

**Projet M1 UFR MIM :**

**Sujet IR**

**Taxinomie MISP**

Marwin NIMESKERN

Grégory GUGGENBUHL

Jules CHARRON

Année 2018 – 2019

**Gratefully**

We are tankful to the MISP team who despite a lack of time have open their door to make some reunion and to present us the MISP’s software.

Summary

[Introduction 6](#_Toc8602804)

[What is a Taxonomy? 6](#_Toc8602805)

[Goal of our taxonomy? 6](#_Toc8602806)

[What is a cooperative taxonomy? 6](#_Toc8602807)

[State of the Art 7](#_Toc8602808)

[MISP: Malware Information Sharing Platform and Threat 7](#_Toc8602809)

[Presentation 7](#_Toc8602810)

[MISP is an open source software on GitHub, allowing the sharing of attacks revealed by the community between MISP instances. 7](#_Toc8602811)

[Main Advantages of MISP 7](#_Toc8602812)

[• One of the largest attack databases in the world 7](#_Toc8602813)

[• Internationally recognized 7](#_Toc8602814)

[• There are multiple data 7](#_Toc8602815)

[• Taxonomies can be used as object typing 7](#_Toc8602816)

[• Open-Source and regularly maintained by a warm community. 7](#_Toc8602817)

[• Can look like SQL (Entity / Relationship) 7](#_Toc8602818)

[• Has a partnership with the University of Lorraine 7](#_Toc8602819)

[Main Drawback 7](#_Toc8602820)

[Conclusion 7](#_Toc8602821)

[The arborescent Taxonomy 8](#_Toc8602822)

[Serving as the basis of scientific classification of species, the term "Taxonomy" refers in computer science to a method of classifying information in a structured architecture so that it can be completed in an evolutionary and generic manner. 8](#_Toc8602823)

[Example of a tree of a botanical taxonomy: 8](#_Toc8602824)

[The MISP taxonomy 9](#_Toc8602825)

[The MISP’s team have of course create an engine to manage their taxonomies. 9](#_Toc8602826)

[Therefore, we have adapted our taxonomy to best fit their system 9](#_Toc8602827)

[Presentation 9](#_Toc8602828)

[Main Advantages 9](#_Toc8602829)

[Main Drawbacks 9](#_Toc8602830)

[Bugyuo Methodology 10](#_Toc8602831)

[Presentation 10](#_Toc8602832)

[Main Advantages 10](#_Toc8602833)

[Main Drawbacks 10](#_Toc8602834)

[Conclusion 10](#_Toc8602835)

[ERCOT Faceted Classification 11](#_Toc8602836)

[Lexique : 11](#_Toc8602837)

[Main Advantages 11](#_Toc8602838)

[Main Drawbacks 11](#_Toc8602839)

[Conclusion: 11](#_Toc8602840)

[Architecting an Enterprise content management Strategy 12](#_Toc8602841)

[Present the strategies of a company for the realization of a project 12](#_Toc8602842)

[The arrow presents the next instance that will be needed to realise the project. 12](#_Toc8602843)

[Main Advantages 12](#_Toc8602844)

[Main Drawbacks 12](#_Toc8602845)

[Conclusion: 12](#_Toc8602846)

[Functional Classification Taxonomies 13](#_Toc8602847)

[Main Advantages 13](#_Toc8602848)

[Main Drawbacks 13](#_Toc8602849)

[Conclusion: 13](#_Toc8602850)

[Sector Taxonomy and definitions 14](#_Toc8602851)

[Main Advantages 14](#_Toc8602852)

[Main Drawbacks 14](#_Toc8602853)

[Conclusion: 14](#_Toc8602854)

[Taxonomy code mapping / professional providers 14](#_Toc8602855)

[ICB Version 2: IPMA Competence Baseline 15](#_Toc8602856)

[Main Advantages 15](#_Toc8602857)

[Main Drawbacks 15](#_Toc8602858)

[Conclusion: 15](#_Toc8602859)

[GEA-NZ: Government Enterprise Architecture New-Zealand 16](#_Toc8602860)

[Description of the main branch 16](#_Toc8602861)

[Main Advantages 16](#_Toc8602862)

[Main Drawbacks 16](#_Toc8602863)

[Conclusion: 16](#_Toc8602864)

[Modification of GEA-NZ 17](#_Toc8602865)

[Risk Governance 17](#_Toc8602866)

[Bugyo 17](#_Toc8602867)

[The multi-level-model (MLM) 18](#_Toc8602868)

[Hierarchical Ascending Classification (HAC) 19](#_Toc8602869)

[Algorithms K-MEANS 20](#_Toc8602870)

[Algorithms Names 20](#_Toc8602871)

[The algorithm K-MEANS to determine new class and there Isobarycenter 20](#_Toc8602872)

[Example of an administrative enterprise 21](#_Toc8602873)

[Annexe 2 : GEA-NZ : Government Enterprise Architecture New -Zealand 22](#_Toc8602874)

[Annexe 2 : Generical Taxinomy in JSON 23](#_Toc8602875)

[Annexe 3 : Modification of GEA-NZ 24](#_Toc8602876)

[An Introduction to Multilevel Modeling - basic terms and research examples - John Nezlek (source youtube) 25](#_Toc8602877)

# Introduction

## What is a Taxonomy?

Nowadays, many scientist projects are founded on operational classification more or less elaborate. We will talk about: taxonomy

It was created by the Swiss botanist Augustin Pyrame de Candolle in 1813, for a species classification in herborize purpose.

