How value flows from health data

The [Open Data Institute (ODI)](#) has been working with WHO for the past month, holding discussions with 23 stakeholders and analysing 56 documents**.**

Data landscaping can help to identify the data stewards responsible for managing and ensuring access to a dataset, the different types of data users, and the relationships between them. This can help to communicate where and how the use of data creates value.

"A data-driven economy also means new information-based health systems. The pooled and analysed data allows personalised decisions concerning prevention or treatment – by healthcare facilities along the care path and the patients themselves. Data is also an essential resource for building new services tailored to current challenges and expectations."

SITRA (2020), 'Towards trustworthy health data ecosystems'
https://www.sitra.fi/en/publications/towards-trustworthy-health-data-ecosystems/

"Creating a map of a data ecosystem can help us to understand and explain where and how the use of data creates value. A data ecosystem map can help to identify the key data stewards and users, the relationships between them and the different roles they play. Representing ecosystems in detailed maps can be particularly useful when contexts are complex, not well understood or not yet fully developed… Mapping requires you to consider different actors, relationships and ideas in the system, and can generate useful insights and talking points. As a collaborative process it can build understanding of a data ecosystem across different stakeholders. The end product is useful as a communication tool to support engagement across the data ecosystem."

GODAN (2019)
'Data Ecosystem Mapping to Support the FAIR Data Principles'
https://www.youtube.com/watch?v=LKiIMTHDpoo

How data, trust and value flow between stakeholders in the data landscape
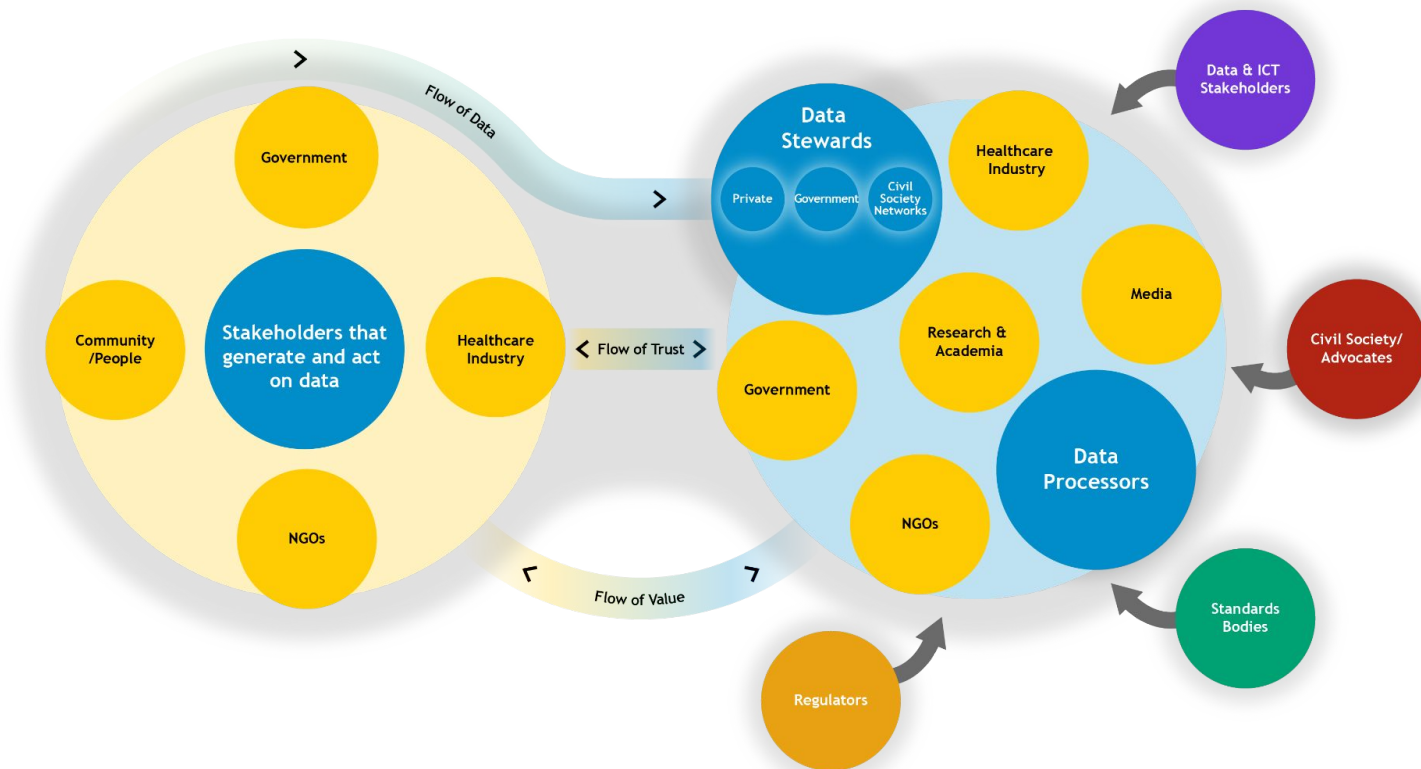
Stakeholders generate and act on data

Data is acted upon by:
- **People,** who manage their own health and wellbeing
- **Health workers, healthcare organisations, and non-government organisations (NGOs),** to support and advocate for better health outcomes
- **The healthcare industry,** which creates innovative healthcare interventions and solutions
- **Decision-makers, governments and regulators,** that develop health policy, optimise healthcare service delivery and spending, and ensure safe and secure health systems.

