Navigation

Document Purpose

This document is to be used with the online CSEA Data Mapping Visualization, the online CSEA Barriers and Challenges, and the online CSEA Data Ecosystem Roadmap. It aims to improve the data literacy of stakeholders working within online CSEA and to provide inspiration from other systems of how online CSEA data barriers and challenges might be overcome.

Together, the four documents act as a foundation to simplify the complexity of the system and enable forward progress towards an ideal state.

Good Practices Bank

This document provides a comprehensive overview of key concepts and best practices in data management, essential for the online CSEA ecosystem aiming to improve its data use. It builds upon the previous work conducted by Safe Online and partners as part of the '*Data for Change*' initiative.

It should be read alongside three associated documents:

1. Online CSEA Data Mapping Visualization
2. Online CSEA Challenges and Barriers
3. Online CSEA Data Ecosystem Roadmap

Together, these four documents provide a simple and broad overview of the state of the online CSEA data ecosystem and a forward view on how to improve the use of data for better outcomes for children, survivors, and society.

Improving system-wide data assets relies on each organization implementing core data knowledge areas well. The core data knowledges areas included in this document are adapted from DAMA guidelines[1], a widely recognized standard across the globe.

[1] DAMA International's Guide to the Data Management Body of Knowledge (DMBOK); https://technicspub.com/dmbok/

Each section explores a critical component of data management, beginning with **Data Governance**, which outlines the policies, roles, and responsibilities that ensure effective data management across the ecosystem. The remaining sections will emphasize how organizational implementation of good data management benefits systemwide data governance.

Data Architecture is examined to describe how data is structured, stored, and accessed, while **Data Security** and **Data Privacy** focus on safeguarding data against unauthorized access and ensuring compliance with regulatory requirements. **Data Resilience** addresses the strategies for ensuring data availability and recovery in the event of disruptions, and **Data Sharing** discusses the mechanisms for securely exchanging data within and between organizations.

The document also covers **Data Ethics**, emphasizing the importance of responsible data use, **Data Integration and Interoperability**, which highlights the need for seamless data exchange across systems, and **Data Analytics**, which provides insights into extracting actionable knowledge from data.

Additionally, the importance of **Data Training and Awareness** is discussed to ensure that all stakeholders are equipped to manage and use data effectively, and **Performance Monitoring and Reporting** is included to guide organizations in tracking and optimizing data management practices. Real-world examples are provided throughout to illustrate best practices and practical applications in each area.

This document is intended to be used by all those connected to online CSEA through data. It is not a complete and exhaustive guide to data management, but an accessible, broad, and relevant overview of good data practices that can be implemented across the online CSEA ecosystem.

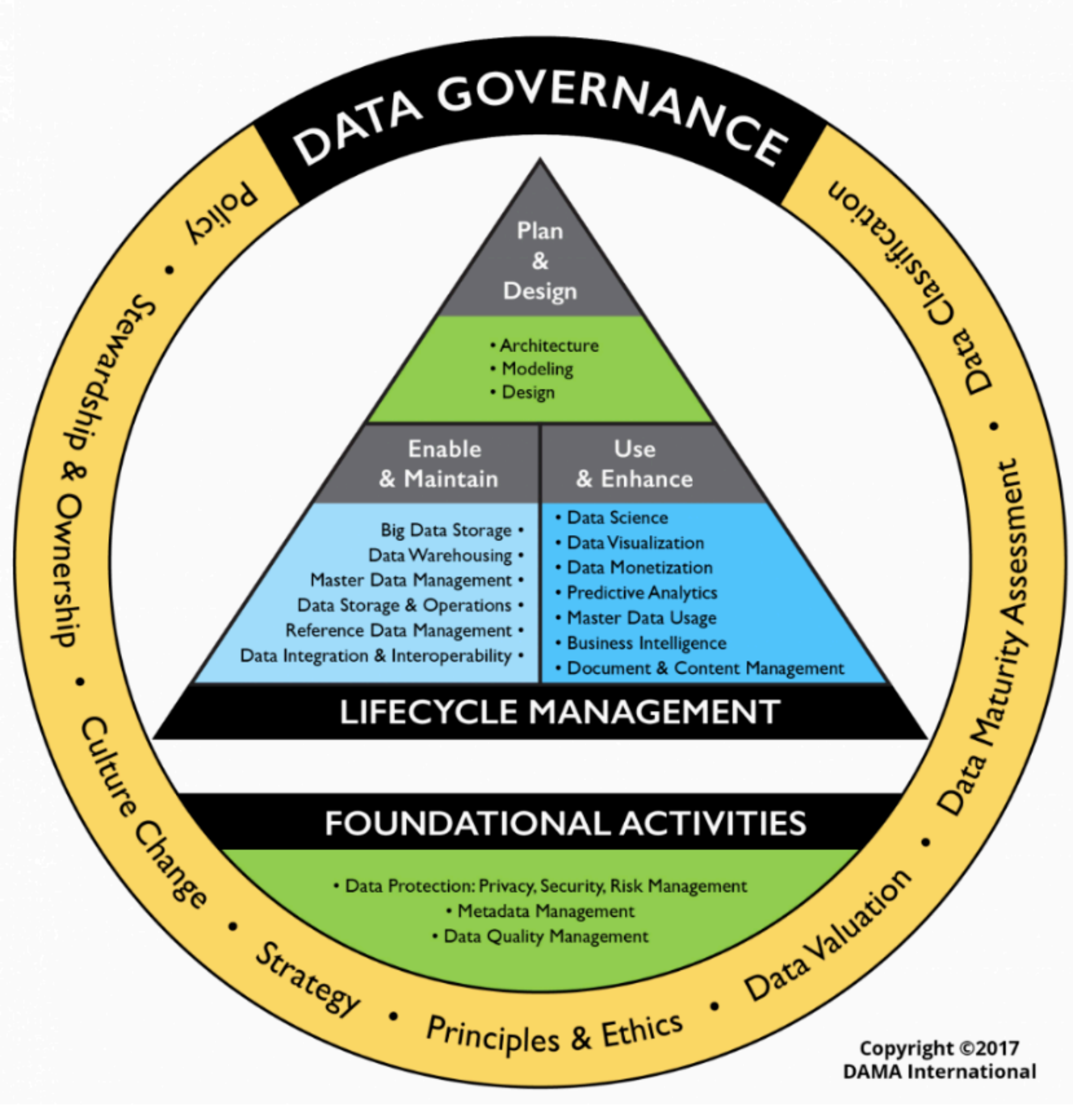


Figure 1: The sections in this report have been adapted from DAMA's Data Management Wheel Evolved

Data Governance

Data governance involves creating and enforcing norms, rules, and infrastructure that build trust by guiding how individuals, organizations, and governments collect, use, and share data responsibly, ensuring sustainable and equitable value creation.

Data Governance Overview

A well-designed data governance framework enables stakeholders to fully harness the economic and social value of both public and private intent data, while leveraging synergies between them. It builds trust in the integrity of the data system and ensures that the benefits are shared equitably. It ensures that infrastructure, laws, economic policies, and institutions work together to support the use of data in a way that aligns with stakeholder values, whilst respecting individual rights.[2]

Although typically considered at an organizational or national level, international multi-stakeholder Data Governance is essential to overcome the global, multi-faceted, pervasive threat of online CSEA.

It has been implemented successfully in other systems, some of which are explored in greater detail in this document.

What is Data Governance?

The following list will explore each element associated with Data Governance shown in Figure 1, to demonstrate **what this means at an ecosystem level**.

Policy: Establish consistent global standards, frameworks, and regulations that guide how data is collected, shared, and used across borders. This includes ensuring data privacy and protection, data quality, or storage standards. It also involves aligning policies to address challenges like data sovereignty, equitable access, recognizing bias, and the ethical use of data. Ultimately, reducing the transactional costs through harmonization and removing barriers to interoperability.

Examples: W3C guidelines for data category vocabularies[3],

Stewardship and ownership: Create clear guidelines on data ownership, determining who has the right to collect, use, and benefit from data, and ensuring that the rights of all stakeholders—including individuals, organizations, and governments—are respected.

At a system level, it promotes shared stewardship models, encouraging collaboration and accountability while ensuring that data is treated as a global public good. Additionally, it works to balance national or organizational interests with international cooperation,

[2] World Development Report 2021, The World Bank (World Development Report 2021: Data for Better Lives | The report (worldbank.org))

[3] https://www.w3.org/TR/vocab-dcat-2/

ensuring that data flows across borders while safeguarding privacy, security, and human rights.

Examples: Better Deal for Data[4], *The Data Economy Lab*[5]

Culture change: Driving culture change by fostering a shift towards responsible, ethical, and transparent data practices. It helps establish norms and values that prioritize data privacy, security, and fairness across borders, encouraging countries, organizations, and individuals to adopt a more unified approach to data management. By promoting education, awareness, and collaboration around data ethics, governance frameworks can influence a culture where trust, accountability, and equity are central to data use and sharing.

This cultural shift involves breaking down silos between sectors and regions, encouraging openness to data sharing for collective benefits while respecting human rights and individual privacy. Data governance also helps cultivate a mindset where stakeholders view data as a shared resource that must be handled responsibly, promoting global cooperation, and embedding these values into policies, systems, and behaviors worldwide.

Examples: FAIR Principles[6], *innovateUS*[7], *Data-Pop Alliance: Mobilize*[8]

Strategy: Create a cohesive, long-term framework that aligns global data practices with overarching goals for responding to online CSEA. This involves developing a unified vision for how data should be managed and utilized across borders, setting clear objectives for cross-national collaboration, and promoting the responsible use of data.

Examples: The Data Values Project[9], *System Stewardship*[10]

Principles and ethics: Create a common ethical framework that all countries and organizations can adhere to, ensuring that data practices are transparent, fair, and respectful of individual rights, such as privacy and consent. Governance should represent the voices and perspective of a diverse range of stakeholders, including marginalized populations, children and youth.

Focusing on principles like data justice, inclusivity, and accountability, ensuring that marginalized or vulnerable populations are not exploited or excluded in the global data economy. It also promotes ethical data use by addressing biases in data systems and ensuring that algorithms and technologies used for data processing uphold fairness and equity.

[4] https://bd4d.org/
[5] https://thedataeconomylab.com/
[6] https://www.go-fair.org/fair-principles/
[7] https://innovate-us.org/
[8] https://datapopalliance.org/mobilize/
[9] https://www.data4sdgs.org/initiatives/data-values-project
[10] https://collaboratecic.com/wp-content/uploads/2023/01/Systems-stewardship-resource-Dec-22.pdf

Examples: Ada Lovelace Institute: Biometrics Council[11], *United Nations: Data Privacy, Ethics and Protection*[12]

Data valuation: Establish frameworks for valuing and verifying data consistently across different nations and sectors, recognizing its economic, social, and strategic importance. It would define criteria for assessing the value of data assets based on factors like accuracy, relevance, and potential impact. Finally, it would strive to distribute the benefits of data proportionally across stakeholders, geographies, and groups.

Data maturity assessment: Create global benchmarks for assessing the maturity of data systems, ensuring that stakeholders can evaluate their capabilities in managing, processing, and leveraging data. It fosters a common understanding of data quality, infrastructure, and governance practices, promoting capacity-building and harmonization across regions.

Examples: UK His Majesty's Government: Cabinet Office[13]

Data classification: Standardize data classification systems to ensure consistency and interoperability between countries and sectors. It would also include the promotion of universal classification schemas or promoting standardized definitions and methodology associated with measuring abuse.

Examples: INHOPE: Global Standard Project[14]

The differences of Data Governance at an organizational, national, or international level is further highlighted in Figure 2. This draws on methodology that splits Data Governance responsibilities into those relating to: *infrastructure, laws and regulations, economic policies, and institutions.*[15]

[11] https://www.adalovelaceinstitute.org/project/citizens-biometrics-council/

[12] https://unsdg.un.org/sites/default/files/UNDG_BigData_final_web.pdf

[13] https://assets.publishing.service.gov.uk/media/64184bccd3bf7f79d9675dbd/Data_Maturity_Assessment_for_Government_-_FINAL_PDF.pdf

[14] https://www.inhope.org/EN/articles/global-standard-project-ontology-launch

[15] World Development Report 2021, The World Bank (World Development Report 2021: Data for Better Lives | The report (worldbank.org))

	Organizational	National	International
Infrastructure policies	● Developing organizational data infrastructure to exchange, store, and process online CSEA data	● Universal coverage of specialist online CSEA officers ● Domestic infrastructure to exchange, store, and process online CSEA data	● Global technical standards for data storage solutions ● Regional collaboration on data infrastructure to achieve scale
Laws and regulations	● Adherence to relevant regulations and laws ● Input into national consultations to improve effectiveness and fairness of laws and regulation	● Safeguards to secure and protect data from the threat of misuse ● Robust data protection regulations and laws.	● Conventions for collaboration on tackling online CSEA ● Interoperability standards to facilitate cross-border data sharing
Economic policies	● Developing innovations with good data practice integrated at outset ● Developing a data strategy to support organizational strategy ● Reduce financial risk associated with data	● Antitrust policies for data platform businesses ● Influencing innovation in data-enabled services ● Taxation of data platform businesses	● Demonstrating the economic value in data sharing ● Championing equitable access for data for LMICs
Institutions	● Achieve relevant industry standards in relation to data management	● Government entities to oversee, regulate and secure data ● Other stakeholders to set standards and increase data access or reuse	● International organizations to support collaboration on data governance and promote standardization ● Cooperation on cross-border regulatory and enforcement issues.

Figure 2: Example differences in governance across organizational, national, and international system views

Governing a global network used to securely transmit information for financial transactions.

The **Society for Worldwide Interbank Financial Telecommunication (SWIFT)** is a global messaging network that facilitates secure and reliable financial transactions between banks and other financial institutions. Broadly, the steps that led to its creation include:

- **Identifying the requirement:** Systems for transferring funds between banks were non-standardized, inefficient, slow, and prone to errors. A standardized, secure, automated messaging system would improve this.
- **Form a network**: Commitment from organizations to developing shared infrastructure governed by a recognized body.
- **Develop standards**: Develop a standardized messaging format, based on international standards that apply to all organizations involved.
- **Gain adoption and expand**: Expand beyond the original founding organizations to become the default standard for interbank communications.
- **Continued governance**: Developing regular updates to remain innovative, maintain regulatory compliance, remain aligned to the needs of stakeholders, and foster collaboration and knowledge sharing.

What does this mean for the online CSEA ecosystem?

This demonstrates how establishing a committed network of organizations is crucial for developing shared infrastructure governed by a recognized body. Standardized messaging and data formats based on international standards ensure consistency across all involved organizations. Any potential solution should start small with a view to scale, with continued governance ensuring ongoing innovation, regulatory compliance, and alignment with stakeholder needs.

Why implement Data Governance?

In simplistic terms, good ecosystem Data Governance seeks to resolve the trilemma of value, trust, and equity between stakeholders and their data.

> *How can people trust that their data will be protected and that they will share in the value that data can produce?*
>
> World Development Report 2021

Data Governance creates the conditions for agreement among participants in the process of collecting, storing, sharing, or using data that fosters trust that they will not be harmed from exchanging data and that part of the value created by data will accrue equitably.[16]

The relationship between Data Governance, other Data Management activities, and value, trust and equity are demonstrated through Figure 3.

