KGE - Knowledge Graph Engineering

iTelos Methodology

General principles

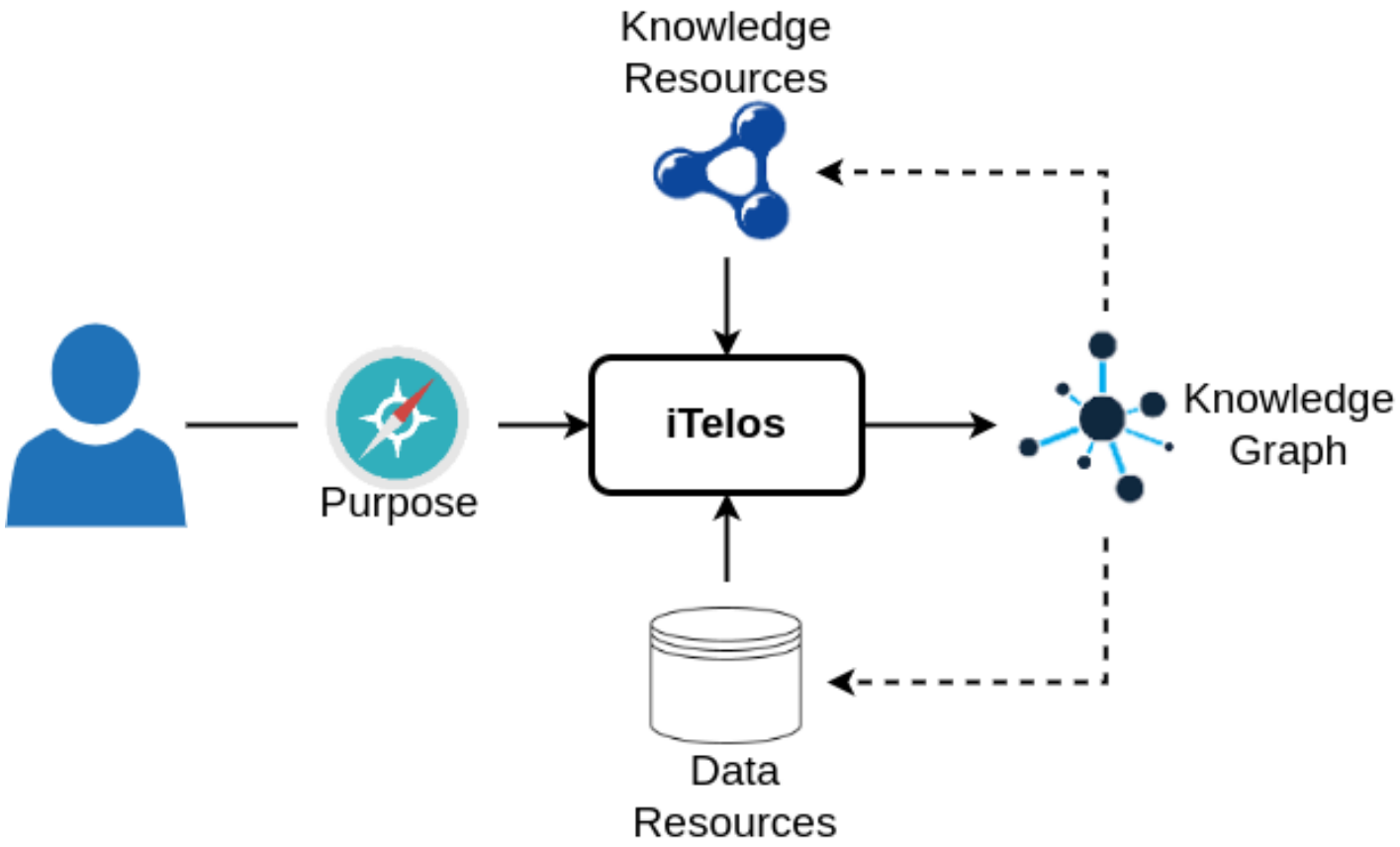
Fausto Giunchiglia

Contents

iTelos - A KGE Methodology

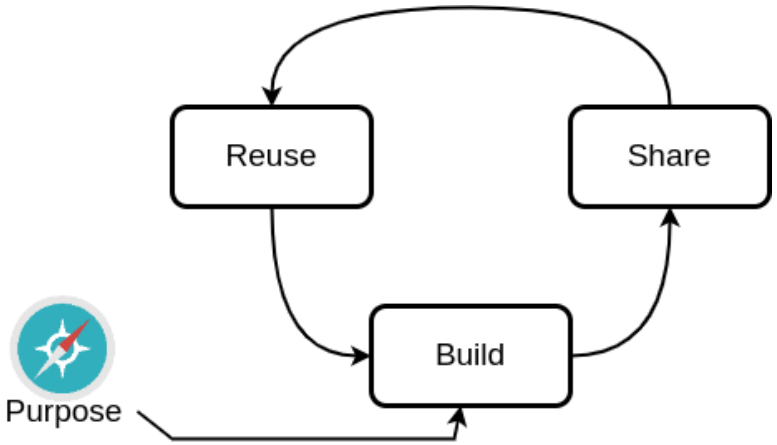
- *iTelos* (from Greek "Telos" = scope) is a structured methodology defined to reduce the effort while developing a KGE process.
- (Explicit goal) The methodology aims to **provide support to the user** in solving all the issues encountered building purpose-specific KGs.
- (Implicit goal) Moreover, iTelos implicitly **produces reusable resources**, enforcing a circular data (re)use, thus reducing the effort in creating new resources.

iTelos - The view



Contents

Reuse, Build and Share



- Is it always possible to reuse resources ?
- Is a Purpose specific resource reusable ?
- Which resources can be reused and which not ?
- How to share resources in order to enhance their reuse ?

Contents

Model-based KGE

- iTelos implements its main principle by adopting a **Model-based** approach to develop KGE processes. It consists in create (and reuse) resources shaped as **information models**.
- An information model represent a **portion of information from the point of view of the Purpose** used to create it.
- The Model-based approach is similar to the **Open Source** idea, where different code libraries are shared and composed to build new programs and applications.
- In the same way, iTelos **creates and/or composes** information models to build new KGs.
- iTelos defines different kinds of models, one for each sub-component of a KG.

(A dedicated lecture will define each kind of models used by iTelos)

Model-based KGE - Reuse models

- An information model represents a portion of information from the point of view of the Purpose used to create it.
- The **more "popular"** is the information carried by the model, the **more reusable** the model will be.
- The information, and the data carrying it, has an high level of *popularity* when it is strongly reused for different purposes.

Model-based KGE - Reuse models - Example

- A model representing *a geographical point in space* (model A) is more reusable than a model representing *an hospital building in a given city* (model B).
- The model A is used in several projects for many different geographical purposes. It models *common*[1] information relative to the purpose considered. It is a popular information model.
- While, model B is used in less projects compared with the previous one. It can model *core*, or even *contextual*, information, depending by the purpose it refers to. It is an information model less popular than A, thus less reusable.