Over time, this term was extended to other areas of science, including the humanities, information sciences or computer science.

The taxonomy designs the practice and science of classification of things or concepts, including the principles that underlie such classification.

A Taxonomy is a classification of information instances that facilitates the discovery and management of information between them.

## Goal of our taxonomy?

Misp is grouping all security incidents that happened in all enterprises which would share information with other enterprises. This information could be important and concerns different level of an enterprise.

The main questions are:

* What are these information and how could it be important for each work of an enterprise?
* How to classify all this information in order to sort the most important?

Our goal is the creation of a cooperative taxonomy, in order to classify all the instances of enterprises. This will allow to make a better search and classification of all the events of MISP.

## What is a cooperative taxonomy?

Cooperative taxonomy is the hierarchical classification of entities of an enterprise, organization or administration, used to classify documents, digital assets and other information. Taxonomies can cover virtually any type of physical or conceptual entities (products, processes, knowledge fields, human groups, etc.) at any level of granularity.

# State of the Art

## MISP: Malware Information Sharing Platform and Threat

### Presentation

 MISP is a Luxembourg government organization that brings together a community of more than 6,000 companies sharing between them information about attacks, frauds, ...

### MISP is an open source software on GitHub, allowing the sharing of attacks revealed by the community between MISP instances.

### Main Advantages of MISP

### • One of the largest attack databases in the world

### • Internationally recognized

### • There are multiple data

### • Taxonomies can be used as object typing

### • Open-Source and regularly maintained by a warm community.

### • Can look like SQL (Entity / Relationship)

### • Has a partnership with the University of Lorraine

### Main Drawback

* Unusual in the foreground, requires a good understanding of the software and its modules
* little concrete application on the data
* Only coded in python that a really slow language because it’s a scripted language

### Conclusion

• The attributes are interconnected and juxtaposed with a generic taxonomy: this will allow typing and classifying all events generated in MISP

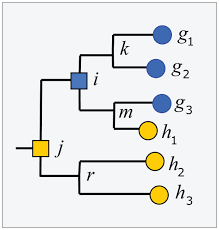
## The arborescent Taxonomy

### Serving as the basis of scientific classification of species, the term "Taxonomy" refers in computer science to a method of classifying information in a structured architecture so that it can be completed in an evolutionary and generic manner.

### Example of a tree of a botanical taxonomy:

#### Phylogenetic tree view

Tree taxonomies are kinds of graphs that help you to better classify an object into several families of class and subclass. These kind of taxonomy have oriented our research to choose our cooperative taxonomy



A tree taxonomy is presented by this generic form

## The MISP taxonomy

### The MISP’s team have of course create an engine to manage their taxonomies.

### Therefore, we have adapted our taxonomy to best fit their system

### Presentation

The taxonomies of MISP are in the form:

**Namespace: predicate = ``value’’**

* Namespace : Name of the taxonomy
* Predicate : Object resource of the taxonomy
* Value  : Value of the resource

As a result of these rules, we make the json in annexe 2 this annexe is here to describe how a generical taxonomy is written in order to be implemented directly in the MISP project.

### Main Advantages

Misp uses taxonomies engine to create statistics, rate and classify events

### Main Drawbacks

* Without predicate parsing, the depth is limited to 1.

To avoid these drawback:

* + A predicate could be parse to create multiple levels of tags and create a deeper taxonomy. Due to our need to depth, our parser will be the: ‘’\_ ‘’
    - Predicate: predicate = ‘’value’’
* Long taxonomy can be problematic to use

To avoid these drawback:

* + A long taxonomy should have an interface for proper use
  + A complex need taxonomy need structuration to be used properly that why we will be inspired by other taxonomy in order to achieve that goal.

Source : <https://www.misp-project.org>

## Bugyuo Methodology

### Presentation

Security Methodology Project for Telecom Operators

The infrastructure is generated by metrics that generate evidence that serves to minimize the risks that give a level of confidence on countermeasures that minimize risk.

See the modification annexe to see the bugyo dendrogram use for this taxonomy.

### Main Advantages

* Gives an interesting approach on the methodology of a security policy, security assurance and their different implementation and operation.
* Created with the collaboration of one of our security teachers who come from Luxembourg.
* Parallel approach to common criteria and deploys an interesting risks approach.

### Main Drawbacks

* Abstract concept which only give indications on how to proceed an engineer to securing a system
* The notions are very high in the methodological approach.
* Even though this representation is powerful, it is unknown to the general public

### Conclusion

To better understand the needs of an ISS, this methodology can help us to see if an event is responding to its needs.

These is a common criterion approach; these criteria have none really main methodology or program to applied them.

In that goal, a methodology like bugyo could be useful to check the advanced of these methodology in their enterprise. Many security events could be marked to follow the proper application of these stuff, furthermore a sharing platform could be useful to see how properly manage it.

The more the number of companies using it and sharing their ways of proceeding will increase,  
the more it will be applied properly and understandable.