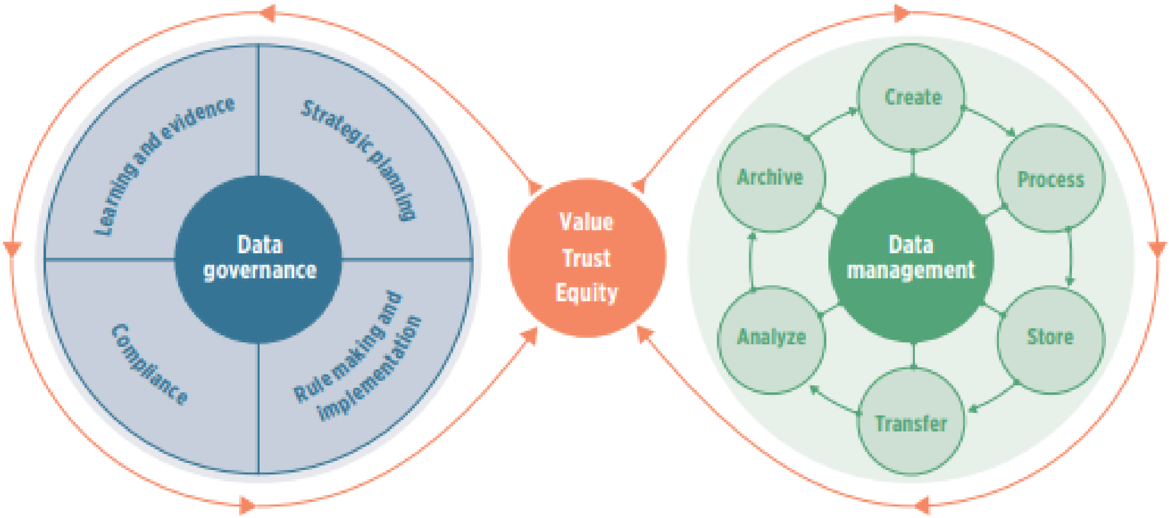


Figure 3: Data Governance is a critical component for establishing value, trust, and equity within a data ecosystem.[17]

Furthermore, as the threat of online CSEA becomes more complex, as do the systems we rely on to respond. Increasingly complex response systems have collaboration, trust, and interoperability as integral components, resulting in governance becoming essential. Stakeholders are becoming more reliant on computational support to deal with data because of the increase in volume, complexity, and creation speed of data. Strong Data Governance improves resilience to this increased reliance, one way it might do that is through the promotion of good data principles such as FAIR Principles[18], shown in Figure 4.

[16] Ibid
[17] Ibid
[18] https://www.go-fair.org/fair-principles/

Findable

- (Meta)data are assigned a globally unique and persistent identifier
- Data are described with rich metadata
- Metadata clearly and explicitly include the identifier of the data they describe
- (Meta)data are registered or indexed in a searchable resource

Accessible

- (Meta)data are retrievable by their identifier using a standardized communications protocol
  - The protocol is open, free, and universally implementable
  - The protocol allows for an authentication and authorization procedure, where necessary
- Metadata are accessible, even when the data are no longer available

Interoperable

- (Meta)data use a formal accessible, shared, and broadly applicable language for knowledge representation
- (Meta)data use vocabularies that follow FAIR principles
- (Meta)data include qualified references to other (meta)data

Reusable

- (Meta)data are richly described with a plurality of accurate and relevant attributes
  - (Meta)data are released with a clear and accessible data usage license
  - (Meta)data are associated with detailed provenance
  - (Meta)data meet domain-relevant community standards

Figure 4: A summary of the FAIR Principles

How to begin the Data Governance Journey

The following steps help establish a governance framework that aligns with international regulations to ensure consistent data management practices. It enhances data security, with self-assessment tools and regular compliance checks. The child-centric model prioritizes individual control over their data, supported by transparency and clear communication. It improves data integration and interoperability through standardized frameworks allowing for seamless data sharing across systems. Finally, system governance can mandate additional measures like secure data sharing, ethics oversight, and national performance monitoring to ensure robust and ethical data practices.

Special consideration should be given to the participation of low and middle-income countries (LMICs), who have more barriers to overcome in terms of technical capacity and resources. Neglecting their voice creates a disparity that impacts both the inclusiveness nature of the process and the quality of its outcomes.

Harnessing local knowledge to improve argi-food data, development, and prosperity

Senegal's Agridata project emphasizes the importance of a coordinated approach to strengthen national data and statistics systems, which can support local knowledge and capacity development in agriculture. It highlights that reliable agri-food data is essential for informed decision-making and to address key challenges like food security, environmental sustainability, and socio-economic growth. The report underscores how innovations in data collection, especially through digital tools, can fill critical gaps and improve the effectiveness of food systems.

Key challenges included:

- **Lack of investment in data systems**, particularly in LMICs, hampers the collection of timely and reliable data.
- **Digitalization barriers** include limited access to technology and insufficient technical capacity.
- **Misalignment of data and policy needs**, where the data produced often does not meet decision-makers' demands.
- **Inconsistent collaboration**, often hindered by a lack of coordination among stakeholders and inadequate feedback mechanisms.

Key drivers for effective system wide collaboration:

- **Clear strategies and roadmaps** with well-defined roles for stakeholders.
- **Capacity-building initiatives** that improve technical skills and data literacy.
- **Innovative technology adoption** to facilitate better data sharing and analysis.

- **Strong multi-stakeholder platforms** that promote dialogue and trust across sectors.
- **Clearly defined responsibilities:** The table below highlights the roles and responsibilities of stakeholders involved in producing agri-food systems data:

National governments	• Provide the enabling environment for data collaboration, such as leadership in developing a national data strategy, guidelines for data sharing and use, directing organizations to lead on data collection and production, and prioritizing areas for collaboration. • Ensure sufficient budget is available and increase investments in surveys and other data production methods. • Provide support in areas such as standards, data anonymization, and security together with NSOs (see below). • Ensure a robust and up-to-date technological infrastructure is available (internet access, data warehouses).
Ministries of agriculture	• Plan data collection needs and activities in collaboration with all relevant stakeholders. • Collaborate closely with the NSOs to ensure the data meets users' needs. • Coordinate among producers and users of data
Ministries of environment	• Produce data and analysis on sustainability and on the impact of agriculture and food systems on the environment
National statistical offices (NSOs)	• Work with ministries to provide common methodologies, concepts, and definitions for collecting indicators specific to policymakers' and other stakeholders' needs. • Coordinate with other stakeholders to produce statistics. • Define standards and set benchmarks for what constitutes good statistical procedures and help other bodies to collect data that meets international standards and identify capacity gaps. • Ensure that country master sampling frames (surveys and census) are available and accessible.
Private-sector actors	• Create the ICT tool to enable easy data collection and sharing.

	• Engage in all the steps of the statistical value chain, from production to dissemination. • Strengthen alignment with government standards so that ICT products and services are usable by governments and fit for purpose. • Whenever possible, share their data with public authorities.
Academia	• Collect data from all stakeholders to carry out analysis and leverage available knowledge. • Provide theoretical knowledge and models which can be used to analyze data as well as knowledge of specific contexts.
Donors and international organizations	• Support governments in implementing data strategies and align priorities with government investments and strategies. • Support in-country data collaboration. • Support investment in digital infrastructure. • Support institutions that generate agriculture and food systems data. • Provide methodological and technological support for capacity building. • Encourage governments to invest additional resources in the production of agricultural statistics. • Ensure coordination among development partners.
Civil society networks and organizations	• Maintain the production of ground truth data (citizen-generated data) and hold governments accountable for the data they produce and publish. • Ensure adequate policy and legal frameworks exist to protect citizens' rights. • Use data provided by official sources and provide feedback to improve its accuracy and utility. • Support capacity building and use of data for their constituencies. • Carry out research and test new approaches for data collection and analysis (as entities that are less risk-averse than governments). • Provide support in areas such as experimentation and interoperability.

What does this mean for the online CSEA ecosystem?

In systems with similar challenges, key ingredients for implementing change include: a strategy and roadmap with clear roles and responsibilities, putting children at the heart of the ecosystem, understanding stakeholder needs, investments in capacity building, and sufficient and suitable technological infrastructure. Adapting their placemat for assigning roles and responsibilities between stakeholders could help clarify how the system works, improving collaboration. These ingredients can be effectively applied to LMICs but requires additional consideration to the financial resources available to secure a generation of more and better online CSEA data.

At a system level:

1. **Establish a governance framework:** Develop policies outlining data management, including roles and responsibilities, aligned to international regulations, setting foundation for consistent data practices.
2. **Enhance data security and protection:** Introduce self-assessment toolkits for stakeholders to assess their data protection performance, mandating compliance with regular assessments.
3. **Put children at the heart of the system: I**mplemented models to ensure individuals (e.g. children) have control over how their data is used, including how consent is collected. Initiate transparency and communication initiatives to further center children within the system, ensuring their voices are considered within the governance framework, and feedback loops are established to ensure sustainable engagement.
4. **Improve data integration and interoperability:** Integrate systems across stakeholders to allow for seamless data sharing, in addition to the adoption of interoperability standards.
5. **Implement additional measures from a central governance team:** These include data security measures, secure data sharing protocols (including those with third-party organizations), collaborative research initiatives, ethics committees, regulatory committees, national performance monitoring and reporting.

An International Decade of Data (IDD)

The call for an International Decade of Data (IDD) emphasizes the crucial role of data in addressing global challenges. As data underpins modern technologies like AI, it is essential to develop a strong foundation of data stewardship, accessibility, and governance to bridge global data gaps, ensure equitable use, and drive socioeconomic progress. The IDD aims to position data as central to global development and foster responsible, inclusive, and sustainable data practices.

Key recommendations include:

- **Uphold human rights in data governance**, prioritizing privacy, transparency, and ethical data usage.
- **Improve data access and sharing** during crises through pre-established frameworks.
- **Professionalize data stewardship** to ensure responsible, effective data collaboration.
- **Leverage innovations** like synthetic data and privacy-enhancing technologies to improve accessibility and inclusivity.
- **Empower individuals and communities** through digital self-determination, ensuring fair and accountable data use globally.

What does this mean for the online CSEA ecosystem?

The call for data to be governed at a system level, reframing who benefits from data, and how individual rights are respected is not just limited to online CSEA. There are initiatives, research, and funds accessible focused on creating this change across sectors that could be applicable to change within online CSEA.

At an organizational level:

Strong systemwide Data Governance will support individual organizations in establishing their own Data Governance.

Implementing data governance needs to start with an understanding of the data needs and aspirations of the organization with the view of getting value from the data assets.

1. Define the organization's Data Strategy and, if required, perform a readiness assessment[19] to understand the gaps between the current and desired governance landscape.
2. Align this to any requirements stipulated at a systemwide governance level aimed at increasing interoperability and harmonization.

Depending on the Data Strategy and the outcome of the readiness assessment, if performed, the following areas may need to be considered:

3. Develop data-related policies and procedures including (but not limited to) information security, data quality, acceptable data usage, metadata management.
4. Develop and implement Data Architecture and Data Quality standards.
5. Ensure that regulatory compliance requirements are understood and met.
6. Ensure that appropriate data governance roles and responsibilities are defined so that appropriate oversight, e.g. audit and assurance, are in place.
7. Put in place a data governance framework to manage data-related risks and issues and provide escalation routes.
8. Set the standards and processes necessary to assess the organizational value of data assets.

Establishing a data governance model that promotes social good through data use

A Better Deal for Data[20], by Tech Matters, aims to establish enforceable commitments to ensure ethical data practices, focusing on privacy, security, and the socially beneficial use of data.

It centers on eight proposed commitments, aimed at being a straightforward set that promote responsible data use, whilst not exploiting the subjects of data collection or their communities. These are:

1. We are using your data to benefit you, your community, humanity, and the planet; not for private gain or profit.
2. We don't claim ownership of your data, it remains subject to your control.
3. We will delete your data, correct it, or transfer it somewhere else if you ask.
4. We will not monetize your data by providing it to third parties for compensation.

[19] https://assets.publishing.service.gov.uk/media/64184bccd3bf7f79d9675dbd/Data_Maturity_Assessment_for_Government_-_FINAL_PDF.pdf

[20] https://bd4d.org/

5. You can decide if you want to make your data open[21], or want to monetize it for your benefit.
6. We will protect and steward[22] your data and comply with applicable privacy laws, but you may have privacy obligations as well.
7. If You allow research with Your Data, we will follow best practices around the anonymization of personal data, and published research results will be made available to You for free.
8. We will be bound by legal agreements implementing these commitments, and anyone we share your data with will be similarly bound.

The effort seeks to build a coalition of like-minded organizations and gather input from diverse stakeholders to refine and expand the model commitments.

The initiative emphasizes the need for standardized legal agreements and explanatory texts to make the commitments practical and enforceable.

The goal is to create a cohesive framework for ethical data use, similar to what the Open Source Initiative has done for software, encouraging broader adoption and trust in data practices.

[21] https://opendatahandbook.org/guide/en/what-is-open-data/#:~:text=Open%20data%20is%20data%20that%20can%20be%20freely,gives%20precise%20details%20as%20to%20what%20this%20means.

[22] https://bd4d.org/resources/2309-brief-data_stewardship.pdf

The following sections describe elements of Data Management, adapted from Figure 1, that are shown to influence the value, trust, and equity equation in Figure 3. The sections will describe what they are, why they are important, and how they could be implemented at an organizational level. However, **each section includes emphasis on what the ultimate benefit to the ecosystem is and how organizational implementation would align to systemwide governance.**

Data Security

Implementing measures such as access control, encryption, and incident response to protect data from unauthorized access, breaches, and other security threats.

Data Security Overview

Data security is critical to the protection of modern digital systems to ensure that an organization's or individual's data is suitably secured in line with both Data Protection legislation and good practice of defending against threat actors that are constantly looking to exploit or compromise private information.

Data security is fundamental to being able to demonstrate that an organization is meeting is Data Privacy requirements, whether those are general ones like the EU's General Data Protection Regulations (GDPR)[23] or more specific such as banking security guidelines for Payment Card Industry Data Security Standards (PCI-DSS)[24]. Understanding what Data Security regulations and standards are necessary for a particular organization is the key starting point coupled with taking industry best practice such as NIST guidance.

NIST National Cybersecurity Centre of Excellence is a recognized authority around which many organizations base their data security actively working with industry experts and technology vendors to ensure that the most pressing data security challenges are being addressed in a practical way. The NIST Cybersecurity Framework[25] provides a core framework around which organizations can plan and assess their approach to managing Data Security through managing and reducing their cybersecurity risks.

[23] https://www.gov.uk/data-protection

[24] https://www.pcisecuritystandards.org/

[25] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1299.pdf

What is Data Security?

Data security is defined by NIST's National Cybersecurity Centre of Excellence as:

> *"The process of maintaining the confidentiality, integrity, and availability of an organization's data in a manner consistent with the organization's risk strategy."*

Other organizations have different descriptions, however, you will find the overall theme is around the protection of an organization's or individual's digital assets from unlawful access and manipulation, whether positively or negatively, throughout the lifecycle of the environment.

