Enterprise Knowledge Graph Foundation

EKG/Manifesto

Manifesto & Principles
for the Enterprise Knowledge Graph

February 2, 2022, Version 0.2.55

Contents

1 Manifesto **2**
- 1.1 Ideals . . . 2
- 1.2 All data connected . . . 2
- 1.3 Standards . . . 3
- 1.4 Data Markets . . . 3
- 1.5 Open World . . . 3

2 Principles **5**
- 2.1 Identity . . . 5
- 2.2 Meaning . . . 6
- 2.3 Distributed . . . 7
- 2.4 Open World . . . 7
- 2.5 Self-describing . . . 7
- 2.6 Measurement . . . 8
- 2.7 Business Orientation . . . 8
- 2.8 Control . . . 8
- 2.9 Ecosystem . . . 8
- 2.10 Standards . . . 9

3 Mapping to FAIR data principles **10**
- 3.1 Findable . . . 10
- 3.2 Accessible . . . 10
- 3.3 Interoperable . . . 11
- 3.4 Reusable . . . 11

Appendices **11**

Acronyms **14**

Glossary **15**

Index **16**

Chapter 1

Manifesto

As a reference, we include the manifesto chapter of the EKG/Manifesto document here as an appendix.

1.1 Ideals

Should we specify higher level ”ideals” to better explain what drives us? This is a section of ”reasons why” in non-data / non-tech terms. This section is very early days and needs much more work.

We welcome your input here.

1. Much higher levels of transparency, sustainability, fairness and accountability are going to be required at all levels in any organization
2. Human Capital needs to be known, valued, leveraged and optimized
3. Data Capital needs to be known, valued, leveraged and optimized
4. Increasing competitiveness depends more and more on having the highest quality and depth of data, information and knowledge
5. One ”censored”, biased, Single Version of (the) Truth (SVOT) is no longer good enough for many — if not most — use cases in most domains
6. The world becomes more and more polarised due to ”information bubbles” that many people are not escaping from, more depth and connectedness is needed
7. The world becomes more and more complex and harder to understand, a holistic view around every given topic, showing all viewpoints, would help people to understand and make better decisions

1.2 All data connected

1. All data will be connected.
    - Information, Knowledge, Meaning: it’s all data.
    - Knowledge & Meaning will be captured as machine-readable executable models.
2. All data will be made available anywhere — within entitlement limits — at any time to any device, node or edge.
3. We embrace the *Open World*[1] and deal with the realities of “multiple versions of the truth (MVOT)”.
4. We combine the digital footprint of activities along with a digital representation of information and knowledge, from which an Enterprise Knowledge Graph (EKG) emerges.
5. All your connected data is an EKG.
6. An EKG is connections of Knowledge Graphs across an Enterprise and beyond.
7. We encourage the use and widespread proliferation of EKG identifiers

[1] Does not mean all data is open to everyone

1.3 Standards

1. EKGs are based on standards and therefore interoperable across boundaries.
2. There are many different types of standards that an EKG needs to be able to deal with.
3. Standards are described as machine-readable models — i.e. ontologies — that EKG/Platforms can execute, interpret or enforce.

1.4 Data Markets

1. Any data source will be turned into a data publisher — or supplier — of one or more self-describing datasets (SDDs).
    - Any data sink will be turned into a data consumer — of one or more datasets.
2. Data suppliers and consumers will find each other via a data market using a standard ”lingua franca” for the data itself, its meaning, all its associated policies and metadata and especially also its use cases as executable models.
3. The data market manages the information supply chains between all the various suppliers and consumers.
4. The global data market will consist of many other more specific data markets e.g. per industry or per enterprise.
5. An EKG is the combination of one or more data markets and the deployment of its use cases.