Data is also used in analysis by:
- **Researchers and academia**, who conduct studies and evaluate impacts
- **The media,** which keeps everyone informed
- **Private companies, civil society networks and advocates,** and **industry and data analysts** to generate new insights.

Use of data is a constant interplay that occurs at a personal level and at a societal level; within local communities, at country and regional levels, and globally.

In some cases, this use of data is explicit: people and institutions understand they are drawing on available health data to make decisions. At other times, the data is a background resource and decisions are automated directly from the data. Data on health services capacity and workforce levels could be used to optimise PPE supply chains, for example, without manual data handling by any one stakeholder.

Sources: Stakeholder groupings were identified from health data ecosystem models developed by: Vayena, Dzenowagis, Langfeld (2016); Serbanati, Ricci, Mercurio & Vasilateanu (2011); Focsa & Mihalas (2012); International Science Council; FCDO; U.S. National Institutes of Health; Digital Enlightenment Forum Europe; GTZ Africa Partnership; Open Data Institute, and Rand Corporation; https://www.healthcareitnews.com/news/cios-seek-analytics-automation-ehr-integration-healthier-supply-chains

Health data can generate value

When data can flow amongst multi-stakeholders in a secure, safe, ethical and equitable manner, trust can be fostered, and value generated.

The impact and outcomes that can be generated from this use of health data can be grouped into six main categories:
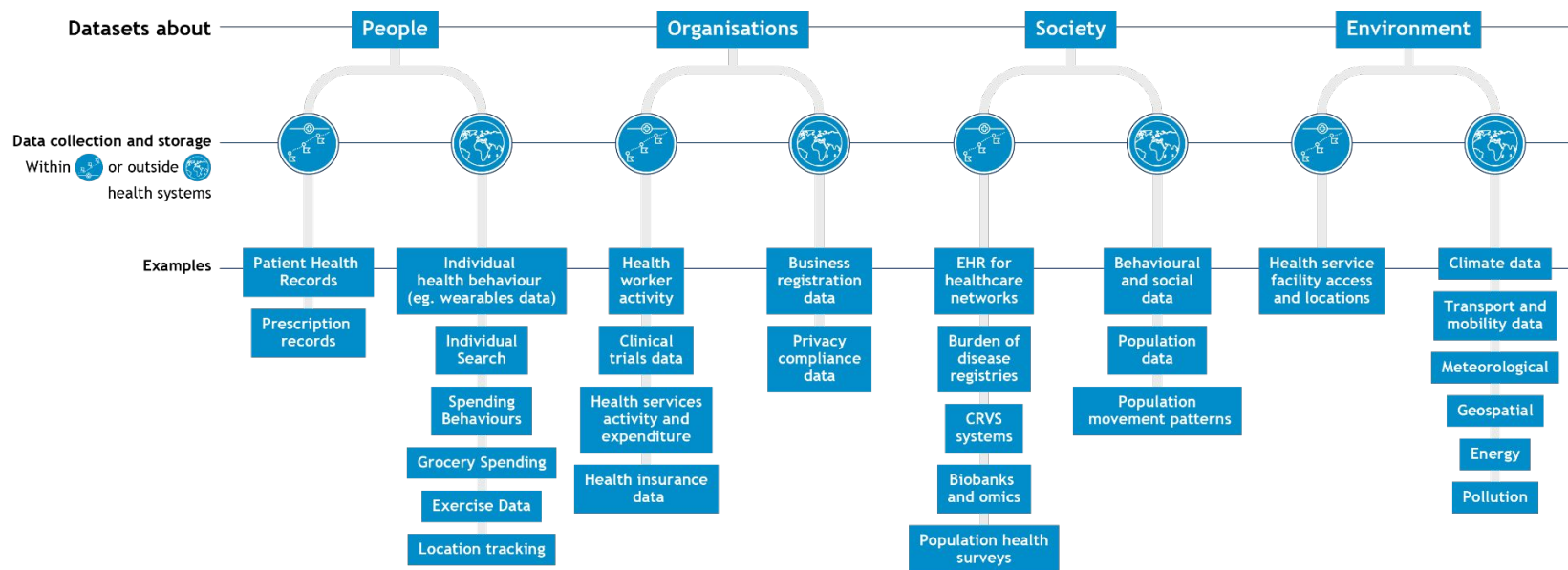
- **Better health outcomes** throughout the patient journey
- **Optimised health service** and resource efficiencies
- **Increased participation**, understanding of connections and collaborations, and sense of engagement and involvement in health decision-making by communities and all stakeholders
- **Groundbreaking innovation**, new industry opportunities and deeper insights into existing systems
- **Reduced inequalities** through using data to better target services and disaggregate data to understand impacts on vulnerable populations
- **Commercial advantage, professional recognition** and support for researchers and other stakeholders that collect, use and share data

Reduced Inequality

Greater patient-public participation

Better health outcomes

Optimised health system

Expanded innovation

Private profits and professional prestige

Stakeholder segments have different use cases and pain points when using health data

