



TelecomValley  
Internet – Mobilité – Usages

3<sup>ème</sup> édition

# SophiaConf 2012

Open Data Hadoop Shinken Grid  
NoSQL Informatique Big Search Open Source  
Administration Communautés Cloud Gestion  
Open Geo Data Supervision Applications

2-4 juillet  
3 conférences

Avec le soutien de



En partenariat avec



## SophiaConf2012 : « Big » et « Open » data mobilisent les foules

La troisième édition de SophiaConf a permis de plonger au coeur de trois thématiques actuelles « Big data », « Open data » et « Open Source : Admin & Supervision ». Des interventions de haute qualité assurées par des experts nationaux et de sociétés locales, un événement à forte valeur ajoutée qui a réussi à captiver près de 700 professionnels Sophilopolitains.

Une troisième édition qui s'achève et toujours autant de succès pour SophiaConf. Témoins de ce succès avéré, les **670 personnes** qui se sont déplacées aux 4 conférences successives et surtout, leur assiduité jusqu'à la dernière minute. « *Certaines personnes sont même venues assister 3 jours d'affilée aux 4 conférences différentes !* » s'exclame Pascal Flamand, animateur de la commission Open Source et organisateur de SophiaConf. La qualité des intervenants et des sujets traités n'y ont certainement pas été étrangers.

### A la une cette année : les data, « Big » et « Open »

Surfant sur les tendances actuelles, SophiaConf s'est intéressée au phénomène des data et a choisi de traiter deux aspects différents. La conférence du 3 juillet intitulée « **Big data, NoSQL, Big search** » avait pour mission d'apporter un éclairage sur **les enjeux et outils de gestion et de recherche de data massives**, dans un environnement Internet de plus en plus saturé en données. Quelques unes des grandes entités et start-ups de la technopôle (SAP Labs France, Orange, France Labs, Gridpocket et Eurecom) ont partagé leur meilleurs pratiques et expliqué les solutions retenues pour répondre à cette nouvelle problématique.

Le 4 juillet, deux conférences complémentaires orientées « **producteurs de données ouvertes** », puis « **transformateurs** » ont abordé le sujet brûlant de l'**Open data**. Les objectifs de ces deux conférences : dans un premier temps, **poser le contexte** avec l'intervention de Charles Népote (FING) sur les origines et enjeux de l'Open data, ainsi que **les aspects législatifs et juridiques** (Jean-François Carrasco, Conseil de développement de la CASA, puis Pascal Agosti, Cabinet Caprioli & Associés). Ont suivi, la présentation de démarches Open data d'institutionnels (Métropole Nice Côte d'Azur, Unité de Santé publique de l'Université de Nice) et notamment l'annonce du **lancement du portail Open data de la Région PACA le 10 juillet**, par Stéphane Martayan, qui comprend la libération de 100 jeux de données. D'autres intervenants (Data publica, Janua) **ont soulevé la question de la possibilité ou non de créer des business models dans le domaine de l'Open data, quelles retombées et surtout quelle réutilisation ?**

Au cours du second volet « transformateurs de données ouvertes », plus technique et destiné plus spécifiquement aux développeurs, les intervenants ont livré **quelques exemples de réutilisation de données ouvertes**, par exemple à des fins de services communautaires à l'image d'OpenStreetMap France représentée par son Président Gaël Musquet. Trois chercheurs d'Inria (Fabien Gandon, Julien Cojan et Patrick Valduriez) ont présenté diverses méthodes et outils facilitant le formatage et l'intégration de données.

### Focus technique sur les outils d'administration et de supervision dans le cloud

La conférence d'ouverture, le 2 juillet, était dédiée à la présentation **d'outils d'aide à l'administration et à la gestion de systèmes informatiques dans le cloud**. Un invité de marque, **Jean Gabès, créateur de l'outil Shinken et principal animateur la communauté Open Spource éponyme**, s'est déplacé depuis Bordeaux pour venir échanger sur le fork de Nagios. Des PME (Janua), de grands groupes (Thalès Alenia Spce, Orange) et centre de recherche (Inria) ont complété cette soirée en partageant leurs expériences d'utilisation des divers outils dans leurs systèmes internes.

