



CENTRE NATIONAL D'ÉTUDES SPATIALES

**PÉRENNISATION, MISE À DISPOSITION ET VALORISATION  
DES DONNÉES SCIENTIFIQUES**

**Paul Kopp – Centre National d'Études Spatiales (CNES)  
Paul.Kopp@cnes.fr**

## Origine de la question au CNES

### ■ Confrontation du CNES à la nécessité de préserver ses données sur le long terme

- ◆ les données sont souvent “uniques” car difficiles à reproduire ou témoins d’un phénomène observé qui ne se reproduira pas
- ◆ les données font souvent partie de séries temporelles nécessaires pour des études à long terme (climatologie, ...)
- ◆ les données anciennes peuvent être utiles, éventuellement associées à d’autres données, à des fins différentes de celles pour lesquelles elles ont été conçues

### ■ Évolution des segments sol de traitement, archivage, diffusion de données

- ◆ fin des systèmes seulement circonscrits aux besoins de communautés bien définies, aux services desquelles ils ont été construits
- ◆ mise en place de systèmes au service de la connaissance
  - gestion du temps : prise en compte du long terme, réutilisation avec la technologie d’aujourd’hui de données créées dans un autre contexte
  - gestion de l’espace : permettre l’utilisation conjointe de données dispersées en divers lieux

## L'analyse du CNES

### ■ Importance du contexte

- ◆ **L'utilisation de données est facile lorsque sont disponibles**
  - leur contexte technique (forme)
  - leur contexte thématique (sens)
- ◆ **Lorsque l'un ou l'autre des contextes est perdu ou s'il n'est plus possible d'en rendre compte, l'utilisation fiable des données est impossible !**

### ■ La menace du temps

- ◆ **Contexte technique et contexte thématique sont toujours disponibles au début**
- ◆ **de manière formelle**
  - dans des documents "papier" ou "numériques"
  - par l'intermédiaire de dispositifs techniques appropriés
- ◆ **de manière informelle, dans la mémoire des individus**

**Au fil du temps, la disponibilité du contexte technique et/ou du contexte thématique s'atténue si ne sont pas prises des mesures en faveur de sa préservation : éviter l'oubli, composer avec les ruptures technologiques.**

## Les actions entreprises par le CNES

### ■ Préserver les données “physiques”

- ◆ opérationnel au CNES
- ◆ pris en charge au Centre Informatique par le Service de Transfert et d'Archivage des Fichiers (STAF)

### ■ Préserver le contexte technique

- ◆ opérationnel au CNES
- ◆ pris en charge par la technologie EAST/OASIS de description des données
  - développée par le CNES, endossée par le CCSDS et l'ISO
  - dotée de logiciels d'implantation

### ■ Préserver le contexte thématique

- ◆ non opérationnel au CNES
- ◆ expérimenté par des travaux technologiques
- ◆ approché dans le cadre de la normalisation CCSDS des dictionnaires de données



CENTRE NATIONAL D'ÉTUDES SPATIALES

**PRÉSERVER LES DONNÉES PHYSIQUES  
LE MODÈLE OAIS**

**Reference Model for an Open Archival Information System  
CCSDS 650.0-B-1  
ISO 14721:2003**

## Le problème à résoudre

### *Constat*

- Un nombre important d'organisations disposent aujourd'hui d'informations sous forme numérique (causes : disponibilité de ressources de calcul, bande passante, connectivité).
- Les transactions avec ces organisations s'appuient sur des formulaires numériques à la place des formulaires "papier" traditionnels.

### *Difficulté*

La préservation de l'information sous forme numérique est beaucoup plus difficile que sous forme traditionnelle :

- ◆ l'information digitale est facilement perdue ou corrompue ;
- ◆ les supports d'information sont rapidement obsolètes à cause des évolutions technologiques ;
- ◆ les structures et formats d'enregistrement de données peuvent ne plus être reconnus comme éléments de représentation de l'information.

⇒ Le stockage "bit à bit" des données ne garantit pas la pérennité de l'information représentée par ces données. Il faut analyser les conditions nécessaires à la préservation à long terme de l'information .

## Reference Model for an Open Archive Information System Objectifs

### *Définition*

Un OAIS est une archive, c-à-d une organisation comportant des personnes et des systèmes ayant accepté la responsabilité de préserver de l'information à long terme et de la rendre disponible à une communauté destinatrice (“Designated Community”).

Cette responsabilité est définie dans le modèle de référence d'un OAIS. Le modèle est “ouvert”, ayant été déterminé au sein de forums ouverts.

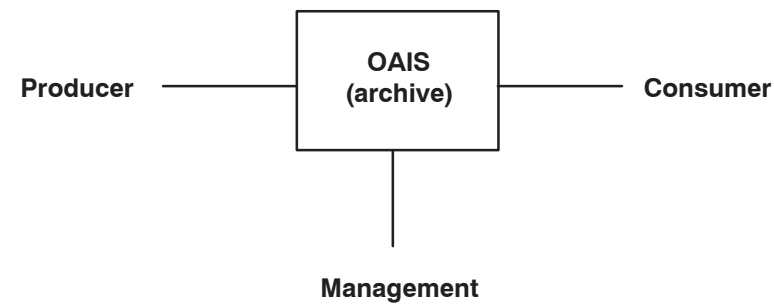
L'expression long terme signifie une durée suffisamment longue pour que l'on ne puisse faire abstraction de changements technologiques ou de changements dans les communautés.

### *Objectif*

Le modèle OAIS fournit un cadre terminologique et conceptuel pour la compréhension des mécanismes de préservation à long terme de l'information.

Le modèle OAIS n'est pas une norme de mise en oeuvre.

## Environnement d'un OAIS



### Rôles

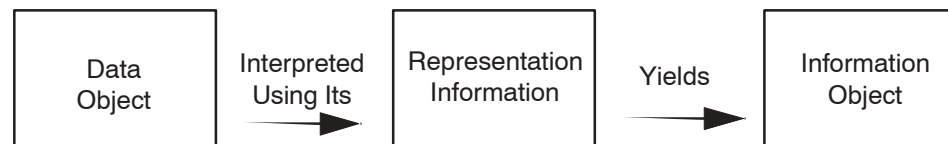
- **Producer** : rôle joué par les personnes ou systèmes qui fournissent l'information à préserver.
- **Management**
  - ◆ rôle joué par ceux qui déterminent la politique de l'OAIS dans un contexte plus large que celui de l'OAIS ;
  - ◆ n'est pas un rôle "au jour le jour"
- **Consumer**
  - ◆ rôle joué par les personnes ou systèmes qui interagissent avec les services de l'OAIS pour trouver et acquérir l'information recherchée ;
  - ◆ la "Designated Community" est le sous-ensemble des consommateurs qui sont capables de comprendre l'information préservée par l'OAIS.

## L'information gérée par un OAIS

Pour comprendre une information reçue sous la forme d'une donnée, il faut disposer de son arrière-plan cognitif. Exemple : comprendre le Français pour lire un texte écrit en Français. À défaut, on utilisera un dictionnaire et une grammaire. Le dictionnaire et la grammaire constituent une information de représentation.

L'information est toujours représentée sous la forme d'une donnée. Exemple : les caractères imprimés d'un livre. Le livre et ses caractères constituent un objet de données.

Un OAIS permet de convertir un objet de données, interprétable à l'aide d'une information de représentation, en un objet d'information.