Stakeholder segment	Why they want to use data						Challenges in using data
	Reduced Inequalities	Greater Patient-Public Participation	Better Health Outcomes	Optimised Health System	Expanded Innovation	Private Profits and Professional Recognition	
Community /People	●	●	●				Data literacy Trust Informed consent
Healthcare industry Research and academia NGOs	●		●		●	●	Management of data across the data journey Data sharing and access Interoperability
Government	●		●	●			Interoperability Timely data access
Media		●					Data literacy Data sharing and access

Datasets are collected about about people, organisations, societies, and the environment

Datasets about

| People | Organisations | Society | Environment |

Data collection and storage

Within or outside health systems

Examples

People:
- Patient Health Records
- Prescription records
- Individual health behaviour (eg. wearables data)
- Individual Search
- Spending Behaviours
- Grocery Spending
- Exercise Data
- Location tracking

Organisations:
- Health worker activity
- Business registration data
- Clinical trials data
- Privacy compliance data
- Health services activity and expenditure
- Health insurance data

Society:
- EHR for healthcare networks
- Behavioural and social data
- Burden of disease registries
- Population data
- CRVS systems
- Population movement patterns
- Biobanks and omics
- Population health surveys

Environment:
- Climate data
- Transport and mobility data
- Meteorological
- Geospatial
- Energy
- Pollution

Data is about people, organisations, societies, and the environment

Datasets are generated through societal and digital infrastructures and through interactions with services.

Core datasets such as the civil registries and vital statistics (CRVS) datasets are minimum requirements for a global data system and are more valuable as digitised assets.

The WHO 'Evolving health data ecosystem' framework, developed in 2016, shows a non-exhaustive list of key datasets (standard and expanded) that are useful to collect, manage and share in comprehensive health data landscapes.
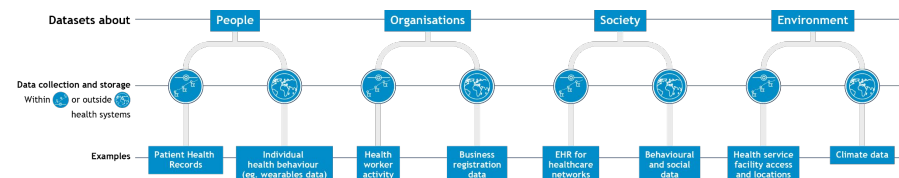
Evolving health data ecosystem
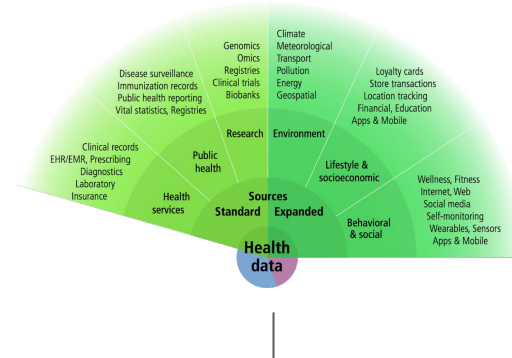
Datasets listed in WHO's 'Evolving data ecosystem model' can be grouped into data about people, organisations, society and environment and grouped by whether they are collected within or outside the health system, as shown above.

Sources: WHO (2020) 'SCORE for health data technical package: global report on health data systems and capacity', https://apps.who.int/iris/handle/10665/339125; WHO (2020) 'Civil registration and vital statistics', https://www.who.int/data/data-collection-tools/civil-registration-and-vital-statistics-(crvs); Vayena E, Dzenowagis J, Brownstein JS, Sheikh A. Policy implications of big data in the health sector (2017), https://www.who.int/bulletin/volumes/96/1/17-197426.pdf

At a personal level, data may link to a personal identifier to reflect and collate data 'about a person'

Various types of health data are collected about individual people and about society as a whole.

Health data is often collected in a clinical setting, such as through an electronic healthcare record. For individuals, this is sensitive data on their health, and their engagement with health services and healthcare workers. Behavioural data can also be sensitive, especially when it relates to health subjects, such as online search behaviour for sexual health information or purchasing behaviour.

At a societal level, this data may be aggregated to collect population-level health data about burdens of disease, disease incidence and prevalence, and health service usage. Because it often originates from sensitive and behavioural data, it can still feel like it is personal data when it is being opened and shared.

Creating and proving trustworthiness that societal data has been de-identified and is safe to share is a prerequisite of being able to collect personal data and generate value at a societal level.
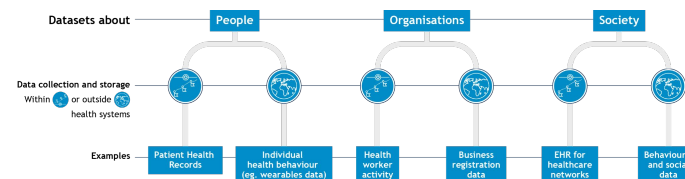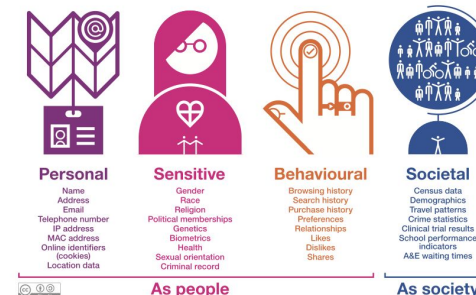
Sources: ODI(2019), 'About Data About Us (case study),
https://theodi.org/article/about-data-about-us-case-study/

At an environmental level, there are a range of influences on health outcomes

WHO has listed 12 environmental health areas where action can be taken to improve health outcomes.