Source : Mr Djamel KHADRAOUI

## ERCOT Faceted Classification

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| **Function** | **Activity** | **Type** | **Entity Type** | **Entity** | **Rule Type** | **System Name** |
| Market Participation | Registration and Qualification | Registration Documents | Market Participant |  | Protocol |  |
| Information Technology | System and Application Development | Revision and Change Request | Market Participant | TDSP |  | MarkeTrak |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  | **Generical description** |  |  |  |  |  |
|  |  | Predicate | ⊥ | Predicate |  |  |
|  |  | Value | ⊥ | Value |  |  |
|  |  | Value | ⊥ | Value |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Statically describes the operation of a business and its activities.

Lexique : A ⊥ B : A independent of B

Each column describes a predicate of an instance from the service of the company for a proper purpose

### Main Advantages

* Briefly presents the operation of a department of a company
* Hardcoded approach for corporate taxonomy
* Representation which strongly resembling a multi-model-level (further details in this document)

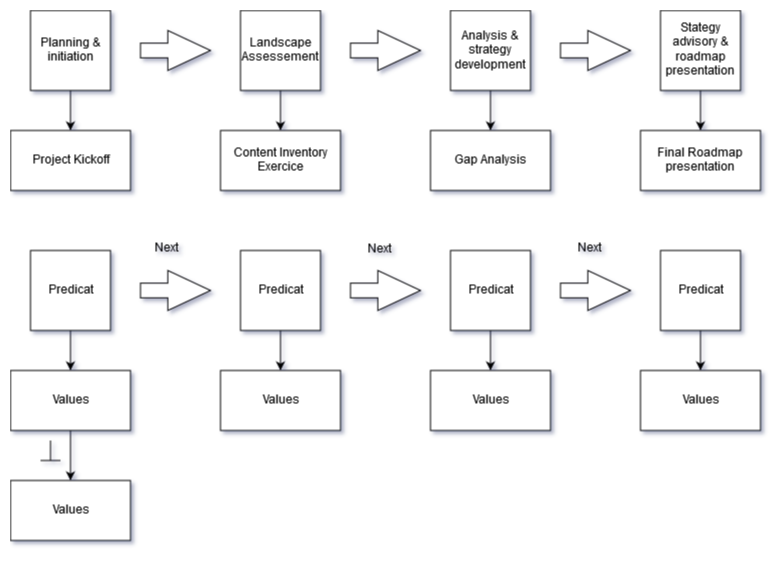
### Main Drawbacks

* Too simple to use alone without another taxonomy
* All links of columns are independent from each other, it a drawback only if no other taxonomy is used to described how to fill the column

### Conclusion:

* Despite these drawbacks these taxonomies could be used as a model for a properly structuration for another taxonomy like the one we decided to choose.
* As announced, the columns independency is only a drawback if no other taxonomy is used, more over it became a strongly advantage in the goal of describe a multi-level-model.

## Architecting an Enterprise content management Strategy



### Present the strategies of a company for the realization of a project

### The arrow presents the next instance that will be needed to realise the project.

Each predicate could have many values, which will be independent between them.

### Main Advantages

* Present a taxonomy, in a timeline manner
* Could be useful to describe instance or event between them in a chronological manner

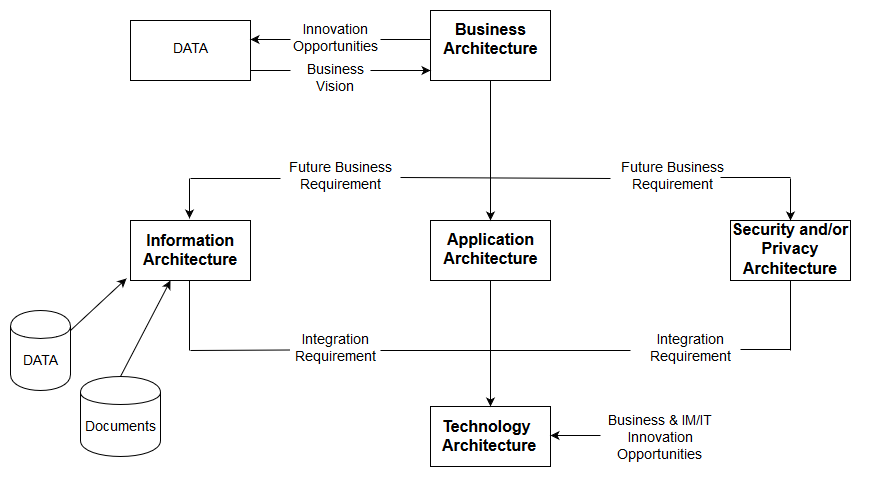
### Main Drawbacks

* Does not in any way present the enterprise infrastructure, it’s useless for present jobs or objects

### Conclusion:

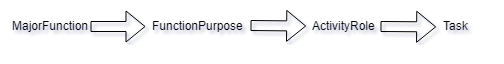
* Our job is not only to describe the infrastructure of an enterprise but all instance of an enterprise therefore it’s could be important to make some way to describe a chronological order for some plans of a company
* It could be an interesting way to structure a task in order to accomplish a company project.

