



**TECHNIQUES
DE L'INGÉNIEUR**

Réf. : H5080 V1

Systèmes d'Information (SI) d'Entreprise et conduite des projets Big Data

Date de publication :
10 novembre 2017

Date de dernière validation :
10 mars 2025

Cet article est issu de : **Technologies de l'information | Documents numériques Gestion de contenu**

par **Pierre DELORT**

Mots-clés

big data | système
d'information | amélioration de
performance

Résumé Cet article débute par les différentes étapes du déploiement des SI en entreprise: automatisation puis reconception, autour du mot «organisation», des process et opérations et enfin intégration de progiciels. Il fait ensuite le point sur les difficultés que connaissent les projets Big Data et conclut sur les éléments de méthode qui permettent de les lever.

Keywords

big data | information system |
performance improvement

Abstract This article begins with the different stages in the deployment of IS in a company: automation and redesign of processes and operations (the operative word being 'organization'), and finally integration of software packages. It then summarizes the difficulties facing big data projects, and concludes with some ways to overcome them.

Pour toute question :

Service Relation clientèle
Techniques de l'Ingénieur
Immeuble Pleyad 1
39, boulevard Ornano
93288 Saint-Denis Cedex

Par mail :

infos.clients@teching.com

Par téléphone :

00 33 (0)1 53 35 20 20

Document téléchargé le : **08/12/2025**

Pour le compte : **7200029588 - ec nat polytechnique enp // samira SAFIR // 41.111.163.4**

© Techniques de l'Ingénieur | Tous droits réservés

Systemes d'Information (SI) d'Entreprise et conduite des projets Big Data

par **Pierre DELORT**

DSI & Président de l'Association Nationale des Directeurs des Systemes d'Information
(ANDSI), Consultant
Professeur Invité Telecom ParisTech, France

1. Historique des SI d'entreprise et évolution des enjeux	H 5 080 – 2
1.1 Automatisation	— 2
1.2 Banalisation des process	— 2
1.3 Hétérogénéité des projets	— 2
2. Difficultés des projets Big Data	— 2
2.1 Problèmes et solutions	— 2
2.2 Exemple : utilisation chez Google de modèles en Ressources Humaines (RH)	— 2
2.3 Difficultés de mise en œuvre des démarches Big Data	— 3
3. Démarches & Méthodes	— 3
3.1 Compréhension commune : le mythe et ses déclinaisons	— 3
3.2 Insérer l'induction	— 3
3.3 Méthodes de conduite des initiatives	— 4
4. Conclusion	— 5
Pour en savoir plus	Doc. H 5 080

L'informatique d'entreprise a connu plusieurs évolutions depuis un demi-siècle. Dévolue initialement à l'automatisation de quelques tâches (les trois premières furent la gestion des stocks, de la paie et de la comptabilité), elle a intégré dans certaines firmes, peut-être depuis le début des années 1990, autour du terme « organisation », une refonte des process et des opérations. Une quinzaine d'années plus tard, la prévalence des progiciels, qui proposent tout autant des outils que des process, a rendu les Directions des Systemes d'Information beaucoup moins actives sur ce sujet, ainsi que sur l'innovation en résultant.

Maintenant d'autres types de projet que ceux d'automatisation (transfert de coûts de personnes vers technologies) sont présents, et pour ceux de transformation ou d'information, le Big Data constitue un relais de valeur qu'apportent les DSI. Relativement à la construction de SI partant de modèles du monde (modèles de traitement, de données...), l'approche Big Data utilise les mathématiques pour trouver des modèles dans les données, modèles qui, sous condition, font prendre de l'avance sur le présent, parfois sur le futur pour y améliorer décisions et opérations.

Nous constatons que si trouver des modèles, par exemple de Machine Learning, plus performants que l'existant (mix de vision, expériences et... modèles) est relativement facile, l'insertion de ces modèles dans les bons mécanismes de décision pour amélioration de performances est moins aisée. Faisant suite à ce constat, nous présentons une démarche générale et des méthodes spécifiques visant à inscrire les projets Big Data dans des objectifs d'amélioration de performances.