The UN Environment Programme's proposed 'Digital Ecosystem for the Planet' outlines key datasets and governance processes that align with this health data landscape model.
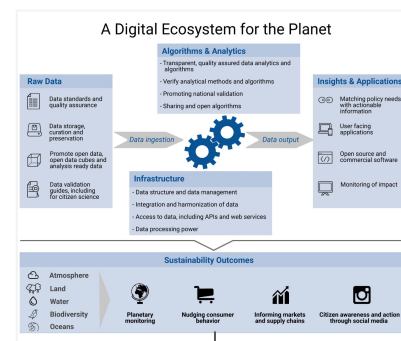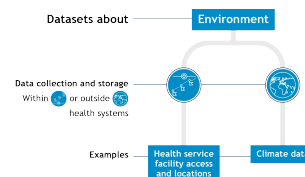
The UN Environment Programme has outlined an environment digital data ecosystem framework that has similar components and information goals as a health data landscape could provide

The WHO global strategy on health, environment and climate change identifies a non-exhaustive list of key sectors that collect data with relevance to health

Sources: UN Environment Programme (2019), 'The Case for a Digital Ecosystem for the Environment', https://www.un-spbf.org/wp-content/uploads/2019/03/Digital-Ecosystem-final.pdf, WHO (2020), 'WHO Global Strategy on Health, Environment, and Climate Change', https://www.who.int/publications/i/item/9789240000377

# Data is managed internally via the data journey

The data journey involves:
- Collection
- Processing
- Use and action

Processing in particular may involve multiple steps, often in a non-linear process where tasks are passed back and forth.

While one organisation may internally manage a dataset across the data journey, more frequently, different steps are carried out by different entities. This creates challenges for each organisation working together with a dataset throughout the data journey:

- **Cultural challenges** and differences in expectations of openness and data access
- **Trust challenges** with the expected validity and legitimacy of the data itself, and in processes to share data between stakeholders
- **Technical challenges** in the interoperability and comparability of data
- **Privacy and security challenges** in how data is stored and access permitted
- **Legal challenges**, where cross-jurisdictional organisations may have differing regulations or requirements for where data can be stored and accessed
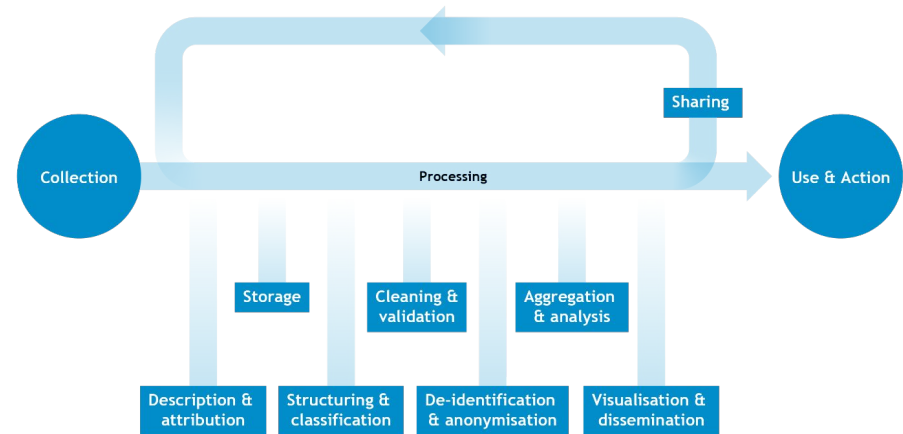
Sources: WHO (2021) ' WHO's data and statistics journey', https://cdn.who.int/media/docs/default-source/world-health-data-platform/consultancies/tor_datagovernancesummit_production_12may2021.pdf?sfvrsn=5fd234ba_11, OpenDataWatch (2019), ``The Data Value Chain: Moving from Production to Impact', https://data2x.org/wp-content/uploads/2019/05/Data_Value_Chain_OnePager.pdf

