

# Plateforme de Connaissance Unifiée (PCU)



PCU

## Présentation et avancement

04/10/2017, Plénière GTLL, Jussieu



proxem



WALLIX  
TRACE, AUDIT & TRUST



bpi**france**  **île de France**



PCU

# Plan

---

- **PCU en résumé**
- **Livrables**
- **Avancement**
- **Questions**

## L'intervenant

- **Marc Dutoo, Responsable projets R&D chez Smile, le premier fournisseur de service Open Source en Europe**
- **Coordinateur du projet PCU, expert Data / API / Cloud**



# PCU en résumé

---

- 6 partenaires, 3 ans (2017-2019), System@tic GTLL / BPI / Région en sponsors
- Afin de démocratiser la Big Data, pour que n'importe quelle entreprise puisse tirer de ses données existantes de la valeur ajoutée pour son coeur de métier :
  - Un module intégré Big Data / Machine Learning pour enrichir toute applications métier
- Démontré dans 2 cas d'usage :
  - E-commerce (jusqu'au magasin physique) & B2B
  - Recherche d'entreprise
- Grâce à:
  - Un processus de traitement de données orienté événement de bout en bout
  - Une fabrique de moteurs de recherche enrichis par Machine Learning et sémantique
  - Des algorithmes d'analyse du comportement utilisateur encore jamais ouverts
  - Une approche open source, "best-of-breed", unifiée, flexible et extensible



PCU

## Partenaires et sponsors



**Smile** : coordinateur, architecture, ecommerce



**Paris 13** : Machine Learning, sémantique



**ESILV** : processus de traitement, sémantique



**Proxem** : fouille de texte et opinions, B2B



**Wallix** : expertise recherche d'entreprise



**Armadillo** : API et interface utilisateur d'intégration et administration

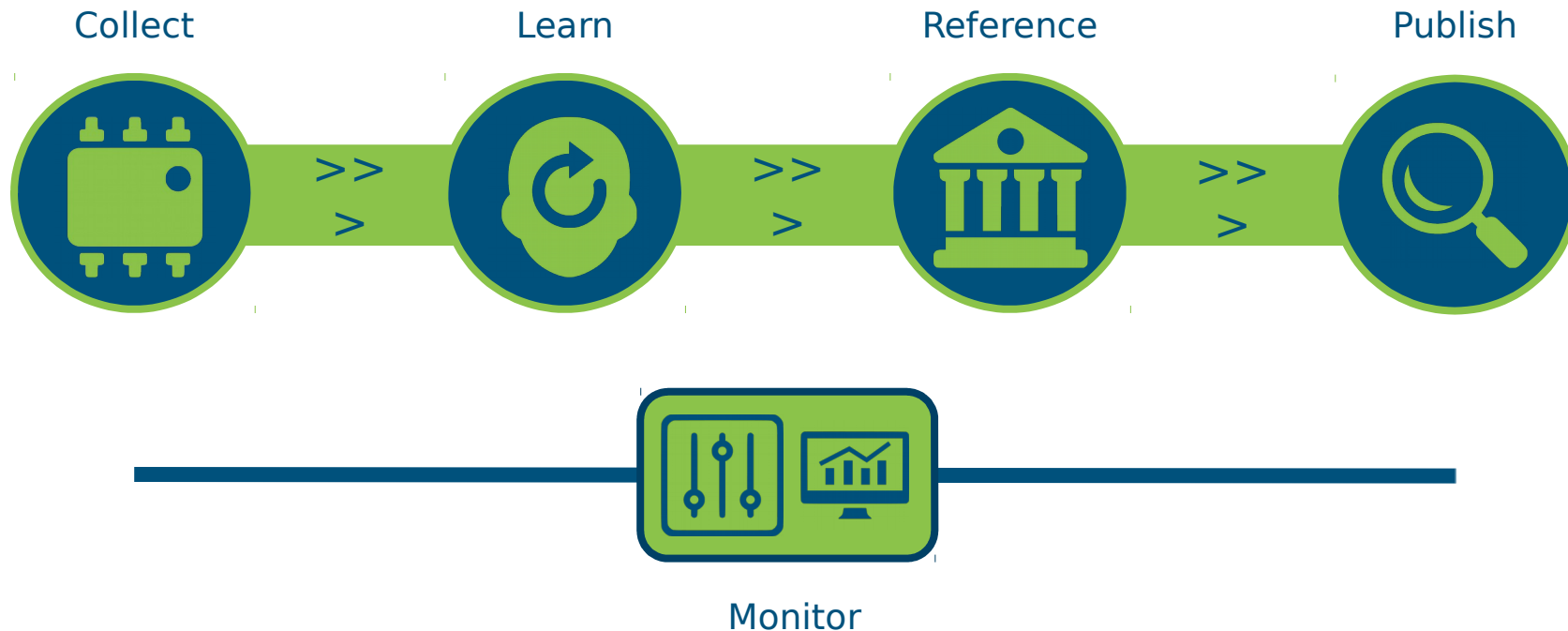
**Sponsors financiers** : BPI, IdF

**Pôle** : **System@tic** (GTLL)





# Architecture d'ensemble





# Livrables

- **Une plateforme générique**



- Unifiée, flexible, au-dessus des meilleurs briques, gérée par API
- Avec un jeu standard de connecteurs, éléments de traitement de données et algorithmes Machine Learning (ML) et fouille de texte

- **Cas d'usage et produits**

- E-commerce (produit, déployé chez les clients innovants de Smile), B2B (déployé chez Smile et Proxem)
- Recherche d'entreprise (produit, déployé chez chaque partenaire)

- **Ecosystème Open Source**

- Liens avec les communautés des composants technologiques intégrés ainsi que des produits dérivés pour chaque métier
- Lieu de la plateforme, ses exemples, de l'essai et adoption



# Avancement – septembre 2017 (T0+11)

## ● **Besoins métier, jusqu'aux prototypes Machine Learning**

- Recherche, Ecommerce (B2C), CRM (B2B), avec 10GB+ de données
- Des prototypes sur Spark + Jupyter : recommandation de produits par collaborative filtering, co-occurrences de mots clés recherchés...

## ● **Architecture and development**

- Etat de l'art, besoins techniques, PoCs (ElasticSearch, Solr, Spark)
- Transformation de données en Spark configurée en YAML (prototype)
- En cours : recherche d'entreprise (API, indexation, connecteur "crawl")

## ● **Organisation du projet**

- Collaboration : documents partagés, audioconférences hebdomadaires, réunions trimestrielles, accord de consortium...
- Github, premier cloud de données et briques Big Data / ML
- Communication : twitter, site web, RISE 2017 (Caen)

<https://pcu-consortium.github.io/>  
<https://twitter.com/PCUConsortium>  
Contact : marc.dutoo@smile.fr



Questions ?

Merci de  
votre  
attention !



proxem



armadillo

WALLIX  
TRACE, AUDIT & TRUST



bpi**france**

★ île de **France**