

Quelle transition digitale pour la recherche et quels impacts sur les métiers du numérique?

pascal.dayre@enseeiht.fr

Résumé

« Aujourd'hui, toutes les activités humaines sont impactées par la transition digitale. Le milieu de la recherche n'y fait pas exception. Nous passerons en revue les technologies actuelles ou en émergence qui vont nous impacter fortement. Puis, nous ferons un parallèle avec les pratiques dans les laboratoires pour essayer de comprendre et mieux appréhender comment nous préparer et être acteurs de ces changements. »

Point de vue

1. Je viens d'un environnement scientifique et technique :

l'IRIT : « le monde numérique au service de l'humain et de son environnement »

2. Bio

3. Mon expérience et du montage des programmes de 5 JDEVs (2011, 2013, 2015, 2017, 2019)

4. Expérience du montage d'un ensemble d'Actions Nationales de formation du CNRS

Historique des JDEVs

<http://devlog.cnrs.fr/jdev2017>
<http://devlog.cnrs.fr/jdev2015>
<http://devlog.cnrs.fr/jdev2013>
<http://devlog.cnrs.fr/jdev2011>



=> environ 200 jours sur les tendances du développement logiciel, des technologies de l'information et leurs utilisations dans la transition digitale des unités de recherche = les ferments du futur

Les réseaux métiers associés au fil des éditions avec DEVLOG : RDEI, Calcul, RBDD, RDM, RESINFO, COGITER, ...

- **Usines logicielles**

=> outils/environnements de développement et déploiement

- 2013.T4 - Les outils de production de code (outils de construction du logiciel)
- 2013.T6 - Les tests logiciel pour aller vers plus de pragmatisme
- 2015.T6 - Les usines logicielles
- 2017.T4 – Usines logicielles

- **Méthodes logicielles et langages**

=> génie logiciel

- 2013.T5 - Méthodes agiles
- 2015.T3 – Données massives scientifiques, recherche par les données
- 2015.T7 - Javascript (paradigmes et éco-systèmes)
- 2017.T6 – Méthodes et techniques pour le développement logiciel
- 2017.T7 – Science des données et apprentissage automatique

- **Modélisation et ingénierie**

=> Programmer des systèmes complexes. Du modèle conceptuel au modèle du logiciel

- 2013.T1 - L'ingénierie des besoins logiciel
- 2015.T2 - Modélisation et ingénierie - (IDM et génération de code)
- 2017.T2 – Ingénierie et web des données

- **Calcul, simulation et logiciels scientifiques**

=> convergence calcul-données, HPC-CLOUD

- 2013.T7 - Développer pour calculer
- 2015.T8 - Logiciels scientifiques et simulation : nouveaux modèles et enjeux
- 2017.T8 – Parallélisme itinérant, virtualisation et reproductivité

- **Science des données**

=> Recherche par les données, ingénierie des connaissances, IA, ...

- 2013.T8 – Le développement web : services web et web sémantique
- 2015.T3 – Données massives scientifiques (big data), recherche par les données
- 2017.T2 – Ingénierie et web des données
- 2017.T7 – Science des données et apprentissage automatique

- **Infrastructures de recherche**

=> infrastructures numériques/de données/du big data

Le déluges de données ne pouvant être traités automatiquement, il est nécessaire de mettre en place des infrastructures numériques. Ces infrastructures permettent de fédérer les développements, de certifier les chaînes de traitement et d'assurer la reproductibilité des expérimentations numériques et leur valorisation.

- 2015.T5 - Infrastructures et interopérabilité: CLOUD et SOA
- 2015.T3 - Données massives scientifiques (Big data), recherche par les données
- 2017.T5 – Infrastructures logicielles et science ouverte

- **Systèmes Embarqués, objets connectés, réseaux de capteurs intelligents**

=> instrumentation, dispositifs expérimentaux ad hoc, collecte de données...

- **2013.T3** - Systèmes embarqués, systèmes mobiles, capteurs et acquisition de données (<http://devlog.cnrs.fr/jdev2013/t3>)
- 2015.T1 - Systèmes embarqués, réseaux de capteurs et objets communicants
- 2017.T1 - Embarqué, réseaux de capteurs et internet des objets (IoT - M2M)

- **Transferts et impacts du numérique**

=> L'apport sociétal et la transformation digitale

- 2013.T2 - Les communautés open-source - comprendre, faire, savoir faire
- 2015.T4 - Transfert marchand et non marchand
- 2017.T3 – Programmation de la matière, fabriques personnelles

Le buzz de l'IA et des données

- Révolution de l'IA et révolution des données ?
- Mythe ou réalité ?
- Conférence de Dartmouth College 1956 : pronostic d'une machine aussi intelligente qu'un être humain en quelques années, traduction automatique, ...
- **IA faible** : mathématique et techniques d'ingénierie , simulation d'intelligence par algorithmes de résolution de problèmes
- **IA forte** : une machine "intelligente" capable d'éprouver une conscience et/ou des sentiments.
- Les comportements intelligents dépendent du traitement de la connaissance
- La difficulté des KMS à prévoir des situations non codées (tsunami, ...)
- Connaissance explicite et connaissance implicite
- Connaissance embarquée dans un SI ou dans la tête des experts...
- Espérances et désillusions... une avancée en dent de scie...
- => **Puissance de calcul et explosion des données**

Coder : c'est quoi ?