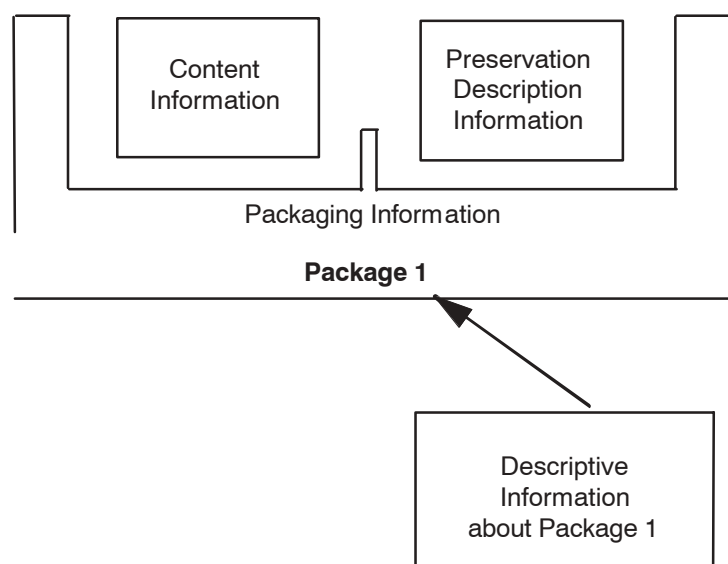


## Les packages d'information d'un OAIS

### Un package d'information se compose

- d'une information de contenu
  - ◆ objet de données et
  - ◆ information de représentation
- d'une information de description relative à la préservation

### Un package d'information est identifiable et décrit par une information descriptive



#### Preservation Description Information

- provenance
- contexte
- référence
- contrôle

Packaging Information : organisation du package

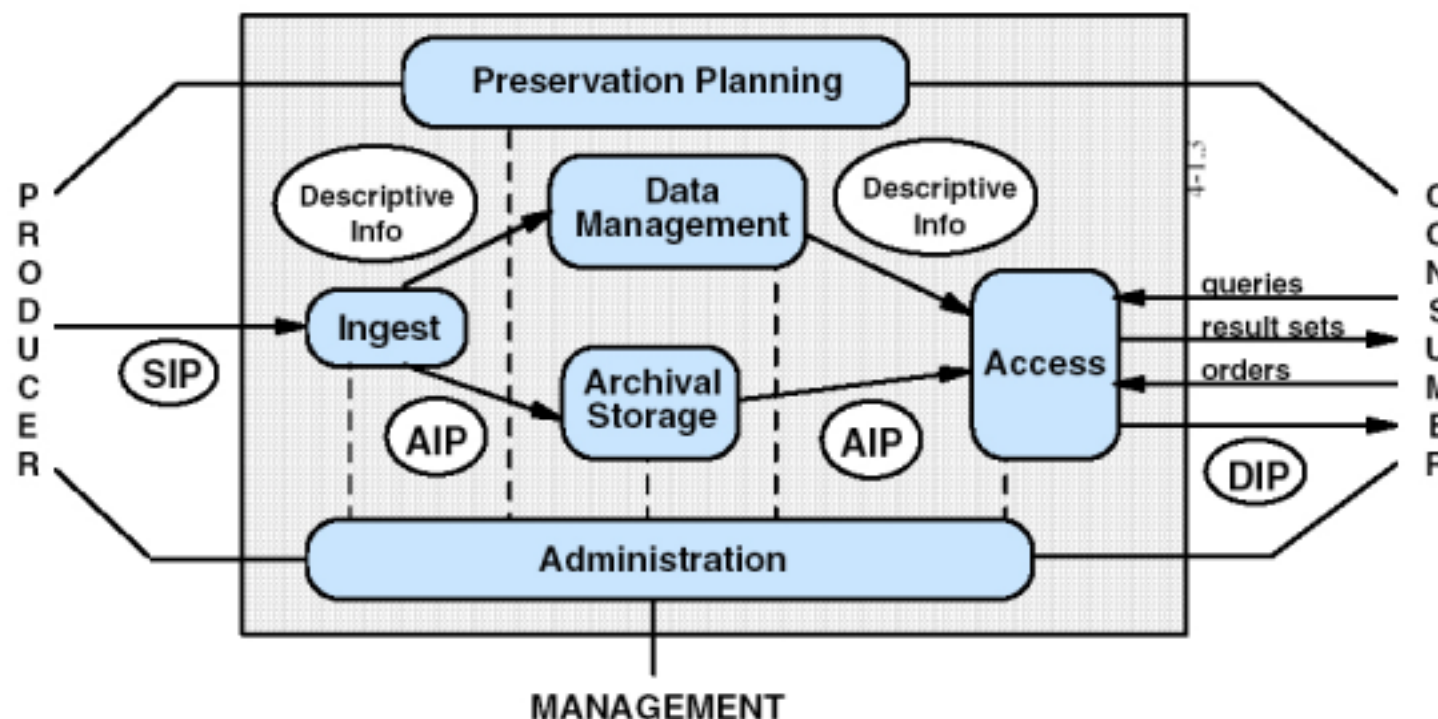
- SIP – Submission Information Package (envoyé par le producteur)
- AIP – Archival Information Package (SIP transformé en vue de l'archivage)
- DIP – Distribution Information Package (AIP transformé en vue de la distribution au consommateur après interrogation de l'OAIS)

## Responsabilités d'un OAIS

Un OAIS doit :

- **négozier avec les producteurs à propos de l'information à recevoir, recevoir l'information**
  - ◆ ex. : sujet et source de l'information, représentation de l'information (supports, formats)
- **obtenir le contrôle sur l'information nécessaire à sa préservation à long terme**
  - ◆ ex. : gestion des copyrights, droit de modifier l'information pour les besoins de la préservation
- **déterminer, en liaison éventuelle avec d'autres, les communautés qui deviendront la communauté destinataire ("Designated Community", dont les membres comprennent l'information fournies par l'OAIS)**
  - ◆ ex. : grand public, spécialistes
- **s'assurer que l'information à préserver est auto-suffisante pour sa compréhension (elle ne doit pas contraindre à l'assistance des producteurs d'information)**
  - ◆ ex. : gestion de l'expertise nécessaire à la compréhension de l'information
- **suivre des règlements et procédures documentés par lesquels l'information préservée l'est en étant protégée des risques ordinaires, et par lesquels l'information distribuée l'est comme copie conforme – ou comme dérivation explicite – de l'information originale.**
  - ◆ ex. : suivi des altérations du contenu de l'information (mise aux normes des formats de date, etc.)
- **mettre l'information préservée à la disposition de la communauté destinataire**
  - ◆ ex. : gestion des transformations de l'information en vue de sa distribution (algorithmes, etc.)

## Le modèle fonctionnel d'un OAIS



- Ingest :** réception des SIP, contrôle qualité des SIP, génération des AIP
- Archival Storage :** archivage des AIP, gestion de l'espace de stockage (sauvegardes, etc.)
- Data Management :** gestion des "Descriptive Information" (schémas et vues)
- Administration :** négociation avec les producteurs, respect des normes, audits
- Preservation Planning :** suivi de l'environnement de l'OAIS, politique d'archivage, choix des normes
- Access :** réception des requêtes des utilisateurs, production des réponses
- Services communs :** systèmes d'exploitation, réseaux, sécurité



CENTRE NATIONAL D'ÉTUDES SPATIALES

**PRÉSERVER LES DONNÉES PHYSIQUES  
LA MÉTHODOLOGIE PAIMAS**

**Producer-Archive Interface Methodology Abstract Standard  
CCSDS 651.0-B-1  
ISO 20652:2006**

## Le problème à résoudre

### *Constat*

**Les relations entre un producteur d'information et une archive sont compliquées :**

- **le producteur peut être une personne physique, une institution (ex. : laboratoire), un centre de traitement de données ;**
- **le producteur traite des informations d'origine généralement unique ;**
- **son horizon temporel est généralement borné (un projet a un début et une fin) ;**
- **l'archive est toujours une institution (généralement lourde) ;**
- **l'archive traite des informations d'origines souvent très diverses ;**
- **son horizon temporel est rarement borné (l'archivage a toujours un début, mais pas de fin).**

### *Difficulté*

**Les exigences et attentes respectives du producteur et de l'archive divergent généralement. Les relations entre producteur et archive peuvent se gripper :**

- **ce que l'archive reçoit n'est pas ce qui est attendu ;**
- **le producteur ne peut se conformer au calendrier de livraison ;**
- **les erreurs de transfert entre producteur et archive ne sont pas détectées facilement.**

## Contrat entre producteur et archive

**Les relations entre producteur et archive doivent être régulées par un contrat (Submission Agreement) régissant un projet d'ingestion d'information du producteur vers l'archive. Ce projet est commun au producteur et à l'archive.**

**Le contrat définit les flux d'information du producteur vers l'archive :**

- **information à préserver**
- **information pour la constitution des AIP**
  - ◆ **information fournie par le producteur au titre du projet en cours**
  - ◆ **information fournie par le producteur au titre du précédent projet**
  - ◆ **information fournie par une autre institution (organisme de normalisation, etc.)**
  - ◆ **information fournie par l'archive**