**2-4 juillet**  
**3 conférences**

- ⇒ **4 conférences techniques, sociétales et méthodologiques**
- ⇒ **4 événements d'animation du territoire**
- ⇒ **3 sponsors publics**
- ⇒ **14 intervenants venus de Paris, Bordeaux, Marseille et du département**
- ⇒ **670 participants aux 4 conférences**
- ⇒ **1 site internet entièrement consacré à l'événement**
- ⇒ **26 articles de presse**

# Open Source : Admin & Supervision



**J. GABES**

### « Shinken, la supervision à l'ère du cloud »

Shinken est un outil de supervision, ré-implémentation de Nagios en Python 'from scratch', qui a pour objectif principal de simplifier la vie des administrateurs de grands parcs et de coller au mieux à l'évolution de l'IT des dix dernières années. Il permet ainsi de répartir la charge de supervision de manière automatique (technique de « sharding »), et de faciliter la tâche des administrateurs dans la description des relations entre les éléments des systèmes d'informations (relations réseaux ou dépendances applicatives). Il offre également des vues synthétiques sur l'état global du SI, en mettant en avant les services importants rendus par le SI, avec de fortes notions de corrélations et d'importance «business ».

### « Administration système, enjeux et problématiques, un métier en pleine mutation »

Les nouveaux usages de l'Internet et du mobile ont provoqué ces dernières années une augmentation exponentielle des ressources de traitement et de stockage des datacenters modifiant ainsi profondément le métier des administrateurs systèmes. Quels sont les enjeux et problématiques de ces nouveaux paradigmes ?

Quelles nouvelles méthodologies mettre en œuvre pour répondre à ces défis ?

Quels sont les nouveaux outils permettant aux administrateurs systèmes de gérer le déploiement d'applicatifs, la gestion du cycle de vie, la surveillance ainsi que l'exploitation de ces nouveaux services ?



**F. AIME**  
(Janua)



**N. HENNION**  
(Thales Alenia  
Space)

### « Déploiement Nagios chez Thales Alenia Space »

Présentation illustrée du déploiement d'une solution de supervision libre (Nagios, Cacti, WeatherMap) sur le réseau interne de Thales Alenia Space.

### « Retour d'expérience Amadeus eTravel: organisation et méthodologie »

Au sein d'Amadeus, la branche eCommerce correspond à la vente de billets d'avions en ligne. Les systèmes Amadeus hébergent les sites de nombreuses compagnies aériennes telles que Lufthansa, Finnair, Qantas, .... Isabelle Fernandez, de l'équipe de support niveau 3, faisant le lien entre les équipes d'exploitation du site de production et les équipes de développement, a présenté les aspects méthodologiques et organisationnels mis en place pour atteindre un taux de disponibilité de 99,9%.

**I. FERNANDEZ**  
(Amadeus)



**B. BRANDL**  
(Orange Portails)



**F. CLEMENT**  
(Orange Portails)

### « CFengine, Puppet, automatisation du provisioning »

Hebex, la direction exploitation d'Orange Portails, exploite depuis plus de 15 ans ses nombreux serveurs en utilisant des FOSS pour accompagner la croissance des services web d'Orange. Si le choix de l'open source a d'abord été économique, c'est vite devenu un principe pour faciliter aussi bien l'intégration des services que leur exploitation et leur montée en charge.

Ainsi, tous nos outils d'exploitation sont des foss ou développés / adaptés en utilisant des technologie open source. Cette présentation avait pour but d'exposer les principaux outils open source que nous utilisons et que nous avons adaptés à notre besoin : cfengine, amanda/bacula, Xen, Big Brother/Xymon, cluster nfs, openldap, sécurité, réseau...