- Quelles définitions :
 - « Transcrire un message en code afin de le transmettre ou de le traiter. »
 - « Procéder au codage d'un message ; encoder. » => nouvelle représentation
 - « Appliquer un code à des données ; écrire un programme pour ordinateur. »
- Décrire/expliciter un comportement, un algorithme
- Quel type de développement ?
 - Technique :
 - Développement système / matériel : SE, contrôleurs
 - Développement réseau
 - Calcul : optimisation machine-architecture (ex :calcul vectoriel avec processeur graphique, calcul matricielle)
 - Infrastructure big data / cloud
 - SI
 - ...
 - « Métier » / domaine :
 - Modélisation
 - Simulation
 - Système expert
 - Expertise
 - Système décisionnel
 - ...
- **Coder, c'est expliciter, représenter, modéliser**
- **2 approches pour programmer sans coder :**
 - À partir de modèles
 - A partir de données

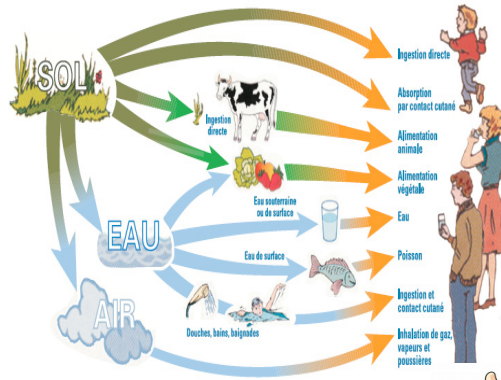


Coder comment ? Avec l'IDM

- **Ingénierie Dirigée par les Modèles**
- Séparation aspects techniques et aspects métier
- Donner la main aux experts métiers (scientifiques, ...) :
Modélisation/simulation d'un domaine => prédictions, validation d'hypothèses
- IDM et modèles scientifiques <http://devlog.cnrs.fr/idm2013>
- De la modélisation scientifique à la modélisation de votre code
<http://devlog.cnrs.fr/idm2014>
- Créez votre langage spécifique de domaine (DSL), votre langage métier <http://devlog.cnrs.fr/idm2015>
=> Programmation projectionnelle ou programmation par langage (MPS – Meta Programming System)

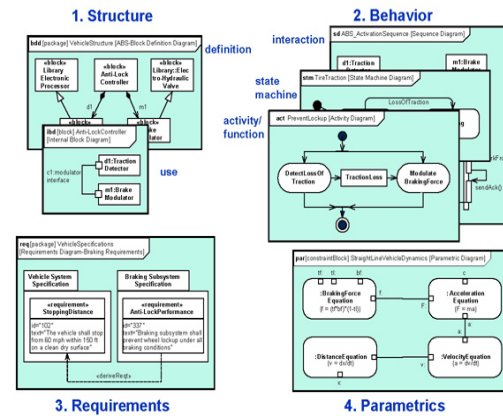
Coder comment ? Avec l'IDM

Modèle(s) conceptuel(s)
d'un système complexe



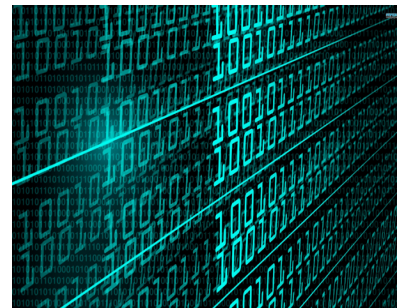
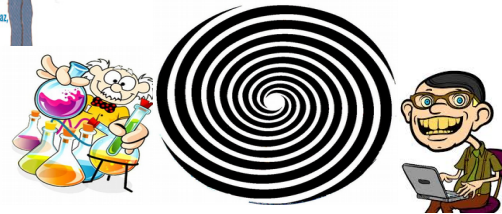
Implémenter

Modèles du logiciel (UML, SySML, GORE, ...)