**Une approche “phasée” est nécessaire pour la définition de ces flux.**

## Les phases

### ■ Phase préliminaire

- ◆ **premiers contacts (qui fera quoi ?, qu'est-ce qui sera transféré ?)**
- ◆ **définitions préliminaires**
  - détermination de l'information à archiver
  - détermination des objets de données et des normes applicables
  - quantification (volumes à transférer, à archiver)
  - exigences de sécurité (confidentialité, etc.)
  - contraintes contractuelles et réglementaires (copyrights, etc.)
  - définition des SIP
  - procédures de validation des SIP
  - calendrier des opérations
  - évaluation des conséquences du transfert sur l'archive
  - Synthèse des coûts, risques
  - Points critiques
- ◆ **Rédaction du projet de contrat**

## Les phases (suite)

### ■ Phase de définition formelle

- ◆ formalisation des éléments de la phase préliminaire
- ◆ détermination et modélisation des objets de données à transférer (dictionnaires de données)

### ■ Phase de transfert

- ◆ test du transfert
- ◆ exécution du transfert
- ◆ vérification de la bonne exécution du transfert (ex. : pas de perte de données), actions correctrices

### ■ Phase de validation

- ◆ test initial de validation
- ◆ exécution des procédures de validation
  - validations systématiques
  - validations à la demande (p.ex. à la fin d'une session de transferts)

**Le chevauchement possible entre phases de transfert et de validation.**

## **Méthodologie d'interface entre producteur et archive**

**La norme “abstraite” d'interface entre un producteur et une archive doit être traduite en norme “concrète” applicable à un problème “concret”.**

**La norme concrète devra prendre en considération :**

- le vocabulaire des communautés concernées ;**
- les modèles d'information qu'elles utilisent habituellement ;**
- les normes en vigueur**
- les outils qu'elles utilisent habituellement (procédures, etc.).**



CENTRE NATIONAL D'ÉTUDES SPATIALES

## PRÉSERVER LE CONTEXTE TECHNIQUE

## Le problème à résoudre

**Un ensemble de données soigneusement archivé peut cependant rapidement devenir inutilisable si l'on n'a pris soin de conserver aussi la description précise de ses formats d'enregistrement.**

**Le format d'enregistrement est le résultat du processus de conception de l'ensemble de données. Il doit être rendu explicite au plus tard à la fin du processus de conception et indéfectiblement lié à l'ensemble de données (sinon, il y a risque de divergence, par suite d'évolutions, entre le format réel et sa description).**

**Un format d'enregistrement se compose de deux parties :**

- **une partie “logique”, donnant le type de données et leur restriction éventuelle à un sous-type, la longueur en bits des données, leur position dans l'enregistrement**
  - ◆ **ex. : WEEK\_End is (type) DAY range SAT..SUN**
- **une partie “physique” donnant l'enregistrement physique des données sur leur support**
  - ◆ **ex. : position du bit de poids fort dans un octet, formes des exposant et mantisse d'un nombre flottant**

## Le langage EAST et les produits associés

**Dialecte du langage ADA, conçu par le CNES pour la description des formats d'enregistrement.**

**Un format d'enregistrement est rendu par un Data Description Record exprimé à l'aide de déclarations EAST.**

**Les déclarations EAST peuvent être produites à l'aide de l'interface graphique OASIS développée par le CNES. À l'aide d'OASIS, il est possible de :**

- **définir l'organisation des données d'un ensemble de données (leur arborescence) ;**
- **définir la syntaxe des données (énumératif, numérique, ASCII, binaire, etc.) ;**
- **définir au niveau d'un champ (noeud ou feuille de l'arborescence) des attributs sémantiques (signification, unités, domaine de validité, etc.) ;**
- **stocker les caractéristiques syntaxiques et sémantiques des données dans un dictionnaire de données réutilisables.**

**Une application produit les données en conformité avec leurs descriptifs EAST grâce à un générateur de données (pas de logiciel d'écriture spécifique).**

**Une application lit les données en conformité avec leurs descriptifs EAST grâce à un interpréteur de données (pas de logiciel de lecture spécifique).**

## Le langage EAST et les produits associés (suite)

Le langage EAST est une recommandation CCSDS : CCSDS 644.0-B-2  
et une norme ISO : ISO 15889:2003.

Le dictionnaire de données généré par OASIS est conforme à la recommandation CCSDS sur la spécification des dictionnaires de données: CCSDS 647.1-B-1 (Data Entity Dictionary Specification Language (DEDSL) – Abstract Syntax).

DEDSL est aussi une norme ISO : ISO 21961:2003.

D'autres informations sur EAST et les produits associés : <http://east.cnes.fr>.



CENTRE NATIONAL D'ÉTUDES SPATIALES

**PRÉSERVER LE CONTEXTE THÉMATIQUE**

## Préserver le contexte thématique

**Enjeu : Un organisme doit être capable de rendre compte sur le long terme du patrimoine de jeux de données issus de la télémessure des instruments qu'il a développés directement ou dans le cadre de coopérations.**

**Pour être recevable par le plus grand nombre, un jeu de données doit être :**

- **identifié** (titre, mots-clés, couverture spatio-temporelle,...)
- **sinon on ne peut pas savoir qu'il existe**
- **convenablement décrit** (caractéristiques, objet, qualité, référentiel d'emploi,...)
- **s'il est mal ou insuffisamment décrit aux yeux de l'utilisateur potentiel, il sera ignoré ou rejeté par celui-ci**
- **utilisable**
  - ◆ **il faut savoir où le trouver**
  - ◆ **il faut savoir y accéder**
  - ◆ **il faut savoir en tirer parti**

## Rendre compte du contexte thématique

### De la donnée à l'information

#### ■ il n'y a pas grande différence entre un jeu de données et un document

- ◆ *Un document est une preuve à l'appui d'un fait (Suzanne Briet, 1951)*
- ◆ *Les données sont une représentation formelle de faits, concepts ou instructions convenant à la communication, l'interprétation ou le traitement par des êtres humains ou des automatismes. (ISO 11179 --Specification and standardization of data elements)*

#### ■ Jeux de données et documents sont porteurs d'information

- ◆ *information : connaissance relative à des objets tels que des faits, événements, choses, processus ou idées, y compris des concepts, laquelle, dans un certain contexte, a une signification particulière (ISO 11179)*
- ◆ *objet : élément de la réalité qui peut être conçu ou perçu (ISO 1087 --Terminology – Vocabulary)*
- ◆ *concept : unité de pensée constituée par abstraction sur la base des caractéristiques communes à un ensemble d'objets (ISO 1087). Un concept peut être superordonné ou subordonné à d'autres concepts dans une relation générique ou partitive, ou coordonné avec d'autres concepts.*

## Rendre compte du contexte thématique

### De l'information au jugement

- *Notre connaissance émane de deux sources principales de notre âme, dont la première est la capacité d'accueillir les représentations (la réceptivité des impressions), la seconde la faculté de reconnaître un objet par ces représentations (la spontanéité des concepts) ; par la première, un objet nous est donné, par la seconde, il est pensé en rapport avec cette représentation (en tant que pure détermination de l'âme). Intuition et concepts constituent ainsi les éléments de toute notre connaissance, de telle sorte que ni les concepts sans une intuition qui leur corresponde d'une certaine façon, ni une intuition sans concept ne peuvent fournir de connaissance. (Emmanuel Kant, Critique de la Raison Pure, Théorie transcendantale élémentaire, 2<sup>ème</sup> partie : La logique transcendantale, Introduction : Idée d'une logique transcendantale, I. : De la logique en général)*
- *Toutes les intuitions, en tant que sensibles, reposent sur des affections, et donc les concepts sur des fonctions. J'entends par fonction l'unité de l'acte par lequel diverses représentations sont ordonnées sous une représentation commune. Les concepts se fondent donc sur la spontanéité de la pensée, comme les intuitions sensibles sur la réceptivité des impressions. De ces concepts, l'entendement ne peut alors faire d'autre usage que de juger par leur moyen. Or comme aucune représentation ne va immédiatement à l'objet, à l'exception de l'intuition, ainsi un concept n'est jamais rapporté immédiatement à un objet, mais à quelque autre représentation de celui-ci (qu'elle soit intuition ou qu'elle soit elle-même déjà concept). Le jugement est ainsi la connaissance médiate d'un objet, par conséquent la représentation d'une représentation de celui-ci. Dans chaque jugement il y a un concept qui compte pour beaucoup d'autres concepts et qui, parmi ceux-là, comprend une représentation donnée, laquelle est rapportée immédiatement à l'objet... Penser, c'est connaître par concepts. Et les concepts sont rapportés, en tant que prédicats de jugements possibles, à quelque représentation d'un objet encore indéterminé. (Emmanuel Kant, Critique de la Raison Pure, Théorie transcendantale élémentaire, 2<sup>ème</sup> partie : La logique transcendantale, 1<sup>ère</sup> division: L'analytique transcendantale, 1<sup>er</sup> livre : L'analytique des concepts)*