1.5 Open World

1. For any given ”Thing” — i.e. an object — there may be many representations in many datasets.
2. An object’s representations may be different in shape, meaning, timeliness, relevance and quality — i.e. any given representation of information about a given object may represent a different version of the truth.
3. All representations of any given object shall be linked via shared identifiers.
    - Identifiers shall be meaningless, opaque, web-resolvable and universally unique. See principle 2.1.
    - An object can have multiple identifiers.
4. Any given object consists of 1 or more Data Points.
5. A Data Point represents a logical property of a given object.
    - The identifier for a Data Point is the identifier of the object it belongs to plus at least one identifier of the axiom that describes its meaning (which is also an object).
6. Data Points for the same object can exist in many different datasets.
    - with potentially multiple versions of the truth in terms of meaning, timeliness, relevance and quality.
7. Any representation for any given Data Point of any data source shall be made available to any device, node or edge in the network within legal, policy and entitlement limits in real-time.
8. Every ”object” that is represented as data in whichever dataset anywhere, shall have an identifier that is universally unique, permanent, meaningless or opaque and therefore shareable, resolvable through the HTTP protocol. See 2.1 Identity.

Work in progress notes: that add more explanation to the above:

- Data will be considered explained when its usage has no misconceptions nor ambiguities.
- The word “data” has a lot of different notions associated with it. We have these statements above, like “all data is connected”, “…data will be made available anywhere…”, which could reinforce a particular notion that data is something that “exists” all around us, like a digital footprint of activities. On top of these statements on “data”, the EKG is introduced as a combination of one or more data markets, which could further reinforce that ”EKG is connected data”. If that’s the notion that the manifesto wants to declare, it may appeal strongly to some sections more than others, like those dealing with consolidation, analysis, reporting, verification etc. If the manifesto’s intent is broader (here, I am very conscious of Carl’s note that it should stay away from any hubris!) it would help if there is a way to declare that information and knowledge is also data. And it is when we combine the digital footprint of activities along with a digital representation of information and knowledge, that a Knowledge Graph emerges. And on top of that notion, an Enterprise Knowledge Graph is connections of Knowledge Graphs across an Enterprise. The

intent of the above is to appeal broadly to different sections of an Enterprise, accommodating different norms and notions (again referring to Carl's insightful comment on Norms and Notions)

Chapter 2

Principles

As a reference, we include the principles chapter of the EKG/Manifesto document here as an appendix.

As a reference, we include the principles chapter of the EKG/Manifesto document here as an appendix.

These principles are intended to provide guidelines for the development and deployment of an Enterprise Knowledge Graph (EKG). The principles emphasize shared meaning and content reuse that are the cornerstone of operating in complex and interconnected environments.

2.1 Identity

Any object in the EKG is identified with at least one universally unique, opaque, permanent, non-reassignable and web-resolvable identifier in the form of an IRI (Internationalised Resource Identifier) for the EKG — i.e. an "Enterprise Knowledge Graph IRI (EKG/IRI)".

The EKG/IRI identifier is permanent, will be proliferated across the company's universe (including ecosystem), and will be used for the expression of facts about the object including relationships between objects.

Additional non-EKG identifiers may also be assigned, and they may be human-readable, "external" to the organization's EKG and be transient and reassignable.

Resolving an identifier can be done in three ways:

1. using it in a transaction — i.e. a query or update statement — submitted or routed via an internet protocol to a "lookup service" that translates one or more given "features" of an object to an EKG/IRI.
2. constructing it via a standardized policy from key components and applying a hash and optionally signing it — where the object represented by the EKG/IRI may or may not already exist.
3. constructing it by giving the object an EKG/IRI based on a random number in case the EKG is the authoritative source for the given object.

Rationale While the semantic web technologies — like Resource Description Framework (RDF) — generally allow for many and varied "Internationalised Resource Identifiers (IRIs)", and this is still encouraged when integrating systems, there is benefit in being able to rely on one canonical and unchanging one, which can for example make the mapping of identities a many-to-one rather than a many-to-many task.

In addition to that, to enhance the likelihood that various EKGs — across departments, organizations or ecosystems — can interoperate easily with each other, the use of standardized EKG/IRIs needs to be encouraged since various EKGs can come to the same identifiers independently, drastically increasing the number of links across EKGs.