The NIST Cybersecurity Framework 2.0 is built around six core functions; **Govern**, **Identify**, **Protect**, **Detect**, **Respond** and **Recover**. An organization that can confidently understand, assess, prioritize, and communicate its risks across these functions and actively works to mitigate those risks will demonstrate that it is managing its Data Security appropriately, thus able to be compliant with Data Protection requirements.

It is important to note from the NIST Cybersecurity Framework that Data Security is more than just about protecting digital assets, it is around the whole lifecycle of the environment, accepting that it will not always be possible to defend against every threat, but how an organization responds and recovers is just as important to ensure that both the organization is able to continue operating but also individuals are able to continue using services and have access to their data.

Improving data security across a global network of law enforcement organizations

INTERPOL, the International Criminal Police Organization, faces unique and significant challenges when implementing data security due to its role as a global law enforcement organization. With member countries spanning the globe, INTERPOL must ensure that sensitive criminal data is securely shared and accessed across different legal jurisdictions, technological environments, and security standards.

- **Global diversity in legal frameworks and standards**: Developed a set of global data protection standards, aligned to international best practices. These applied uniformly across member countries creating a baseline level of security and legal compliance. There were instances where INTERPOL had to work closely with member states to ensure local laws were still respected.

- **Secure communications across borders**: Established the I-24/7 global police communications system, allowing law enforcement to share sensitive data in real-time using advanced encryption protocols.
- **Data integrity and authenticity**: Introduced strict data validation and verification processes for all submitted data. It also introduced data signature technology to authenticate the origin of the data and protect against data tampering.
- **Cultural and technological differences**: INTERPOL provides extensive capacity-building programs to improve member cyber security infrastructure and practices.
- **Collaboration between members**: Some collaboration initiatives include developing global security standards that are robust and adaptable to different legal jurisdictions, or the sharing of resources and expertise between member organizations to enhance national data security measures.

What does this mean for the online CSEA ecosystem?

This example highlights the development of global data security standards aligned with international best practices, ensuring a baseline level of security and legal compliance across member countries while respecting local laws. Data integrity and authenticity are safeguarded by strict validation processes and data signature technology. It also highlights the importance of addressing cultural and technological differences by providing capacity-building programs to enhance digital proficiency, including data security. Collaboration between members is key, with success partially due to initiatives focusing on developing adaptable global security standards and sharing resources to strengthen national data security.

Why implement data security?

Data security has become critical in the modern age with the rise of hackers and cyber criminals that seek to profit from compromising an organization's or individual's data security. Without an appropriate level of effort being applied to data security organizations of all sizes can become a soft target of hackers leading to theft of data or a victim of ransomware. There are many organizations that track events and trends around the globe for cyber threats.

The NCC Group publishes periodic reports, the most recent June 2024[26], that highlighted that phishing attacks remain the most prevalent attack vector, affecting 84% of businesses, with 50% of these businesses and 32% of charities reporting such incidents in the last 12 months. The most notable trend of 2024 thus far has been the continued rise in

[26] https://www.nccgroup.com/uk/resource-hub/cyber-threat-intelligence-reports/

ransomware attacks that remains one of the top cybersecurity challenges for organizations across all industries.

To manage the cybersecurity threat, it is essential that organizations of all sizes implement good data security foundations as a minimum, which can then be built on over time to improve an organizations maturity in the data security area. It should be noted though, that some data protection standards will mandate a minimum level of data security to meet those standards.

Furthermore, a baseline level of Data Security is essential for promoting trust between stakeholders within a data sharing system, particularly for the sensitive nature of the data used in countering online CSEA. This sensitive nature extends beyond just data, it also includes techniques, technology, and tradecraft employed by law enforcement and industry (among others); therefore, data security is paramount with the online CSEA ecosystem.

How to begin the data security journey?

The essential first step is to understand the region(s) that the organization will be operating within and what legislation (if any) is relevant to those operations. A good starting point is the NIST Cybersecurity Framework[27], regardless of where in the world the organization is operating, however there are other schemes to consider such as Cyber Essentials that is backed by the UK Government through the National Cyber Security Centre (NCSC) or ISO27001 the international standards for information security management systems (ISMS).

Using the NIST Cybersecurity Framework 2.0[28] as a baseline from which to work, starting from the six functions at a high-level will enable an organization to build maturity in how it manages cybersecurity risks around a well-established structure:

1. **Govern** – The organization's cybersecurity risk management strategy, expectations, and policy are established, communicated, and monitored.
    - Understand and assess specific cybersecurity needs.
    - Develop a tailored cybersecurity risk strategy.
    - Establish defined risk management policies.
    - Develop and communicate organizational cybersecurity practices.
    - Establish and monitor cybersecurity supply chain risk management.
    - Implement continuous oversight and checkpoints.

2. **Identify** – The organization's current cybersecurity risks are understood.
    - Identify critical business processes and assets.
    - Maintain inventories of hardware, software, services, and systems.

[27] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1300.pdf
[28] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1299.pdf

    - Document information flows.
    - Identify threats, vulnerabilities, and risk to assets.
    - Lessons learned are used to identify improvements.

3. **Protect** – Safeguards to manage the organization's cybersecurity risks are used.
    - Manage access.
    - Train users.
    - Protect and monitor your devices.
    - Protect sensitive data.
    - Manage and maintain software.
    - Conduct regular backups.

4. **Detect** – Possible cybersecurity attacks and compromises are found and analyzed.
    - Monitor networks, systems, and facilities continuously to find potentially adverse events.
    - Determine and analyze the estimated impact and scope of adverse events.
    - Provide information on adverse events to authorized staff and tools.

5. **Respond** – Actions regarding a detected cybersecurity incident are taken.
    - Execute an incident response plan once an incident is declared, in coordination with relevant third parties.
    - Categorize and prioritize incidents and escalate or elevate as needed.
    - Collect incident data and preserve its integrity and provenance.
    - Notify internal and external stakeholders of any incidents and share incident information with them – following policies set by your organization.
    - Contain and eradicate incidents.
6. **Recover** – Assets and operations affected by a cybersecurity incident are restored.
    - Understand roles and responsibilities.
    - Execute your recovery plan.
    - Double-check your work.
    - Communicate with internal and external stakeholders.

There are many resources that will support an organization working through the framework as well as external support that can be brought in to help (same applies for Cyber Essentials and ISO27001). However, it is important to remember that the size of the organization will drive the initial depth/breadth of implementation, with the option to scale the assessment/implementation as the organization grows.

It is also critical to remember that Data Security is as much about the security culture of the people within the organization and the individuals, as it is about the technical controls that can be implemented to help mitigate cybersecurity risks.

A strong security culture means that employees and individuals at all levels are consistently aware of and committed to best practices in safeguarding data. This includes adhering to protocols, such as using strong passwords, avoiding phishing scams, and regularly updating software, but also goes beyond compliance to cultivate a mindset where data protection is a shared responsibility.

When security culture is prioritized, individuals understand the importance of their role in preventing breaches—whether it's through careful handling of sensitive information, reporting suspicious activity, or recognizing the potential impact of small oversights. Encouraging ongoing training, awareness programs, and open communication about risks helps foster a proactive environment. In such a culture, data security becomes ingrained in daily operations, with individuals recognizing how their behavior influences the overall cybersecurity posture of the organization.

Data Protection

Data Protection is about ensuring that an individual's rights are protected when they share their personal data while allowing for the responsible use of that data. It is about organizations collecting only the personal data they need, processing it lawfully, only keeping it while they need it, and being transparent about the purposes for collecting it. Data protection law enables the secure sharing of data when it is fair and proportionate to do so.

Data Protection Overview

Data protection globally is underpinned by certain key principles. These principles broadly are:

- Lawfulness, fairness and transparency;
- purpose limitation;
- data minimization and necessity;
- accuracy and data quality;
- storage limitation;
- integrity, confidentiality and security; and
- accountability.

The aim is to respect the privacy and data protection rights of individuals while enabling the responsible use of that data and fostering innovation, whilst at the same time safeguarding the data.

> *While there are certain risks and costs associated with using mobile data to produce social good insights, there may also be risks and harms associated with a failure to include this and other new data sources to inform policy, humanitarian response and other development interventions.*
>
> *The State of Mobile Data for Social Good Report*[29]

[29] https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/06/Mobile-Data-for-Social-GoodReport_29June.pdf.

Restoring the trust of a customer base by implementing robust data privacy policies

Following the Cambridge Analytica scandal, where it was revealed that data from up to 87 million Facebook users had been improperly harvested and used for political advertising without their explicit consent, Facebook undertook a series of significant measures to improve data privacy, enhance user control over personal information, and restore trust.

- **Increased user control and transparency**: Simplifying privacy settings to make them easier to find and use, including bringing them into a single space. Introducing a "Access your information" tool, allowing users to easily manage (and delete) the data that has been collected about them.
- **Third-party app restrictions**: Restricting the data that third-part apps could access and introducing a mandatory review process to ensure compliance. Automatic expiry of data access was also introduced based on time since a user had granted permission for data access.
- **Enhanced data privacy policies**: Aligning their global corporate policies with GDPR and increasing transparency to users of what data they are sharing and how it would be used.
- **Independent oversight board**: Establishing an oversight board with binding powers on matters concerning content moderation or privacy issues.
- **Training and awareness**: Implementing a more rigorous data privacy training for employees.

What does this mean for the online CSEA ecosystem?

This example demonstrates the importance of increasing user control and transparency by simplifying privacy settings and providing tools for managing personal data. Introducing compliance reviews ensure better protection of user data, with automatic data access expiry enhancing security. Aligning global policies with GDPR and improving transparency about data usage builds user trust. Establishing an independent oversight board with binding powers strengthens governance on content and privacy issues. Finally, rigorous data privacy training for employees is essential for maintaining high standards of data protection.

What is Data Protection?

Data protection is about protecting the data and information relating to identified or identifiable individuals. The notion of data protection originates from the right to privacy, and both are instrumental in preserving and promoting fundamental values and rights. Data protection aims to ensure fair, lawful and transparent processing of personal data by both the public and private sectors. In so doing, organizations demonstrate respect for the rights of the individuals whose data they are processing, build trust with those individuals, and comply with data protection legislation.

Implementing privacy by design to create conditions that protect individuals and promotes innovation

Data-Pop Alliance[30] have developed four key principles for technical solutions to data sharing, aimed at maintaining the necessary conditions for data privacy and individual protection. These are:

1. **Distributed data repositories**: Data are stored in separate repositories, with tools deployed from an external or remote human query. This allows tracking and auditing of metadata associated with patterns of communications between repositories and queries.
2. **Move the algorithm to the data**: Keeping data at its repository means raw data does not have to be exposed, allowing the repository owner to control degree of privacy by controlling granularity of query answers. Each repository can provide local query processing and computational capabilities.
3. **Data always in encrypted state: at rest and in computation**: Raw data remains in encrypted state during storage and transit, with new cryptographic algorithms employed to allow operations on encrypted data without the need to decrypt first.
4. **Encode data usage agreements in legal trust networks**: Develop and employ trust network models for large scale data sharing with an ecosystem, combining computer network tracing user permissions for each piece of data within a legal framework, specifying actions and violations of data use. Ultimately, allowing a high degree of interoperability.

What does this mean for the online CSEA ecosystem?

[30] https://datapopalliance.org/wp-content/uploads/2020/09/How-to-use-Big-Data_VFI_Data-Pop-Alliance-Paper.pdf

This example demonstrates an infrastructure that would allow multiple stakeholders within an ecosystem to data share on a large scale, whilst maintaining protection their data and the individual who the data concerns. It creates the condition for flourishing trust between stakeholders and creates more equitable access to computational resources.

How does this differ from Data Privacy?

In the EU, human dignity is recognized as an absolute fundamental right. In this notion of dignity, privacy or the right to a private life, to be autonomous, in control of information about yourself, to be let alone, plays a pivotal role. Privacy is not only an individual right but also a social value. Historically, in other parts of the world, such as the U.S.A., privacy has often been regarded as an element of liberty, the right to be free from intrusions by the state.

The right to privacy is not absolute. Public authorities can override the right to privacy when it is lawful, proportionate, and necessary to do so to protect national security, public safety, the economy, health or morals, the rights and freedoms of other people, and to prevent disorder or crime.

Driving decision making in humanitarian aid and development efforts using big data

Global Pulse is an initiative launched by the United Nations (UN) in 2009 to leverage big data and real-time analytics for sustainable development and humanitarian action. It seeks to use data from various digital sources to gain real-time insights into global trends, predicting and responding to crises more effectively.

- **Data privacy and ethics**: Global Pulse developed a comprehensive data privacy and protection tool[31] to guide its use of data. This built on the UN Principles on personal Data Protection and Privacy.[32] This emphasized the principles of informed consent, data minimization, and anonymization. They also established an advisory group to provide oversight.
- **Data access and partnerships**: Partnerships were built on trust and mutual benefit, with clear data sharing practices.
- **Network of Pulse Labs**: Global Pulse set up a network of innovation hubs where data scientist, engineers, and policy experts collaborate to develop and test new

[31] https://www.unglobalpulse.org/document/risks-harms-and-benefits-assessment-tool/

[32] https://unsceb.org/privacy-principles

tools and technologies. They also served as capacity building to other agencies or partner organizations. The hubs serve to bring UN stakeholder challenges to experts in data, strategic foresight, digital, behavioral science, and innovation. Experts are drawn from within the UN, but also through collaborative partnerships across government and industry.

What does this mean for the online CSEA ecosystem?

This example highlights the importance of establishing a strong data privacy and protection framework, emphasizing informed consent, data minimization, and anonymization, with oversight from an advisory group. Building partnerships based on trust and mutual benefit, with clear data sharing practices, is crucial. Additionally, setting up a network of innovation centres can foster collaboration among data scientists, engineers, and policy experts, while also serving as capacity-building hubs for other agencies and partner organizations.

Why implement Data Protection?

Data protection law sets out what should be done to make sure everyone's data is used properly and fairly.

Organizations likely have personal data about customers and clients such as names, addresses, contact details. This might even have sensitive information such as medical data. They may need this to deliver goods or services, but organizations shouldn't use it in ways people wouldn't expect. Organizations have a duty to protect data.

Data protection law applies to all workplaces, business ventures, societies, groups, clubs and enterprises of any type. The rules are the same for all sizes of organization because if personal data falls into the wrong hands, it makes no difference where the error came from.

There are many benefits to complying with data protection law. As well as being the law, good data protection also makes good economic sense because it saves time and money. It also shows people that you care about their information, which is good for reputation and brand.

Rolling out a privacy-compliant at-home medical device

A US pharmaceutical company was developing a medical device for at-home use by cancer patients. The device and incorporated software needed to hold patient personal data to allow Health Care Practitioners to understand how patients are responding to the treatment.

A clear understanding of the scope of the processing allowed them to be transparent with test subjects during clinical trials and patients after the commercial launch. This enabled participants to provide *informed consent*[33].

Considering privacy during product development meant that remediations and mitigations for privacy risks were built into the product from the outset.

The outcome was assurance that the device was compliant with privacy laws and regulations in all the relevant geographies from launch.

What does this mean for the online CSEA ecosystem?

