GREI Workshop 2023

Supporting Research Data Sharing at the Organizational Level: A Wholistic Mindset

JANUARY 24, 2023

Robin Champieux, MLIS
Interim University Librarian
Oregon Health & Science University (OHSU)

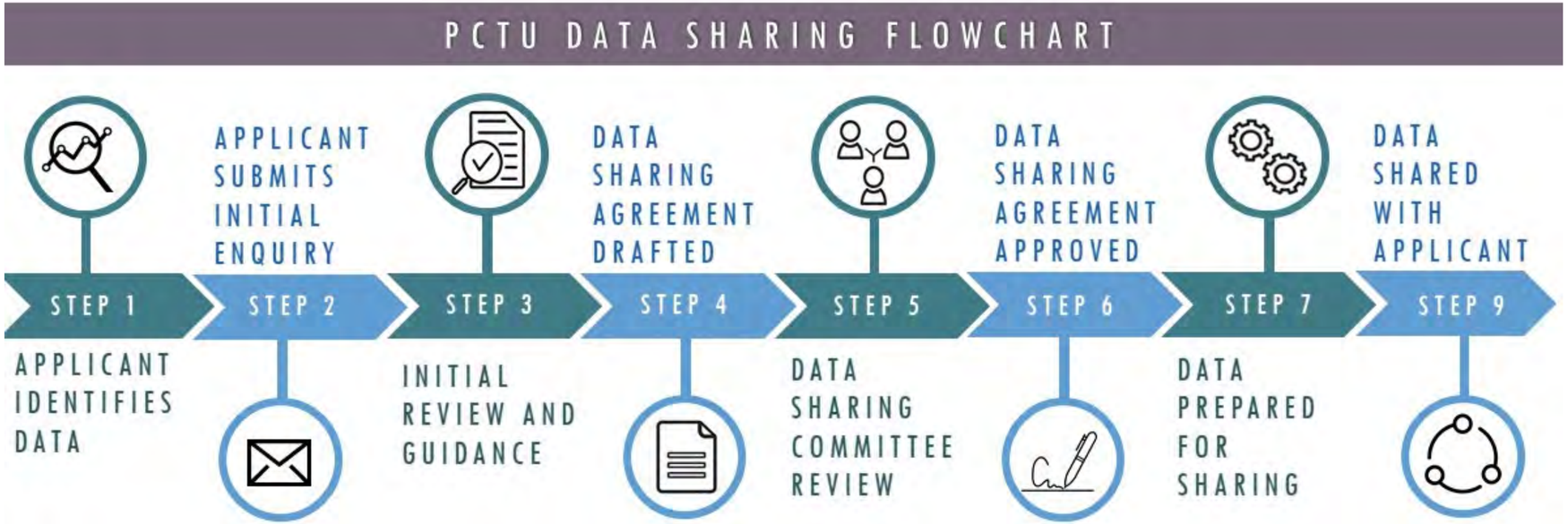


Image source: Queen Mary, University of London, Pragmatic Trials Unit
https://www.qmul.ac.uk/pctu/collaborate-with-us/data-sharing/make-a-data-sharing-request/