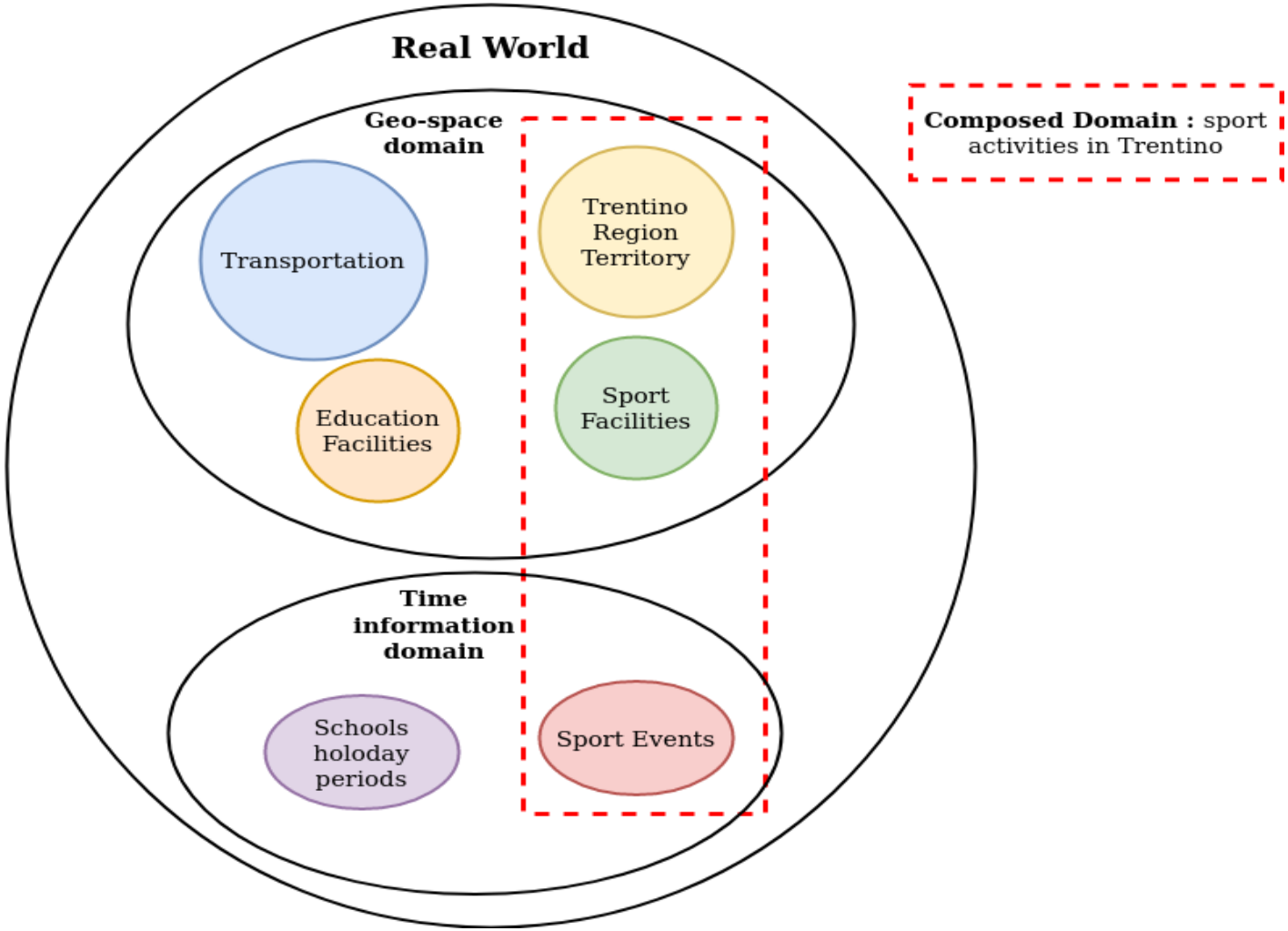
[1] see Lecture on "Reuse Problem"

Model-based KGE - Domain building

- ***Domain of Interest*** (DoI): A classification of all the information elements used to satisfy a specific purpose.
- Therefore, we can define a **DoI as a set of information models**. More precisely, a **set of KGs created by models composition**.
- Extending the Model-based KGE popularity principles over domains, we can define more and less popular (reusable) DoIs.
- DoIs are modeled as KG sets, which in turn can be grouped together to create new Purpose specific DoIs.

iTelos, at different levels of granularity, allows to **create and compose KGs** representing popularity-based DoI, defined by the purpose used to develop the KGE process.