Note that the Package and Use Case diagrams are not shown in this example, but are respectively part of the structure and behavior pillars

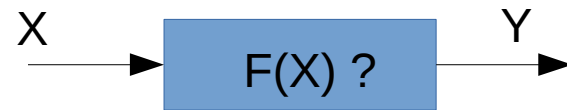
Explorer
et
adapter



Simulation(s)

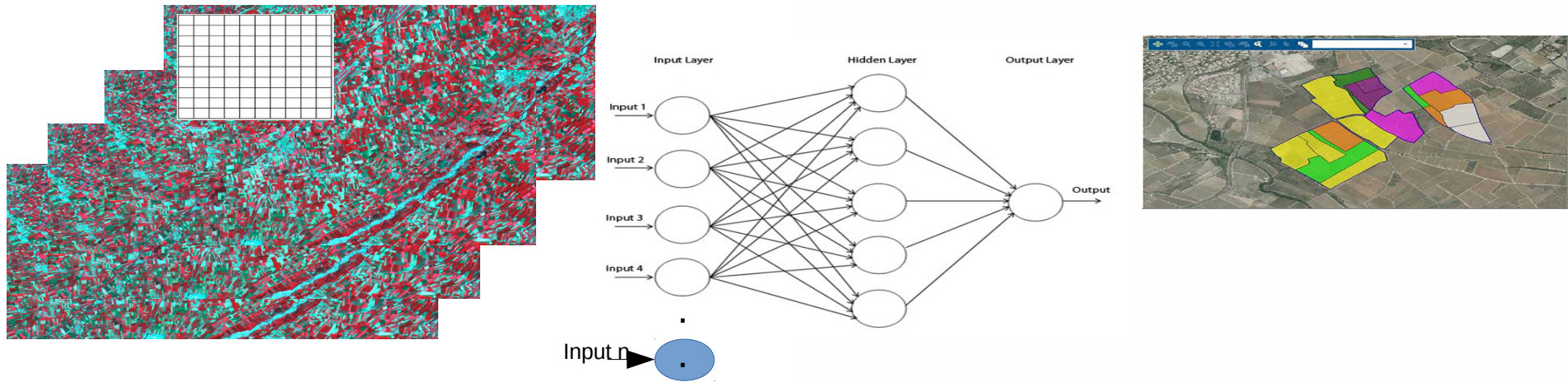
Générer

Coder comment ? à partir des données

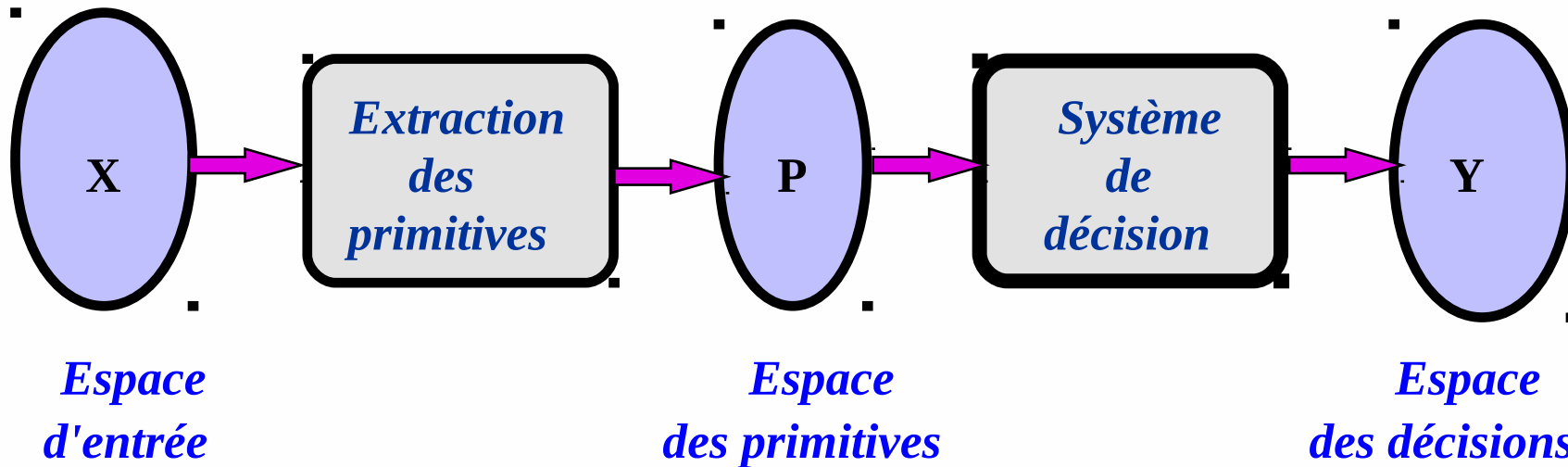


- De l'algorithme aux statistiques descriptives
- Recherche par les données pour extraire un modèle/une fonction
- Comprendre et prédire un phénomène observé à partir d'un modèle
- Une observation représentée par le tuple $\langle X, Y \rangle$:
 - X , un vecteur de variables/attributs/paramètres prédictifs
 - Y , une variable/un vecteur cible dont on souhaite prédire la valeur pour des événements non observés
- On recherche $F(X)$, tel que $Y = F(X) + E(X)$
 - **$F(X)$, la fonction de prédiction modélisant notre phénomène**
 - $E(X)$, l'erreur de prédiction (bruit, approximation)
- Données d'apprentissage → algo d'apprentissage automatique → $F(X)$
ens. d'observations $\{\langle X, Y \rangle\}$
- $F(X)$ est recherchée dans un espace de fonction/modèle

Coder comment ? à partir des données



Utilisation d'un perceptron pour apprendre une fonction de classification du sol à partir d'un ensemble de vérité terrain $\{<X, Y>\}$



Du perception au l'apprentissage profond : reconnaissance automatique de formes dans un grand vecteur d'entrée X
 => nécessite puissance de calcul et volume de données

Quel besoin d'infrastructure ?