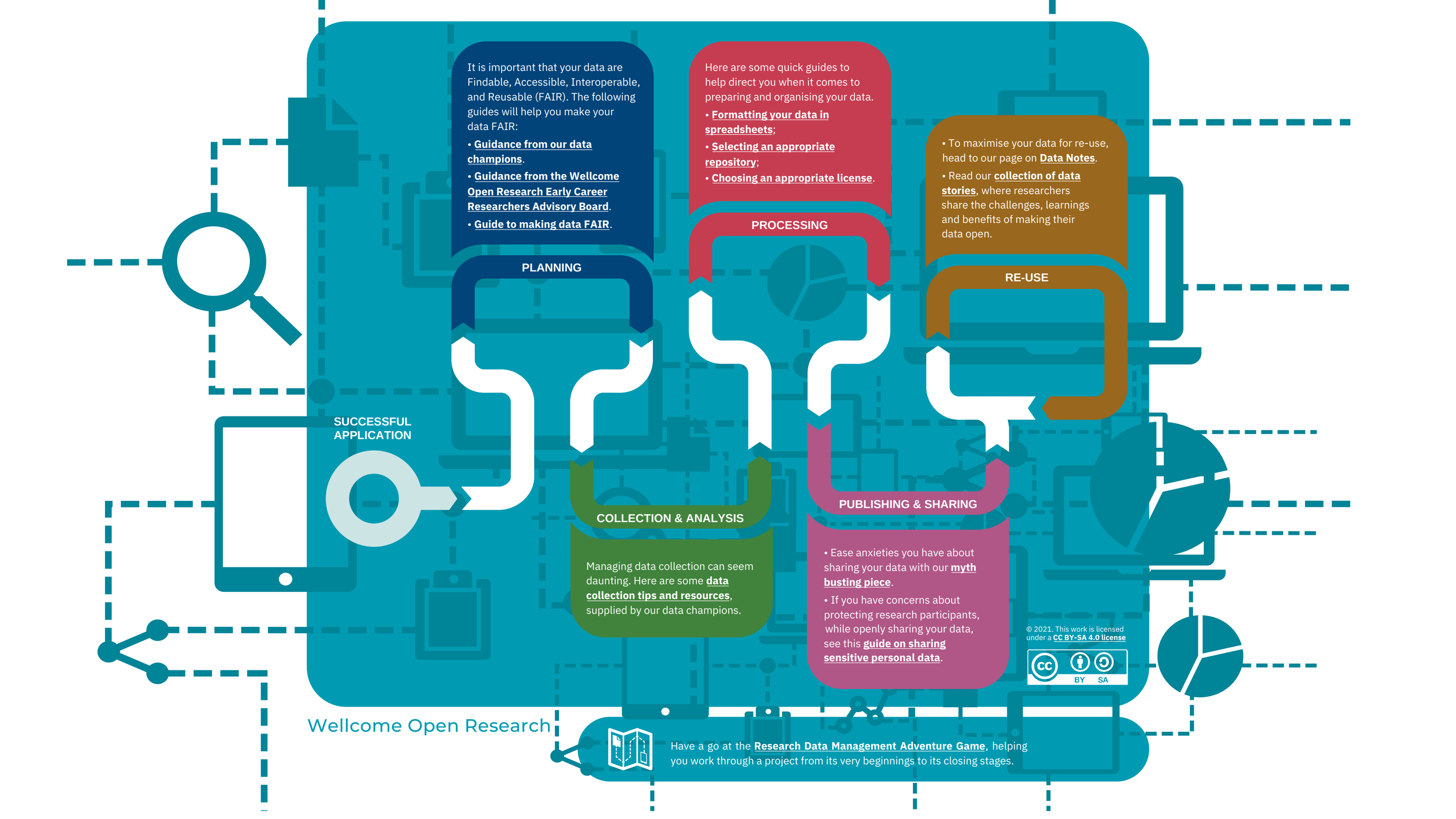

It is important that your data are Findable, Accessible, Interoperable, and Reusable (FAIR). The following guides will help you make your data FAIR:
• Guidance from our data champions.
• Guidance from the Wellcome Open Research Early Career Researchers Advisory Board.
• Guide to making data FAIR.
PLANNING
SUCCESSFUL APPLICATION
Here are some quick guides to help direct you when it comes to preparing and organising your data.
• Formatting your data in spreadsheets;
• Selecting an appropriate repository;
• Choosing an appropriate license.
PROCESSING
COLLECTION & ANALYSIS
Managing data collection can seem daunting. Here are some data collection tips and resources, supplied by our data champions.
PUBLISHING & SHARING
• Ease anxieties you have about sharing your data with our myth busting piece.
• If you have concerns about protecting research participants, while openly sharing your data, see this guide on sharing sensitive personal data.
• To maximise your data for re-use, head to our page on Data Notes.
• Read our collection of data stories, where researchers share the challenges, learnings and benefits of making their data open.
RE-USE
© 2021. This work is licensed under a CC BY-SA 4.0 license
Wellcome Open Research
Have a go at the Research Data Management Adventure Game, helping you work through a project from its very beginnings to its closing stages.