Model-based KGE - Domain building example



Model-based KGE - Summary

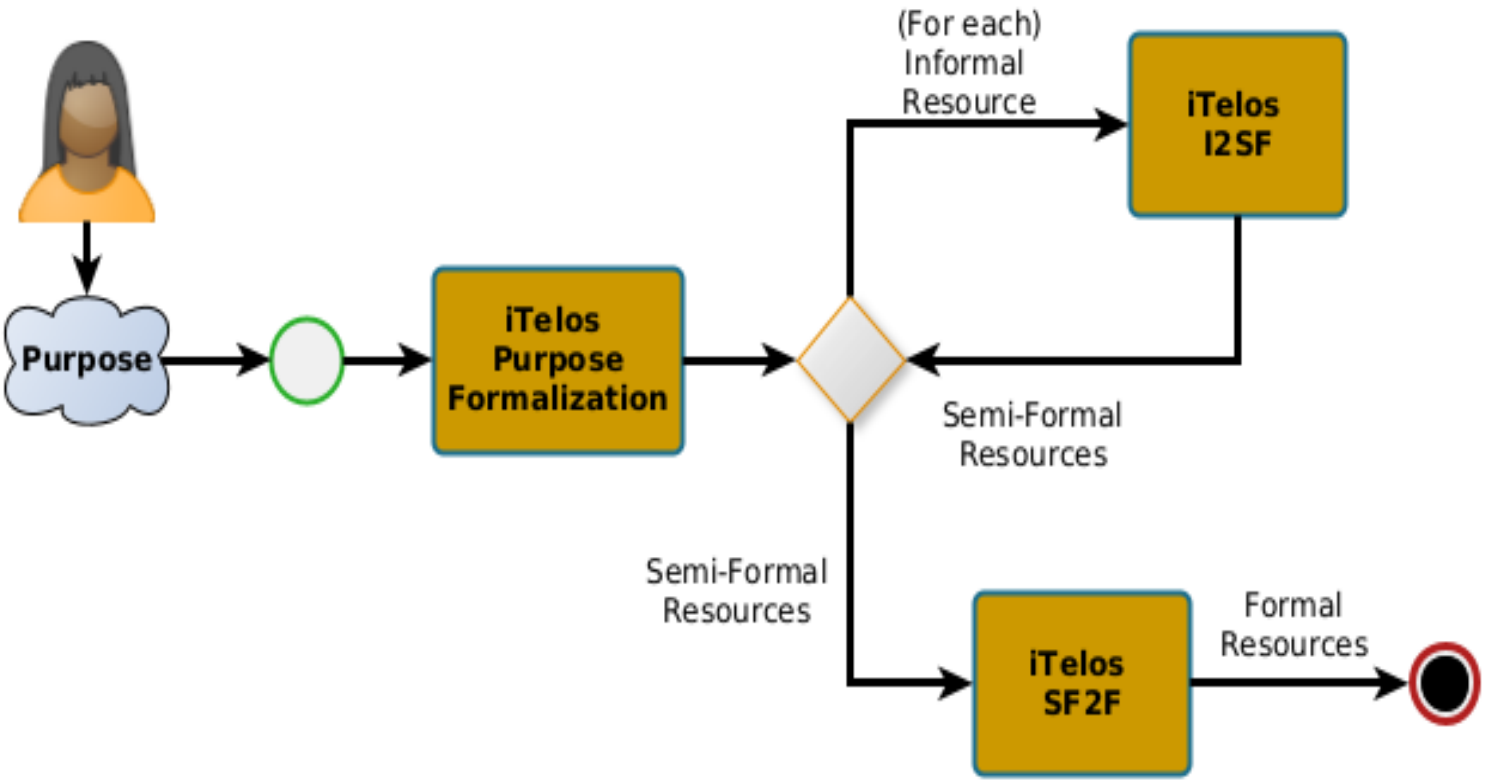
1. iTelos produces and composes information models.
2. Information models are used to compose KGs representing DoI (Purpose specific).
3. iTelos allows KGs creation and composition, thus enabling DoI building.
4. The creation of DoIs, built using the Model-based approach, enforce the models reuse, thus reducing the effort in building KGs for new DoIs.

Model-based KGE - Implementation

How iTelos is able to **create** and **compose** information models ?