Implications

- There should be a mapping or service to resolve other names, keys or IRIs to the EKG/IRI.
- Since it is immutable, the EKG/IRI will have to be *opaque* i.e. not be a human-readable since even human names, company names, customer numbers, Social Security Numbers can change over time.

- Objects that already have a well-established RDF-compliant and Linked Data compliant identifier may not necessarily need an additional EKG/IRI. In fact, they may already have one that is external to the company's EKG. It is in many cases recommended to even give those well-established objects from well-established external datasets a company EKG/IRI. Examples of such objects are Legal Entity Identifiers (LEIs) and Financial Instrument Global Identifiers (FIGIs). The Enterprise Knowledge Graph Foundation (EKGF) will maintain a list of these for convenience.
- The use of multiple identities generally means that an EKG should *not* use the Unique Name Assumption (UNA) (where use of a different identifier would imply a different object). However the UNA *would* apply specifically to the EKG/IRI, and this should be ensured by any service.

2.2 Meaning

The meaning of every Data Point must be directly resolvable to a machine-readable definition in verifiable formal logic.

A Data Point combines an object — using its EKG/IRI as its identifier — with the value of a *property* in some context. Hence data is expressed at its most granular level for both data at rest and data in motion.

The property itself is always an IRI, often called "predicate-IRI", that refers to an object[1] that represents "the meaning" of the given Data Point. This object has its own identity and is defined through further properties based on logic that allow information to be rigorously combined, queried and inferred. These properties that define properties — also called "axioms" — are standardized by means of the Web Ontology Language (OWL2) by the World Wide Web Consortium (W3C) and are grouped into *"OWL ontologies"* for management purposes.

Rationale Expressing data at a granular level allows ultimate flexibility for it to be sliced, diced, combined and aggregated. This capability to combine and infer information is further enhanced by the use of property definitions built on logic. Having the properties themselves be objects that can be looked up means that all data is self-defining and carries its meaning with it. Since the information is self-defining there is no fixed schema for the EKG as a whole and it can non-disruptively incorporate additional knowledge.

Implications Some further discovery, with subject matter experts and creators of source systems, is often needed to truly understand what a given set of data really means and what can be inferred from it. In other words you cannot rely on the name of a column in a spreadsheet. A deceptively simple column name such as "number of European customers" leaves open the meaning of "European" and "customer" and timing (when does one start and stop being a customer?). And different sources could have different interpretations of that same name. The benefit is consistency, accuracy and the ability to make sound business decisions.

Advanced at higher levels of EKG/Platform maturity the term Data Point may in fact become a more complex data structure that is used "on the wire" that represents the Data Point at a more "holistic" level, supporting "multiple versions of the truth (MVOT)". Since an EKG supports many datasets that have overlapping information coming from multiple sources, there could be:

1. multiple EKG identifiers (EKG/IRIs) for the same object
    - One object can have multiple identifiers that can be linked together [2] and therefore be rightfully addressable with any of these identifiers.
2. multiple definitions of meaning
    - One property of an object can have multiple definitions of meaning, for instance "legal name" can be defined in multiple ontologies and be semantically equivalent [3] or one property can be defined as a subproperty of another but broader semantic definition [4].
3. multiple equal or different values coming from multiple sources
4. multiple versions over time of those values (temporality)

For each of these four "axes" — identity, meaning, source, temporality — you could have multiple options to choose from even while logically, from a user perspective, it's the same data point. Advanced client applications, services or AIs can use

[1] the official term in the RDF standard for this object is "resource"
[2] for instance via owl:sameAs i.e. "Individual Equality"
[3] See Equivalent Object Properties or Equivalent Data Properties
[4] See Data Subproperties

these Data Points to perform last-minute "at the edge" computations around finding the right value from the right timeline and source with the right quality for the given context.

2.3 Distributed

Data Points for any object may be stored in many physical stores. Any access point provides connectivity to all EKG content regardless of where it resides.