## Rendre compte du contexte thématique

### La modélisation conceptuelle

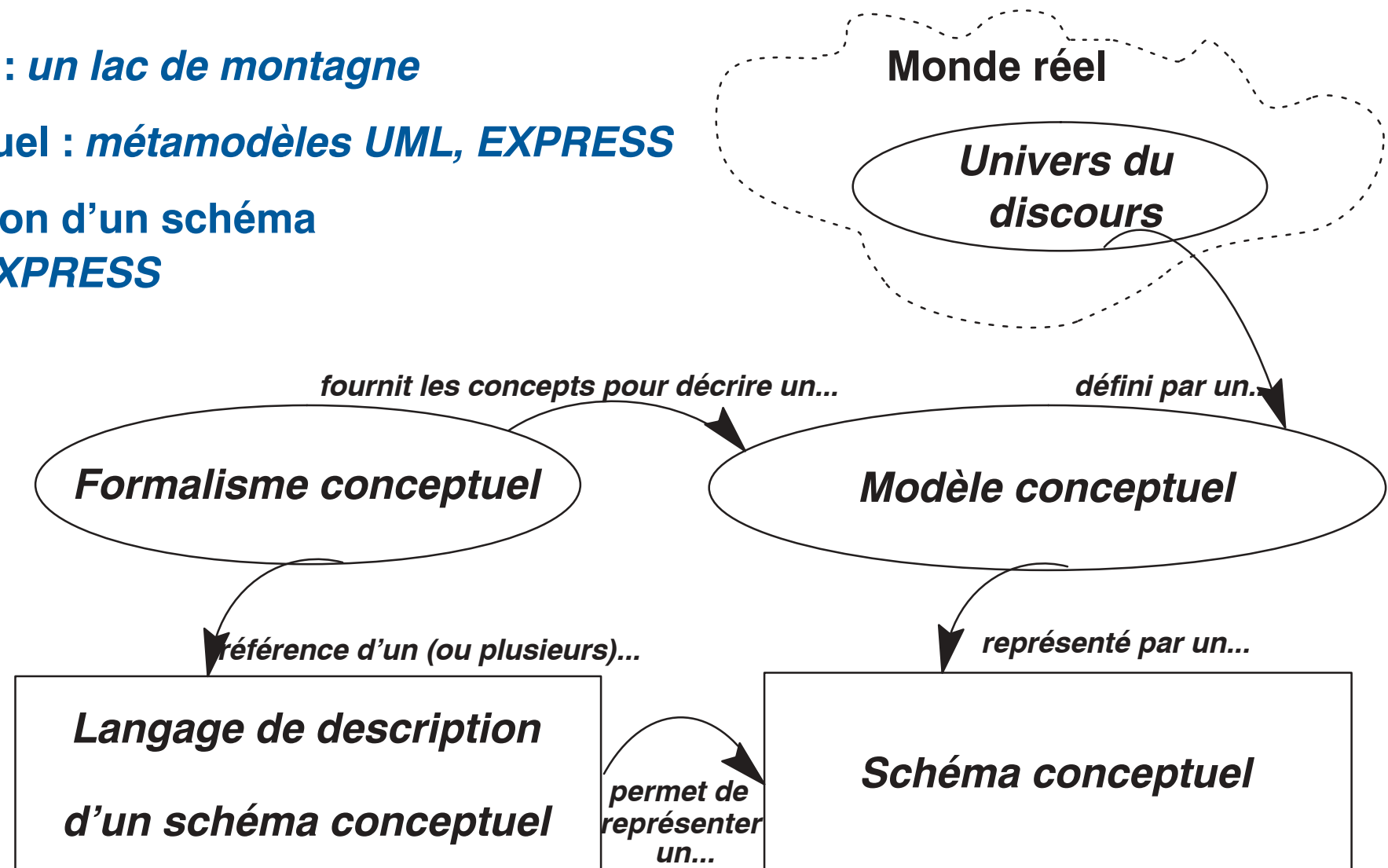
- rendre compte du contexte thématique d'un jeu de données, c'est rendre explicites les concepts qui en constituent l'armature
- tout jeu de données ou document manifeste un certain regard sur le monde réel ; ce regard dépend du modèle conceptuel appliqué au monde par l'observateur

## La modélisation conceptuelle

Source : ISO19103 *Geographic information – Conceptual schema language*

### Exemples

- univers du discours : *un lac de montagne*
- formalisme conceptuel : *métamodèles UML, EXPRESS*
- langage de description d'un schéma conceptuel : *UML, EXPRESS*



## La modélisation conceptuelle

**L'univers du discours est un fragment du monde réel ou d'un monde hypothétique que l'on souhaite modéliser. Le résultat de la modélisation est un modèle conceptuel.**

**L'univers du discours ne contient pas seulement des entités "physiques". Il contient aussi, éventuellement, leurs propriétés, leurs fonctions, les relations qui peuvent exister entre plusieurs entités.**

**Le modèle conceptuel est construit à l'aide des concepts proposés par un formalisme conceptuel (il peut y avoir plusieurs formalismes conceptuels pour le même univers du discours).**

**Le modèle conceptuel peut rester intellectuel. Il peut aussi être représenté par un schéma conceptuel exprimé à l'aide d'un langage de description spécifique, capable de rendre compte du formalisme conceptuel.**

*Cette représentation d'un procédé général de la faculté d'imagination, servant à procurer d'un concept son image, je l'appelle le schéma appliqué à ce concept.* (Emmanuel Kant, Critique de la Raison Pure, Théorie transcendantale élémentaire, 2<sup>ème</sup> partie : La logique transcendantale, 1<sup>ère</sup> section : L'analytique transcendantale, Livre deuxième : L'analytique des principes, Chapitre 1 : Du schématisme des concepts purs de l'entendement).

## La modélisation conceptuelle

Lorsque l'univers du discours est la terre (et son environnement), nous pouvons utiliser un formalisme conceptuel comme celui-ci :

- une entité est désignée par son nom (ex. : maison) et caractérisée par
  - ◆ sa définition (ex. : bâtiment destiné à l'habitation d'une famille)
  - ◆ ses propriétés
  - ◆ ses contraintes (ex. : hauteur maximale)
  - ◆ ses relations avec d'autres entités (ex. : cas particulier d'habitation)
- une propriété peut être
  - ◆ une opération applicable (ex. : élévation)
  - ◆ un attribut propre (ex. : hauteur)
  - ◆ une association (ex. : mitoyenneté avec un immeuble)
- un attribut est caractérisé par son type
  - ◆ thématique (ex. : style de la maison)
  - ◆ temporel (ex. : date de construction)
  - ◆ localisation géographique (ex. : nom de la ville)

## La modélisation conceptuelle

**Ce formalisme permet de modéliser l'entité géographique constituée (comme suggéré plus haut) d'une maison.**

**Mais le même formalisme peut s'appliquer à quelque chose de plus immatériel comme l'entité géographique constituée par le vent, au-dessus d'une zone particulière de l'océan.**

**En pratique, nous avons accès à l'univers du discours par l'intermédiaire de perceptions, observations, mesures (cf. la citation d'E. Kant sur l'intuition) filtrées conformément à un modèle conceptuel. L'enregistrement de celles-ci n'abolit pas ce modèle conceptuel mais l'enregistrement est lui-même tributaire d'un modèle conceptuel qui peut, à son tour, être rendu par un schéma conceptuel. Ce schéma porte le nom de schéma d'application (ISO 19101).**

**Un schéma d'application est donc le schéma d'un jeu de données. Il rend compte de façon complète et précise du contenu et de la structure d'un jeu de données.**

## La modélisation conceptuelle

**Modèles et schémas conceptuels sont des appareils qui nous permettent d'exprimer des jugements**

*(les concepts sont des prédicats de jugements possibles).*

**Un jugement est valable s'il est recevable par d'autres** (sinon, il est seulement l'expression d'une opinion). **Pour cela, modèles et schémas conceptuels à partir desquels le jugement est rendu doivent être ceux d'une communauté. Cette adhésion s'exprime par des normes** (de droit ou de fait).