- The first phase of iTelos is dedicated to the **Purpose formalization**. The objective of this phase is to extract from the initial Purpose, all the required information which will lead the following activities of the KGE process.
- Then the methodology structure is divided in two macro components:
  - Model creation: **Informal to Semi-Formal** (I2SF) resource conversion.
  - Model composition: **Semi-Formal to Formal** (I2SF) resource conversion.

Model-based KGE - iTelos Implementation



Contents

Informal resources

- iTelos recognizes as informal, all those resources which:
  - are not **machine processable**;
  - are **unstructured** or **semi-structured**, and/or not associated with a **reference schema** representing the information they are carrying;
  - are **syntactically not standardized**. They do not adopt standards for data types, or, if they adopt standards, there is an **high level of syntactic heterogeneity** in the resource set collected for a specific Purpose.

Informal resources

- Informal resources are those **less reusable**.
- Unfortunately, they are, currently, the **most available online**.
- A KGE process, most of the times, in the beginning has to deal with informal resources.
  - When a new Purpose comes, it is impossible to reuse at 100% already existing formal resources.

Semi Formal resources - Pros

- iTelos recognizes as Semi-Formal, all those resources which:
  - are **machine processable**;
  - are **structured**, and associated with a **reference schema** representing the information they are carrying;
  - are **syntactically standardized**. They adopt standards for data types, thus reducing the **syntactic heterogeneity** in the resource set collected for a specific Purpose.

Semi Formal resources - Cons

- Nevertheless, Semi-Formal resources:
  - have to be integrated/composed, considering a specific purpose (a unique schema is required);
  - can be expressed in multiple languages, thus a language alignment is required;
  - are not associated to unique identifiers and metadata, used to link multiple (kinds of) resources to a specific Purpose.

Semi Formal resources

- Semi-Formal resources are semantically and syntactically improved.
- But they still do not fully comply with the principles of FAIR [2].
- They are more reusable respect the Informal ones, but still not enough to be used in order to significantly reduce the effort in KGE process.

[2] Findable, Accessible, Interoperable, and Reusable

Informal to Semi-Formal (I2SF) process

- The iTelos I2SF macro-phase receives in input the *formalized Purpose*, which is used, in the beginning of I2SF phase, to identify which resources have to be collected.
- Considering all the available data sources, the resources collected can be informal, semi-formal or even already fully formalized.
- The I2SF phase has the objective to transform the informal resources into semi-formal.
- The I2SF phase **acts over each single resources** collected. It is responsible to **create** single semi-formal models (always considering the initial Purpose), but it does not compose them.

Contents

Formal resources

- iTelos recognizes as Formal all those resources which:
  - are generated from Semi-Formal resources;
  - are semantically aligned on a unique knowledge resource (Schema, Ontology, Teleology, Teleontology, see next lecture) defined over a specific Purpose;
  - are associated to unique identifiers and metadata, used to link multiple (kinds of) resources to a specific Purpose.