The physical stores could include traditional databases as well as varied Platforms specifically designed for hosting EKGs. And some which may be hosted outside the enterprise including by information services or governments. All seamlessly accessed using W3C (internet-based) protocols which also allow for browser or API *linked data* traversal and query.

Rationale Federating different systems allows the knowledge available to an enterprise to be accessed as a whole. It allows best of breed systems to be used for specific purposes, and evolved over time without disrupting the EKG as a whole. It allows scalability through adding new hardware, swapping out systems, and optimizing access through moving data closer to where it needs to be processed.

Implications The potential for a loose and evolving nature does mean some degree of monitoring of access patterns and performance; and service level agreements for vital information access.

2.4 Open World

Information can vary over time, come from many internal and external sources, and be based on different identifiers and models. These *multiple versions of the truth* need to be reconciled on access by context.

Different sources might not always be consistent about the same Data Point. And that may be legitimate for various business, geographical, privacy, legal or timing reasons. Rather than try to force everything into a single overruling set of facts, an EKG allows them to coexist, with the access context used to make coherent selections which are consistent within themselves. In order to make those selections multiple people and systems must have transparent access to all facts (including source, identity, meaning and value-at-time information) about all objects. Machine-executable business rules may be used at query time to join instances of data and e stablish quality rankings.

Rationale This approach allows the EKG to represent the sometimes messy reality, which encompasses not only a variety of different orgarnizations and systems within the enterprise, but also external sources.

Implications Attention needs to be paid to maintaining sufficient context with each data set, and considering what is needed for each data usage use case or context, including quality.

2.5 Self-describing

An EKG is composed of a set of *self-describing datasets* that provide information about lineage, provenance, pedigree, maturity, quality, licensing and governance.

The properties in each data point are linked to their definition so the meaning is not in doubt. A Dataset definition supplements this with management information such as its pedigree (how/when was it derived/sourced?) and its provenance (where/who did it come from?). This applies whether the information is maintained in the EKG itself or accessed/loaded from existing enterprise systems (data at rest); or received as data streams/messages (data in motion).

Rationale This information is essential for data selection, enforcing policy and management of the ecosystem as a whole. As well as being essential for management, the definitions taken together comprise a knowledge *catalog* for the content of the EKG.

Implications The information needs to be maintained and made available on an ongoing basis. It also needs to be sought out for external sourced data, whether accessed in place or loaded into an EKG platform.

2.6 Measurement

The quality and characteristics of the managed knowledge must be measurable and measured. Measurement criteria are used to designate fitness-for-defined-purpose and must be actionable.

Quality comprises multiple dimensions including accuracy, timeliness, completeness, relevance, conformity, integrity. EKG technology allows greater power and flexibility for expressing quality rules and metrics, but existing data quality tooling could also be used.

Rationale This information is needed to allow the right information to be used for each use case.

Implication The information needs to be maintained and made available on an ongoing basis.

2.7 Business Orientation

All information in the EKG, and associated artifacts, are linked to defined and prioritized *use cases*. Nothing in the EKG exists without a known business justification and purpose.

An EKG use case encompasses a business narrative and outcome expressed in business terms and links to relevant ontologies and Datasets.

Importantly, the use cases for an EKG can make use of lower level use cases, thus forming a *use case tree*, though it is not a strict hierarchy since common use cases can be reusable. At an implementation level a use case may be associated with user stories that form the points of interctaion with end users or client systems. The vision is that a fully fleshed out and implemented use case can be deployed as a reusable component.

Rationale Use cases anchor the EKG to real business needs, and allow it to evolve incrementally while delivering business value. Without this, the tendency is often to focus on information modeling for its own sake, without a focus or rationale for what to include or exclude. The use cases provide the basis for associating users with EKG functionality and provide the context for information selection.

Implications The use cases themselves need to be developed and managed as part of the EKG development method, and ideally as part of the EKG itself.