**La normalisation est l'activité de synthèse par lequel une communauté exprime consciemment les concepts qui la fondent comme communauté :**

*Seule, la liaison (conjunctio) d'une diversité d'éléments en général ne peut jamais venir en nous par l'intermédiaire des sens et ne peut donc pas non plus être contenue en même temps dans la forme pure de l'intuition sensible car elle est un acte de la spontanéité de la faculté de représentation; et, comme il faut l'appeler entendement, pour la distinguer de la sensibilité, ainsi toute liaison, que nous en ayons conscience ou non, qu'elle soit liaison d'une diversité d'éléments de l'intuition ou de concepts et, s'agissant de l'intuition, que celle-ci soit sensible ou non, est un acte de l'entendement que nous désignerons du nom commun de synthèse, afin de faire remarquer ainsi que nous ne pouvons alors rien nous représenter comme lié dans l'objet si nous ne l'avons pas lié nous-même auparavant, et que, de toutes les représentations, la liaison est la seule qui ne puisse nous être fournie par des objets, mais seulement rapportée par le sujet lui-même, parce qu'elle est un acte de sa spontanéité. On remarquera sans peine ici que cet acte doit être originellement unique et équivalent pour toute liaison, et que la décomposition, l'analyse, qui semble être son contraire, la présuppose toujours ; car où l'entendement n'a rien lié auparavant, il ne saurait non plus rien décomposer, puisque c'est par lui seul que ce qui a été lié a pu être donné comme lié à la faculté de représentation. (Emmanuel Kant, Critique de la Raison Pure, Théorie transcendantale élémentaire, 2<sup>ème</sup> partie: La logique transcendantale, 1<sup>ère</sup> division : L'analytique transcendantale, 1<sup>er</sup> livre : L'analytique des concepts, 2<sup>ème</sup> section : De la déduction des concepts purs de l'entendement, 2<sup>ème</sup> paragraphe : De la déduction transcendantale des concepts purs de l'entendement : De la possibilité en général d'une liaison).*

## Les métadonnées

**Les métadonnées (*metadata*) sont des données particulières. Elles sont habituellement agencées en jeux de métadonnées (*metadata datasets*).**

**Un jeu de métadonnées rend compte d'un ou plusieurs jeux de données, en tant que jeu de données. La manière dont un jeu de métadonnées rend compte d'un jeu de données en tant que jeu de données est déterminée par le modèle conceptuel (et le schéma conceptuel) dont le jeu de métadonnées est tributaire.**

**Ce modèle conceptuel est l'appareil qui permet d'exprimer un jugement sur un jeu de données en tant que jeu de données. Un tel jugement permet éventuellement de distinguer un jeu de données parmi d'autres jeux de données. Un jeu de métadonnées est ordinairement un instrument de discernement.**

**Comme tout jeu de données, un jeu de métadonnées est régi par un schéma d'application.**

## La normalisation des jeux de métadonnées

Pour que le jugement rendu par un jeu de métadonnées soit valable, c'est-à-dire recevable par d'autres, il faut que son modèle conceptuel (et donc son schéma d'application) soit celui d'une communauté. Cette adhésion s'exprime, comme pour les autres schémas d'application par des normes de droit ou de fait.

Quelques normes de métadonnées (par complexité croissante)

- **Dublin Core**      **Dublin Core Metadata Initiative (DCMI)**
- **BIB-1**            **norme bibliographique internationale (Library of Congress)**
- **GILS**             **Government Information Locator Service (USA)**
- **DIF**              **Directory Interchange Format (CEOS/GCMD)**
- **CSDGM**         **Content Standard for Digital Geographic Metadata (FGDC)**
- **ISO 19115**      **Geographic Information Metadata (ISO)**

## La norme ISO 19115

- publiée par l'ISO en 2003
- dévolue à la description de l'information géographique
- plus de 400 éléments descriptifs pour un jeu de données, ainsi classés :
  - ◆ **identification**
    - références et description synthétique du jeu de données
    - couverture spatio-temporelle
  - ◆ **contenu**
    - entités géographiques
    - attributs
  - ◆ **distribution**
  - ◆ **qualité**
    - indicateurs de qualité
    - historique du jeu de données
  - ◆ **représentation spatiale**
  - ◆ **systèmes de référence**
  - ◆ **présentation**
  - ◆ **schéma d'application**

Le langage de description de schéma conceptuel est composé de :

- UML (partie statique)
- IDL (types de base)
- OCL
- dictionnaire ISO 11179

## Le Bureau des Métadonnées du CNES Point de vue technique

**Le Bureau des Métadonnées du CNES est un système d'information composé**

■ **d'une ou plusieurs bases de jeux de métadonnées**

- ◆ réparties en plusieurs lieux
- ◆ chaque base est alimentée au niveau local
- ◆ toutes les bases peuvent être interrogées en même temps

■ **d'interfaces d'accès pour :**

- ◆ l'interrogation par l'utilisateur
- ◆ l'administration des jeux de métadonnées
  - ingestion de métadonnées
  - retrait de métadonnées
- ◆ le management global
  - détermination des normes de métadonnées acceptées
  - détermination des profils d'utilisation

## Le Bureau des Métadonnées du CNES

### Point de vue technique (suite)

La conception du Bureau des Métadonnées suit les recommandations de l'OAIS (Reference Model for an Open Archival Information System du [CCSDS](#))

Ce modèle définit un système d'archivage comme étant la composition d'un ensemble d'unités fonctionnelles pouvant avoir chacune son autonomie :

- ◆ ingest *ingestion des métadonnées*
- ◆ archival storage *stockage des métadonnées*
- ◆ data management *gestion des métadonnées en base*
- ◆ administration *contrôle au quotidien du Bureau des Métadonnées (ex. :acceptation des métadonnées)*
- ◆ preservation planning *observation de l'environnement du Bureau des Métadonnées, audits*
- ◆ access *accès aux métadonnées par les utilisateurs*

Les métadonnées sont confectionnées à l'extérieur du système à l'aide d'un éditeur structuré. Les fichiers résultant de l'édition structurée sont ingérés dans le système.

## Le Bureau des Métadonnées du CNES

### Point de vue thématique

Le Bureau des Métadonnées sert les besoins de l'observation de la terre. L'analyse du besoin a été conduite par la méthode des interviews : des producteurs potentiels de métadonnées comme des utilisateurs potentiels du Bureau des Métadonnées ont été interrogés.

#### *Conclusions de l'analyse du besoin*

##### ■ approche du Bureau des Métadonnées par profil d'utilisation

- ◆ un profil d'utilisation rend compte d'un domaine (métier, thématique,...) ;
- ◆ un profil d'utilisation est déterminé par un ensemble de mots clés représentatifs du domaine ; ces mots clés sont contrôlés par des thésaurus ;

##### ■ faculté d'interroger le Bureau des Métadonnées par mots clés thématiques

##### ■ faculté d'interroger le Bureau des Métadonnées par sélection géographique

- ◆ noms géographiques
- ◆ rectangle

##### ■ faculté d'interroger le Bureau des Métadonnées par sélection temporelle

- ◆ dates de début et de fin
- ◆ saison

##### ■ faculté d'accès à des représentations graphiques (*browse, quicklooks,...*)

## Le Bureau des Métadonnées du CNES

### Point de vue thématique (suite)

A chaque profil sont associés également :

- les critères de sélection pertinents pour le profil
- les règles de présentation à l'utilisateur
  - ◆ composition et présentation des *shortlists* résultant d'une interrogation de l'utilisateur
  - ◆ présentation des métadonnées à l'utilisateur

**Le nombre de profils n'est pas limité. Il est même préférable d'en créer beaucoup pour réduire le bruit** (ratio des métadonnées indésirables / total des métadonnées restituées).