Entities that take on the role of collecting data and making it available are often referred to as 'data stewards'
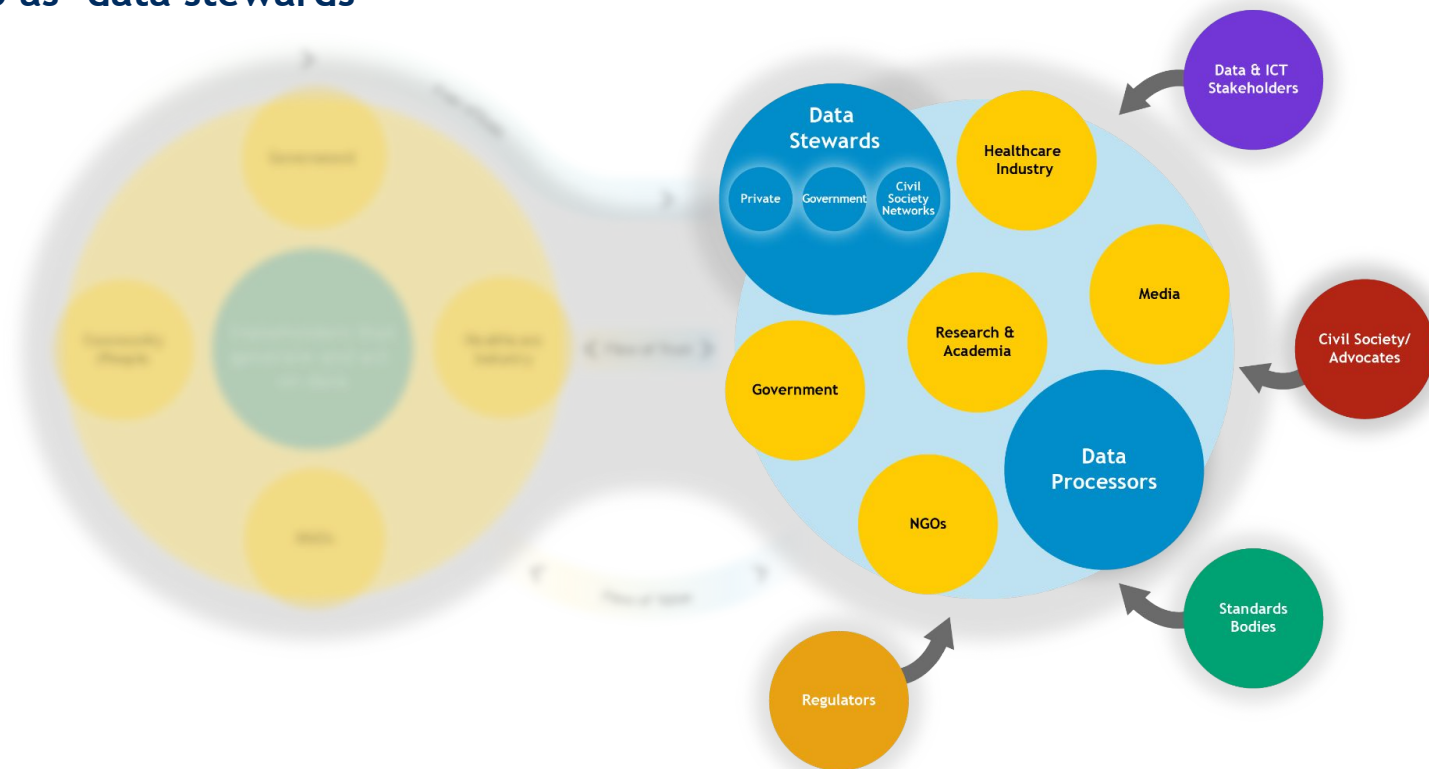
Data stewards are entities that collect, manage, and share data

Data stewards:

- Collect data, either directly or from other data stewards
- Manage data, in particular cleaning it and creating statistics, and storing it
- Share data with other data stewards and data processors who perform further analysis, and create useful analyses, visualisations, products and services

Data stewards may be state actors, such as governments; private companies, such as those that collect and manage wearable health and fitness data or social media; or civil society organisations, such as social justice networks that conduct slum dweller censuses to ensure visibility of marginalised communities.

The United Nations Economic and Social Council (ECOSOC)'s Report on the Friends of the Chair Group on Economic Statistics recommends 'national statistical offices should adopt the role of data stewards, in addition to being data producers, and actively establish partnerships across the public and private sectors'.
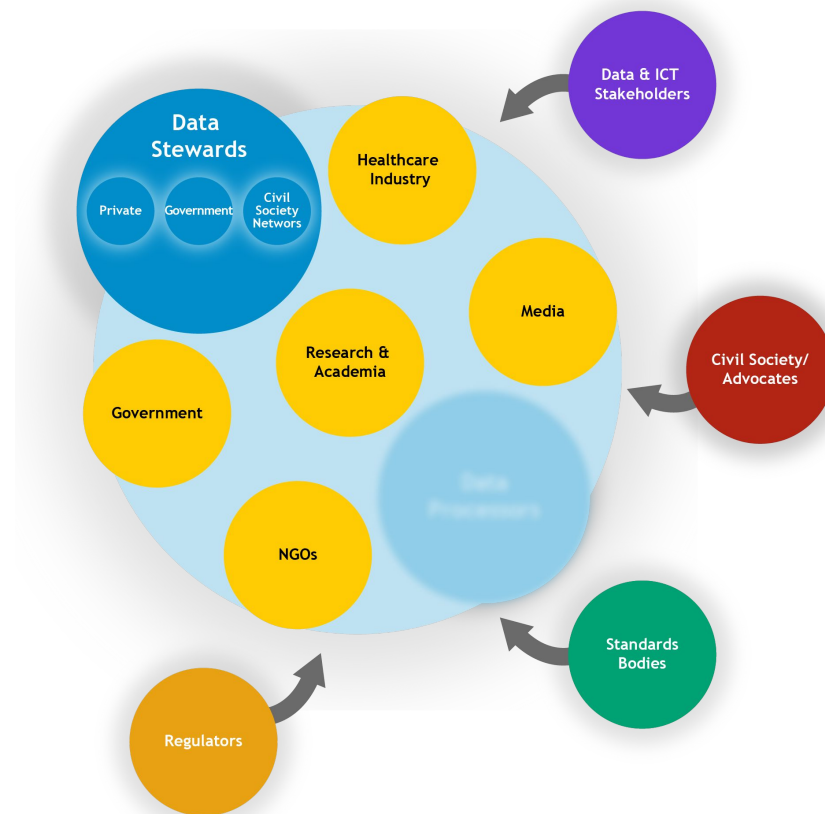
Sources: UN (2020) 'Fifth Meeting of the Friends of the Chair Group on Economic Statistics', https://unstats.un.org/unsd/statcom/groups/focg_econ/meetings/5th/default.cshtml