2.8 Control

Entitlement, privacy and business policies will be modeled in the EKG and automatically executed, enforced and audited at the Data Point level.

The EKG can use enterprise and organization knowledge to express access not only in terms of access control lists, but in terms of business rules, policies, logic and information content.

Rationale Use of the EKG itself to control and enforce access allows more power and conciseness of policy expression and execution while linking to existing enterprise directories.

Implications Appropriate enterprise directories should be integarted in the EKG. It can take some thought to design what the policies should be at the business level.

2.9 Ecosystem

An enterprise will use a heterogenous set of technologies and data sources which will be incorporated into the EKG over time. All data entering the ecosystem are subject to service level agreements.

A true EKG is a federation of multiple datasources and systems both within and external to the enterprise; seamlessly knitted together with standard protocols and APIs. An EKG needs to be managed and evolved as a fairly complex system of systems. To provide the flexibility and agility needed, its management needs to be automated, and linked to development processes, through the discipline of *data operations*.

Rationale As an ecosystem, an EKG can non-disruptively evolve over time in technology, scalability and information and business needs addressed.

Implications The management of the EKG needs to be planned for and resourced. It's important to coordinate with the owners of existing systems being used to provide capability under the EKG umbrella.

2.10 Standards

Data, ontologies, definitions and business rules should be compliant with open standards and transparent governance procedures as defined by recognized standards bodies.

Where necessary, EKGF will work with relevant standards bodies or projects to propose new standards or enhance existing ones.

Rationale It's important for enterprises to have trust in the quality, interoperability and stability of the structures and interfaces used in such a strategic investment as an EKG.

It provides freedom of choice in being able to mix and match different technologies and models, either in a federated EKG or over time. That includes standards being able to respond to new advances in both technology and techniques.

Implications It can sometimes take longer to work with industry partners to reach the consensus needed for standardization, and to use change control/governance procedures. Release cycles may be phased to avoid constantly changing interfaces. In order to achieve agility some work products may be deployed on standards that are have a *provisional* status — these need to be carefully identified.

Chapter 3

Mapping to FAIR data principles

3.1 Findable

F1 (Meta)data are assigned a globally unique and persistent identifier

F2 Data are described with rich metadata (defined by R1 below)

F3 Metadata clearly and explicitly include the identifier of the data they describe

F4 (Meta)data are registered or indexed in a searchable resource

Differences EKG Principles are slightly more specific or prescriptive:

- Metadata (objects) and data (objects) can potentially have multiple identifiers (but at least one).
- Those identifiers do not necessarily be ”persistent” as long as they are (always) resolvable (through HTTP).
- The EKG identifiers (EKG/IRIs) of data objects (but not necessarily metadata objects) should be ”opaque” as in ”meaningless”, (relatively) safe to be emailed around, stored in other platforms, maximising ”proliferation”.
- FAIR principle F3 would be phrased the other way around: the data described by metadata refers to it via the metadata identifier (i.e. the predicate-IRI).
- FAIR principle F4 slightly differs as well, the EKG Principles require metadata to be directly resolvable (via HTTP) machine-readable definitions of the semantics in verifiable formal logic (preferably OWL 2).

3.2 Accessible

Once the user finds the required data, they need to know how they can be accessed, possibly including authentication and authorisation.

A1 (Meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2 Metadata are accessible, even when the data are no longer available

Differences EKG Principles are slightly more specific or prescriptive:

- *”using a standardized communications protocol”* would be explicitly the HTTP protocol (or actually HTTPS/TLS) as a minimum requirement and in addition to that any other protocol, standardized or not.
- *”metadata are accessible”* would be more explicit for the EKG: all metadata has to be accessible through IRIs that are always ”resolvable” via the HTTP protocol. In other words, make sure that all your OWL 2 ontologies or RDFS schema vocabularies are placed on highly available durable infrastructure that can always be accessed via HTTP.