Image source: UC Berkeley, Berkeley Research Data Management Program
https://researchdata.berkeley.edu/service-areas

Productive DMS is supported by an interdependent set of factors

TECHNOLOGIES

POLICIES

PRACTICES

VALUES

PEOPLE

DMS is harder and less worthwhile when we don't understand and nurture this ecosystem

More Labor

Making decisions and engaging with best practices and available resources is harder

More Risk

Lots of bespoke solutions and a lack of assessment can lead to mistakes

Hard to prioritize

Good things things aren't scaled

The labor of doing and supporting DMS is greater

Less impactful results

Less Innovation

Less ROI

Maturity models can help us understand & grow organizational qualities that impact DMS.

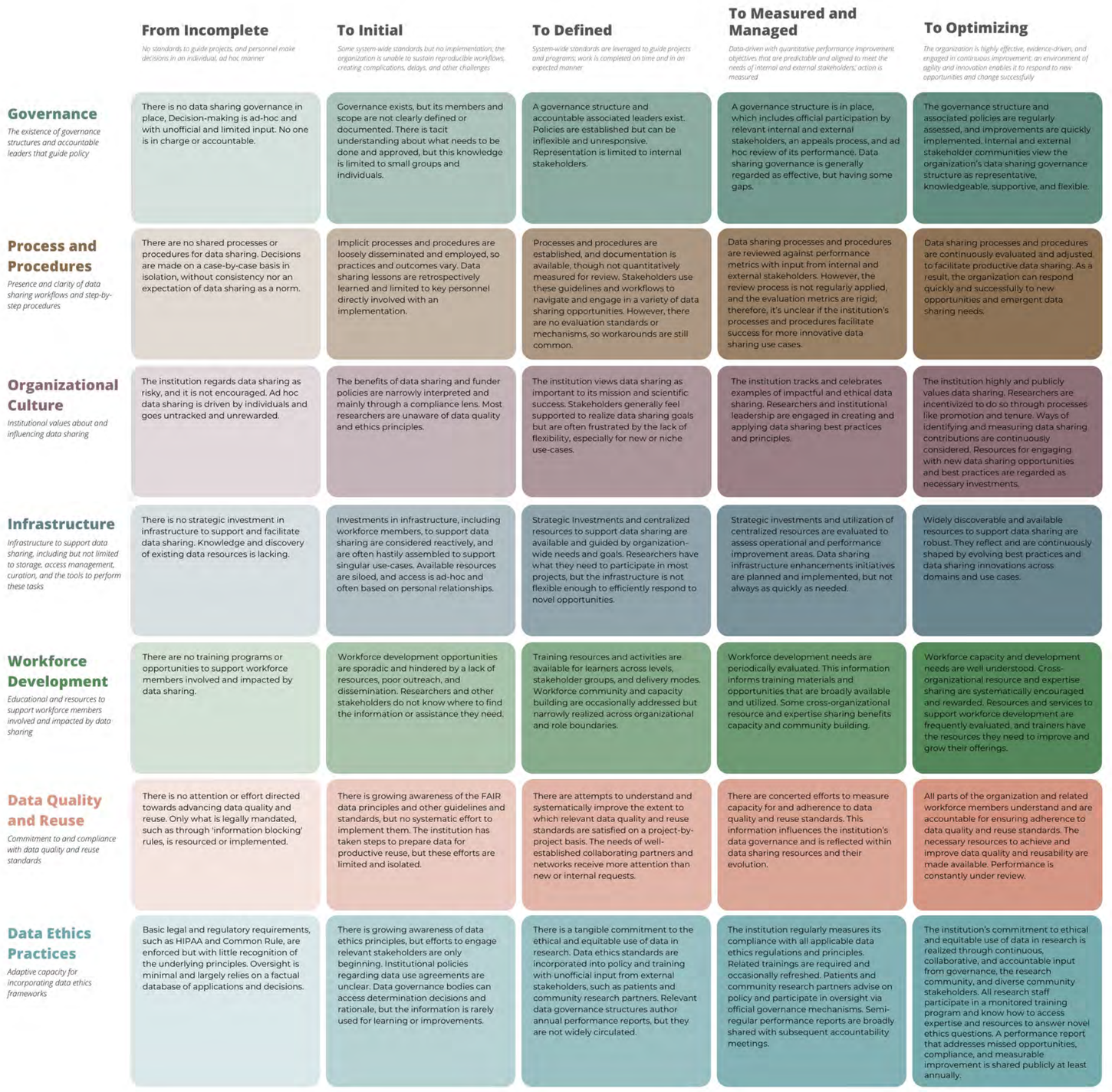
- A framework for describing and evaluating the processes, structures, technology, culture, and people associated with and enabling growing effectiveness in an area of focus,
- Used to identify strengths and weaknesses and to generate improvement plans.
- Describe the factors or domains associated with an area of focus along a scale of increasing capacity.