- Infrastructure numérique = réseau + cycle d'horloge pour le calcul ?
- Evolution du GENCI vers une offre de service
- **Convergence calcul-données**
- **Convergence HPC-Cloud**
- Architectures ad hoc pour le Big data
- Software Defined Network (SDN)
- Réseaux de données

=> Infrastructure numérique =

Infrastructures logicielles « métier »
Services de ressources (calcul, données)
Ressources informatiques : serveurs de calcul, de stockage

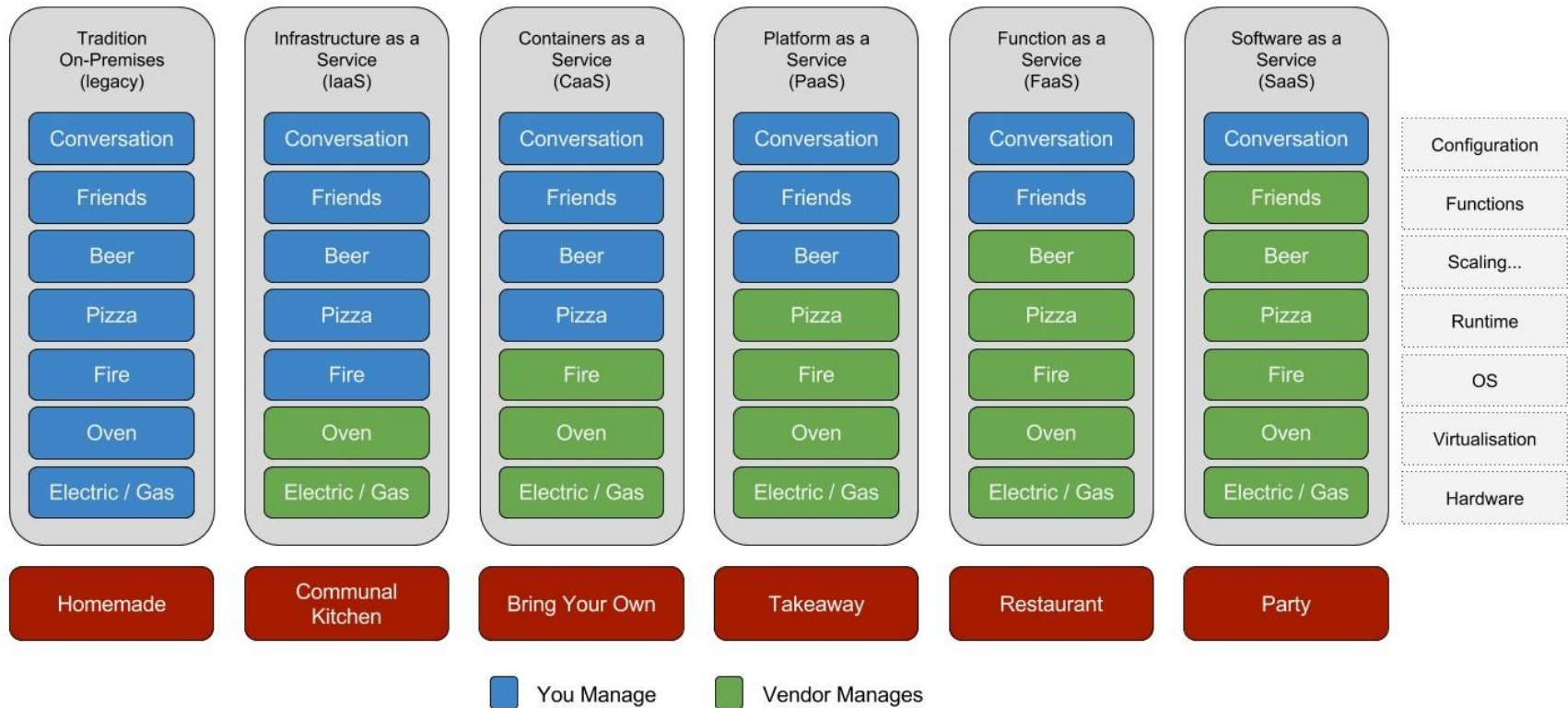
Quelle approche Devops ?