Data processors take data from data stewards for two main reasons

- To process it (this may involve cleaning and validation, statistical analysis, and securing data storage)
- When using data to create solutions for stakeholders, including reports, visualisations, dashboards and digital applications

Making data available requires robust data sharing agreements as part of an overall data governance approach that builds on best practices in managing the data journey.
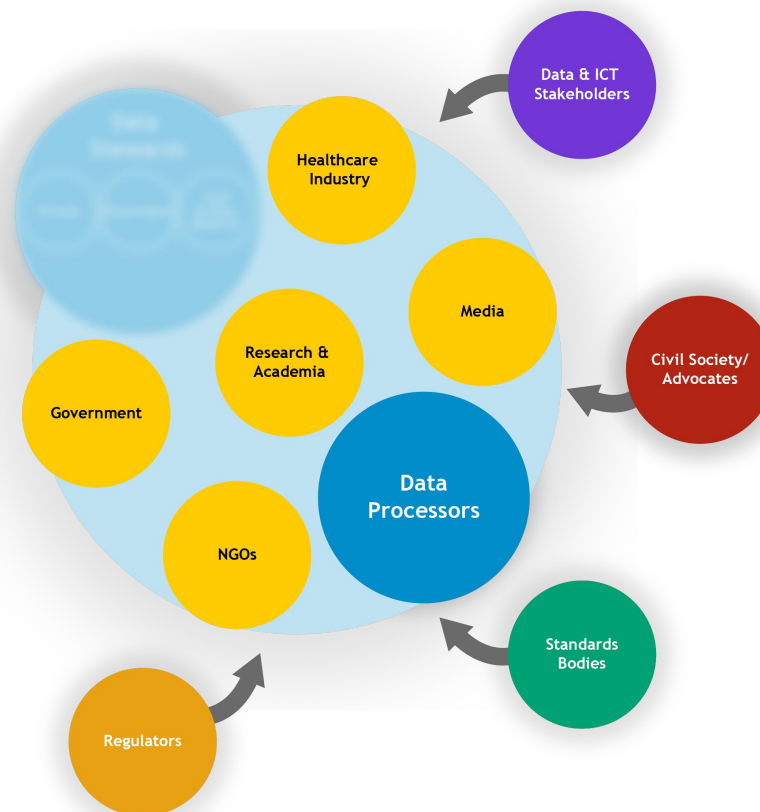
Underpinning these relationships between data stewards, processors and data users are data governance processes and promising practices that help ensure that personal data is respected and protected, and that data is able to generate the expected value

Data governance processes ensure that data stewards and other stakeholders can ensure data collection, management and use is ethical, and that data helps reduce inequalities in access to health and the resources for healthy living.
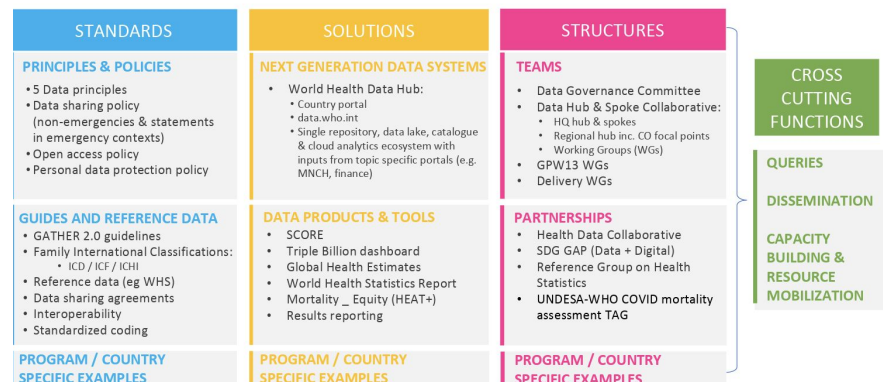
See the pre-read on Data Governance maturity and best practices for discussion of the key processes and components.