3.3 Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1 (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2 (Meta)data use vocabularies that follow FAIR principles

I3 (Meta)data include qualified references to other (meta)data

Differences EKG Principles are slightly more specific or prescriptive:

- In the EKG the metadata that describes meaning i.e. the semantics (there are also many other types of metadata) that *formal and broadly applicable language for knowledge representation* has to be preferably OWL 2 or at least RDF Schema or SHACL.

3.4 Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1 Meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1 (Meta)data are released with a clear and accessible data usage license

R1.2 (Meta)data are associated with detailed provenance

R1.3 (Meta)data meet domain-relevant community standards

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).

Differences EKG Principles are slightly more specific or prescriptive:

- *optimization of the reuse of data is the ultimate goal*. It's an important goal for EKG as well but reuse of knowledge and whole use cases — with everything that comes with it — is an even higher level goal. Furthermore, overall connectedness of all data and knowledge is an equally important goal.
- *metadata and data should be well-described* but should also be as "unbiased" as possible, not designed for one particular (set of) use case(s) but designed to represent the version of the truth that a given data source provides with the highest level of integrity. See principle 2.5 Self-describing.
- *Meta(data) are richly described with a plurality of accurate and relevant attributes*. "*Richly described*" would not be specific enough for EKG. It would have to be *directly resolvable to a machine-readable definition in verifiable formal logic*.
- For all types of recognized metadata, which is metadata that services of the various EKG platforms can recognize, the EKGF will specify accepted standards or define standards itself or in collaboration with partners like the Object Management Group® (OMG). Since an EKG is a collection of SDDs, each dataset will have its various types of metadata organized in a structured way.

Appendices

Acronyms

EKG Enterprise Knowledge Graph 2, 3, 5–11, 15, Glossary: EKG
EKG/Platform Enterprise Knowledge Graph Platform General Terms: EKG/-Platform
EKGF Enterprise Knowledge Graph Foundation 6, 9, 15
FIGI Financial Instrument Global Identifier 6
IETF Internet Engineering Task Force 15
LEI Legal Entity Identifier 6
MVOT multiple versions of the truth 2, 6
OMG Object Management Group® 11
OWL2 Web Ontology Language 6
RDF Resource Description Framework 5, 6, 15
S3 Simple Storage Service Glossary: object store
SDD self-describing dataset 3, 11, 15, Glossary: self-describing dataset (SDD)
SVOT Single Version of (the) Truth 2
UNA Unique Name Assumption 6
W3C World Wide Web Consortium 6, 15