Pizza as a Service 2.0

Source : Paul Kerrison

<http://www.paulkerrison.co.uk>



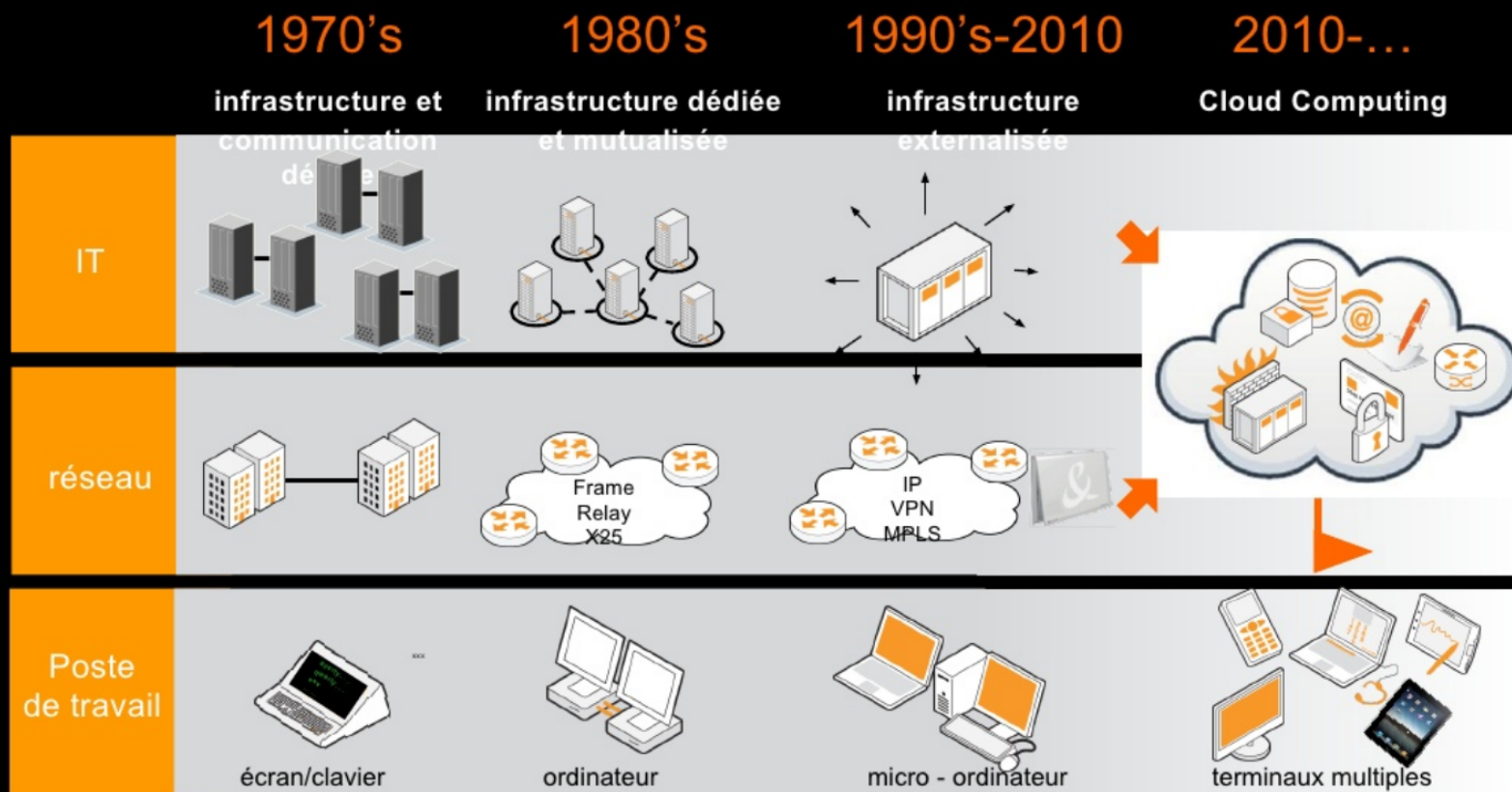
- Cloud et virtualisation
- Environnement de développement en ligne
- Infrastructure as Code (cf. demande de ressource)
- Quelles Offres De Service ? Clouds privés de la recherche ou clouds public ?
- Libérez les développeurs !
- Cf. <http://devlog.cnrs.fr/jdev2015/t5>

Décentralisé versus centralisé



évolution de l'IT et du réseau

Clip slide



Du dev collaboratif au dev coopératif

- Gestionnaire de version : centralisé (SVN) puis décentralisé (Git)
- Développement de standards ouverts d'architecture SOA pour garantir l'interopérabilité

Ex : OGC (Open Geospatial Consortium)