### « Grid'5000 : Administration d'une infrastructure distribuée et développement d'outils de déploiement et d'isolation réseau »

Grid'5000 est une plateforme de 1500 machines distribuées sur 10 sites et dédiée à la recherche en informatique et réseau. Elle peut être vue comme un cloud HaaS (Hardware As A Service), permettant à ses utilisateurs de déployer des OS sur des machines physiques en quelques minutes via des outils comme OAR, kadeploy3 et kavlan. Une fois ce déploiement effectué, les chercheurs peuvent alors lancer leurs expérimentations scientifiques sur la plateforme. Cet exposé a décrit ces logiciels ainsi que les outils utilisés par les administrateurs de la plateforme (git, puppet, capistrano), et a donné quelques exemples d'utilisation (déploiements de Nimbus, OpenNebula, OpenStack, gLite ...).



**N. NICLAUSSE**  
(Inria)

### Lieu :

Polytech'Nice Sophia  
Site des Lucioles  
Sophia Antipolis



**2-4 juillet**  
**3 conférences**

### Big data, NoSQL, Big search



**D. TRASTOUR**  
(SAP Labs  
France)

#### « Big enterprise data : opportunités et enjeux »

L'approche Big Data a son origine principalement dans l'analyse de web logs ou de réseaux sociaux. Cependant son domaine d'application s'étend de plus en plus, et le Big Data est en train de conquérir le monde de l'entreprise et de l'informatique décisionnelle. David Trastour de SAP Research a décrypté certains des enjeux autour du Big Data, et les challenges tant au niveau de la gestion des données qu'au niveau de l'exploitation et consommation de ces données.



**O. VARENE**  
(Orange Portails)

#### « Calcul distribué avec Hadoop MapReduce au sein du moteur de recherches Orange »

Dans le cadre de la mise en œuvre du moteur de recherches d'Orange, Olivier VARENE nous a fait partager le contexte et les motivations du projet initial, une overview des technologies mises en œuvre autour de Hadoop, le résultat des investigations menées ainsi qu'une estimation du retour sur investissement.



**A. MAZOYER**  
(France Labs)



**O. TAVARD**  
(France Labs)



**C. ULMER**  
(France Labs)

#### « Big Search »

Avec l'arrivée du big data se pose le problème de la recherche d'informations pertinentes. Aurélien MAZOYER, Olivier TAVARD et Cédric ULMER de France Labs ont présenté comment les technologies open source proposent de répondre à ce défi. Leur intervention s'est notamment intéressée au moteur de recherche Lucene/Solr (déjà adopté par IBM, Twitter ou encore Apple), avant d'expliquer comment interfacier ces technologies avec Hadoop, afin de pouvoir effectuer efficacement des recherches de données sur le big data.

#### « Big data et smart grids - projet BigFoot »

BigFoot est un projet de recherche portant sur une nouvelle infrastructure logicielle de stockage et traitement de grosse quantités de données. La plate-forme BigFoot sera construite avec et contribuera au développement d'outils open source, notamment Hadoop. Elle a pour l'objectif de permettre à la fois le traitement de données en mode "batch" et en temps réel, de simplifier le déploiement de la plate-forme à l'aide de machines virtuelles et de fournir un ensemble d'outils optimisés. La plate-forme BigFoot permettra l'exécution de nombreux applications liées au "data mining" dans les domaines de la sécurité informatique et du smart grids: par exemple l'observation de l'état de santé du réseau et ses composants, la segmentation et optimisation de l'offre énergétique, et pour finir une aide contextualisée aux consommateurs.



**F. GLUSZAK**  
(Gridpocket)



**P. MICHIARDI**  
(Eurecom)



Les professionnels sophilopolitains venus en nombre.



Cédric ULMER (France Labs) a introduit le sujet « Big search ».

