociology 8, p. 1173155, 2023, https://www.frontiersin.org/articles/10.3389/fsoc.2023.1173155/full, accessed: 04/21/2026.

[Pa15] Pasquale, F.: The black box society: the secret algorithms that control money and information. Harvard University Press, Cambridge, 2015.

[Pf24] Pfuhl, H. et al.: Aufbau einer überregionalen Data-Science-Community. ISBN: 9783885797463, 2024, https://dl.gi.de/handle/20.500.12116/45170, accessed: 04/28/2026.

[Ri19] Ribes, D.: STS, Meet Data Science, Once Again. Science, Technology, & Human Values 44 (3), pp. 514–539, 2019, https://journals.sagepub.com/doi/10.1177/0162243918798899, accessed: 04/21/2026.

[RJ20] Radford, J.; Joseph, K.: Theory In, Theory Out: The Uses of Social Theory in Machine Learning for Social Science. Frontiers in Big Data 3, p. 18, 2020, https://www.frontiersin.org/article/10.3389/fdata.2020.00018/full, accessed: 04/24/2026.

[Ru13] Ruppert, E.: Rethinking empirical social sciences. Dialogues in Human Geography 3 (3), pp. 268–273, 2013, https://journals.sagepub.com/doi/10.1177/2043820613514321, accessed: 04/24/2026.

[Sa18] Salganik, M. J.: Bit by bit: social research in the digital age. Princeton University Press, Princeton, 2018.

[Sc26] Schaare, H. L. et al.: Meet the Northwest Data Space, 2026, https://doi.org/10.5281/zenodo.18953220.

[SG89] Star, S. L.; Griesemer, J. R.: Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social Studies of Science 19 (3), pp. 387–420, 1989, https://journals.sagepub.com/doi/10.1177/030631289019003001, accessed: 04/28/2026.

[Sm20] Smith, P. L. et al.: Building Socio-technical Systems to Support Data Management and Digital Scholarship in the Social Sciences. In (Crowder, J. W. et al., eds.): Anthropological Data in the Digital Age. Springer International Publishing, Cham, pp. 31–57, 2020, http://link.springer.com/10.1007/978-3-030-24925-0_3, accessed: 04/21/2026.

[St23a] Steinmann, L. et al.: Das Data Science Center an der Universität Bremen: Interdisziplinärer Knotenpunkt und Service-Infrastruktur für die datenintensive Forschung. In: E-Science-Tage 2023. heiBOOKS, 2023, https://books.ub.uni-heidelberg.de//heibooks/catalog/book/1288/chapter/18083, accessed: 04/30/2026.

[St23b] Steinmann, L. et al.: Workshop: „Aktuelle Entwicklungen und Perspektiven (an Hochschulen) im Bereich Data Science". ISBN: 9783885797319, 2023, https://dl.gi.de/handle/20.500.12116/43170, accessed: 04/30/2026.

[St25] Steinmann, L. et al.: Bremen's Interdisciplinary Data Competence Center DataNord: An Integrative Approach for Data Literacy in Research. In: E-Science-Tage 2025. heiBOOKS, 2025, https://books.ub.uni-heidelberg.de//heibooks/catalog/book/1652/chapter/23928, accessed: 04/28/2026.

[TB51] Trist, E. L.; Bamforth, K. W.: Some Social and Psychological Consequences of the Longwall Method of Coal-Getting: An Examination of the Psychological Situation and Defences of a Work Group in Relation to the Social Structure and Technological Content of the Work System. Human Relations 4 (1), pp. 3–38, 1951, https://journals.sagepub.com/doi/10.1177/001872675100400101, accessed: 04/28/2026.

[Tr81] Trist, E.: The Evolution of Socio-Technical Systems: A Conceptual Framework and an Action Research Program. 1981.

[VBH17] Van Der Aalst, W. M. P.; Bichler, M.; Heinzl, A.: Responsible Data Science. Business & Information Systems Engineering 59 (5), pp. 311–313, 2017, http://link.springer.com/10.1007/s12599-017-0487-z, accessed: 04/29/2026.

[Wi16] Wilkinson, M. D. et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1), p. 160018, 2016, https://www.nature.com/articles/sdata201618, accessed: 04/29/2026.