<http://www.opengeospatial.org/>

cf. <http://devlog.cnrs.fr/ids2014>

La donnée comme infrastructure

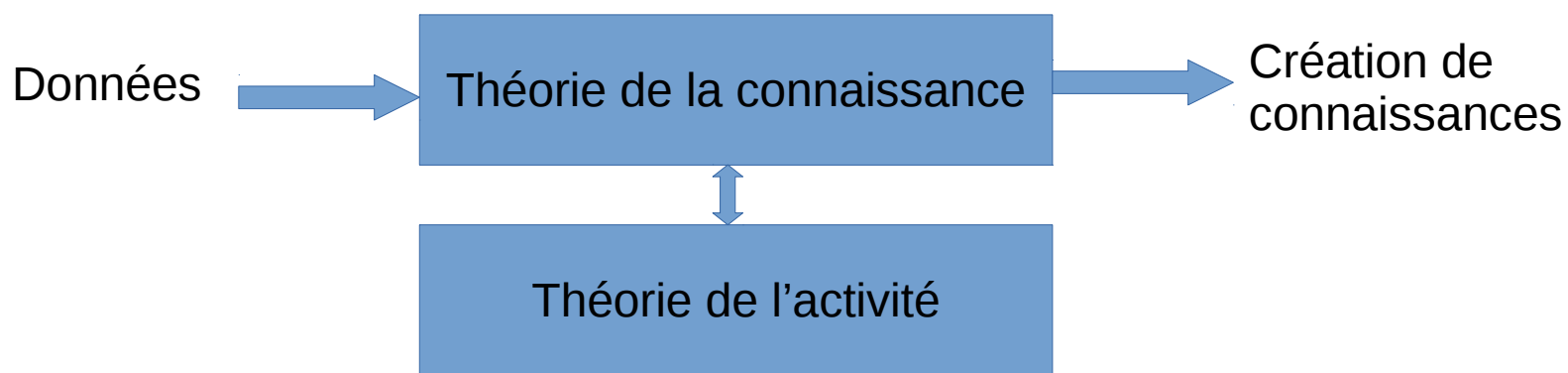
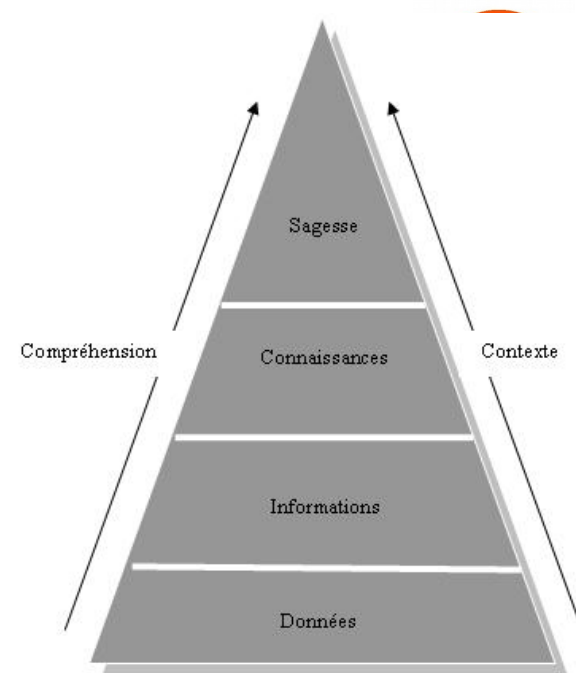
- Des communautés métier s'organisent autour de standards SOA :
 - OGC pour la géographie
 - HL7 pour les hôpitaux
- Objets métiers bien définis. Quels sont les objets métier de la recherche ?
- Prétentions des années 90 d'informatiser le monde par des schémas XML...
- Recherche par les données
- Peu propice à la recherche !
- SOA : des services (playmobil) aux micro-services (lego)
- Du XML au web des données : de l'arbre au graphe
- La connaissance est universelle ou contextuelle/culturelle ?
- => création des représentations de connaissance contextuelle/ points de vue
- => nécessité de réifier les représentations de connaissance
- cf. rapport « La donnée comme infrastructure essentielle » de l'ADG 2016-2017

Infrastructures de recherche

- Enjeu d'aujourd'hui : les communs de la donnée
- Plateforme versus infrastructure !
- Plateforme (mode silo) = backend + frontend (ENT)
- SI= backend (infra) + frontend (ENT)
- Méta-infrastructure pour ouvrir les plateformes
- => méta-infrastructure pour fédérer les différentes couches de services des plateformes
- => des ENTC plus en plus génériques
- **Les communs de la donnée**
- 'The Internet of (FAIR) Data & Services' (IFDS) :
Findable, Accessible, Interoperable, Re-usable
 - <https://www.go-fair.org/go-fair-initiative/>
 - <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>
- Permettre l'utilisation, la circulation et l'interopérabilité des données
- La gestion des lots de la recherche et de leurs données

Vers une transition digitale : comment et pourquoi utiliser les données les contenus dans l'activité de recherche