Glossary

CoE for the EKG the group of people that is overseeing the EKG 15
EKG use case TODO 8
Canonical Identifier A permanent identifier for an object, distinguished within an EKG. 15
data point a fact representing the value(s) of a property for an object in some context. Each value may be a simple data (e.g. a number, string or date) or another object. The combination of object, property and value is a *triple*. 3, 6–8
dataset a collection of data, published or curated by a single agent, and available for access or download in one or more serializations or formats. 7, 8
EKG much like "the web" is a virtual concept, an EKG is a virtual concept that combines all information and knowledge of an enterprise — at any level in the organization — or ecosystem. 2, 5,
EKG system architecture the logical system architecture of EKG is divided into multiple layers or environments such as EKG/Platform, EKG/Storage and EKG/DataOps environment. 15
EKG/DataOps environment a logical system architecture component, the environment where EKG/DataOps pipelines run. Since EKG/DataOps pipelines can run anywhere and any given data provider can publish their own data as a SDD in any technical or physical environment, the actual DataOps "environment" is just a logical concept that does not necessarily map one-to-one to a particular physical environment. 15
EKG/DataOps pipeline a DataOps pipeline in the EKG/DataOps environment is a series of programs, called "steps", that are run in sequence — hence the name pipeline — where the first step captures the data from a given source and the last step produces an SDD (for "inbound pipelines") or any other type of artifact (for "outbound pipelines"). Each DataOps pipeline step takes an input and produces an output based on its configuration which usually refers to a model that instructs the DataOps pipeline step how to process the data. Crafting these models is the main activity in the EKG/DataOps Practice. 15
EKG/IRI an IRI that forms the identity of an object in the EKG. Any given object in an EKG has an EKG/IRI for which special rules are defined by the EKGF. Not to be confused by Canonical Identifiers. 5, 6, 10, 15
EKG/Method a methodological approach to the development of an Enterprise Knowledge Graph (EKG), covering all practices from the definition of desired business outcomes to deployment of EKG use cases that deliver on these outcomes and beyond. 15
EKG/Platform a logical system architecture component (see EKG system architecture), the layer of software services that provide and serve the EKG to end-users and other systems. The platform logically is a set services that enforce any of the specified policies in the SDDs that have been published in the EKG. 3, 6, 14, 15
EKG/Storage a logical system architecture component, the layer of data-storage services such as a "Triplestore" or an "Object Store" that serve the various other layers in the EKG system architecture. 15
IRI a standard defined by the Internet Engineering Task Force (IETF) in 2005 in RFC 3987. An Internationalized Resource Identifier (IRI) is defined similarly to a , but the character set is extended to the Universal Coded Character Set. Therefore, it can contain any Latin and non-Latin characters except the reserved characters. Instead of extending the definition of , the term IRI was introduced to allow for a clear distinction and avoid incompatibilities. IRIs are meant to replace URIs in identifying resources in situations where the Universal Coded Character Set is supported. By definition, every URI is an IRI. Furthermore, there is a defined surjective mapping of IRIs to URIs: Every IRI can be mapped to exactly one URI, but different IRIs might map to the same URI. Therefore, the conversion back from a URI to an IRI may not produce the original IRI. The IRI standard is a superset of the older "Uniform Resource Identifier (URI)" standard (IRI ⍰ URI). See also EKG/IRI. 5, 6, 10, 15
linked data the collection of interrelated datasets on the Web. Linked Data is a concept defined by Tim Berners-Lee and the W3C that is part of what they call "the Semantic Web" or "the Web of Data". Tim Berners-Lee started the idea in 2006 by defining 4 simple rules: `https://www.w3.org/DesignIssues/LinkedData.html`. 6
practice the DataOps practice, — as defined by the EKG/Method — is one of the practices that the CoE for the EKG executes. 15
property an identified and fully-defined linkage between an object and values. Or in RDF terminology, a relation between subject resources and object resources. See `https://www.w3.org/TR/rdf-schema/#ch_property`. 6
RDF Schema TODO RDF Schema 11
self-describing dataset (SDD) an EKG is logically composed of a set of *self-describing datasets* that provide information about lineage, provenance, pedigree, maturity, quality, governance, entitlement policies, retention policies, security labels, IP policies, pricing policies, caching policies, organizational ownership, accountabilities, data quality feedback loop, issue management policies and so forth. Any given system owner that connects their data to the EKG becomes in fact a publisher of a self-describing dataset and can therefore control how their data is handled by the EKG/Platform. The EKG/Platform consists of services that enforce any of the specified policies in the self-describing dataset. At the technical level, datasets can be files or streams or API definitions etc. See also principle *2.5 Self-describing* on page 7. 3,
system architecture the conceptual model that defines the structure, behavior, and more views of a system. 15
URI a Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or physical resource. See `https://www.ietf.org/rfc/rfc3986.txt` and IRI. 15

Index

accountability, 2, 15

censoring, 2
competitiveness, 2

data
    capital, 2
    quality, 2
        tooling, 8
    quality feedback loop, 15

FAIR, 10
fairness, 2

human capital, 2

identifier resolution, 5
issue management, 15

opaque, 5
organizational
    ownership, 15

principles, 5
    Business Orientation, 8
    Control, 8
    Distributed, 7
    FAIR, 10
    Identity, 3, 5
    Meaning, 6
    Measurement, 8
    Open World, 7
    Self-describing, 7, 15
    Standards, 9

sustainability, 2

transient, 5
transparency, 2

universally unique, 5

web-resolvable, 5