## Functional Classification Taxonomies



This graphics explain how for them a data is share in some architecture:

* Business: harvest the data information
* Information: store the data in some disks
* Application: Use this data for the purpose of the company
* Security / privacy: Secure the data harvested
* Technology: has been used by the application part to process the data



Business / Application / Technical / Information

Define links between the major instances of an archetype company.

Example: Education **=>** Teach **=>** Teacher **=>** Network course

### Main Advantages

* The major instances of a company could be linked thanks to that

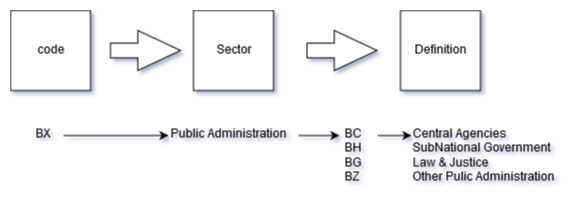
### Main Drawbacks

* Doesn’t explain what is an instance of an enterprise

### Conclusion:

* The main idea of linking could be used to linked other instances from a taxonomy more complex like the GEA-NZ taxonomy

## Sector Taxonomy and definitions



Interesting mapping of instances of jobs from any sector of activity

### Main Advantages

* Professional directory of classification
* Easy way to research an instance once the code is know

### Main Drawbacks

* More a directory of job than a real taxonomy

### Conclusion:

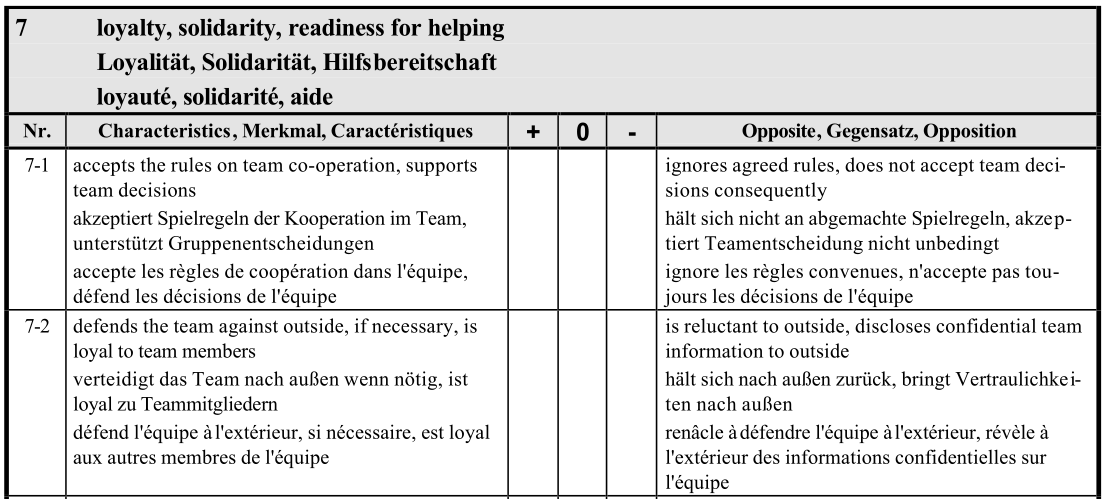
* It could be interesting, once our taxonomy operational to map our instance realised thanks to that way.

## Taxonomy code mapping / professional providers

It’s a directory of instance, consequently it’s presented the exact same advantages and drawbacks than the previous Sector Taxonomy and definitions.

## ICB Version 2: IPMA Competence Baseline

Example of ICB-Version 2: page 74



A taxonomy asking a company of their competence.

Present a really interesting way to make a survey in order to create an instance from this taxonomy.

### Main Advantages

* The survey is a really interesting and easy way to asking a thing to a customer