#### Lieu :

Inria  
2004, Route des Lucioles - Sophia Antipolis

# Open data : des paroles aux actes

## « Producteurs de données ouvertes »



C. NEPOTE  
(FING)

### « Open Data, genèse et bilan de deux années d'expérimentation en France et dans le monde »

Qu'est-ce que l'open data ? D'où vient ce mouvement et pourquoi maintenant ? Quels en sont les principaux enjeux et qu'est-ce cela produit ? Cette intervention a présenté les principes de base de l'open data ainsi qu'une synthèse des deux premières années d'expérimentations en France.

### « Le nouveau marché des données »

L'ouverture des données bouscule l'industrie de l'information et contribue à faire émerger un nouveau marché, celui de la donnée en tant que telle, indépendamment de son usage. Ce marché émergent voit apparaître de nouveaux acteurs et se repositionner les anciens.



F. BANCILHON  
(Data Publica)



S. MARTAYAN  
(Région PACA)

### "Présentation du programme d'ouverture des données publiques de la Région PACA »

La Région Provence-Alpes-Côte d'Azur s'est engagée lors de l'[Assemblée plénière](#) du 16 décembre 2011 dans un projet ambitieux de libération de ses propres données publiques, de mise en place d'un portail et d'une animation régionale avec l'ensemble des acteurs de son territoire (Hackathons, concours d'applications innovantes, etc.). Le programme régional s'appuie également sur les données publiées par les partenaires de la Région. Ce chantier spécifique avec les partenaires de la Région est actuellement décliné dans le cadre de Marseille-Provence 2013, la Capitale européenne de la Culture avec comme priorité les thématiques Transports, Culture, Tourisme. Il sera courant juillet ouvert à l'ensemble des collectivités de la région.

### « Cartographie Libre et Collaborative »

La cartographie a longtemps été réservée aux élites politiques, militaires ou scientifiques. Internet a bouleversé le rapport aux données cartographiques notamment grâce aux solutions de Google. OpenStreetMap a franchi une autre barrière, celle de la production des données par les citoyens. Comment cette communauté s'est-elle imposée en moins de 10 ans ? Comment s'organise l'Open Data citoyen ?



G. MUSQUET  
(OpenStreetMap)



JF. CARRASCO  
(CD CASA)

### « CADA 1978, LOADDT 1999, deux lois d'ouvertures se parlent sur le terrain des Conseils de Développement »

Si la mise à disposition des Données Administratives n'est pas si jeune, l'arrivée dans les faits et les usages a dû attendre des influences étrangères, souvent anglo-saxonnes pour ouvrir localement les yeux et les pratiques. Comment les lieux de Démocratie Participative que sont les Conseils de Développement au sein des Communautés d'Agglomérations prennent ce train et pour quoi en faire ?

### « Ville intelligente et Open Data : l'habitant au cœur de la démarche »



G. RABIN  
(NCA)



P. STACCINI  
(UNS)

### « Data Publica Sanitaria : de l'Open Data à l'Open Report en santé publique au service du chercheur, du professionnel, du citoyen et du décideur »

Les données de santé sont-elles éligibles au monde de l'Open data ? La question a été posée. La réponse est OUI. Le site du ministère du travail, de l'emploi et de la santé met à disposition dans un format réutilisable un ensemble de bases de données géolocalisées qui décrivent principalement la structure et le fonctionnement du secteur hospitalier. Le site [data.gouv.fr](#) propose une suggestion de recherche intitulée : Financement et dépenses de la sécurité sociale. Le site [data-publica.com](#) a identifié une entrée "Santé & nutrition". Eurostat, Ameli.fr, Afssaps, Insee, OCDE, INVS figurent parmi les principaux producteurs de données. Au-delà de ce recensement, d'autres questions surgissent : quelles données ? quels traitements ? pour quels usages ?