This example shows how integrating privacy considerations early in solution development, including a clear understanding of data processing scope, ensures transparency with users and allows for built-in mitigations against privacy risks. This proactive approach results in a solution that is compliant with privacy laws and regulations across all relevant geographies from the outset, enhancing user trust and regulatory assurance. In this example, transparency and clarity of the scope and process allowed patients to provide informed consent. This is an ongoing topic of debate in terms of best practice, however some examples of current industry practice include the GDPR checklist[34], or the below areas that must be covered for creating specific and informed consent[35]:

- **The controller's identity**: the individual should know the identity of the controller. This means you need to identify yourself, and also name any third-party controllers who will be relying on the consent. If you buy in 'consented' data, that consent is only valid for your processing if you were specifically identified. You don't need to name your processors in your consent request (although you do need to comply with separate transparency obligations).
- **The purposes of the processing**: separate consent will be needed for different processing operations wherever appropriate – so you need to give granular options to consent separately to separate purposes, unless this would be unduly

[33] https://bd4d.org/resources/2407-brief-consent.pdf

[34] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/lawful-basis-for-processing/consent/

[35] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/consent/what-is-valid-consent/#what3

disruptive or confusing. And in every case, a consent request must specifically cover all purposes for which you seek consent.

- **The processing activities**: again, where possible you should provide granular consent options for each separate type of processing, unless those activities are clearly interdependent – but as a minimum you must specifically cover all processing activities.
- **The right to withdraw consent at any time**: we also advise you should include details of how to do so.

How to begin the Data Protection journey?

At an organizational level:

1. The first step in the data protection journey is usually by assessing an organization's current state against the key areas of data protection law including:
   - Responsibility and accountability (governance);
   - individual rights requests and privacy notices (transparency);
   - maintaining a processing inventory and risk assessing processes;
   - data security;
   - data sharing; and
   - vendor assurance, breach handling, change assurance, data retention, training, and demonstrating compliance (policies and procedures; auditability).

   A list of worldwide Data Protection and Privacy legislation has been compiled UN Trade and Development[36] and the International Association of Privacy Professionals.[37]

Once complete, the following steps are likely to be:

2. Structuring privacy governance and operating model, then assigning role and responsibilities.
3. Developing individual rights procedures including subject access requests and developing privacy notices.
4. Creating a record of processing/data inventory, capturing and risk assessing the processes.
5. Ensuring data security including cyber security, physical security and security relating to individuals

[36] https://unctad.org/page/data-protection-and-privacy-legislation-worldwide
[37] https://iapp.org/resources/global-privacy-directory/

6. Assessing third-party risk including where we give access to our systems, those we share data with, and vendors.
7. Developing and implementing personal data breach procedure (to align with wider security breach procedures)
8. Embedding processes to ensure data protection considerations are managed as part of transformation (organizational or systems changes). Applying privacy by design principles.
9. Ensuring that the data lifecycle is managed including acquisition, management, retention and ultimately deletion/archiving of data.
10. Delivering privacy training and workshops.
11. Ensuring privacy artefacts are maintained including a process inventory, an information asset register, a record of roles and responsibilities, maintain policies and procedures, and privacy assurance records.

Data Sharing

Data sharing is the seamless, secure, and timely exchange of data that provides value to all stakeholders involved while maintaining data privacy, security, and compliance. It enables stakeholders to collaborate effectively, access relevant insights, and make informed decisions, ultimately enhancing the ecosystem's overall functionality and innovation. Efficient sharing also reduces redundancy, improves resource utilization, and fosters trust, ensuring that data flows freely and responsibly across the ecosystem to maximize impact for children.

Data Sharing Overview

Data sharing refers to the practice of making data available to others, typically across organizations, departments, or sectors, to improve collaboration, efficiency, and innovation. It involves the exchange of data between different entities while ensuring that the data is used in a manner that is secure, compliant with legal and regulatory frameworks, and aligned with the privacy expectations of individuals.

> *"I have seen first-hand how proportionate, targeted data sharing, delivered at pace between organizations in the public, private and voluntary sectors, has been crucial to supporting and protecting the most vulnerable during the response to the COVID-19 pandemic. Be it through the shielding program for vulnerable people or sharing of health data in the Test and Trace system. On a local and national level, data sharing has been pivotal to fast, efficient and effective delivery of pandemic responses."*
>
> *Elizabeth Denham CBE, UK Information Commissioner, May 2021*

What is Data Sharing in an ecosystem

Data can be accessed, used, and shared for various purposes, often unlocking new benefits or insights that are unattainable with a single dataset. For example, online CSEA data from one source is frequently combined with data from other sources to generate valuable insights. This data can be shared with external partners, such as researchers, government bodies, or law enforcement agencies, enabling them to create new insights and inform effective interventions.

Data sharing relationships can take multiple forms[38]:

[38] https://open-data-institute.gitbook.io/data-governance-playbook/play-nine-how-to-set-up-successful-data-sharing-partnerships/understanding-how-data-sharing-occurs-in-the-health-sector

- **Openly licensed**: Third parties can access the data without having to be known to the data provider, or to enter into a specific agreement, other than having to follow open data licenses or terms of use requirements.
- **Non-openly licensed**: Subscription model to accessing data under established terms of use.
- **Alliances and networks**: Data partners agree to make data available in standardized formats that can be used by everyone in the network. Common infrastructure may be used to pool data, and a group-level data sharing agreement may be put in place that describes how all partner members contribute data and how they access it.
- **Research hubs**: Data may only be available in limited formats, or secured on specific platforms where analysis can occur, but data cannot be removed from the platform. These relationships often require a greater level of oversight of the security measures in place by each partner to ensure data is protected.
- **1:1 partnerships**: Enables two organisations to collaborate on a specific project, for a specific period of time, working together on well-defined datasets.
- **Transactional**: Acquiring data from external sources in return for a payment for primary data use to optimize healthcare delivery.

These forms will be employed along a continuum, depending on the 'openness' of the data itself. This continuum is shown, through the context of telecommunications, in Figure 5.

Several international frameworks exist to build data sharing agreements upon, these include:

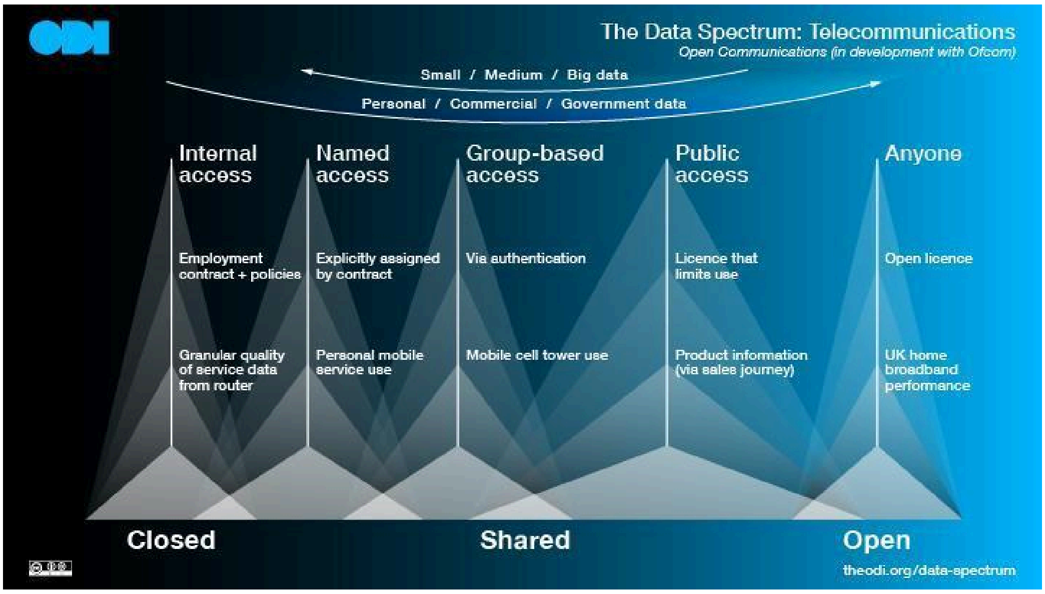
The Five Safes Framework[39]	An internationally recognised approach to managing risk from sharing data
FAIR Principles[40]	A framework that ensures that data is Findable, Accessible, Interoperable and Reusable.
CARE Framework[41]	These principles complement the existing FAIR principles encouraging open and other data movements to consider both people and purpose in their advocacy and pursuits.

[39] https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/

[40] https://www.go-fair.org/fair-principles/

[41] https://www.gida-global.org/care

WHO Data Principles[42]	The principles are intended primarily for use by WHO staff across all parts of the Organization in order to help define the values and standards that govern how data that flows into, across and out of WHO is collected, processed, shared and used.



The Data Spectrum: Telecommunications

Open Communications (in development with Ofcom)

Small / Medium / Big data

Personal / Commercial / Government data

Internal access — Employment contract + policies — Granular quality of service data from router

Named access — Explicitly assigned by contract — Personal mobile service use

Group-based access — Via authentication — Mobile cell tower use

Public access — Licence that limits use — Product information (via sales journey)

Anyone — Open licence — UK home broadband performance

Closed / Shared / Open

theodi.org/data-spectrum

Figure 5: The Data Spectrum[43]

How to begin the Data Sharing journey

At a system level: Data Sharing can be promoted through system governance. One way in which it might do this is by establishing Data Sharing principles. Below are example principles from Chatham House Model Principles for Public Health[44]:

- Building Trust
- Articulating the Value
- Planning for Data Sharing
- Achieving Quality Data
- Understanding the Legal Context
- Creating Data Sharing Agreements
- Monitoring and Evaluation

[42] https://www.who.int/data/principles

[43] https://theodi.org/insights/tools/the-data-spectrum/

[44] https://chathamhouse.soutron.net/Portal/Public/en-GB/RecordView/Index/169144

At an organizational level:

Sharing data relies on trust between organizations. Whilst there are many benefits to sharing data, there is also risk of harm. Trust and trustworthiness are crucial not only for unlocking the societal and economic value of data but also for helping companies and organizations realize the value of their services, products, and ecosystems. By acting in a trustworthy manner, organizations can create value while minimizing potential harms and are more likely to earn the trust of the individuals, organizations, and ecosystems they engage with and depend on.

Assessing your own, and other organization's trustworthiness is therefore a critical prerequisite before entering data sharing agreements. The below list provides ten elements of trustworthiness in relation to data sharing.

Active and positive impact	Ensuring that an organization has a positive impact on society and others, and its data practices uphold norms of equity and fairness.
Engagement and accountability	Ensuring that the interests and concerns of stakeholders are actively sought and addressed, and that the organization can and will be held to account by external parties if necessary.
Ethics and transparency	Ensuring that an organization states and adheres to its ethical principles, and is as open to public examination as possible.
Financial sustainability and revenue generation	Ensuring that an organization has business and revenue models that can support the activities necessary to sustain it.
Governance and strategic oversight	Ensuring that an organization and its data practices are robust, in line with stated organizational purpose, meet requisite standards and are subject to internal oversight and correction.
Legal standing and compliance	Ensuring that an organization and its data practices abide by relevant laws and regulations.
Privacy and security	Ensuring that an organization protects the privacy of its employees, customers, and partners; and that any risks to data it collects, uses or shares are understood and managed.
Quality and accuracy	Ensuring that any data the organization collects, uses or shares meets requisite standards and is fit for purpose.
Readiness and mitigation	Ensuring that an organization is ready to evolve, scale and adapt to changing circumstances, and that it is capable of

	doing so in a manner that limits harms and respects existing stakeholders.
Skills and knowledge	Ensuring that an organization's staff have the necessary skills and knowledge to perform their roles and deliver on organizational objectives.

Figure 6: The 10 elements of trustworthiness concerning data sharing[45]

Organizations can use these elements[46]:

- As a framework to develop a more detailed, holistic assessment of the trustworthiness of an organization as a steward of data. This might be:
- To perform internal assessment of an organization, across the various areas of activity within the organization, or
- to assess the trustworthiness of external stakeholders or potential partners.
- to spotlight specific areas within an organization where trustworthiness can be built further.
- to demonstrate an organization's trustworthiness to others, by using the elements as a framework for highlighting and communicating the range of systems, processes and structures it has in place in order to be a trustworthy steward of data.
- To better understand organizational priorities, assess the degree to which those priorities align with organizational goals or principles and see how their priorities compare with those of partners and stakeholders.

The first step to sharing data is to have a clear purpose for sharing the data. There should be clarity on the roles of the various parties and a clear description of what happens to the data at each stage.

The broad steps that are then followed when initiating data sharing agreements are usually[47]:

1. **Application:** registration or application depending on 'openness' of data
2. **Review:** internal or external review by the data institution depending on the complexity of the request. It may also include audit of the requestor's facilities or

[45] https://open-data-institute.gitbook.io/p22-trustworthy-data-stewardship-guidebook/-MW92wuAXMrYPE7sgA-M/assess/how-to-assess-trustworthiness/activity-1-determine

[46] Ibid

[47] https://theodi.org/insights/reports/how-do-data-institutions-facilitate-safe-access-to-sensitive-data/

competencies. Training may also be required, or the acceptance of relevant accreditations.

3. **Agreement:** agreement of a contract, agreement, or license to legally define use case and access of data. It may also include pricing or other commercial terms.
4. **Minimizing sensitivity:** by modifying the data to reduce the risk of data sharing, or only providing aggregated results or insights built upon the datasets.
5. **Technical access:** controlling how data is transferred, whether by direct transfer, data stream, an interface access, or research environment.

A data or information sharing agreement (or equivalent) is considered good practice and should include details including:

1. What data is being shared, the purpose for sharing the data, the benefits of the sharing to the organizations, the individuals whose data is shared (if it includes personal data), and more broadly, and why the sharing is necessary to achieve those benefits.
2. Whether the data includes more sensitive personal data (e.g. medical data, protected characteristics) or commercially sensitive data, including any limitations on how the data may be used and how the recipient will demonstrate that they are abiding by the limitations.
3. How the data needs to be governed ensuring that only relevant data is shared, that data quality is maintained, that data is provided in an agreed format, and that retention and deletion of data is agreed.
4. What appropriate technical (information security) and organizational (physical i.e. entry gates to premises; and human i.e. training) controls need to in place both at the recipient and during data transfer.
5. A timescale for the agreement including processing terminating the agreement and a clear understanding of what happens to the data, for example, deletion or returned to the original provider.

A key point to make is that any controls, restrictions, limitations, expectations that an organization has about sharing their data will most likely be the same concerns that other organizations have when sharing with them.

Other reference documents for Data Sharing include:

- Support Centre for Data Sharing[48]
- NYU GovLab data responsibility journey[49]

[48] https://eudatasharing.eu/sites/default/files/2019-10/EN_Report%20on%20Model%20Contract%20Terms.pdf
[49] https://dataresponsibilityjourney.org/sharing

- Protection Information Management[50]

Sharing datasets to counter global health challenges

The Global Health Data Exchange (GHDx) is a comprehensive and open-access repository of health-related data. It was launched to address the need for a centralized, accessible platform for global health data, GHDx plays a critical role in public health research, policy-making, and the monitoring of health trends worldwide.

- **Open access**: it is freely accessible on the internet, promoting transparency and democratizing access to global health data.
- **Standardized**: Data is standardized to ensure consistency and interoperability, critical for temporal or geographic studies.
- **Meta-data standards**: Each dataset is accompanied by detailed meta-data, including source, methodology and context of data. This helps analysts understand and correctly use the data for their own outputs.
- **Scalable**: The technical infrastructure efficiently hosts large datasets with high accessibility, catering for growing number of users.