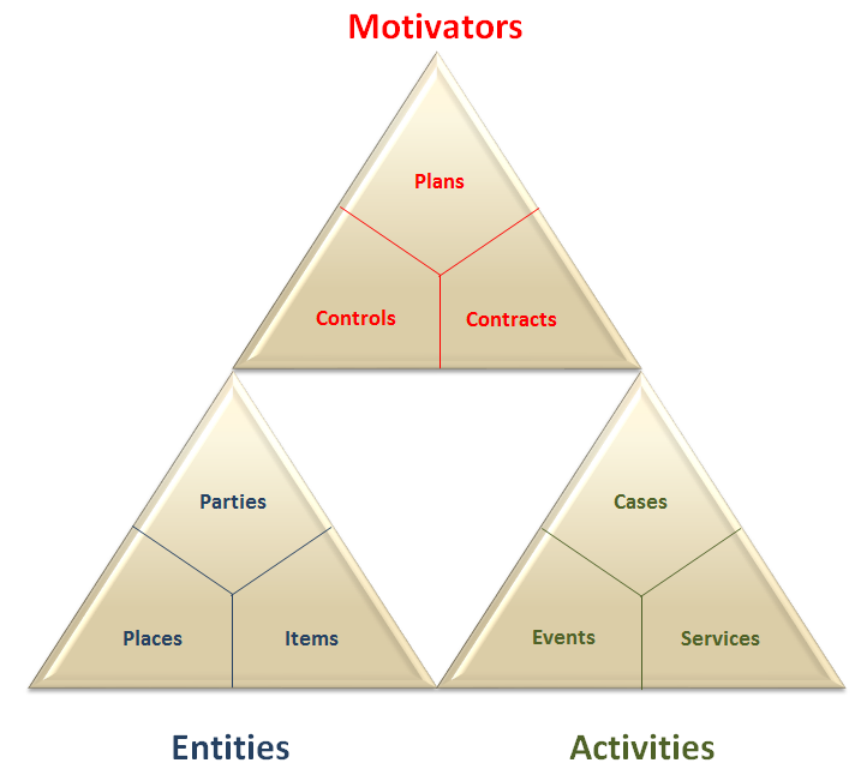
### Main Drawbacks

* The taxonomy itself isn’t really interesting to make a corporate taxonomy

### Conclusion:

* Once our taxonomy is designed we could use this way to ask for create dynamically some instance of our taxonomy.

## GEA-NZ: Government Enterprise Architecture New-Zealand



The taxonomy that has been choose by us.

Thanks to three complete tables also divided in three branches  
Note : See the annexe 2 to see the main branch presented without their leaf for more readability  
See the NGA-NZ for properly instructions about their leaf, the taxonomy being used as it is been , the rewrite here would have no interest except for the add we made of it.

### Description of the main branch

* **Motivator:** Search
  + **Plans:** actions accomplished for choose a direction
  + **Controls:** the constraints of actions
  + **Contracts:** agreement between different parties
* **Entities:** all possibly instances in an enterprise
  + **Places:** Workplace and any location
  + **Items:** All items buyable or not
  + **Parties:** Staff entities
* **Activities:** all possibly interaction between parties
  + **Cases:** Special and normal event managed
  + **Events:** Planning and / or Spontaneous event
  + **Services:** Offer of the service

### Main Advantages

* A really complete cooperative taxonomy for each instances of an enterprise

### Main Drawbacks

* It’s as complete as it is complex to use, that’s why It’s need a structuration for a properly usage.

### Conclusion:

This taxonomy is powerful but it does not force any structuring to make some object like : jobs.

This structure is crucial to make some classification furthermore.

# Modification of GEA-NZ

Note: Please see the Annexe 3 for the picture of adaptation into GEA-NZ taxonomy.

In purpose to make the control region more accurate we add two more sub-category.

## Risk Governance

* Risk Governance in order to evaluate the risk taken with the object
* **Residual Risk:** This element is a remaining risk in the enterprise
* **Risk Analysis:** Risk tolerate by the enterprise
* **Risk Assessment:** General process of analysis and risk assessment
* **Risk Management:** Control and organize the activity taking into account the risks
* **Risk Treatment:** Measures selections and implements the chosen metrics

## Bugyo

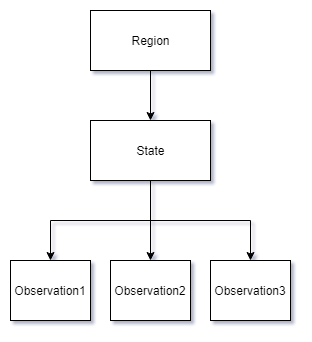
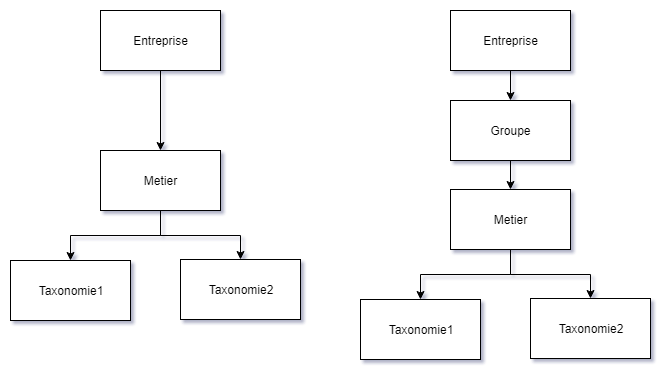
* Bugyo in order to add some security methodology in the code to verify directly a component
  + **Service model:** applied to the model as a whole
    - **Absence of relevant vulnerability**
      * Is a vulnerability present? yes / no?
    - **Unmanaged/Managed Ratio**
      * Less Unmanaged / Higher confidence about the ratio?
  + **Maintenance Management:** applied to the model to see the maintenance of it
    - **Probe Maintenance:** Is the maintenance
      * ***Reactive:*** Ad-hoc maintenance / maintenance policy present?
      * ***Pro-active:*** Regular interval to prevent / Documentation and log
    - **Infrastructure object model maintenance:** maintenance for each object?
      * ***Reactive:*** Ad-hoc maintenance / maintenance policy present?
      * ***Pro-active:*** Regular interval to prevent / Documentation and log
  + **Metric Construction:** is used to measure the system's security policy
    - **Scope:** request the number of elements measured by the system
      * Are there countermeasure? How much it’s goes around the system?
    - **Depth:** How for in details it’s goes?
      * Is there an interface for?
        + A single object? The whole chain?   
          The whole chain and all the values passing through it.
    - **Rigor:** How is the formalism of the measures? Is it strict?
      * None verification / informal-verification / semiformal verification
    - **Reliability of metric:** Are the measure repeatable?
      * None verification / reproductible / informally / semiformal
    - **Timeliness:** How recent are the software?
      * None verification / Version prior the newest (if retro compatible)   
        Most recent version (Avoiding flaws with patch)
    - **Frequency:** How fresh is the evidence of measure?
      * Year / Month / Week / Day / Hours
    - **Stability:** How stable is the measure
      * Is there a human avoiding False/True positive?   
         Or / And are there statistics algorithms?