Thus, any organization can identify itself in a model and grade its current capabilities in each domain.

Crowston K, Qin J. A capability maturity model for scientific data management: Evidence from the literature. Proceedings of the American Society for Information Science and Technology. 2011;48(1):1–9.

Research Data Sharing: A Maturity Model for Organizational Capacity

https://doi.org/10.5281/zenodo.7369811



	From Incomplete *No standards to guide projects, and personnel make decisions in an individual, ad hoc manner*	To Initial *Some system-wide standards but no implementation, the organization is unable to sustain reproducible workflows, creating complications, delays, and other challenges*	To Defined *System-wide standards are leveraged to guide projects and programs; work is completed on time and in an expected manner*	To Measured and Managed *Data-driven with quantitative performance improvement objectives that are predictable and aligned to meet the needs of internal and external stakeholders; action is measured*	To Optimizing *The organization is highly effective, evidence-driven, and engaged in continuous improvement; an environment of agility and innovation enables it to respond to new opportunities and change successfully*
Governance *The existence of governance structures and accountable leaders that guide policy*	There is no data sharing governance in place, Decision-making is ad-hoc and with unofficial and limited input. No one is in charge or accountable.	Governance exists, but its members and scope are not clearly defined or documented. There is tacit understanding about what needs to be done and approved, but this knowledge is limited to small groups and individuals.	A governance structure and accountable associated leaders exist. Policies are established but can be inflexible and unresponsive. Representation is limited to internal stakeholders.	A governance structure is in place, which includes official participation by relevant internal and external stakeholders, an appeals process, and ad hoc review of its performance. Data sharing governance is generally regarded as effective, but having some gaps.	The governance structure and associated policies are regularly assessed, and improvements are quickly implemented. Internal and external stakeholder communities view the organization's data sharing governance structure as representative, knowledgeable, supportive, and flexible.
Process and Procedures *Presence and clarity of data sharing workflows and step-by-step procedures*	There are no shared processes or procedures for data sharing. Decisions are made on a case-by-case basis in isolation, without consistency nor an expectation of data sharing as a norm.	Implicit processes and procedures are loosely disseminated and employed, so practices and outcomes vary. Data sharing lessons are retrospectively learned and limited to key personnel directly involved with an implementation.	Processes and procedures are established, and documentation is available, though not quantitatively measured for review. Stakeholders use these guidelines and workflows to navigate and engage in a variety of data sharing opportunities. However, there are no evaluation standards or mechanisms, so workarounds are still common.	Data sharing processes and procedures are reviewed against performance metrics with input from internal and external stakeholders. However, the review process is not regularly applied, and the evaluation metrics are rigid; therefore, it's unclear if the institution's processes and procedures facilitate success for more innovative data sharing use cases.	Data sharing processes and procedures are continuously evaluated and adjusted to facilitate productive data sharing. As a result, the organization can respond quickly and successfully to new opportunities and emergent data sharing needs.
Organizational Culture *Institutional values about and influencing data sharing*	The institution regards data sharing as risky, and it is not encouraged. Ad hoc data sharing is driven by individuals and goes untracked and unrewarded.	The benefits of data sharing and funder policies are narrowly interpreted and mainly through a compliance lens. Most researchers are unaware of data quality and ethics principles.	The institution views data sharing as important to its mission and scientific success. Stakeholders generally feel supported to realize data sharing goals but are often frustrated by the lack of flexibility, especially for new or niche use-cases.	The institution tracks and celebrates examples of impactful and ethical data sharing. Researchers and institutional leadership are engaged in creating and applying data sharing best practices and principles.	The institution highly and publicly values data sharing. Researchers are incentivized to do so through processes like promotion and tenure. Ways of identifying and measuring data sharing contributions are continuously considered. Resources for engaging with new data sharing opportunities and best practices are regarded as necessary investments.