What does this mean for the online CSEA ecosystem?

This example demonstrates it is possible to design an infrastructure or platform that can support data sharing at scale by providing a common legal foundation, parameters for responsible data use, common protocols for data access and exchange and a pathway to sustainable business models.

Alternatives methods of Data Sharing

There are several alternatives to formally sharing data, some of which are summarized below:

1. Having a central 'golden source' (i.e. master copy) of data that can be accessed but not copied with access controls in place to limit what data can be viewed depending on role. This is most often used for intra-organizational sharing but stops copies of datasets proliferating across a network.

[50] http://pim.guide/wp-content/uploads/2018/05/Framework-for-Data-Sharing-in-Practice.pdf

The following are some examples of Privacy Enhancing Technologies that allow access to data without formally sharing it such that privacy or commercial sensitivity is protected:

2. Differential Privacy[51] is a method for releasing statistical information about a dataset without revealing whether a particular individual's data is part of the dataset. As an example, the US Census Bureau began using it with the 2020 Census data to protect the privacy of individual's data.
3. Homomorphic Encryption[52] is a technique that allows computations to be performed on encrypted data without decrypting it first. This permits work on personal or commercially sensitive data without revealing the underlying data.
4. Secure Multi-Party Computation[53] is a protocol that allows different parties to process their combined data without any party needing to share all its data with the other parties. It allows them to share computing tasks without revealing each other's data.
5. Sharing by way of common systems, this might be by a data host organization publishing a portal that offers limited functionality or access to a dataset with appropriate access controls. It could also be achieved by allowing people from other organizations inside the host organization's security regime, giving them appropriate access to their systems.

Enabling global environmental protection and research through open access datasets

The Global Biodiversity Information Facility (GBIF) is an international network and data infrastructure that provides open access to data about life on Earth, primarily focusing on biodiversity. GBIF aims to facilitate the free and open access to biodiversity data, enabling scientists, researchers, policymakers, and the public to better understand and conserve the natural world.

- **Open access**: Unrestricted access to a vast repository of biodiversity data, this approach fosters transparency, collaboration, and knowledge sharing across disciplines and borders.
- **Harmonizing taxonomies**: Within biodiversity there are different naming conventions, synonyms, and classifications across regions and disciplines. GBIF developed standardized taxonomies to improve interoperability.

[51] https://digitalprivacy.ieee.org/publications/topics/what-is-differential-privacy
[52] https://www.splunk.com/en_us/blog/learn/homomorphic-encryption.html
[53] https://digitalprivacy.ieee.org/publications/topics/what-is-multiparty-computation

- **Diverse data sources**: GBIF aggregates data from a wide range of sources, including museum collections, field observations, and citizen science projects. This is integrated into a unified system.
- **Incentivizing data sharing**: Encouraging organizations to share their data, particularly where this is considered proprietary or where there was concern about misuse. GBIF built trust between organizations and demonstrated the value of open access to system and organizational purposes. Incentivization models included providing unique identifiers for each dataset to allow for attribution during research breakthroughs and international policy change.

What does this mean for the online CSEA ecosystem?

This example demonstrates the value of open access to data, which promotes transparency, collaboration, and cross-border knowledge sharing. Harmonizing taxonomies across different naming conventions and classifications is crucial for improving data interoperability. Aggregating diverse data sources into a unified system enhances the comprehensiveness of the data. Additionally, incentivizing data sharing by building trust and demonstrating the benefits of open access helps overcome concerns about proprietary data and potential misuse.

Improving educational outcomes by seamlessly sharing data on students and research

The Erasmus+ program is a flagship initiative of the European Union that promotes educational exchange, cooperation, and mobility among European universities and educational institutions. Erasmus+ is one of the largest and most ambitious education programs in the world, involving over 4,000 institutions across 33 countries. The program offers mobility grants and supports strategic partnerships and cooperation projects between institutions.

- **Strategic partnerships**: Supporting partnerships between educational institutions, businesses, and other organizations to develop innovative practices and share knowledge.
- **Open access to educational resources**: Provides access to online platforms, digital tools, and databases that support teaching and learning.

- **Student data exchange**: Transferring student data in streamlined ways with standardized protocols and digital tools. Doing so in a way that complied with GDPR and other data privacy regulation.
- **Monitoring performance**: Tracking outcomes at a system level was essential for the program's success. These frameworks included regular surveys, feedback mechanisms, and impact assessments[54].

What does this mean for the online CSEA ecosystem?

This example demonstrates the importance of fostering strategic partnerships between institutions, businesses, and organizations to drive innovation and knowledge sharing. Providing open access to resources, such as digital tools, supports capacity building. Streamlining data exchange using standardized protocols while ensuring compliance with GDPR and other data privacy regulations is crucial. Monitoring performance through regular surveys, feedback mechanisms, and impact assessments is essential for evaluating and ensuring success.

[54] https://commission.europa.eu/strategy-and-policy/eu-budget/performance-and-reporting/programme-performance-statements/erasmus-performance_en#budget-performance--outcomes

Accelerating global development through a unified data ecosystem

The World Bank Development Data Partnership is a collaborative initiative led by the World Bank that brings together international organizations, private sector companies, and academic institutions to leverage data for development purposes. The partnership aims to address global development challenges by facilitating access to and sharing of data, particularly from private companies, that can be used to generate insights, inform policy decisions, and drive sustainable development. Key features include:

- **Facilitating access to private sector data**: Facilitates access to data that is often held by private companies, such as telecom operators, satellite imagery providers, and social media platforms. It does this by negotiating formal data-sharing agreements between stakeholders, outline how data will be shared, used, and protected.
- **Building capacity and knowledge sharing**: The partnership provides capacity building initiatives and facilitates the exchange of knowledge and best practice among members.
- **Data for good**: The partnership removes barriers associated with 'ownership approach' to data, focusing instead on creating outcome-driven change to those who are vulnerable.
- **Data-driven policy making**: With a focus on low-resource geographies, the partnership drives informed decisions by providing greater access to data. Allowing policies to be tailored to the specific needs and conditions of populations.

What does this mean for the online CSEA ecosystem?

This example highlights the importance of facilitating access to private sector data through formal data-sharing agreements that define data use and protection. Building capacity and promoting knowledge sharing among members enhances overall effectiveness of a solution. Shifting from an ownership approach to a focus on outcome-driven change helps address the needs of vulnerable populations. Additionally, leveraging data to drive policymaking in low-resource areas enables more informed and tailored decisions that address the specific needs and conditions of different communities.

Data Ethics

Data ethics involves not just what is possible or legal with data, but what is right for an organization to do. It means recognizing and adhering to moral guidelines about what should and shouldn't be done with data.

Data Ethics Overview

Data is used to drive insights and to deliver purpose. The online CSEA's reliance on data will continue to increase; however, increased use of data and data driven technologies presents increased risk of harm to individuals through breaches of privacy or unjust decision making.

> *All of [the human] rights are important and commitment to one right should not detract from the importance and protection of another right. Taking rights in conjunction wherever possible is healthier than taking rights in opposition to each other.*[55]

Organizations taking an ethical approach to data go beyond what is legally permissible or technically possible to consider what is good for their partners and stakeholders, for the organization, and for society.

Broadly speaking, Data Ethics can be considered in four themes[56]:

1. Societal benefits and value
2. Distribution of risks, benefits and burdens
3. Respect for individuals and groups
4. Public trust and engagement

[55] Report of the Special Rapporteur on the right to privacy, Joseph A. Cannataci. 8 March. A/HRC/31/64, pp. 6, 10. Annex II. A more in-depth look at Open Data & Big Data.

[56] https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-019-0359-9/tables/4

Lessons from medical research: Abuses of human subjects in research experiments[57]

Following a series of highly publicized abuses of human subjects in research experiments in the United States, the medical field recognized the need to incorporate ethics into data use frameworks.

The Common Rule was implemented, it incorporates three central ethical principles, which underlie ethical frameworks in many countries around the world. The first, respect for persons, pertains to individual dignity and autonomy and is applied through the concept of informed consent. The second principle, beneficence, requires weighing the risks to people associated with a research activity against the benefits to people and society. The third principle, justice, considers who gets chosen to be a research subject and what population segments stand to benefit — or be harmed — by the research results.

By adopting these ethical pillars, the medical field in the U.S. addressed a major problem afflicting human subject research at the time, namely that engagement with research subjects had shifted away from regarding them as humans capable of suffering harm to viewing them as mere objects or data points. Because data analytics today may avoid any personal interaction with the data subjects themselves, many big data research projects present similar risks. And because data analytics is now deployed ubiquitously across industries — not just in health care — opportunities for abuse have multiplied exponentially.

What is Data Ethics?

Data Ethics[58] is an emerging branch of ethics that studies and evaluates moral problems related to:

- Data

  generation, recording, curation, processing, dissemination, sharing and use
- Algorithms

  artificial intelligence, artificial agents, machine learning and robots
- Corresponding practices

  responsible innovation, programming, hacking and professional codes

It does this to formulate and support morally good data-driven solutions to problems.

[57] https://iapp.org/resources/article/building-ethics-into-privacy-frameworks-for-big-data-and-ai/

[58] https://royalsocietypublishing.org/doi/epdf/10.1098/rsta.2016.0360

Strengthening the cross-platform response to online CSEA using signal data

Lantern is a program run by the Tech Coalition, it is a cross-platform signal sharing program for companies to strengthen enforcement of their child safety policies.

- **Safety and privacy by design**: Developing a cross-platform program means it is necessary to take proportionate measures to facilitate sharing whilst maintain data privacy. Lantern established clear guidelines and rules for appropriate data sharing among participating companies; performs ongoing reviews of its policies and practices; and conducts mandatory training and routine check-ins with members.
- **Pilot and scale**: Lantern began with a two-year pilot program, investigating whether sharing particular data led to positive outcomes. After a successful pilot, it is now being scaled up to include new data from other sectors, including the financial sector.
- **Trust and independence**: Tech Coalition were identified as a suitable organization to host Lantern, as it was a place where collaboration already occurred, they had the required trust of it's members for sharing sensitive data, they were able to perform an oversight capacity as they didn't use or produce signal data themselves, and they are independent of government.
- **Setting entry standards engenders trust**: Having minimum, stringent eligivility requirements and applications processes engenders trust between members that facilitates sensitive data sharing. The vetting process and ongoing compliance checks ensure that the fundamental question of "can they use the data appropriately" is continually reviewed.

What does this mean for the online CSEA ecosystem?

It is possible to create data sharing solutions across platforms, strengthening the response to cross-platform offending. A balance between data privacy and data sharing can be found, a necessary step in fostering a secure, safe, and privacy-conscious environment for exchanging CSEA data. Setting entry standards for membership organizations builds trusts between members, as does the independent position of the hosting organization themselves.

Why implement Data Ethics?

Individuals expect organizations to be transparent and accountable in how they use data and reward those with good practices with loyalty. Ethical practices also support strong organization to organization relationships as partners seek reassurances that organizations are implementing best practice when it comes to data.

To mitigate risks associated with the privacy and ethical challenges of using data, considerations should be given during early in data collection or project design. Simultaneously, any risks to individual privacy should be considered against the societal benefits of using data.

How to begin a Data Ethics journey

At a system level:

Incorporating frameworks at a system level can guide individual organizations' decision-making processes to address the newly emerging field of data ethics. Frameworks offer governance options to organizations with weighty data ethics questions or large scale processing of sensitive data.

Establishing an external ethics review board may be appealing for small and medium sized organizations, providing ethical standards and best practices, creating an accessible knowledge base. These also protect vulnerable populations, protect the reputation and trust of the ecosystem by providing a basis for public expectations, [and allow for an] evaluation of the profession.[59]

An external ethics review board may provide resource such as Ethical Data Impact Assessment templates[60], or other examples of best practice.

At an organizational level:

There are four key steps in implementing a data ethics framework:

1. **Develop the core ethical principles**, which are the building blocks of a data ethics framework. Common core principles include that the use of data shall be beneficial, fair, and transparent. Additional principles should align with an organization's core values.
2. **Understand the ethical risks including addressing privacy risks**, understanding the limitations of data and addressing bias, and consider the potential negative outcomes from the use of the data.
3. **Implement the appropriate ethical controls including the policies and processes** to ensure adherence to ethical principles through the data lifecycle, the technical

[59] http://bdes.datasociety.net/wp-content/ uploads/2016/10/ EthicsCodes.pdf.

[60] https://secureservercdn.net/192.169.221.188/b1f.827.myftpupload.com/wp-content/uploads/2020/04/Model-Ethical-Data-Impact-Assessment-January-2019-002.pdf

and organizational measures to mitigate ethical risk, the procedures to ensure that the organization is transparent and accountable. There are existing tools that describe processes for doing so, for example: Ethical Data Impact Assessments and Oversight Models[61]

4. **Implement the appropriate governance structure** by ensuring that senior leadership takes responsibility for data governance and ethics, leveraging existing data governance structures where possible, and ensuring that stakeholders receive the training they need to understand their responsibilities to use data ethically.

Empowering individuals by giving them more control over their data

The Commons Project is a non-profit organization that develops technology platforms and digital solutions aimed at improving public health and empowering individuals with greater control over their personal health data. Their flagship product is CommonHealth, which allows individuals to collect, manage, and share personal health data securely from their devices. It is an open-source, interoperable solution that integrates with many EHR systems (see example within Data Integration & Interoperability section). Key features include:

- **Health data aggregation**: CommonHealth aggregates health data from multiple sources, to provide a single comprehensive overview of a single individual.
- **Access and empowerment**: Users can choose specific data points or health records to share, rather than entire medical histories. Providing easy access to personal health data empowers individuals to have more informed conversations, leading to improved health outcomes.

What does this mean for the online CSEA ecosystem?

This example shows that aggregating data from multiple sources into a comprehensive overview can significantly enhance individual empowerment by allowing users to selectively share specific data points. This approach provides easier access to personal information, facilitating more informed discussions, in this example with healthcare providers and potentially leading to better health outcomes.

[61] Ibid

Data Architecture

Data architecture is the design and structure of how data is stored, organized, and accessed within an organization. It involves setting up systems and databases to ensure that data is efficiently collected, managed, and used for decision-making. Essentially, it's like creating a blueprint for how data flows and is handled throughout the organization.

Data Architecture Overview

Data architecture provides a strategic framework for managing data within an organization. It outlines how data is collected, stored, integrated, and accessed across various systems, ensuring consistency and efficiency. This architecture includes data models, databases, data flow diagrams, and governance policies that define data relationships, standards, and processes. By aligning data management with organizational goals, data architecture supports effective decision-making, enhances data quality, and ensures compliance with regulations. It also facilitates scalability and adaptability, enabling organizations to efficiently handle growing data volumes and complex data environments while maintaining a unified view of their data assets.

What is Data Architecture?

Data architecture is the blueprint for managing and organizing data within an organization. It defines the structure, storage, integration, and accessibility of data across various systems. Data architecture encompasses data models, standards, policies, and governance frameworks to ensure data is consistently organized and used efficiently. It plays a critical role in supporting business operations, analytics, and decision-making by providing a coherent view of data flow and relationships. By aligning data with business needs, data architecture helps in optimizing data assets, improving data quality, and ensuring compliance with regulations.