# The multi-level-model (MLM)

Multilevel models (also known as hierarchical linear models, nested data models, mixed models, random coefficient, random-effects models, random parameter models, or split-plot designs) are statistical models of parameters that vary at more than one level. (Citation Wikipedia)

A multi-level-model is created when a complex structuration is used and require different cutting levels.

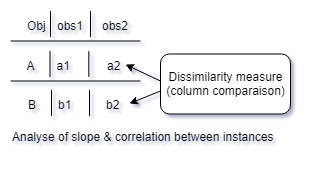
Each region could have many states and each state could have many observations and so on.



***Generical Taxonomy enterprise job enterprise service of job***

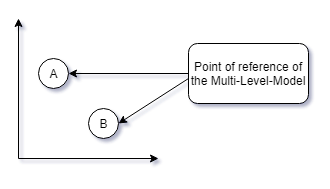
***Generical Form MLM: Taxr,s,o1 = Taxinomy region, state, observation1***

For each object created we have a level and for each object many stuffs of the taxonomy could be used



In order to make possible this stuff we could make for each type of object a table in order to compare an object between another object of the same type. (like job, service,)

Once a table is made for an object, we can observe a dissimilarity measure, reporting differences between the different columns.



In the end, we can make a graphics thanks to the dissimilarity measure to make a graphics of point of reference for each object.

As a result, with these graphics we can make an HAC (Hierarchical Ascending Classification

# Hierarchical Ascending Classification (HAC)

In data mining and statistics, hierarchical clustering (also called hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters.

An HAC is created when:

* a taxonomy is needed
* **dissimilarity is measured**
* for clustering many objects in order to make cluster analysis and segmentation.

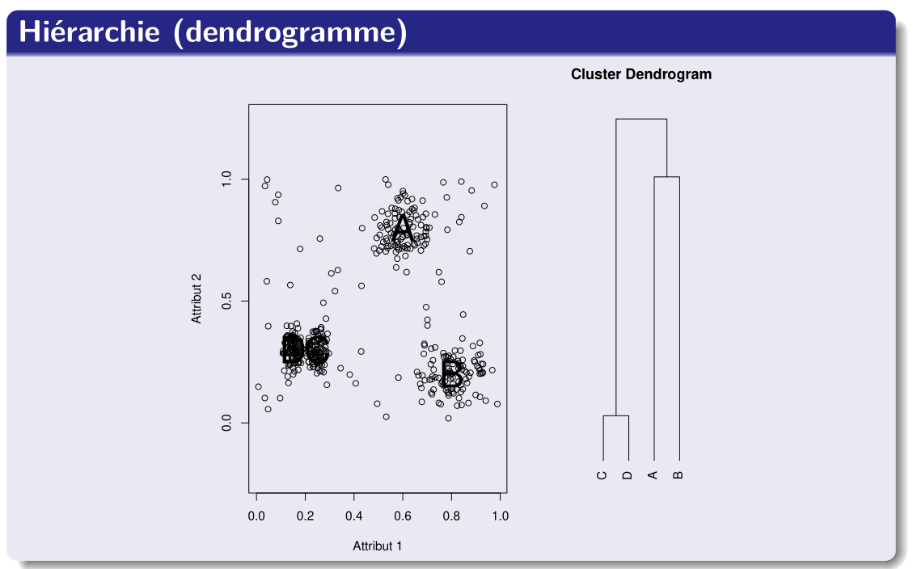
**Once created, this cluster will make it possible to present wiser choices to the client.**

Point Descriptor:

* As see precedingly each point will present an object created with the taxonomy

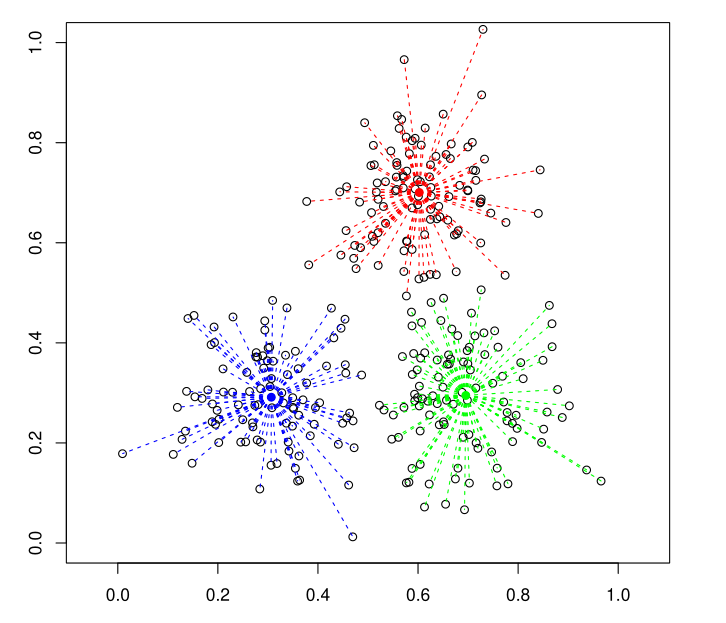
Strategies for hierarchical clustering generally fall into two types:

* **Agglomerative:** This is a "**bottom-up**" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
* **Divisive:** This is a "**top-down**" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.