- Une autre approche que l'IA: les pratiques de la connaissance
- Connaissance : explicite 10 %, implicite 90 % ...
- La connaissance est embarquée dans la tête des experts, les artefacts et les pratiques (KMS)



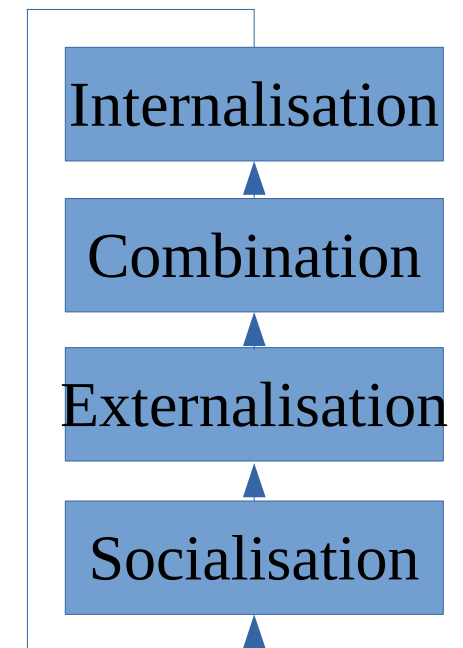
Pratiques de la connaissance =>

- Quels artefacts
- quelles solutions

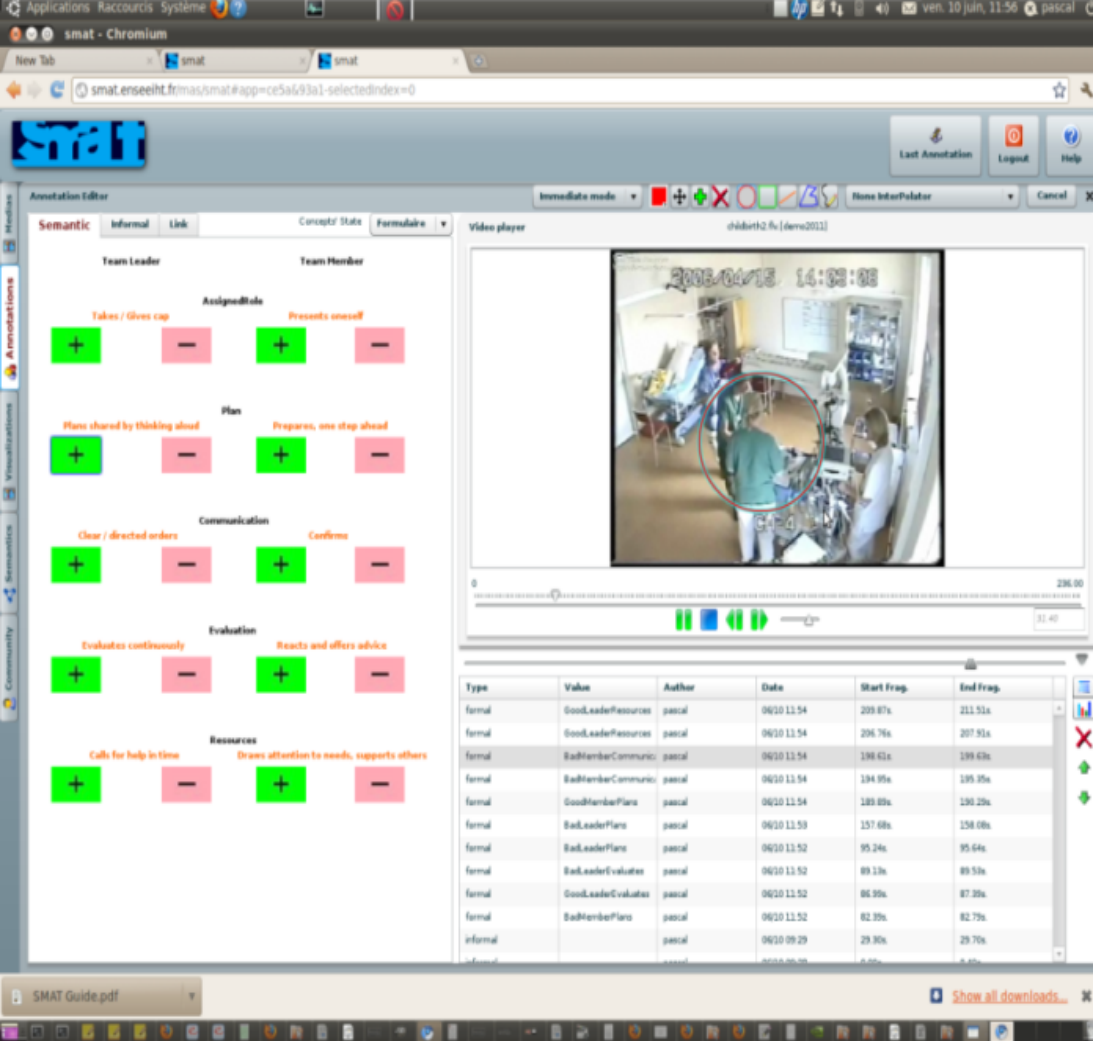
Document / donnée / méta-donnée

- Document / contenu / média (texte, image, vidéo, son)
- Quelles utilisations ? Capitalisation des savoir-faire, traçabilité, science reproductible, ...
- Approche constructiviste et collaborative de la production de connaissance
- Annotation : association d'une valeur (sémantique) à un fragment de document
 - Annotations automatiques : cf. TAL (SIIM2017) pour indexer les comptes rendus d'examens
 - Annotations manuelles par les experts : détournage d'une zone cancéreuse pour la radio-thérapie (SMAT)
- Indexation/référencement => Recherche d'Information
- Annotation = création de méta-données
- Annotation => création de contextes de connaissances (Ontologie => graphe de contexte)
- Plusieurs experts => plusieurs graphes de connaissance et confrontation d'expertise

Nonaka :
organisational
Knowledge creation



Apprendre par simulation réflexive



Trophée francophone du numérique pour l'éducation dans la Catégorie « Développement pédagogique innovant ». Nov 2016. Serious Game « Prévention des risques liés à la communication au bloc opératoire »

MODÉLISER ET MONITORER L'ACTIVITÉ HUMAINE ?

Objectif : représentation d'une situation professionnelle ciblée



KPLAB / SMAT - Collaboration hôpital Stocklrom
Réduire de 50 % la mortalité au bloc opératoire :
Co-construction d'une conceptualisation commune de l'activité collaborative

* Piombo C., Batatia H., Dayre P., Ayache A., (2007). An ontology based Web Annotation System to create new learning practices. Proceedings of EC-TEL'2007

* Dayre P. (2012). Un outil d'aide à la transformation du travail. SELF2012

JDEV2019 ?

<http://devlog.cnrs.fr/jdev2019>

- T1 - Systèmes autonomes et capteurs intelligents
- T2 - Apprentissage et Web sémantique
- T3 - La programmation des objets dématérialisés
- T4 - Usine logicielle et reproductibilité
- T5 - Infrastructure de recherche
- T6 - Génie logiciel: Devops
- T7 - Langages, Data sciences et ENTC
- T8 - HPC , IA (symbolique) et machine learning (quantitative)