Why implement Data Architecture?

Implementing data architecture is essential for organizations to effectively manage and leverage their data assets. In today's data-driven world, organizations generate vast amounts of data from various sources, including user interactions, operational processes, and external environments. Without a structured approach, this data can quickly become fragmented, inconsistent, and difficult to access, leading to inefficiencies and missed opportunities.

Data architecture provides a blueprint that ensures data is systematically organized, stored, and accessed. It enables the creation of standardized data models, policies, and governance frameworks that ensure data consistency, quality, and security across the organization. This structure allows for better data integration, enabling different stakeholders to share and utilize data more effectively.

Moreover, a well-implemented data architecture supports better decision-making by providing a clear and unified view of the organization's data. It allows for advanced analytics, helping organizations identify trends, optimize responses, and innovate. Additionally, it simplifies compliance with regulatory requirements by managing data privacy and security.

Ultimately, implementing data architecture helps organizations harness the full potential of their data and partner stakeholder, data driving efficiency, innovation, and enhanced response to online CSEA.

Robust data architecture underpins successful space missions

The European Space Agency (ESA) manages and processes vast amounts of data generated by its various space missions, making data architecture a critical component of its operations.

- **Data federation**: The ESA use a federated approach, meaning it integrates and manages data from multiple sources, both within the agency and from partners. It utilizes a common data pool, allowing the integration of multiple data types..

What does this mean for the online CSEA ecosystem?

This example describes an approach to data architecture that allows multiple stakeholders to collaborate on multi-source data types, which could be employed in a potential CSEA solution.

What are the core components of Data Architecture?

The core components of data architecture are essential for structuring and managing data effectively within an organization. These components include:

1. **Data Models**: Define how data is structured, including entities, relationships, and attributes, providing a blueprint for database design.
2. **Data Storage**: Encompasses databases, data lakes, and data warehouses, where data is stored and managed.
3. **Data Integration**: Facilitates the movement and transformation of data between different systems, ensuring data consistency and availability across the organization.
4. **Data Governance**: Involves policies, standards, and procedures that ensure data quality, security, and compliance with regulations.
5. **Data Security**: Protects data from unauthorized access, breaches, and threats, ensuring its confidentiality and integrity.

6. **Metadata Management**: Provides information about data, such as its source, usage, and structure, enabling better data discovery and usage.

These components work together to create a cohesive data architecture that supports business needs, decision-making, and operational efficiency.

What should you consider when designing scalable data architecture?

Designing a scalable data architecture involves best practices that ensure the system can grow and adapt with the organization's needs.

- **Modular Design**: Break down the architecture into manageable, independent modules, allowing for easier updates and scalability.
- **Data Modeling**: Use flexible data models that accommodate future changes in data types and structures without extensive rework.
- **Cloud Integration**: Leverage cloud services for their elasticity, enabling dynamic scaling based on demand without the need for heavy upfront investments.
- **Data Partitioning**: Implement techniques like sharding and partitioning to distribute data across multiple servers, improving performance and scalability.
- **Automation**: Automate data processes, such as ETL (Extract, Transform, Load), to ensure efficiency as data volumes increase.
- **Monitoring and Analytics**: Continuously monitor system performance and optimize based on usage patterns and emerging requirements.
- **Data Governance**: Implement strong data governance to maintain quality and consistency as the architecture scales.

These practices ensure that the data architecture remains robust, efficient, and capable of handling growing data needs. There are several models available that utilize cloud computing. Google[62], Microsoft[63], and Amazon[64] (among others) offer solutions that architect data infrastructure in such a way that protects privacy, is scalable, and allows multi-stakeholder collaboration on datasets whilst retaining any necessary control to the data steward.

[62] https://cloud.google.com/blog/topics/developers-practitioners/what-data-pipeline-architecture-should-i-use/

[63] https://azure.microsoft.com/en-gb/solutions/data-lake

[64] https://aws.amazon.com/ec2/nitro/nitro-enclaves/

How can you assess the effectiveness of your data architecture?

Organizations can assess the effectiveness of their data architecture by focusing on several key metrics and practices.

- **Data Quality**: Evaluate the accuracy, consistency, and completeness of the data across the architecture. High data quality indicates a well-structured architecture.
- **Performance Metrics**: Monitor the speed and efficiency of data processing, retrieval, and integration. Faster performance under varying workloads suggests effective scalability and optimization.
- **User Accessibility**: Assess how easily users can access and utilize data for decision-making. Effective architecture provides intuitive and secure access to relevant data.
- **Data Integration**: Check how seamlessly data flows between different systems and stakeholders. Effective integration supports business agility and collaboration.
- **Compliance and Security**: Regularly audit for adherence to data governance policies, security protocols, and regulatory requirements.
- **Scalability**: Review the system's ability to handle increased data volumes and new requirements without performance degradation.

By regularly measuring these aspects, organizations can ensure their data architecture aligns with organizational goals and adapts to evolving needs.

Data Integration and Interoperability

Standardizing data and using APIs or other technical solutions to enable seamless data sharing and integration across different systems and platforms, ensuring compatibility and accessibility.

Data Integration and Interoperability Overview

Data integration is the process of bringing data together to provide a consistent, assured source that can be used by analysts, data users and data systems. Data interoperability – clear, shared expectations for the contents, context and meaning of data across systems and services that create, exchange and consume data – can significantly accelerate and provide greater certainty of the data by increasing the compatibility and consistency of data and how it is obtained.

Good data integration enables a broader set of users and organizations to make greater use of the data available. Conversely, poor data integration will lead to the organization having many disparate data sets, duplicated effort to process and manage that data and significant variability in the outputs and outcomes of different data users.

These two concepts go hand-in-hand because effective data integration often relies on data interoperability. Without interoperability, integrated data might not be consistent or easily usable, limiting its effectiveness in providing comprehensive insights.

Creating an open, competitive, and innovative financial services system

Open Banking is a financial services innovation that allows third-party providers to access financial data and initiate transactions on behalf of customers, leveraging standardized APIs. Four key steps in its development were:

- **Adoption of Open APIs**: Open Banking mandates standardized APIs, such as those defined by EU's PSD2[65] to ensure secure and seamless data exchange between banks and third-party providers.
- **Enhancing data sharing**: Gaining explicit customer consent for data sharing, improving data portability[66] and service provider choice.

[65] https://www.ecb.europa.eu/press/intro/mip-online/2018/html/1803_revisedpsd.en.html

[66] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/right-to-data-portability/#:~:text=The%20right%20to%20data%20portability,way%2C%20without%20affecting%20its%20usability.

- **Regulatory initiatives**: Leveraging regulatory powers to drive standardization, such as seen in the UK Open Banking Standards[67].
- **Driving ecosystem innovation**: Open Banking enables fintechs to develop new services and fosters competition, leading to better financial products and services for consumers.

What does this mean for the online CSEA ecosystem?

This example highlights the importance of adopting standardized Open APIs to ensure secure and efficient data exchange between stakeholders. Enhancing data sharing by obtaining explicit customer consent and improving data portability allows for greater service provider choice. Regulatory initiatives can be used to drive standardization and compliance. Lastly, Open Data principles can foster ecosystem innovation.

What is Data Integration and Interoperability?

Data integration is the process of bringing together data from different sources to build a unified and comprehensive view of it. It enables data from many disparate sources to be processed and analyzed more accurately, based on a range of sources.

Data interoperability is the ability to create, exchange and consume data to create clear shared expectations for the content, context and meaning of the data. It enables the correct interpretation of data that crosses system and organizational boundaries.

Providing countries with a globally standardized tool to capture and categorize violence against children

Collecting high quality data on children's experience with violence is limited by the varying notions of violence across the globe. The International Classification of Violence Against Children (ICVAC) developed a comprehensive set of operational definitions of violence and categories for statistical analysis. The project followed a process that built a global standard that recognized local sensitivities:

[67] https://www.openbanking.org.uk/

- **Call to action**: Stakeholder meetings to discuss challenges, barriers, and best practice.
- **Stakeholder agreement**: Accepted recommendation to build a comprehensive set of operational definitions, focusing on acts of violence.
- **Formation of a task force**: Gathering experts and developing a first draft for review.
- **Global consultation**: A large scale review by a plethora of globally representative stakeholders, followed by multi-country testing. The testing focused on comparing ICVAC definitions to national laws, and the feasibility of gathering data on ICVAC variables.

What does this mean for the online CSEA ecosystem?

This example highlights that it is possible to foster the production of accurate data from globally diverse cultures, that can empower action and accountability from governments and other stakeholders.

Why implement Data Integration and Interoperability?

Data integration is a prerequisite to good analytics and use of data. It provides strong assurance and controls that the data being brought together is correct and consistent, empowering analysts, data scientists and technologies such as AI to place trust in the data. It manages and deals with inconsistency across sources and can work centrally to ensure that the act of bringing data together, consolidating and unifying it, and identifying discrepancies between sources is done efficiently once for all consumers of that data.

Having good data interoperability can significantly simplify the work of data engineers and consumers when they are integrating and using data. It increases trust in the source data, enables a better and more accurate understanding of what the data contains, means and the context of that data, and can significantly reduce the latency / increase the velocity of data flows. This both leads to more accurate and faster results when using the data.

Improving healthcare outcomes with seamless data interoperability

Electronic Healthcare Records (EHRs) enable seamless and secure exchange of patient health information across US healthcare providers and systems. Achieving this required five key changes within the system:

- **Creation of standards**: such as HL7[68], DICOM[69] and CCDA[70].
- **Health Information Exchanges[71]**: Creation of platforms that act as intermediaries, aggregating and normalizing data from diverse sources to secure and standardize data transmission.
- **Standardized APIs**: Using standardized APIs[72] enables healthcare application to access and exchange patient data securely across different EHR platforms.
- **Data governance & security**: Establishing methods for gaining patient consent to data sharing and securing that data during transmission and storage
- **Regulatory initiatives**: Leveraging regulatory powers to incentivize adoption of standards, including Meaningful Use[73].

What does this mean for the online CSEA ecosystem?

Key lessons from this example include the importance of creating and adopting standards to ensure interoperability. Data exchanges play a crucial role in aggregating and normalizing data from various sources for secure and standardized transmission. Standardized APIs are essential for secure data access and exchange across different platforms. Regulatory initiatives can drive the adoption of these standards.

How to begin the Data Integration and Interoperability journey

Organizations should consider these twelve activities:

1. Design the end-to-end data journey; from acquisition or ingest, through data storage, usage and outbound sharing. Ensure that ingested data attributes are

[68] https://rhapsody.health/blog/what-is-hl7/
[69] https://www.dicomstandard.org/about
[70] https://www.particlehealth.com/blog/what-is-ccda-consolidated-clinical-document-architecture
[71] https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/what-hie
[72] https://hl7.org/fhir/http.html
[73] https://www.ama-assn.org/practice-management/medicare-medicaid/meaningful-use-electronic-health-record-ehr-incentive

mapped through to business questions and outcomes to understand the most important data attributes enabling effective prioritization.

2. Evaluate and understand the data: Understand whether the data will add value prior to working on it. This evaluation should include input from legal and policy review, not just data specialists.
3. Measure the value obtained from data: Measure which data sets are adding value, which are not in use, which are duplicate or other data sets fully provide as much insight. Reduce effort, time, spend on low value data sets (up to and including full decommissioning them).
4. Recruit specialist Data Engineers who can assure data quality and automate pipelines.
5. Provide controlled or managed access from your data platform to your data users. Use a subscription model where appropriately authorized users or applications can use data sets and apply strong access and audit controls to data usage.
6. Create feedback loops from analysts and users, to continually improve the shared data core. Integrating analysts and users' findings back into data can help to optimize other data ingested or encourage reflection and change in how data is ingested.
7. Build APIs from source systems and automate data flows to reduce data latency and increase consistency and accuracy. Consider building outbound APIs for onward sharing.
8. Define and publish data standards for key data sources and regularly assure compliance. Standards can take many forms, such as ISO standards or semantic data models.
9. Build a data model for core terms across all data sources to encourage consistency in field names and the format and values held within them to enable joins between data sets. This can be coupled with Authoritative Lists for key business terms.
10. Publish Terms of Use for sectors, segments, or jurisdictions that allow organizations to sign up and provide and received shared data. Avoid bilateral agreements as these will not scale.
11. Consider Cloud as providers, given their excellent large-scale data platforms, and integration of capabilities is the default. Beware of the costs of high data volumes and data egress.
12. DevOps and Automation are key supporting technologies that will enable rapid, consistent data handling, and rapid evolution. Manual processing of data will significantly increase latency and reduce consistency.

Facilitating the automated translation between different CSAM categorization schemas

Different countries, sectors and organizations use distinct categorization schemas to process CSAM reports. INHOPE's Global Standard project aims to develop a common ontology of CSAM categorizations.

The four main stages of this project are[74]:

- **Design of the Ontology**: This will be done through Focus Group meetings of the International Working Group.
- **Technical integration**: Once developed, the ontology will be integrated into ICCAM - INHOPE's secure platform for the exchange of CSAM reports between hotlines.
- **Hotline analysts training and implementation**: Hotlines will categorize CSEM/CSAM according to the new ontology for a duration of approximately 1 year, which will create a sufficient volume of material to refine the ontology and create an annotated CSAM hash- set that can be used by hotlines within the existing ecosystem.
- **The final products**: The Ontology and the basis for a Translation Matrix will be shared with the relevant stakeholders in the sector, i.e. hotlines, the technology industry, law enforcement agencies and other relevant partners.

Key concepts in Data Integration and Interoperability

- **API (Application Programming Interface)**: This is a published interface that enables data to be ingested or egressed from a system. Most legacy systems do not have APIs as they were mostly concerned with data within the system, not with sharing data in or out of systems. Modern systems will increasingly provide some APIs to their data, often following the REST pattern.
- **Data standards**: These can provide a powerful binding force for data sharing between organizations. However, they are time consuming to agree, often reflect a compromise between organization's needs, are rarely complied with inside an organization (as they will have specific implementation needs), and translations between internal data structures and shared data may be inconsistent. Data Validation against standards is an important assurance step.
- **Data pipeline**: A structured, often automated process that allows for the efficient and consistent movement, transformation, and combination of data from disparate

[74] https://www.inhope.org/EN/articles/the-global-standard-project

sources, ensuring that it is accessible and usable across different systems and for various purposes.

- **Data mesh**: A decentralized approach to data architecture that emphasizes domain-oriented decentralized data ownership and access. It shifts the traditional centralized data architecture, where data is collected and managed in a single repository, towards a distributed model where data ownership and management are decentralized across different domains or organizations.
- **Data lake**: A data lake is a centralized repository that allows you to store a vast amount of structured, semi-structured, and unstructured data at scale. Unlike traditional data warehouses that store processed data in a structured format optimized for querying, data lakes store raw data in its native format until it's needed for analysis.
- **Data subscription model**: This is a process or mechanism where data consumers (e.g. applications, analytics tools, or users) subscribe to receive data updates or access to specific datasets from a data provider.

Creating the conditions for cross-country and sector collaboration

The EU Data Strategy and the concept of the Single Data Market are central components of the European Union's broader vision to harness the power of data for economic growth, innovation, and societal benefit while ensuring data privacy, security, and ethical use.