**Il est possible de créer des profils génériques englobant tous les autres profils.**

**Le Bureau des Métadonnées gère à part les informations nominatives** (ex. : adresse d'un organisme). **Une modification dans une information nominative est répercutée dans toute présentation de métadonnées.**

**Le Bureau des Métadonnées donne accès en ligne à des documents adventices** (thesaurus, etc.)

## **Le Bureau des Métadonnées du CNES**

### **Point de vue thématique (suite)**

**Le Bureau des Métadonnées ne reconnaît qu'une norme à la fois : toutes les métadonnées devant être ingérées par le Bureau des Métadonnées doivent être conformes à cette norme.**

**Toutefois, le Bureau des Métadonnées peut représenter une partie seulement des éléments de métadonnées restituées à l'utilisateur, par application d'un filtre. En particulier, l'application du filtre peut convertir, sur demande de l'utilisateur, la norme de référence du Bureau des Métadonnées en une norme plus simple (ex. : Dublin Core)**

**A ce jour, la norme de référence du Bureau des Métadonnées est la norme**

**ISO 19115:2003 Geographic data – Metadata**

**Elle a été choisie car elle est la plus riche de toutes les normes connues au CNES, dans le domaine de l'observation de la terre.**

**Le Bureau des Métadonnées peut gérer simultanément plusieurs versions de la norme.**

## Le Bureau des Métadonnées du CNES

### Point de vue thématique (suite)

**Nous assimilons un jeu de métadonnées à un *document*, c'est-à-dire au support d'un enseignement (cf. l'étymologie du mot *document*).**

**Un jeu de métadonnées doit être pensé en fonction de ses destinataires : les utilisateurs du Bureau des Métadonnées. Il devrait se couler dans le moule des catégories intellectuelles selon lesquelles les utilisateurs du Bureau des Métadonnées conçoivent leur approche du réel.**

**Il faut se souvenir que l'utilisateur du Bureau des Métadonnées est en situation d'ignorance. Cette situation d'ignorance est le plus souvent une situation pénible. Le caractère pénible de l'ignorance est le principal moteur d'une recherche d'information (Michael Buckland).**

#### ***Méthode :***

**La composition d'un jeu de métadonnées devrait suivre les préceptes de l'analyse documentaire : observer, comprendre, structurer, formuler, résumer, indexer**

**(cf. *L'analyse documentaire*, Suzanne Waller, ADBS Editions, 1999).**

## Le Bureau des Métadonnées du CNES

### Point de vue thématique (suite)

- **observer** : survoler le jeu de données, le situer dans son environnement (autres jeux de données, lieux de stockage, questionnaires, etc.), identifier sa bibliographie (auteurs, date, langues, citations, etc.), en bref tous les éléments factuels ;
- **comprendre** : discerner la problématique du jeu de données et les catégories d'intérêt pour lesquelles il peut être une réponse ;
- **structurer** : discerner la logique qui a conduit le jeu de données à ce qu'il est, avec ses qualités et ses limites ;
- **formuler** : identifier les éléments de métadonnées qu'il conviendra d'enrichir ; identifier le vocabulaire adéquat ;
- **résumer** : *Le résumé est un texte concis reflétant fidèlement, sans interprétation ni critique, le contenu d'un document. Il a pour but d'aider le lecteur à cerner la pertinence du document, vis-à-vis de l'information recherchée (AFNOR Z 44-004) ;*
- **indexer** : trouver les mots qui inspireront la requête de l'utilisateur, choisis dans un document de contrôle utilisant un langage documentaire : thésaurus (ISO 2788), catalogue de noms géographiques (ISO 19112), catalogue d'entités géographiques (ISO 19110).

## Le Bureau des Métadonnées du CNES

### Point de vue thématique (suite)

L'expérimentation thématique du Bureau des Métadonnées par le GIP MEDIAS France a mis en lumière la difficulté à composer des jeux de métadonnées porteurs de sens pour les utilisateurs :

- que renseigner dans les métadonnées ?
- quel titre choisir ?
- comment rédiger un résumé ?
- quels sont les mots clés pertinents ? dans quels thésaurus ?
- comment rendre compte de la qualité d'un jeu de données ?

Remarque de l'un des *cobayes* : *la restitution de l'information devrait être en "langage humain"* (cf. le rapport bilan de l'expérimentation).

A notre sens, ces difficultés ont pour origine l'insuffisante modélisation conceptuelle au sein des communautés.

## Le Bureau des Métadonnées du CNES

### Point de vue thématique (suite)

**Le Bureau des Métadonnées est un service d'information. Pour être reconnu comme tel, il devra surmonter six obstacles :**

- **identification** : une source d'information doit être identifiable comme réponse possible à une série de questions ;
- **disponibilité** : une source d'information doit être communicable sous une forme appropriée ;
- **prix payé par l'utilisateur** : le prix (en argent, temps, effort, inconfort) doit être acceptable pour l'utilisateur ;
- **prix payé par le service d'information** : le prix (en argent, temps, effort, aménagement) doit être acceptable pour la tutelle du service ;
- **compréhension** : la source d'information communiquée à l'utilisateur doit être comprise de ce dernier ; sinon, il faut lui fournir l'expertise nécessaire ;
- **acceptabilité** : une autorité reconnue doit endosser la source d'information communiquée à l'utilisateur pour que celle-ci soit acceptée.

*(cf. Information and Information Systems, Michael Buckland, Greenwood Press, 1991)*

## Retour sur le Bureau des Métadonnées du CNES

### *Point de vue technique*

- la technologie est globalement maîtrisée par le CNES

### *Point de vue thématique*

- une phase d'expérimentation thématique du Bureau des Métadonnées par MEDIAS France a conduit à des recommandations d'amélioration du Bureau des Métadonnées, prises en compte dans la version actuelle
- une phase de composition opérationnelle de jeux de métadonnées pour le Bureau des Métadonnées par MEDIAS France a confirmé la difficulté de l'entreprise

### *Point de vue institutionnel*

- plusieurs rapports ont été produits en interne et en coopération avec d'autres organismes sur l'intérêt, pour l'établissement, d'une politique "métadonnées"
- pour l'instant, la greffe n'a pas pris
  - ◆ mais pas d'obstacle de principe à la poursuite de l'activité
- sans doute parce qu'il s'agit d'une activité *transverse*, plutôt en décalage par rapport à la culture *projet* du CNES

## Un point de vue de l'étranger

### Executive Order 12906

#### “Coordinating Geographic Data Acquisition and Access : The National Spatial Data Infrastructure”

*Beginning nine months from the date of this order, each agency shall document all new geospatial data it collects and produces, either directly or indirectly, using the standard under development by the FGDC, and make that standardized documentation electronically accessible to the Clearinghouse network. Within one year of the date of this order, agencies shall adopt a schedule, developed in conjunction with the FGDC, for documenting, to the extent practical, geospatial data previously collected or produced, either directly or indirectly, and making that data documentation electronically accessible to the Clearinghouse network.*

**(signé en 1994 par le Président Clinton)**



CENTRE NATIONAL D'ÉTUDES SPATIALES

**DE L'INFORMATION À LA CONNAISSANCE**

## Le problème à résoudre

L'information restituée à l'utilisateur par une archive est utile parce qu'elle permet à son destinataire de modifier son capital de connaissances.

Selon le cas, l'information

- augmente le capital de connaissances ;
- appauvrit le capital de connaissances
  - ◆ p. ex. : lorsque l'information restituée contredit une autre information.

L'utilisateur ne peut pas toujours modifier son capital de connaissances lorsqu'il reçoit une information. Parfois l'information n'est pas lisible. Il faut alors la transformer.

Exemple : un tableau de chiffres n'est pas lisible à l'oeil ; il peut le devenir après application d'une palette de couleurs.

La transformation de l'information pour que l'on puisse en mesurer son impact sur un capital de connaissances est appelée **service**.

Les services sont un moyen de valoriser l'information.