Formal resources

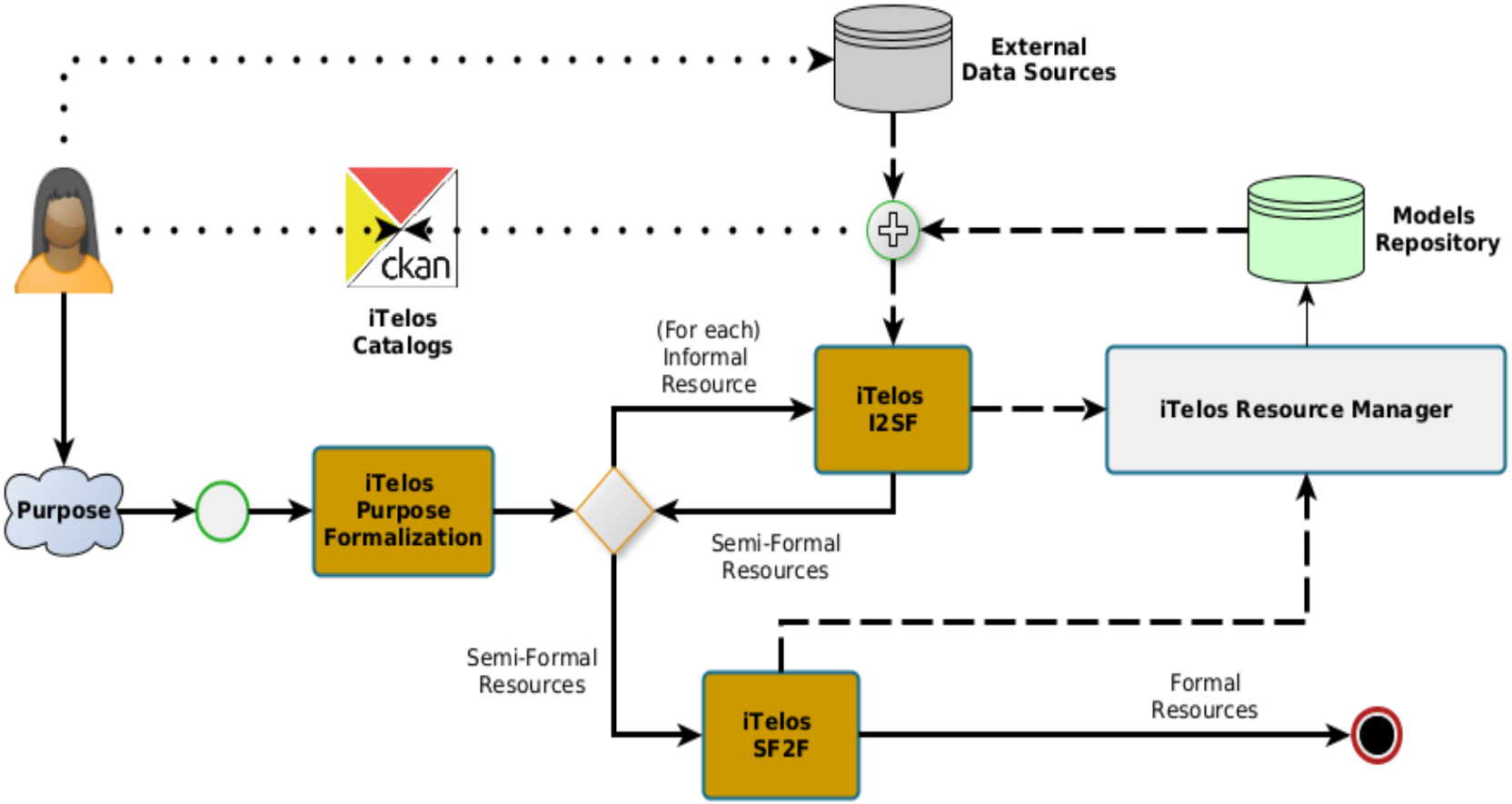
- Formal resources fully comply with FAIR principles
- Thanks those features they are available to be shared, and thus reused.
- Formal resources are composed together in order to form a **unique single resource** (KG) satisfying the initial Purpose.
- Nevertheless, thanks to their features, they can be seen also as **separated resources** (formal models) representing the different sub-components.

Semi-Formal to Formal (SF2F) process

- The iTelos SF2F macro-phase receive in input the *formalized Purpose* too, which is used to lead the composition and transformation of semi-formal resources into formal ones.
- The I2SF phase has the objectives to:
  - transform the semi-formal resources into formal ones;
  - compose the formal resources created in order to create a single resources able to satisfy the Purpose.
- Unlike the I2SF phase, the SF2F one acts considering **all the input semi-formal resources together**. It is responsible to create formal information models, as well as to compose them into a unique purpose specific resource (final KG).

Contents

Sharing Formal resource



iTelos KG sub-components & Models

- Which are the KG's sub-components that can be represented as models ?
- Which models (semi-formal and formal resources) are produced and shared by iTelos ?

Follow the next lecture !

KGE - Knowledge Graph Engineering

Fausto Giunchiglia

iTelos Methodology

General principles