These algorithms are usually based on distance (prototype or density) and / or Grid-based like Dendrogram (discretization of data)

These methods can make their own concept thanks to cloud of points



This picture represent a method based on prototypes

A prototype is:

A representative individual

A seed (called an Isobarycenter\*)

The center of each class is also called an isobarycenter

Finally, once this step arrived we must minimised by calculating the intra class inertia

## Algorithms K-MEANS

The K-means method is an optimization algorithm of intra-class inertia

### Algorithms Names

* K-means, K-moyennes, K-noyaux
* Algorithm of dynamic clouds, mobile centers
* Lloyd’s algorithm

### The algorithm K-MEANS to determine new class and there Isobarycenter

#### Description

* Alternation of two stages:
  + Grouping objects into classes around centers
  + Adjustment of centers according to the component objects

classes

* Stopping the algorithm when there is no more change (or after a fixed number of iterations)
* Convergence of the algorithm in a guaranteed finite time

(usually correct result in 10 iterations)

#### Realization

* Random initialization of centers
* As long as the result varies
  + For all objects
    - Calculate the distance to all centers
    - Assign the object to the class most close
  + For all classes
    - Calculate the center of gravity of the objects assigned
    - Assign the center of gravity as new center of the class

(Source Wikipedia / Mr Blansche M1 AI Course 04 / data search course 05)

# Example of an administrative enterprise

•Financial service

o Accountant

o Expert Accountant

o Tax lawyer

•Administrative department

o Administrative Tactical Consultant

o Criminal lawyer

• Secretary

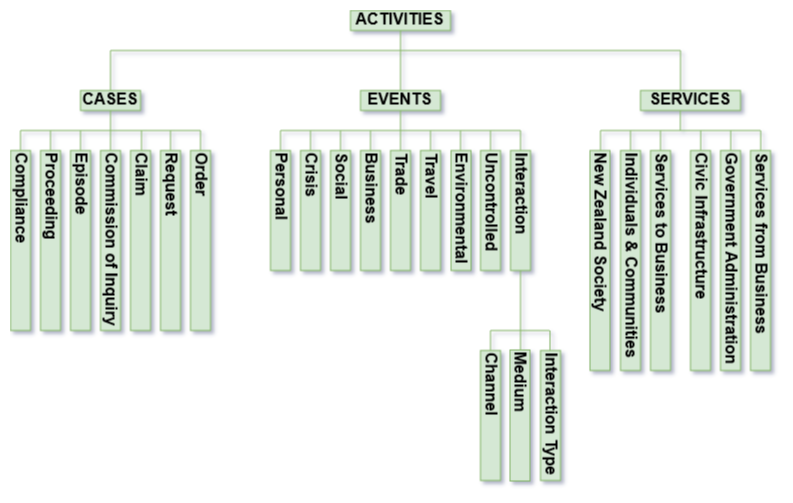
* Accountant :
  + It ensures the translation of the flow of economic and financial activity of an organization by keeping accounts, in compliance with the general chart of accounts and international standards.
  + It ensures the control of the accounts and their analyzes.
  + It may be specialized in revenue tracking (customer accounting) or expenditure

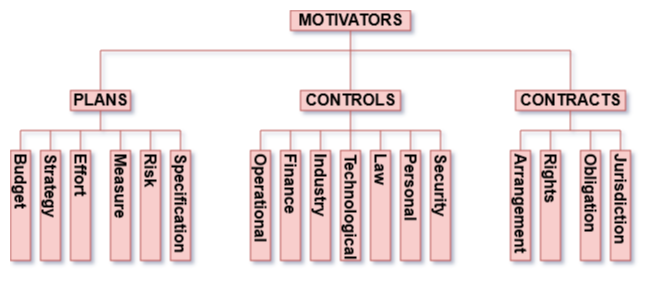
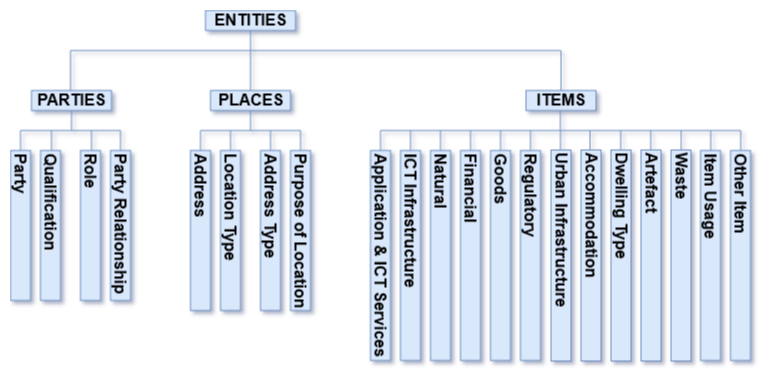
(supplier accounting).