Infrastructure *Infrastructure to support data sharing, including but not limited to storage, access management, curation, and the tools to perform these tasks*	There is no strategic investment in infrastructure to support and facilitate data sharing. Knowledge and discovery of existing data resources is lacking.	Investments in infrastructure, including workforce members, to support data sharing are considered reactively, and are often hastily assembled to support singular use-cases. Available resources are siloed, and access is ad-hoc and often based on personal relationships.	Strategic investments and centralized resources to support data sharing are available and guided by organization-wide needs and goals. Researchers have what they need to participate in most projects, but the infrastructure is not flexible enough to efficiently respond to novel opportunities.	Strategic investments and utilization of centralized resources are evaluated to assess operational and performance improvement areas. Data sharing infrastructure enhancements initiatives are planned and implemented, but not always as quickly as needed.	Widely discoverable and available resources to support data sharing are robust. They reflect and are continuously shaped by evolving best practices and data sharing innovations across domains and use cases.
Workforce Development *Educational and resources to support workforce members involved and impacted by data sharing*	There are no training programs or opportunities to support workforce members involved and impacted by data sharing.	Workforce development opportunities are sporadic and hindered by a lack of resources, poor outreach, and dissemination. Researchers and other stakeholders do not know where to find the information or assistance they need.	Training resources and activities are available for learners across levels, stakeholder groups, and delivery modes. Workforce community and capacity building are occasionally addressed but narrowly realized across organizational and role boundaries.	Workforce development needs are periodically evaluated. This information informs training materials and opportunities that are broadly available and utilized. Some cross-organizational resource and expertise sharing benefits capacity and community building.	Workforce capacity and development needs are well understood. Cross-organizational resource and expertise sharing are systematically encouraged and rewarded. Resources and services to support workforce development are frequently evaluated, and trainers have the resources they need to improve and grow their offerings.
Data Quality and Reuse *Commitment to and compliance with data quality and reuse standards*	There is no attention or effort directed towards advancing data quality and reuse. Only what is legally mandated, such as through 'information blocking' rules, is resourced or implemented.	There is growing awareness of the FAIR data principles and other guidelines and standards, but no systematic effort to implement them. The institution has taken steps to prepare data for productive reuse, but these efforts are limited and isolated.	There are attempts to understand and systematically improve the extent to which relevant data quality and reuse standards are satisfied on a project-by-project basis. The needs of well-established collaborating partners and networks receive more attention than new or internal requests.	There are concerted efforts to measure capacity for and adherence to data quality and reuse standards. This information influences the institution's data governance and is reflected within data sharing resources and their evolution.	All parts of the organization and related workforce members understand and are accountable for ensuring adherence to data quality and reuse standards. The necessary resources to achieve and improve data quality and reusability are made available. Performance is constantly under review.
Data Ethics Practices *Adaptive capacity for incorporating data ethics frameworks*	Basic legal and regulatory requirements, such as HIPAA and Common Rule, are enforced but with little recognition of the underlying principles. Oversight is minimal and largely relies on a factual database of applications and decisions.	There is growing awareness of data ethics principles, but efforts to engage relevant stakeholders are only beginning. Institutional policies regarding data use agreements are unclear. Data governance bodies can access determination decisions and rationale, but the information is rarely used for learning or improvements.	There is a tangible commitment to the ethical and equitable use of data in research. Data ethics standards are incorporated into policy and training with unofficial input from external stakeholders, such as patients and community research partners. Relevant data governance structures author annual performance reports, but they are not widely circulated.	The institution regularly measures its compliance with all applicable data ethics regulations and principles. Related trainings are required and occasionally refreshed. Patients and community research partners advise on policy and participate in oversight via official governance mechanisms. Semi-regular performance reports are broadly shared with subsequent accountability meetings.	The institution's commitment to ethical and equitable use of data in research is realized through continuous, collaborative, and accountable input from governance, the research community, and diverse community stakeholders. All research staff participate in a monitored training program and know how to access expertise and resources to answer novel ethics questions. A performance report that addresses missed opportunities, compliance, and measurable improvement is shared publicly at least annually.