### « Eléments de réflexions sur l'Open Data »

Le déploiement de l'économie numérique va de pair avec la mise en exergue de l'exploitation de l'information. Celle-ci est par excellence financièrement valorisable et dès lors, de nombreuses sociétés privées développent des modèles commerciaux autour de l'accès, de la gestion et du recoupement d'informations. Celles-ci doivent être fiables (ou considérées comme telles). Cette fiabilité mais aussi la rareté des informations sont souvent reconnues aux informations détenues par les collectivités publiques. Pascal AGOSTI et Ilène CHOUKRI ont rappelé le cadre juridique applicable en matière d'Open Data (réutilisation des données publiques) et ont effectué un focus sur les règles applicables aux collectivités publiques tant en termes d'obligations et d'opportunités à saisir que de gestion des données à caractère personnel.



I. CHOUKRI  
(Caprioli & Associés)



P. FLAMAND  
(Janua)

### « Business models : enjeux, opportunités et risques »

Il semble que le marché de l'open data soit déjà une réalité, mais que les bénéfices ne sont pas forcément là où l'on pourrait les attendre. Pour comprendre la dynamique qui se met en œuvre, Pascal FLAMAND a commencé par cartographier les différents acteurs du marché; définir les grands enjeux, les opportunités et les risques et s'est essayé à modéliser une typologie des usages, du cycle de vie de la donnée et enfin d'identifier les différents flux financiers en présence. Cette démarche a permis de déduire différents business models, existants ou à inventer.

### Lieu :

Inria  
2004, Route des Lucioles - Sophia Antipolis

# Open data : des paroles aux actes

## « Transformateurs de données ouvertes »



**P. FLAMAND**  
(Janua)



**C. NEPOTE**  
(FING)

### « Business models : enjeux, opportunités et risques »

Il semble que le marché de l'open data soit déjà une réalité, mais que les bénéfices ne sont pas forcément là où l'on pourrait les attendre. Pour comprendre la dynamique qui se met en œuvre, Pascal FLAMAND a commencé par cartographier les différents acteurs du marché; définir les grands enjeux, les opportunités et les risques et s'est essayé à modéliser une typologie des usages, du cycle de vie de la donnée et enfin d'identifier les différents flux financiers en présence.

### « Cartographie Libre et Collaborative : utilisation des données »

La cartographie a longtemps été réservée aux élites politiques, militaires ou scientifiques. Internet a bouleversé le rapport aux données cartographiques notamment grâce aux solutions de Google. OpenStreetMap a franchi une autre barrière, celle de la production des données par les citoyens. Comment cette communauté s'est elle imposée en moins de 10 ans? Comment s'organise l'Open Data citoyen?



**G. MUSQUET**  
(OpenStreetMap)



**F. GANDON**  
(Inria)

### « DataLift : Un catalyseur pour le web de données »

Financé par l'ANR, le projet Datalift s'intéresse à faciliter le passage de données brutes structurées disponibles dans plusieurs formats (bases de données, CSV, XML) vers des données sémantiques interconnectées sur le Web de données. Le projet cherche notamment à proposer des outils assistant des étapes cruciales: sélection des vocabulaires, schémas ou ontologies pouvant décrire les données; conversion des données en RDF en suivant les vocabulaires sélectionnés; publication et contrôle d'accès aux données sur le web; interconnexion des données avec d'autres jeux de données. Nous cherchons non seulement à concevoir et assembler une plateforme pour publier et interconnecter des jeux de données mais aussi à valider les techniques et méthodes choisies en publiant des jeux de données provenant d'un réseau de partenaires. Voir aussi <http://datalift.org>

### « Technologie des données ouvertes »

L'ouverture des données met à disposition de tous un grand nombre de données peu ou mal structurées. Le challenge de la communauté scientifique et industrielle est d'utiliser ces données éparses et mal structurées en les intégrant. Nous décrirons certaines des approches pour intégrer et présenter ces données de façon utilisables.