## Les services

### *Définitions*

- **service** : partie distinguée d'une fonctionnalité fournie par une entité via des interfaces
- **interface** : ensemble nommé d'opérations qui caractérisent le comportement d'une entité
- **opération** : spécification d'une transformation ou d'une requête qu'un objet peut être appelé à exécuter ; une opération a un nom et une liste de paramètres

**Un service fournit une fonctionnalité offrant à son utilisateur une valeur ajoutée.**

**Une interface est une spécification abstraite des opérations mises en oeuvre et des restrictions éventuelles à l'ordre de leur mise en oeuvre.**

**Une opération est la spécification du changement d'état de l'objet auquel elle s'applique. Elle est aussi la spécification abstraite d'une action supportée par l'interface. Une opération peut comporter des paramètres.**

**La mise en oeuvre concrète d'une interface s'appelle un port. Elle comprend la spécification propre à la plateforme de mise en oeuvre et une méthode d'identification du service (p. ex : une adresse).**

**(source ISO 19119)**

## Chaînage des services

Lorsque les entrées d'un service dépendent d'un autre service, les deux services sont considérés comme étant chaînés. Une chaîne de services peut être représentée par un graphe orienté (cyclique ou acyclique). Des chaînes avec des séquences de services en parallèle sont concevables.

La gestion d'une chaîne de services est confiée à un mécanisme de **workflow**, dont la complexité dépend du type de chaînage.

On distingue dans la norme ISO 19119 trois types de chaînage :

- chaînage transparent (tout est défini par l'utilisateur qui "a la main" sur la chaîne)
- chaînage translucide (la chaîne est gérée par un workflow mais l'utilisateur "voit" les services de la chaîne")
- chaînage opaque (la chaîne est gérée par le workflow ; l'utilisateur ne voit qu'un seul service, agrégat des services de la chaîne).

On peut concevoir une arborescence de chaînes : un service d'une chaîne peut lui-même être constitué d'une chaîne de services.

## Taxinomie des services

### *Taxinomie proposée par la norme ISO 19119 pour les données géographiques*

- **Geographic human interaction services** (ex. : services de présentation de l'information)
- **Geographic model/information management services** (ex. : services de catalogage, services d'accès à l'information)
- **Geographic workflow/task management services** (ex. : activation d'un workflow)
- **Geographic processing services**
  - ◆ **Geographic processing services – spatial** (ex. : service de rectification d'image)
  - ◆ **Geographic processing services – thematic** (ex. : service de “subsetting”, service de classification thématique)
  - ◆ **Geographic processing services – temporal** (ex. : service d'échantillonnage temporel)
  - ◆ **Geographic processing services – metadata** (ex. : service d'annotation d'une image)
- **Geographic communication services** (ex. : service de codage de l'information géographique en vue de son transport)

**Pour l'ensemble de ces catégories, la norme ISO 19119 identifie environ 80 types de services.**

## Les métadonnées de service

**Les métadonnées décrivent des ressources, généralement des ensembles de données.**

**Un service peut être considéré comme une ressource. On peut donc légitimement lui associer aussi des métadonnées. Ces métadonnées sont des métadonnées de service.**

**La norme ISO 19119 est une norme de métadonnées pour des services relatifs à l'information géographique.**

**Le serveur de métadonnées "International Directory Network" ([IDN](#)) propose plusieurs centaines de métadonnées pour des services relatifs à l'observation de la terre. Il s'agit de métadonnées très simples régies par le format IDN de métadonnées de service "Service Entry Resource Format" ([SERF](#)).**

## **Le Bureau des Métadonnées et des Services du CNES**

**Le CNES a étendu en 2006 le Bureau des Métadonnées aux métadonnées de service.**

### ***Gestion des métadonnées de service***

- **choix de la norme ISO 19119 et de la taxinomie ISO 19119 pour le classement des services**
- **les métadonnées de service sont traitées selon le même principe que les métadonnées de données (ingestion, gestion en base, interrogation)**

### ***Lancement des services par l'utilisateur***

- **visualisation des métadonnées de service disponibles**
- **présentation à l'utilisateur des paramètres associés à chaque opération**
- **fourniture par l'utilisateur des valeurs de paramètres nécessaires**
- **choix d'un port**
- **lancement du service**
- **restitution du résultat**

**Le système résultant est encore à l'état de prototype.**

## Les services OGC

L'Open Geospatial Consortium (OGC) a spécifié plusieurs interfaces de services relatifs à l'information géographique ainsi que leurs mises en oeuvre

### ■ Catalogue Service Implementation Specification (CAT)

- ◆ interface pour l'interrogation de catalogues de métadonnées,
- ◆ interface pour la publication de métadonnées

### ■ Web Map Service (WMS) Implementation Specification

- ◆ interface pour l'interrogation de serveurs de cartes géographiques

### ■ Web Feature Service (WFS) Implementation Specification

- ◆ interface pour l'interrogation de serveurs d'entités géographiques
- ◆ interface pour la publication d'entités géographiques

### ■ Web Coverage Service (WCS) Implementation Specification

- ◆ interface pour l'interrogation de serveurs de données (tables de valeur sur un domaine spatio-temporel)

Les ports sont généralement des mises en oeuvre http (parfois Z39.50).

## Interopérabilité des services

### *Définition*

**Deux mises en oeuvre (deux “implémentations”) différentes d’un même service sont interopérables si l’invocation de l’une entraîne l’invocation de l’autre. Les résultats de l’une et l’autre invocation peuvent être fusionnés.**

### *Conséquences*

**La requête adressée lors de la première mise en oeuvre doit être transmise pour exécution lors de la seconde mise en oeuvre.**

- **une liaison doit exister pour véhiculer les messages d’une mise en oeuvre à l’autre ;**
- **la requête adressée lors de la première mise en oeuvre doit être codée par celle-ci sous une forme interprétable par le seconde mise en oeuvre ;**
- **réciroquement, les réponses de la seconde mise en oeuvre doivent être codées par celle-ci sous une forme interprétable par la première mise en oeuvre.**

**L’interopérabilité des services suppose qu’ils soient interconnectés, qu’ils partagent une même sémantique et une même syntaxe.**

**Rendre un service existant interopérable avec d’autres services identiques est toujours très difficile.**

**L’interopérabilité facilite la pérennisation des services.**

## Les niveaux d'interopérabilité du CEOS

**Le Committee on Earth Observation Satellites (CEOS) a défini quatre niveaux d'interopérabilité :**

- **Level 1.** A user accessing one service is routed directly to another related service. For example a user identifying a relevant data set in a directory is routed directly to the associated inventory service, or a user of one inventory service is routed automatically to another related inventory to extend a search. At this level, the user interface differences between the routed services are not hidden from the user.
- **Level 2.** This is similar to level 1 except that a user accessing one service is routed directly to another related service, and context information is passed at the same time. The context information typically includes information concerning the user, their interests, and the activity on the original service (for example a search query) which can be interpreted by the new service to assist the user in making further queries. For example search query criteria entered at one site can be passed to another, translated and loaded as a search query for the user to activate. At this level the user is still exposed to the differences in operation between the two services, but related information can be shared.
- **Level 3.** At level 3, the user is hidden from the specific operations of each service. A service request is routed automatically to one or more services, and the results from each returned to the original access point. At this level (and levels 1 and 2) each service is likely to use a different data model and the query may need translation to reflect these differences. Similarly, the query results may also reflect the individual service data models and may or may not be translated before they are presented to the user. At this level the user is no longer exposed to the differences between service operation.
- **Level 4.** At the highest level of interoperability a single data model is assumed to apply to all services within an interoperable system. Thus a query entered by a user can be executed using a single distributed database. This level of interoperability permits direct database operations such as union and joins between metadata held by different catalogue services and for operations. The interoperable system is thus seen as a single database system by the user.

**Il est admis que le niveau d'interopérabilité optimal est le niveau 3.**



CENTRE NATIONAL D'ÉTUDES SPATIALES

## LA PRÉSERVATION DES CONNAISSANCES

## Le problème à résoudre

### *Règle*

**L'usage commun que le Philosophe conseille de suivre, quand il faut donner des noms aux choses, veut généralement qu'on appelle sages ceux qui ordonnent droitement les choses et les gouvernent bien. C'est pourquoi, de toutes les idées que les hommes se font du sage, le Philosophe retient que le propre du sage est de mettre de l'ordre.**

*(Saint Thomas d'Aquin, Somme contre les Gentils)*

**Pour rendre compte à long terme de son patrimoine de connaissance, un organisme doit identifier, expliciter, collecter, ordonner le vocabulaire normatif de l'ensemble de ses systèmes de données.**

**Pour le CNES, la question se pose avec la sonde Rosetta conçue il y a plusieurs années pour rejoindre une comète en 2013. Le vocabulaire en vigueur au moment de la conception de la sonde devra pouvoir être compris en 2013.**

## Définitions

**Un vocabulaire normatif est constitué de termes dont l'ensemble constitue une terminologie. Une terminologie s'exprime dans la langue de l'intellection, liée à la connaissance intellectuelle des objets du monde, distincte de la langue d'usage, liée à l'usage pratique des objets du monde, et n'est pas un langage de représentation des objets du monde en vue de leur manipulation** (Christian Roche, Université de Savoie).