Expert Accountant :

* + It stops the accounts of a company and ensures the regularity and reliability of the accounts for which it is responsible.
  + He is consulted in matters of management, organization, tax law and social law.
* Tax lawyer :
  + He advises on all issues of interest to the law and the economy.
  + He practices as a liberal or as an employee in the legal department of a company or a specialized firm.
* Criminal lawyer :
  + He exercises before the criminal courts the defense of natural or legal persons for offenses relating to the penal code.
* Secretary (Assistant) :
  + He works for one or more service managers, and is the internal interlocutor of the company and its external partners.
  + Makes, formats mail and documents.
  + Provides a hotline.
  + Manage the agenda and appointments.
  + Ensures the reception.
  + Organize the meetings and seminar.

# Annexe 2 : GEA-NZ : Government Enterprise Architecture New -Zealand





# Annexe 2 : Generical Taxinomy in JSON

{ namespace": "Title of your Taxonomy",

"description": "Taxonomy to classify the information of ... .",

"refs": [

"url ... "

],

"version": 1,

"predicates": [

{

"value": "Name1 ",

"expanded": "Expanded Name 1 ",

"description": "TDesciption of the name 1"

},

{

"value": "Name2 ",

"expanded": "Expanded Name 2 ",

"description": "TDesciption of the name 2"

}

],

"values": [

{

"predicate": "Name1",

"entry": [

{

"value": " value1 ",

"expanded": "Def Value1"

},

{

"value": "value2",

"expanded": "Def Value2"

}

]

},

{

"predicate": "Name2",

"entry": [

{

"value": " value1 ",

"expanded": "Def Value1"

},

{

"value": "value2",

"expanded": "Def Value2"

}

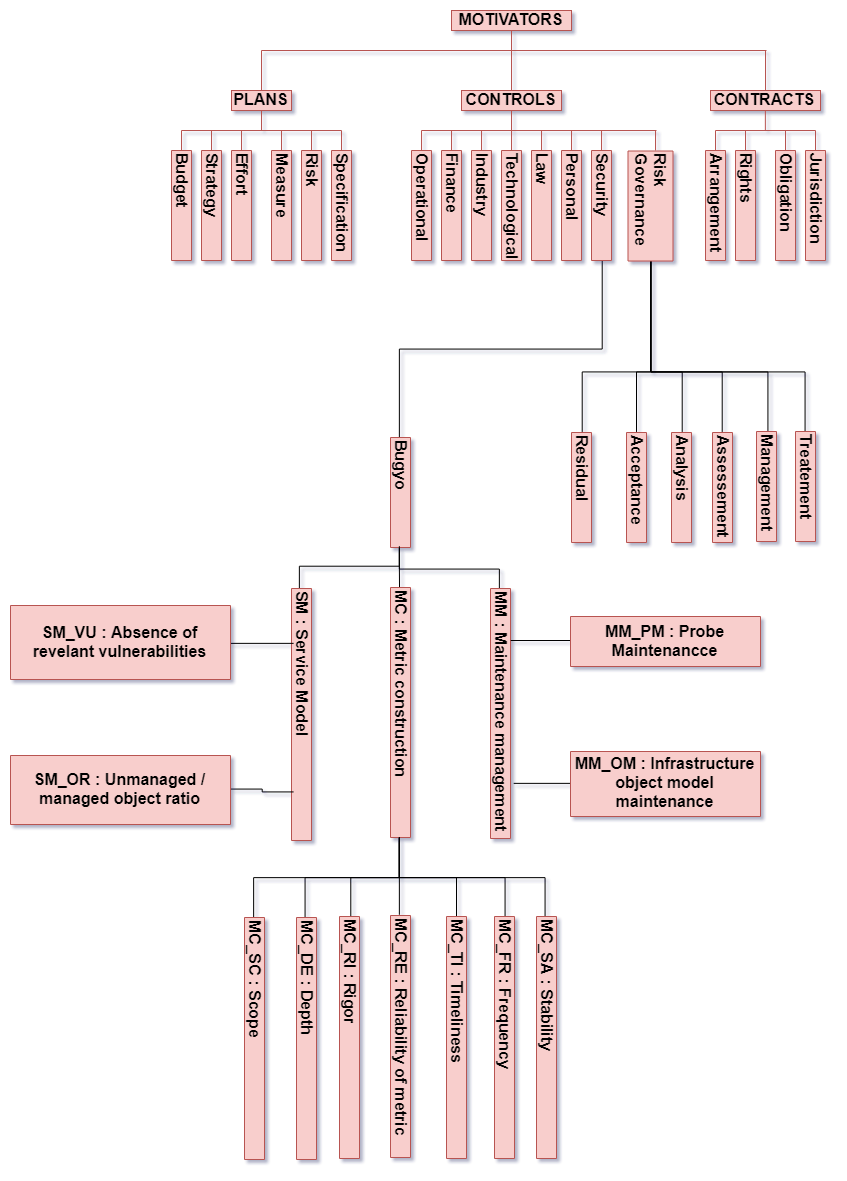
]

}

]}

# Annexe 3 : Modification of GEA-NZ

Bugyo and a risk gouvernance have been added



Bibliographie :

# An Introduction to Multilevel Modeling - basic terms and research examples - John Nezlek (source youtube)