Research Data Sharing: A Maturity Model for Organizational Capacity

	From Incomplete	To Initial	To Defined	To Measured and Managed	To Optimizing
	No standards to guide projects, and personnel make decisions in an individual, ad hoc manner	*Some system-wide standards but no implementation; the organization is unable to sustain reproducible workflows, creating complications, delays, and other challenges*	*System-wide standards are leveraged to guide projects and programs; work is completed on time and in an expected manner*	*Data-driven with quantitative performance improvement objectives that are predictable and aligned to meet the needs of internal and external stakeholders; action is measured*	*The organization is highly effective, evidence-driven, and engaged in continuous improvement; an environment of agility and innovation enables it to respond to new opportunities and change successfully*
Organizational Culture *Institutional values about and influencing data sharing*	The institution regards data sharing as risky, and it is not encouraged. Ad hoc data sharing is driven by individuals and goes untracked and unrewarded.	The benefits of data sharing and funder policies are narrowly interpreted and mainly through a compliance lens. Most researchers are unaware of data quality and ethics principles.	The institution views data sharing as important to its mission and scientific success. Stakeholders generally feel supported to realize data sharing goals but are often frustrated by the lack of flexibility, especially for new or niche use-cases.	The institution tracks and celebrates examples of impactful and ethical data sharing. Researchers and institutional leadership are engaged in creating and applying data sharing best practices and principles.	The institution highly and publicly values data sharing. Researchers are incentivized to do so through processes like promotion and tenure. Ways of identifying and measuring data sharing contributions are continuously considered. Resources for engaging with new data sharing opportunities and best practices are regarded as necessary investments.
Infrastructure *Infrastructure to support data sharing, including but not limited to storage, access management, curation, and the tools to perform these tasks*	There is no strategic investment in infrastructure to support and facilitate data sharing. Knowledge and discovery of existing data resources is lacking.	Investments in infrastructure, including workforce members, to support data sharing are considered reactively, and are often hastily assembled to support singular use-cases. Available resources are siloed, and access is ad-hoc and often based on personal relationships.	Strategic Investments and centralized resources to support data sharing are available and guided by organization-wide needs and goals. Researchers have what they need to participate in most projects, but the infrastructure is not flexible enough to efficiently respond to novel opportunities.	Strategic investments and utilization of centralized resources are evaluated to assess operational and performance improvement areas. Data sharing infrastructure enhancements initiatives are planned and implemented, but not always as quickly as needed.	Widely discoverable and available resources to support data sharing are robust. They reflect and are continuously shaped by evolving best practices and data sharing innovations across domains and use cases.

Support for DMS must be met with continuous investment in technical and social infrastructure.

- Infrastructure capacity differs by organization, and investment will reflect local priorities and means.
- Shared and reusable data sharing infrastructure allows institutions to leverage economies of scale and access infrastructure that would otherwise be out of reach or peripheral compared to other needs.
- The benefits are greater than the cost savings. Community-based governance and management that underpin shared resources drive innovation and engagement with best practices.

We can leverage GREI and generalist repositories to grow institutional DMS capacity.

There's assurance (and success) in meeting people where **they and we** are at.

Recent reflections on taking heart

Standards, what standards?

Let's partner on this.

Someone else has got this covered.

Contact Me
champieu@ohsu.edu

Research Data Sharing: A Maturity Model for Organizational Capacity

https://doi.org/10.5281/zenodo.7369811