**F. BANCILHON**  
(ISM)



**J. COJAN**  
(Inria)

### « DBpedia en français: extraction de données structurées depuis Wikipédia en français »

DBpedia est un projet communautaire dont le but est d'extraire des données structurées des pages Wikipedia et de les rendre disponibles sur le Web. La version originale parcourt Wikipedia à partir des pages anglophones, perdant ainsi les données propres aux pages des autres versions linguistiques. Le projet Semanticpedia a pour but de mettre place une version francophone de DBpedia en extrayant les données des pages francophones de Wikipedia.

### « Eléments de réflexions sur l'Open Data »

Le déploiement de l'économie numérique va de pair avec la mise en exergue de l'exploitation de l'information. Celle-ci est par excellence financièrement valorisable et dès lors, de nombreuses sociétés privées développent des modèles commerciaux autour de l'accès, de la gestion et du recoupement d'informations. Celles-ci doivent être fiables (ou considérées comme telles). Cette fiabilité mais aussi la rareté des informations sont souvent reconnues aux informations détenues par les collectivités publiques. Pascal AGOSTI et Ilène CHOUKRI ont rappelé le cadre juridique applicable en matière d'Open Data (réutilisation des données publiques) et ont effectué un focus sur les règles applicables aux collectivités publiques tant en termes d'obligations et d'opportunités à saisir que de gestion des données à caractère personnel.



**I. CHOUKRI**  
(Caprioli & Associés)



**P. VALDURIEZ**  
(Inria)

### « Intégration de données ouvertes avec WebSmatch »

Les initiatives de type open data produisent de très grands nombres de sources de données, dans divers domaines. Ces sources sont très précieuses mais très hétérogènes. Bien qu'elles soient relativement faciles à trouver avec un bon moteur de recherche, il est très difficile de les intégrer, selon des dimensions communes comme l'espace ou le temps, afin de produire de l'information de qualité, cruciale par ex. pour l'analyse de situation ou la prise de décision. Pour cela, il faut d'abord comprendre de quoi parlent les sources, ce qui nécessite de découvrir leurs méta-données. C'est facile lorsque les sources viennent avec leurs schémas, mais particulièrement difficile lorsque ces méta-données sont mélangées avec les données, comme par ex. dans des fichiers excel. Ensuite, il faut mettre en correspondance les schémas des différentes sources, par une tâche de schema matching qui détermine les éléments de données d'une source reliés avec ceux d'une autre. Le schema matching est difficile à cause de l'hétérogénéité structurelle (différences de formats) et sémantique entre données et méta-données. Pour répondre à ces défis, nous développons WebSmatch (un environnement ouvert et flexible d'outils d'intégration de données du Web. WebSmatch est accessible via des Web services, pour pouvoir être utilisé directement par nos partenaires.

### Lieu :

Inria  
2004, Route des Lucioles - Sophia Antipolis

### La commission Open Source de Telecom Valley

Créée en 2007, la commission est conduite par Pascal FLAMAND, fondateur de la société Janua, elle fait partie des 6 commissions thématiques de [Telecom Valley](#). Les enjeux de la commission sont multiples, tels que permettre aux acteurs du logiciel libre de se rencontrer et d'échanger, organiser des conférences pointues spécifiques à l'Open Source et aider les acteurs de Telecom Valley concernés par l'Open Source.



La commission est **membre fondateur du Cominé National du Logiciel Libre (CNLL)**. Elle est également à l'origine du Livre blanc "Intégration et publication de logiciels sous licences Open Source : Mode d'emploi". Enfin elle a activement participé à la mise en œuvre de la première édition de SophiaConf en 2010.

### Telecom Valley

**Association professionnelle à but non lucratif**, [Telecom Valley](#) regroupe les principaux acteurs IT de Sophia Antipolis et des environs.



Fondée en 1991, l'association a fêté en 2011 ses **20 ans de service pour l'innovation technologique et l'accélération de l'adoption et usages des TIC** en région PACA.

Telecom Valley regroupe aujourd'hui **130 membres**, constituant ainsi un grand réseau entre PME, grands groupes, instituts de recherche et formation.

[www.telecom-valley.fr](http://www.telecom-valley.fr)