Data & Health Statistics Governance Framework
The necessary standards, solutions and structures that ensure quality and integrity of all aspects of WHO's data & Health Statistics

| STANDARDS	SOLUTIONS	STRUCTURES	CROSS CUTTING FUNCTIONS

PRINCIPLES & POLICIES
• 5 Data principles
• Data sharing policy (non-emergencies & statements in emergency contexts)
• Open access policy
• Personal data protection policy
	NEXT GENERATION DATA SYSTEMS
	• World Health Data Hub:
	  • Country portal
	  • data.who.int
	  • Single repository, data lake, catalogue & cloud analytics ecosystem with inputs from topic specific portals (e.g. MNCH, finance)
		TEAMS
		• Data Governance Committee
		• Data Hub & Spoke Collaborative:
		  • HQ hub & spokes
		  • Regional hub inc. CO focal points
		  • Working Groups (WGs)
		• GPW13 WGs
		• Delivery WGs
			QUERIES
			
			DISSEMINATION
			
			CAPACITY BUILDING & RESOURCE MOBILIZATION

GUIDES AND REFERENCE DATA • GATHER 2.0 guidelines • Family International Classifications: • ICD / ICF / ICHI • Reference data (eg WHS) • Data sharing agreements • Interoperability • Standardized coding	**DATA PRODUCTS & TOOLS** • SCORE • Triple Billion dashboard • Global Health Estimates • World Health Statistics Report • Mortality _ Equity (HEAT+) • Results reporting	**PARTNERSHIPS** • Health Data Collaborative • SDG GAP (Data + Digital) • Reference Group on Health Statistics • UNDESA-WHO COVID mortality assessment TAG	
PROGRAM / COUNTRY SPECIFIC EXAMPLES	**PROGRAM / COUNTRY SPECIFIC EXAMPLES**	**PROGRAM / COUNTRY SPECIFIC EXAMPLES**	

Sources: WHO (2021) ' WHO's data and statistics journey',
https://cdn.who.int/media/docs/default-source/world-health-data-platform/consultancies/tor_datagovernancesummit_production_12may2021.pdf?sfvrsn=5fd234ba_11,

Data and ICT providers offer technologies that data stewards can use to more easily fulfil their role in managing and making data available for ethical, equitable and effective use

Tools can be open source, created for specific use cases, or can be proprietary, tech startup tools that are purchased.

Technology solutions can also help address security concerns and create granular permission controls and defined 'sandboxes' where data can be analysed in a controlled environment. Technologies can enable anonymisation and de-identification of data, speeding up opportunities to share and access data.

Managing the technology to meet data governance needs and to make data available can be costly to maintain and has implications for ongoing resourcing.
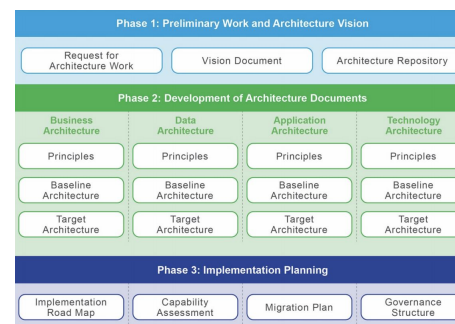
Tanzania's Health Enterprise Architecture aligns principles, governance structures, and health systems service delivery together to ensure technology supports public health goals through appropriate use and sharing of data.

"WHO recognizes Enterprise Architecture as a method of assessing or describing technology adoption in the health sector. For member countries to achieve health information services improvement, WHO recommends the following development steps: architecture vision, business architecture, information systems architecture, technology architecture, opportunities and solutions, migration planning, implementation governance, and architecture change management."

Tanzania Health Enterprise Architecture
version 1, September 2020

Sources: http://hidl.afya.go.tz/#/dashboard/initiatives?gn=7&ans=11;
http://api-hidl.afya.go.tz/uploads/library-documents/1617092274-3I9GuToO.pdf;
http://api-hidl.afya.go.tz/uploads/library-documents/1621600291-eciyv3jK.pdf

Regulators also play a key role ensuring that data stakeholders abide by privacy regulations and use data in an ethical manner aligned with the legislative context

An independent regulatory authority can assess and ensure accountability in addition to any accountability and auditing processes that data stewards and other data entities may implement internally.

In low- and middle-income countries, there are limited governance frameworks in place that enable protection of personal and community data.

STRUCTURAL

Global principles
and norms

STRUCTURAL

Data trusts,
data intermediaries
and data institutions

STRUCTURAL

Public-patient
participation
processes

STRUCTURAL

Ethics and
data governance
committees

Regulators

LEGAL

Data sharing
contracts

LEGAL

Global IP and
data governance
frameworks

LEGAL

Data audit,
certification and
assurance

LEGAL

Open data
licensing

GISAID is a non-profit led, public-private partnership with clear data licensing arrangements to ensure attribution is given to those researchers contributing data

Genome sequencing platform GISAID allows variant tracking, and is used by governments to inform their outbreak response, and by the healthcare industry to shape vaccine design and therapeutic interventions.

The value of health data
- Over 180 countries have shared genomic sequences of COVID-19 variants on the GISAID platform. To date 2 million genomes have been catalogued.

- The platform is the most trusted for data sharing of COVID-19 genome data, with African and South American scientific contributions to the platform more than doubling between January - April 2021.

- Interactive real-time platforms and variants are able to be tracked for policy and therapeutic intervention planning. Data has been used to assess vaccines and diagnostic medical devices.