- T9 - Ingénierie cognitive
- T10 - Systèmes complexes et logiciel

Conclusion

Quelles évolutions des métiers vers un monde de plus en plus intégré?

- Faire évoluer l'infrastructure numérique
- Besoin d'une bonne offre d'infrastructure (CLOUDs + infra recherche)
- Besoin de bonnes bases : DEVOPS et le besoin de virtualisation : Infrastructure as code = libérez les développeurs !
- Développer les communs de la données
- **Développer d'applications, d'ENTC spécifiques pour la transdisciplinarité : médiatiser et capitaliser les données et les pratiques**
- Développer des Web API
- Les sciences des données
- Le cadre législatif...RGPD, DPO, ...
- **Opportunité de la RGPD pour la recherche**
- Valorisation et solutions opendata. Artefact numérique ?
- Solutions pour les sciences participatives (ouverture sociétale)
- Faire l'évolution du logiciel / des SI : monolithique vers une urbanisation SOA
- **Développer des standards d'interopérabilité : passer des plate-formes au infrastructures !**
- Développer des langages spécifiques
- Contribuer à la modélisation des connaissances (les ontologies, le RDFS après les modèles entités-associations)
- Faire l'évolution des méthodes logicielles : dev isolé / dev collaboratif / dev coopératif ; méthode orienté utilisateur
- Interroger nos métiers et le besoins des laboratoires : c'est quoi le logiciel dans nos laboratoires
- Partager et utiliser les Offres De Service, les Très Grandes Infrastructures de Recherche (ex : huma-num)
- Améliorer la maturité des logiciels : certifié/validé/qualifié ; tâche mono-utilisateur / multi-utilisateur
- Maîtriser l'IoT de l'acquisition à sa gestion (CLOUD)
- Maîtriser les objets dématérialisés
- Contribuer à la transition digitale des laboratoires

Merci !

- **Participez à la prochaine ANF APSEM2018, Apprentissage et SEMantique :**

l'apport croisé et les nouvelles perspectives des technologies du web des données et de la recherche par les données.

Nous nous intéresserons donc:

- à la structuration de l'information à partir du web sémantique pour améliorer l'apprentissage et l'analyse des données.
 - à l'apprentissage pour aider à la structuration de l'information.
 - aux nouvelles approches basées analyse des graphes.
 - à des cas d'étude.
- Plus d'information sur <http://devlog.cnrs.fr/apsem2018>
 - Contact : pascal.dayre@enseeiht.fr