**Un terme est la désignation (c-à-d. la représentation) d'un concept au moyen d'une unité linguistique.**

(ISO 1087 – Terminologie – Vocabulaire)

## Un exemple simple de terminologie : le thésaurus

**Un thésaurus est une liste d'autorité organisée de descripteurs et de non-descripteurs obéissant à des règles terminologiques propres et reliés entre eux par des relations sémantiques (hiérarchiques, associatives, ou d'équivalence).**

**Cette liste sert à traduire en un langage artificiel dépourvu d'ambiguïté des notions exprimées en langage naturel.**

**Un descripteur est un mot ou un groupe de mots retenu dans un thésaurus et choisi parmi un ensemble de termes équivalents pour représenter sans ambiguïté une notion apparaissant dans un document ou dans une demande de recherche documentaire.**

(source : AFNOR NF Z 47-102 Principes généraux pour l'indexation des documents)

**Exemple de thésaurus thématique : GEneral Multilingual Environmental Thesaurus (GEMET)**

<b>ATMOSPHERE (air, climate)</b>	
atmosphere	S:14
. air	S:15
. . atmospheric composition	S:16
. . . atmospheric component	S:17
. . . . atmospheric ozone	S:18
. . . . . tropospheric ozone	S:19
. atmospheric process	S:20
. . air movement	S:21
. . air-water interaction	S:22
. . . ocean-air interface	S:23
. . atmospheric circulation	S:24
. atmospheric structure	S:25
. . atmospheric layering	S:26
. . . boundary layer	S:27
. . . inversion layer	S:28
. . . ozone layer	S:29
. . ionosphere	S:30
. . stratosphere	S:31
. . troposphere	S:32
. Earth-Sun relationship	S:33
. . season	S:34
. . solar radiation	S:35
. . . duration of sunshine	S:36
climate	S:37
. climate type	S:38
. . continental climate	S:39
. . . mountain climate	S:40
. . desert climate	S:41
. . equatorial climate	S:42
. . oceanic climate	S:43

. . temperate climate	S:44
. . . Mediterranean climate	S:45
. . tropical climate	S:46
. climatic change	S:47
. climatic factor	S:48
. . air temperature	S:49
. . atmospheric humidity	S:50
. meteorological phenomenon	S:51
. . atmospheric inversion	S:52
. . atmospheric precipitation	S:53
. . . rain	S:54
. . . snow	S:55
. . cloud	S:56
. . cyclone	S:57
. . drought	S:58
. . fog	S:59
. . frost	S:60
. . glaze	S:61
. . hail	S:62
. . haze	S:63
. . hurricane	S:64
. . mist	S:65
. . storm	S:66

Extrait du Thesaurus GEMET  
de l'Agence européenne de  
l'environnement

5539 descripteurs

## L'ontologie

L'identification des notions relatives aux objets du monde et celles de leurs relations relèvent de l'ontologie, définie comme l'étude de l'être en tant que tel, distinguée de l'étude de tel être en particulier (du grec  $\text{ον, οντος}$ , participe présent du verbe être).

Une ontologie célèbre est celle d'Aristote qui définit l'être par 10 catégories :

### ■ substance

- ◆ substance première (matérialité – ex. : Paul)
- ◆ substance seconde (essence – ex. : homme)

### ■ qualité, quantité, relation, action, passion, lieu, temps, disposition, possession (ces 9 catégories sont les “accidents”)

Pour Aristote, tout ce qui se dit d'un être se dit par rapport à l'une ou l'autre de ces catégories sous la forme de propositions, ou jugements ( $\lambda\omicron\gamma\omicron\varsigma$  - logos). Une proposition est constituée d'un sujet, d'un prédicat et d'un opérateur qui les relie. Les syllogismes constituent les règles de combinaison de propositions (prémisse majeure, prémisse mineure, conclusion) et établissent une logique. Les substances secondes conduisent (après Aristote) à la classification complexe des êtres (cf. arbre de Porphyre au III<sup>e</sup> siècle : détermination des genres par différenciation).

## La démarche ontologique

### *Approche “phénoménologique”*

- C'est une approche utilitariste, fondée sur l'observation des objets tels qu'ils sont perçus.

### *Approche “métaphysique”*

- C'est une approche rationnelle, fondée sur l'analyse conceptuelle.
- Les concepts sont pensés par subsomption.

(Source : Professeur Christian Roche, Terminologie et Ontologie, Larousse – Revue « Langages » – n° 157 – mars 2005)

**Les deux approches sont légitimes. L'approche phénoménologique est la plus immédiate. Son risque est de confondre l'ontologie avec les propriétés de l'outil d'expression de l'ontologie (ex. d'outil : l'éditeur “protege” de l'université de Stanford).**

## De l'ontologie à la terminologie

**L'ontologie doit être vue comme une méthode de modélisation des objets perçus ou conçus du monde réel qui a pour résultat une terminologie.**

**Or un organisme comme le CNES a besoin d'une terminologie pour ses centres de données, comme expression de leur pouvoir cognitif à l'intention de leurs utilisateurs potentiels, donnée sous la forme d'un ensemble.**

- **de termes et de leurs domaines de validité,**
- **de relations par lesquelles certains termes sont valablement liés,**
- **des propositions vraies dans un contexte donné.**

**En cela, la terminologie est un élément qui peut contribuer à une "meilleure" vision du CNES**

- **par les utilisateurs potentiels de ses données,**
- **par les autres agences dans la perspective de futures coopérations car toute terminologie a une portée normative et favorise l'interopérabilité,**
- **par lui-même comme élément constitutif de son identité.**

## Constituer une terminologie ?

### ■ pas de (vraie) difficulté technique

- ◆ le fondement théorique est solide (Logique descriptive)
- ◆ il existe des langages de spécification d'ontologies (notamment par le W3C : OWL)
- ◆ il existe des langages d'interrogation d'ontologies (notamment par le W3C : SPARQL)
- ◆ il existe des éditeurs d'ontologies (ex. : Protege, Swoop) et des “raisonneurs” (ex. :Pellet)

### ■ mais des conditions initiales

- ◆ volonté d'Établissement
- ◆ mise en perspective dans une politique de données
- ◆ une certaine directivité

### ■ et l'exigence de procéder avec méthode, par exemple :

- ◆ organisation et définition du champ d'application
- ◆ collecte de l'information connue des experts du domaine
- ◆ analyse de l'information collectée en vue d'une ontologie
- ◆ développement ontologique initial et constitution d'une terminologie préliminaire
- ◆ raffinement / validation

(source : Information Integration for Concurrent Engineering – IDEF5 Ontology Capture Method Report)



CENTRE NATIONAL D'ÉTUDES SPATIALES

**PÉRENNISATION, MISE À DISPOSITION ET VALORISATION  
DES DONNÉES SCIENTIFIQUES**

## Conclusion

**La pérennisation, la valorisation, la mise à disposition des données désignent l'ensemble des activités qu'il faut mener pour que des données, porteuses d'une information collectée, mise en forme et documentée dans un contexte déterminé, puissent restituer cette même information dans un autre contexte.**

**Pour cela, il nous faut comprendre la nature profonde des relations entre un système de données et son utilisateur, c'est-à-dire entre un gisement de savoir et un être humain en quête de savoir, doté à cette fin, d'une faculté de connaissance ou intellect "en lequel les formes des réalités sensibles ont une existence plus parfaite que dans les réalités sensibles ; elles sont alors plus simples et ont plus d'extension"<sup>1</sup>. De ce point de vue, le problème se ramène à celui d'une herméneutique, c'est-à-dire de la lecture d'un recueil porteur de savoir et de son interprétation par le lecteur. Cette aptitude du système à l'interprétation de son contenu est manifestée, dès lors qu'il s'établit, entre le système et son utilisateur, une connivence durable dont il revient au concepteur du système d'élaborer l'alchimie.**

1. *Saint Thomas d'Aquin, Somme contre les Gentils, II, c. 50.*