Stakeholders
Research and academia, governments, country health ministries, policy makers, technology partners (Seqiris, a7digital, QIAGEN)
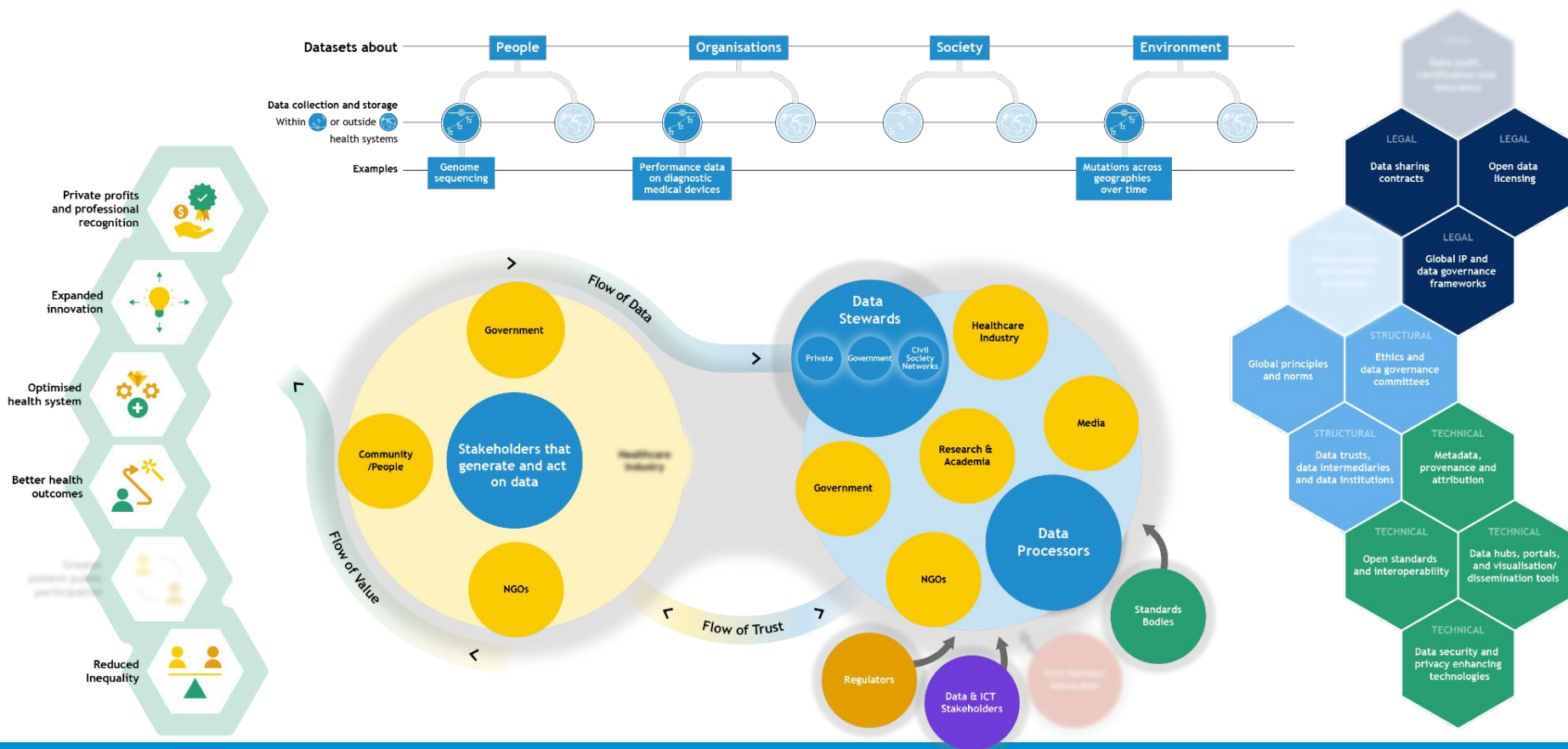
Governance processes
Structural: The GISAID Initiative receives administrative support from Freunde von GISAID e.V. a registered non-profit association, organised and operated exclusively for charitable, scientific and educational purposes. The GISAID Initiative's activities are governed by several organisational bodies that operate independently of each other. One critical element to the success of the GISAID Initiative is its ability to provide for a fair and transparent, as well as a verifiable and unbiased mechanism not only to govern, but to take measures and guard against bias in decision-making, preserving scientific independence.

Legal: GISAID uses licence conditions that ensure any research publications give credit to the data contributors, reducing international imbalances in recognition and participation in scientific advancement that impact on willingness to share data.

Technical: The GISAID platform uses standards to ensure data can be compared and aggregated and provides a range of data visualisation and dissemination tools.

Sources: https://www.gisaid.org/hcov19-variants/;
https://www.gisaid.org/collaborations/enabled-by-hcov-19-data-from-gisaid/;
https://www.nature.com/articles/d41586-021-00305-7;
https://www.nature.com/articles/d41586-021-01194-6

Key elements of GISAID's health data landscape

As the value from health data collection and use is generated, it is recognised more widely and generates further trust amongst people, communities and societies. In turn, this encourages increased data sharing so that value can be further increased in a virtuous cycle

We can use this model to show:
- Opportunities to collaborate
- Elements that can be strengthened and matured
- Implementation examples from best practices

Thank you to all informants and participants
for sharing your time and expertise