- **Sectoral data spaces**: This strategy aims to create nine common European data spaces in strategic sectors, such as health or public administration. Each sector will have nuance specific to the types of activities undertaken.
- **Data sovereignty**: A core principle is ensuring individuals and organisations have control over their data; empowering users to decide how and where their data is used.
- **Data altruism**: Encouraging organisations to voluntarily share data for the public good, rather than seeing data only as an organisational asset.
- **AI ethical guidelines**: Promotes the development of ethical guidelines for data use, particularly in areas like AI, to ensure that data-driven technologies are aligned with European values and fundamental rights.

- **Legal frameworks**: Harmonizing regulations across the EU to reduce barriers to data movement whilst respecting individuals rights and privacy, for example this includes GDPR

What does this mean for the online CSEA ecosystem?

The EU Data Strategy could help combat child sexual abuse by creating specialized data spaces for child protection, which improve information sharing and coordination among agencies. Ensuring data sovereignty allows for controlled and secure data use, protecting sensitive information. Promoting data altruism encourages voluntary sharing of crucial data for public good, aiding in prevention and intervention efforts. Implementing ethical guidelines for AI ensures that technologies used in detecting abuse respect privacy and rights.

Data Analytics

Making data accessible to stakeholders for informed decision-making, while using advanced analytics responsibly to gain insights and maintain ethical standards.

Data Analytics Overview

Data analytics is the systematic process of examining and interpreting complex data to extract valuable insights and inform decision-making. It involves collecting data from diverse sources, cleaning and preparing it for analysis, and applying statistical models and algorithms to identify patterns, trends, and correlations. Through exploratory data analysis and visualization, analysts can uncover actionable insights and present findings in a clear, accessible manner. Data analytics encompasses various techniques, including predictive and prescriptive analytics, to forecast future trends and recommend optimal actions.

Why implement Data Analytics?

Implementing good, effective data analytics is crucial because it transforms raw data into actionable insights that enhance preventive measures and response strategies. By analyzing data from various sources, such as reported incidents, case records, and social media activity, organizations can identify patterns, detect emerging threats, and allocate resources more effectively to high-risk areas. Effective data analytics supports the development of targeted prevention programs, improves coordination among stakeholders, and helps in evaluating the success of intervention strategies. Additionally, it allows for better tracking of cases and trends, facilitating more informed decision-making and strategic planning to safeguard children and support survivors.

Data Analytics can be combined with compelling narratives to inform, engage, and influence the intended audience. It is critical in turning knowledge into actionable recommendations within the online CSEA ecosystem.

Generally, a good data story should[75]:

- **Be relevant for the audience (including children, parents, and care-givers)**. It should match the current levels of data literacy and knowledge on the topic. It shares what the data says but highlights what matters for the audience.
- **Be based on good data**. The analyst should have confidence in the integrity of their data, understanding the context in which it was collected, recognising bias that exists and including that for others to verify.
- **Deliver a clear, compelling narrative**. It should include a clear, powerful headline and well-constructed key messages that brings out the key meaning of the data for the audience.

[75] https://unstats.un.org/sdgs/data-storytelling/documents/Practical_Guide_to_Data_Storytelling_in_VNRs_and_SDG_Reporting.pdf

- **Use effective and suitable visuals**. All visuals should serve the purpose of helping the audience understand what the data means.

Further guidance on creating effective data storytelling can be found in the UN's Practical Guide to Data Storytelling report[76].

Creating the conditions for cross-country and sector collaboration

The UN Global Platform is a cloud-based collaborative environment. It is designed to support international collaboration in the field of statistics and data science by providing a space where national statistical offices, international organizations, academic institutions, and the private sector can work together on data-related projects.

- **Enhancing statistical capabilities**: Provides enhanced data capabilities, particularly to low-resource countries, including tools, novel data types, technologies, and methodologies. It offers a range of user-friendly data science services, including data cleaning, processing, analysis, or visualization.
- **Promoting data science**: Integrates big data sources with traditional statistical methods. For example this includes satellite imagery, social media data, or mobile phone data.
- **Fostering international collaboration**: Facilitates shared projects and knowledge sharing on specific data challenges. Contains virtual workspaces where teams can communicate, manage projects, and share data.

What does this mean for the online CSEA ecosystem?

This example highlights the importance of enhancing statistical capabilities for low-resource countries by providing advanced tools and user-friendly data science services, such as data cleaning and analysis. Integrating big data sources with traditional methods, like satellite imagery and social media data, enriches the data analysis process. Promoting international collaboration through shared projects and virtual workspaces supporting effective communication and knowledge sharing, enabling teams to address specific data challenges collectively.

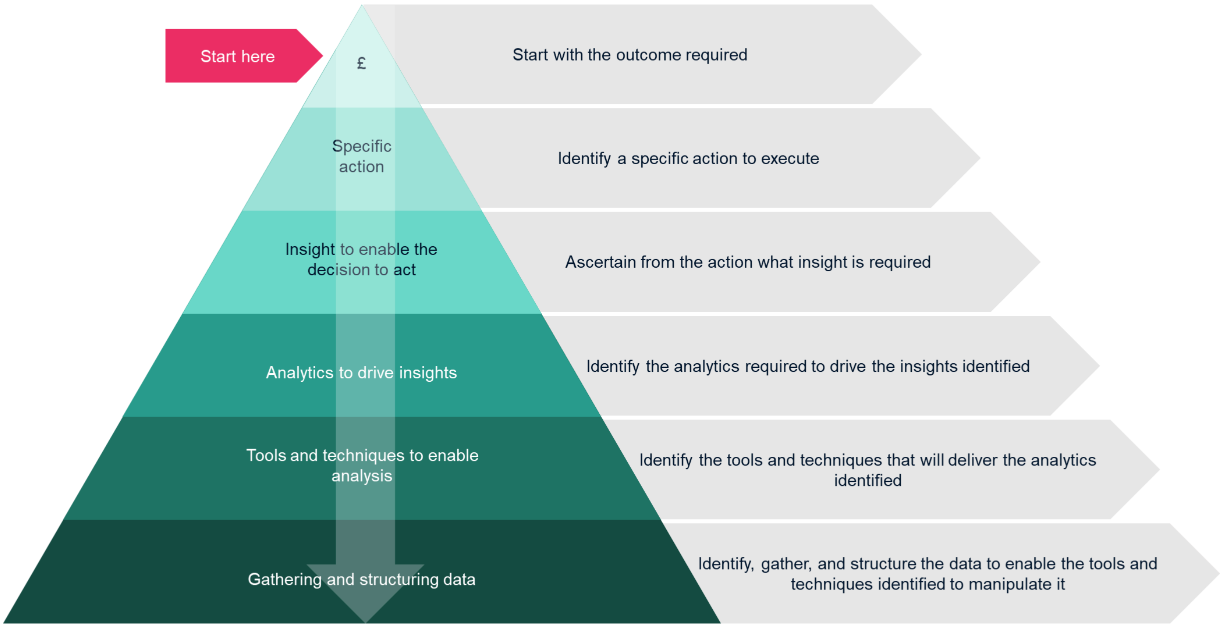
[76] Ibid

How to begin the Data Analytics journey

At an organizational level:

Orient analytics towards outcomes

Focus on the desired result by taking a top-down outcome focus when exercising data-curiousness.



Start here
£
Start with the outcome required
Specific action
Identify a specific action to execute
Insight to enable the decision to act
Ascertain from the action what insight is required
Analytics to drive insights
Identify the analytics required to drive the insights identified
Tools and techniques to enable analysis
Identify the tools and techniques that will deliver the analytics identified
Gathering and structuring data
Identify, gather, and structure the data to enable the tools and techniques identified to manipulate it

Figure 2: The value pyramid for maintaining a top-down outcome focus balanced with data curiousness.

An example of how this is achieved in practice might be through the development of a data dashboard. Prior to its development, start with a table, such as Figure 3, considering the business outcomes and questions that need answering. This ensures that any subsequent inclusion is aligned to an outcome and has a specific purpose.

Business Outcome	**Question to answer**	**Data that answers this**	**Visualization**	**Rationale**
More reliable data sharing	How much of our data was discarded over the past 6 months	Number of discarded files compared with number of sent files	Line chart showing both % and total discarded files.	The number of discards only makes sense when compared with the volume we send, a % is also useful as it could highlight systemic issues

Figure 3: Example table to help align dashboard elements to outcomes

Build solid foundations

Without agreed terms and good quality data, people won't trust the analytics, or if they repeat it, they'll come up with a slightly different answer (e.g. for 'number of employees', some people might include students, some might not). Results should be based on immutable sources, that are transparent and include clear rationale.

An example of this is to define and agree upon universal terms (e.g., "number of employees") or Authoritative Lists for key business terms.

Prioritize data accessibility and understanding

Ensuring that the data ownership and decision-making authority is clear in governance plans to reduce time spent finding the data and the right person to ask for permission on using or sharing it.

Furthermore, it's crucial when doing analysis that the business context and provenance of the data is known, if analysts aren't domain experts, they should consult them. For example, within employee data at an organization an analyst found that the "leaver's date" data looked unusual, with the occasional instance of somebody leaving the organization several weeks after they had passed away. It turns out that, in the business process, that was the date the leaving form was filled in by HR, not when the person actually left.

Implementing fast decision-making on data usage through good governance to empower analytics and providing clear context and metadata for datasets are key foundational components of data analytics.

Build trust through transparent methodologies

To ensure the effectiveness of data analysis in combating child sexual abuse, it is crucial to clearly document the assumptions and limitations of analysis and provide access to the underlying data and calculations when appropriate.

Being transparent about assumptions, potential biases, sensitivities, and uncertainties helps stakeholders understand the context and reliability of the results. Additionally, generating logs at key transformation steps, such as during data ingestion, can help track data integrity by revealing discrepancies between the number of lines in the source file and the records added to the database. This practice aids in identifying and resolving issues promptly, thereby maintaining the accuracy and reliability of the analysis.

Structure and comment for future use

When considering the whole life of a data analytical process, it should be assumed that ownership will transfer to new individuals or entities. As a result, following these steps will help preserve continuity during that transfer:

- Use clear, consistent naming conventions;
- Break complex processes into modular, reusable functions;
- Include inline comments explaining the logic behind key decisions;
- Document data sources, including version and date of extraction; and
- Where possible make use of standard packages and tools such as PowerBI or Kibana for visualization rather than building something bespoke.

Artificial Intelligence (AI) Governance

AI Governance seeks to ensure that AI is developed and implemented in an ethical, fair, secure, and legal way such that it builds trust, supports compliance, and drives innovation.

AI Governance Overview

Governing AI presents some significant additional challenges above and beyond more standard data governance challenges. These include, for example, the risks of creating outcomes that are damaging to individuals or society, of building bias into AI systems, and of infringing human rights such as privacy and property rights. AI Governance seeks to ensure that these risks are managed or mitigated and that the use of AI is human-centric, responsible, transparent and explainable, secure, and those that create and use AI are accountable.

> *AI holds the potential to address complex challenges from enhancing education and improving health care, to driving scientific innovation and climate action. However, AI systems also pose risks to privacy, safety, security, and human autonomy. Effective governance is essential to ensure AI development and deployment are safe, secure and trustworthy, with policies and regulation that foster innovation and competition.*
>
> *Organization for Economic Co-operation and Development (OECD)*[77]

What is AI Governance?

AI Governance seeks to ensure that the development and use of AI is principled and ethical. The OECD published its AI Principles in 2019 (updated in 2024). Other organizations ,including UNESCO Principles[78], have produced similar sets of principles. Organizations typically base their AI governance frameworks on the common principles which broadly are[79]:

1. **Inclusive growth, sustainable development and well-being** which highlights the potential for trustworthy AI to contribute to overall growth and prosperity for all – individuals, society, and planet – and advance global development objectives.
2. **Human-centered values and fairness** which highlights that AI systems should be designed in a way that respects the rule of law, human rights, democratic values

[77] https://www.oecd.org/en/topics/policy-issues/artificial-intelligence.html

[78] https://unsceb.org/sites/default/files/2023-03/CEB_2022_2_Add.1%20%28AI%20ethics%20principles%29.pdf

[79] https://oecd.ai/en/ai-principles

and diversity, and should include appropriate safeguards to ensure a fair and just society.

3. **Transparency and explainability** is about transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes.
4. **Robustness, security, and safety** emphasizes that AI systems must function in a robust, secure and safe way throughout their lifetimes, and potential risks should be continually assessed and managed.
5. **Accountability** stresses that organizations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the OECD's values-based principles for AI.

Why implement AI Governance?

The possible risks associated with using AI are easily overlooked and may be created inadvertently. The possible harms are potentially significant including harms to people, groups, and even society more widely. From an organizational perspective, reputational risk, cultural and societal risk (the assumption that the technology must be right and is therefore not challenged), acceleration risks (that the AI changes in unanticipated ways or too quickly for risks to be foreseen), economic risks (the cost of remediating a large-scale error) and legal risks (breaching regulatory requirements or a class action lawsuit). Appropriate AI Governance is intended to address and manage these risks.

How to begin the AI Governance journey?

At a system level: A well-designed AI governance model should offer tailored incentives and safeguards that align with the unique characteristics of various AI systems and applications, promoting ethical and human rights-based governance. This approach maximizes AI's positive societal impact while mitigating its risks.

Effective AI governance can be achieved through coordinating an ecosystem of key functions - supported by a central organization already managing broader data governance - such as technology development, consensus building through research, stakeholder engagement, standards setting, capacity building, and ongoing monitoring and accountability.

At an organizational level:

The initial steps on the AI Governance journey involves:

1. **Understanding the organizational culture and structure**. This includes understanding the existing governance landscape in, for example, information security and data protection / privacy. It also requires consideration of what sort of AI the organization is considering deploying and the type of organization (i.e. health care, technology, law enforcement).

2. **Developing a proposed AI governance structure** which aligns to the organization's plans and strategy.
3. **Engaging with the relevant stakeholders** including senior management, those actively involved in developing and deploying AI, related governance functions such as privacy, and the users of the AI whether internal, e.g. HR's recruitment team or external, e.g. customers.
4. **Developing the AI Governance Strategy and plan** including establishing AI risk management processes, developing a repository of AI algorithms and applications, and fostering responsible AI accountability processes and procedures.
5. **Determining AI maturity levels** across the organization and uplifting deficient areas.

Operational Resilience

The purpose of Operational Resilience is to ensure that organizations can prevent, respond to, recover, and learn from disruptions and to minimize the risks and impacts from external threats while managing internal pressures. Data resilience is a subset of operational resilience focusing on data-related disruptions such as recovering from data breaches, recovering lost or compromised data assets, and restoring availability of data.

Operational Resilience Overview

Operational resilience helps organizations to thrive in uncertain environments while protecting their capacity to provide their services to stakeholders and partners. This is achieved by developing the capability within an organization to prevent, adapt and respond to operational disruption. From a data perspective, this involves the ability to recover from relatively simple disruptions such as accidently or inadvertently deleted files to moderate disruptions such as a failed server to major disruptions such as a ransomware attack.

What is Operational Resilience?

Operational resilience is the ability of organizations to absorb and adapt to shocks and disruptions. It extends beyond business continuity and disaster recovery. Organizations must have robust plans in place to perform their functions, no matter what the cause of the disruption.

Operational resilience thinks about all the critical dependencies needed to perform their functions - this is beyond just IT and should consider people, process, third parties, premises. Resilient organizations understand all these dependencies and manage them to avoid vulnerabilities, for example, single points of failure.

Why implement Data Resilience?

As internal pressures and external threats continue to disrupt organizations, such as cyber incidents, complex supply chain risks as well as increasing demand from customers and other stakeholders; it's becoming increasingly important for organizations to protect their people, processes, and technologies.

Organizations need to adopt operational resilience at an operational and strategic level to put the organization in the best position to mitigate, react to, and recover from major incidents by delivering rehearsed strategies and effective leadership. This ultimately leads to a robust and adaptive organization.

How to begin the Data Resilience journey?

Embedding operational resilience within an organization require a holistic approach:

1. Policy and strategy need to be updated to incorporate resilience.
2. Governance and operating models need to be aligned so that different segments of an organization work together as seamlessly and as consistently as possible.

3. Disaster recovery and business continuity planning needs to have been incorporated into the organization before disruptions happen.
4. Scenario testing and exercising should take place regularly to ensure that the key players know and have practiced their roles and responsibilities, to surface potential hurdles to managing disruptions so that they can be addressed or mitigated, and to challenge the organization to explore possible sources of disruptions.
5. Consideration should be given to what tooling and reporting are needed to have the visibility needed to identify disruptions and be able to contain them.
6. Appropriate third-party risk management processes need to be in place to identify vulnerabilities and potential sources of disruption which are outside of the organizations immediate control.
7. The policies, standards, processes and procedures need to be in place to ensure that the organization meets (and can demonstrate that it meets) its regulatory requirements.

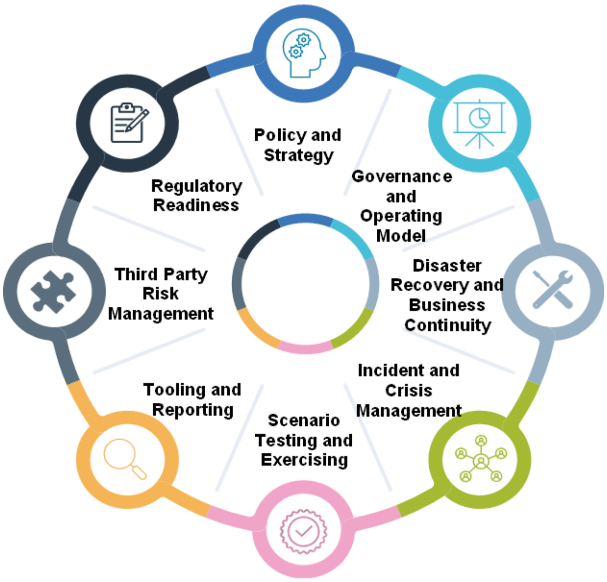

Policy and Strategy
Governance and Operating Model
Disaster Recovery and Business Continuity
Incident and Crisis Management
Scenario Testing and Exercising
Tooling and Reporting
Third Party Risk Management
Regulatory Readiness

Figure 4: A holistic approach to implementing organizational Data Resilience

From a data perspective, this implies, for example, operational policies and strategies include provisions for data, data breaches management aligns information security and privacy requirements, business continuity and disaster recovery plans consider loss of confidentiality, integrity and availability of data, there is a complete data inventory including an evaluation of the value of the data, and the right artifacts are in place to meet data protection and other data-related regulatory requirements.

Training and Awareness

Providing regular training and awareness programs to ensure all employees understand and adhere to data management, security policies, and best practices.

Before you get to training

In terms of adhering to the law, policies, and best practices it is worth thinking about other steps you can take that don't rely on people remembering the right way to do things.

1. **Make the compliant or correct process the easiest one**. Ideally, engineer-in the process electronically. This could mean having a single form or electronic process to follow that is easy to find, easy to fill in and gets responded to quickly. If staff are searching for paper forms and navigating complex procedures, they might seek simpler alternatives.
2. **Consider turning processes into a flow chart rather than pages of text**. This is a good test to ensure the process can be followed and is explicit enough. Poor policy documents can be contradictory and therefore impossible to know how to follow.
3. **Implement just-in-time reminders**. Individuals are likely to overlook annual mandatory training; therefore, it is more effective to provide reminders of the appropriate actions at the moment they are being performed.
4. **Tell people why something is important**. This is especially important with cyber security training; without an appreciation of the threat, data security measures can seem cumbersome and unnecessary.
5. **Reward the right behaviors** – rather than punishing non-compliance.

What is training?

It can be helpful to think of training and awareness on data as improving your organization's ability to manage data as an asset (this is distinct from data science, data literacy or analytics training). Good asset management focuses on the three pillars of **cost**, **risk**, and **performance**. Training should help staff optimize all three regarding the data they handle.

Risk

Most mandatory or annual training focuses on risk reduction. There are legal and policy requirements for handling data in the online CSEA domain (as well as risks faced by every organization) to ensure:

- Evidence is handled appropriately to maximize any conviction.
- Personal data is handled appropriately (e.g. GDPR, Data Protection Act).
- Staff operate in accordance to specific laws related to Child Protection (e.g. Online Safety Act, The Protection of Children Act).
- Data is held securely, and computer networks are well protected.

Risk training should focus on what relevant laws and policies apply to data – and what are the main risks an organization is trying to mitigate?

Performance

This should cover good data management practices (for example, data is easy to find, well documented and organized), resulting in minimal delays in finding, using, or understanding data. This is where organizations can communicate their data governance approach, supported by robust and accessible documentation.

Cost

The most pertinent element of cost from an individual's perspective relates to cloud computing environments, as storage and compute costs are significant to organizations. Training should include raising awareness of the incurred costs of how the handle and process data, maximizing the efficiency of how these resources are used.

Building a learning community

To ensure best practice is shared within (and across) organizations, organizations should consider setting up data communities. This not only provides opportunities to learn from others, but it also provides a support network people can turn to answer questions and solve problems improving system capacity.

Performance Monitoring and Reporting

Performance monitoring and reporting in data management involves the continuous tracking and assessment of data-related processes to ensure they meet defined standards and objectives. This includes analyzing metrics such as data accuracy, consistency, and timeliness, and generating reports to provide insights into the effectiveness of data management practices. These insights guide decision-making and drive improvements in data quality and operational efficiency.

What is high-performance for a non-profit?

How do you measure performance when you can't use profit? How do you ensure what you are measuring doesn't lead to undesirable behaviors – 'gaming the system' for better metrics?

Online CSEA is a complex system, a state with many parts that are interrelated, with no single organization controlling the outcomes, and there are unpredictable feedback loops; you can't predict what will happen with any intervention.

What should you be measuring?

At a system level: There are several conditions that are prerequisites for producing healthy systems, these are[80]:

- **Shared purpose and principles**: partners in a place are aligned around a common purpose that cuts across and provides the motivation for their work.
- **Trusting relationships**: people and organizations are connected and develop honest authentic relationships as a foundation for working together.
- **Collaborative behaviors**: people across the system value collaboration, and work in a connected way.
- **Sharing power**: actions are taken to address imbalances of power and gain diverse perspectives. Decisions are devolved as close to the ground as possible.
- **Systems infrastructure**: processes and structures shift from an organizational to systems focus e.g. workforce, commissioning, governance and data.
- **Enabling leadership**: leaders see their role as creating enabling conditions for collaborative approaches.
- **Learning and insight**: a learning culture focused on experimentation, convening and collective sense-making as a driver of improvement and building trust.
- **Embedding and influencing**: people and partners are motivated to improve, embed and influence for the adoption of these practices more widely.

[80] https://collaboratecic.com/insights-and-resources/transforming-local-places-the-potential-of-human-learning-systems/

Any efforts to measure success should start with purpose: monitoring (any metrics and qualitative reporting) should help an organization make decisions and answer questions with their purpose in mind.

What organizations measure should help them learn and adapt, not meet targets. Targets can take the focus away from purpose, and the people behind it, driving the wrong behaviors and emphasizing achieving a number. A well-documented example of targets driving the wrong performance is the use of targets in UK policing[81].

> *When a measure becomes a target, it stops being a useful measure.*

Broadly speaking there are three overarching questions to answer when considering an organization's impact within a complex system.

Is the system achieving its purpose?

As well as metrics for an individual organization's performance, metrics should consider how the system is performing as a whole.

Is the system healthy and resilient to change?

This could describe overall levels of motivation and empowerment, or collaboration levels between stakeholders, levels of shared learning, or measure of adaptability.

Is your organization delivering value for money?

Recognizing there is limited funding available within the online CSEA system to share between stakeholders, is an organization achieving its purpose efficiently and therefore offering value for money?

Further detail on these broad questions includes:

1. Outcomes
   - Is purpose being achieved? These should be high level and enduring outcome-orientated aims both for organizations and the wider system. These may be proxy measures that an organization is unlikely to have complete control over but should be contributing towards them.
2. System Health
   - What is the health of relationships between organizations and within individual organizations? Are there high levels of collaboration, sharing, and trust?
   - Are workforces motivated, content and do they understand their organization's purpose?

[81] https://assets.publishing.service.gov.uk/media/5a8165bfed915d74e33fdfc7/Review_Targets_2015.pdf

3. Inputs
    - Monitoring an organization's finances and resources consumed, tracking whether it is operating sustainably and efficiently.
4. Outputs
    - Is an organization achieving what it said it would?
    - What is the ratio of output to input, is it delivering value for money?

Prioritizing learning rather than measuring and monitoring

In the UK, East Sussex's initiative to tackle loneliness demonstrated how monitoring could extend beyond metrics. They focused on trying to learn why loneliness was an issue rather than measure it. This helped them uncover systemic issues around awareness and trust between organizations that were inhibiting progress[82].

What does this mean for the online CSEA ecosystem?

This example highlights the importance of incorporating systems thinking and root cause analysis

[82] https://collaboratecic.com/case-studies/east-sussex/

Conclusions

This document has explored key concepts and best practices in data management. Each example shares lessons of what could be implemented within the online CSEA system, or increased in scale, to potentially mitigate the barriers and challenges this system faces. This section will summarize key themes from the examples, these range from the establishment of system governance, new technical intermediary data platforms, to placing children further into the heart of the system.

Establishing a system governance body is essential for collectively advancing outcomes for children by better use of data. Other sectors have had significant success in leveraging data after creating an independent, global system governing body, that oversees, provides guidance, or otherwise advises on:

- Data security
- Data ethics, particularly concerning the use of AI
- Standardizing and simplifying legal frameworks
- Regulatory development, leveraging these developments to drive further standardization
- System level performance monitoring and reporting
- Innovation and technical development, particularly reducing the risk of deduplication
- Capacity building across stakeholders and geographies
- Inter-stakeholder collaboration, including facilitating the sharing of resource, challenge, or expertise
- Unifying taxonomies and definitions

A system governing body should be continually responsive to the evolving needs of stakeholders, ensuring that their interests are recognized and addressed. It must be trusted by all parties, earning confidence through transparency and consistent actions. Members should be fully committed to the body's mission, actively supporting its objectives. The governing body should also maintain a neutral stance, avoiding any conflicts of interest by not acting as a data consumer, producer, or owner. This impartiality is crucial for maintaining its integrity and the trust of its stakeholders.

Implementing shared stewardship models such as the Better Deal for Data, or driving cultural change like Data-Pop Alliance's Mobilize programme, should improve the way data is understood and valued across the online CSEA ecosystem. Doing this whilst developing a system wide data strategy (e.g. The Data Values Project), principles of ethical frameworks (e.g. Ada Lovelace Institute: Biometrics Council) will help create a more equitable distribution of value from data - including for children.

Children should be kept at the heart of the system, empowering them through increased control and transparency over the collection and use of their data, leveraging the benefits of informed and age relevant consent[83], and being mindful of the impact on outcomes for children during decision making and in the design of data infrastructure.

Leveraging the momentum seen in other sectors in putting individuals at the center of data ecosystems, utilizing initiatives such as the UN's International Decade of Data, should improve the value of data for children.

Data should be seen less as an organizational asset, but as a resource for social good. Proportionate data sharing can be critical for protecting those that are vulnerable, but the ethics privacy and independent rights can be circumstantial. These barriers have been shown to be surmountable, with strong data literacy and an awareness of ethics within the system increasing the likelihood of appropriate and proportional requests for data sharing. Once global access to data is democratized, innovation, insights, and outcomes flourish. This can be replicated for online CSEA, focus should be on creating value from data for all stakeholders, in a way that creates an intrinsic link between social good and individual value.

Creating a minimum standard for entry into a data sharing system creates trust between members and creates more opportunities for collaboration. Systems that have set robust eligibility criteria demonstrate the necessary trust between members that is essential for sharing, increasing interoperability and increasing the ease at which partnerships can scale, in terms of data volume, types or new members. This can require working closely with individual countries to ensure that local laws are respected.

Initiatives such as the Tech Coalition's Lantern programme is a good example of how minimum barriers to entry can engender trust between organizations within online CSEA.

Integrity, privacy, and authenticity of data are foundational for forming trusted relationships between stakeholders. Being able to demonstrate high standards surrounding data management, particularly transparency and accountability, is necessary to engender confidence in your data and in your outputs based on other's data. This includes metadata standards, which might include descriptions of sources, methodology, or data context, ultimately helping analysts understand and correctly create their own outputs. Successful systems have validation and verification processes for data and outputs, demonstrating clear and appropriate use of the data. This can include the use of independent oversight boards, increased training and awareness for staff, or technical solutions like data signature technology.

Utilizing existing collaborations, such as those like the World Bank Development Data Partnership, to facilitate the relationships between different stakeholders will help reach common understandings that emphasize data sharing for social good whilst building trust.

[83] Age relevant consent accounts for a child's developmental stage and ability to make informed choices, in some instances this might mean consent mediated by parents or caregivers rather than the child themselves.

Technical solutions are already available that can enhance interoperability. The examples from successful systems included intermediary platforms for aggregating and normalizing data from diverse sources or using standardized APIs to access and exchange data across different platforms.

Implementing technical principles, such as those by Data-Pop Alliance, increases interoperability in the design of solutions within online CSEA for data sharing. Encoding data usage agreements in trust networks will unlock barriers currently preventing sharing at a large scale.

Developing comprehensive data pictures from multiple, diverse, data sources provide a more accurate and holistic understanding of needs, behaviors, and circumstances. By integrating data from various sources, it also helps identify patterns and insights that might be missed when considering data in isolation, leading to better decision-making and outcomes. This can lead to more personalized and effective services.

Open-access repositories of data are one mechanism that could be employed within the online CSEA ecosystem, mirroring their success in democratizing health (Global Health Data Exchange) or environmental (Global Biodiversity Information Facility) data collected by private industry.

Develop a bounded, focused solution that is scalable, then grow it. The examples demonstrate the requirement to bound your initial solution, for example to a single nation, set of core members, or data type. This should be focused on demonstrating value to all stakeholders, with well-designed data architecture that can scale. Once this has been successfully implemented, then the solution can iterate and scale.

Annex A: A template for capturing best practice during stakeholder engagements

Name:		Organization:	
Title of good practice:		Useful links:	
Description:			