1. Historique des SI d'entreprise et évolution des enjeux

1.1 Automatisation

Dans les entreprises, l'accent a été mis, initialement et longtemps, sur l'**automatisation**, c'est-à-dire un transfert de coûts d'hommes vers machines. Ceci a, au tout début, concerné des îlots d'activité dans les fonctions stock, paie et comptabilité. Avec la baisse des coûts de l'informatique, cette automatisation s'est propagée à des îlots adjacents et s'est accompagnée de recherche d'amélioration de temps de cycle, de qualité, de traçabilité et également de meilleure information pour de meilleures décisions, par exemple concernant l'affectation de ressources, l'utilisation d'actifs, la tarification en temps réel, etc.

À la fin des années 1980, la naissance d'une mode managériale, le **Re engineering** rendit populaire la notion de process, ainsi que la nécessaire réflexion préalable à l'automatisation d'activités. Autour du mot **organisation**, les Directions des Systèmes d'Information (DSI) se dotèrent de nouvelles compétences utiles aux projets d'importance : analyse de process, conception d'opérations, gestion du changement, formation...

1.2 Banalisation des process

En 2003, un article influent de Nicholas Carr dans la Harvard Business Review [6] annonça, si ce n'est la fin de l'informatique, du moins celle de son importance en tant qu'avantage concurrentiel. Il affirma une saturation prochaine de l'équipement en TI des entreprises et le fait que lorsqu'une firme achète un progiciel, elle achète peut-être d'abord un corpus de process qui, de ce fait, se banalisent.

Dès lors, les process et les technologies deviennent des commodités pour lesquelles le risque de panne est prévalent.

Un **exemple** historique est l'utilisation d'énergie électrique, centrale dans la première caractéristique du système Ford, la production en flux continu, mais totalement banalisée quelques décennies après sa première mise en œuvre en 1908. Ainsi, une centrale électrique de 8 000 chevaux se trouvait au centre de l'usine de Highland Park et fit l'objet, de la part de Henry Ford, d'autant de fierté que la chaîne d'assemblage, lorsqu'il fit visiter le site à Charles Chaplin en 1923 [1].

Les conclusions de Carr sont que les DSI d'entreprise doivent s'éloigner de l'innovation et se focaliser sur les coûts et les vulnérabilités (le risque n'est pas tant de perdre des opportunités de gains, mais de perdre par défaillance), dépenser moins, suivre et ne pas précéder (les évolutions de performances pénalisant les pionniers)...

1.3 Hétérogénéité des projets

Carr a principalement focalisé son analyse sur les projets d'automatisation, laissant de côté les cinq autres types de projets qui sont concernés par les TI, et que nous pouvons présenter sous une forme simple, mais parlante, en projet :

1. **obligatoire**. Cela peut être une contrainte réglementaire (Euro...) ou un minimum de performance sur un marché ;
2. **d'infrastructure**, car nécessaire à des TI apportant, eux, des services ;
3. **d'information** afin de procurer de meilleures informations (plus riches, fraîches, plus fiables...) ;

4. **majeur** (ou stratégique) par la création d'un avantage concurrentiel sur les process/métiers de base... ;

5. **de transformation** en ré-inventant du business impactant, en sus des process/métiers de base, l'offre de service.

En laissant de côté les deux premiers (obligatoire et d'infrastructure), la nouveauté apportée par les technologies et pratiques du Big Data concerne les projets d'information, et principalement les projets majeurs et de transformation.

2. Difficultés des projets Big Data

2.1 Problèmes et solutions

Le Big Data, en ce que les mathématiques trouvent des modèles dans les données, suscite en France encore peu de projets majeurs ou de transformation. Les cas de succès concernent principalement les sociétés « Web », sans doute à cause d'un mix d'abondance de données, qui plus est centrales à l'activité, et de la jeunesse des structures et personnes, favorable à l'adoption de mécanismes de décision reposant davantage sur le quantitatif que sur l'expérience. Maintenant, avant que d'exposer les méthodes que l'on constate significativement fonctionner, nous allons exposer un cas issu de Google qui rendra concret l'ampleur des transformations ouvertes.

2.2 Exemple : utilisation chez Google de modèles en Ressources Humaines (RH)

Traditionnellement, la gestion des RH se base sur des relations de confiance plutôt que sur des chiffres et des graphiques (et de nombreuses personnes y sont « sévèrement déficientes en mathématiques, analyse prédictive et statistiques » [2]). Google, dont le coût de personnel représente peut-être 60 % de ses OPEX (coûts de fonctionnement) et où chaque personne réalise en moyenne par an 1 million de dollars de chiffre d'affaires et 200 000 dollars de profit, procède différemment :

- une équipe d'analystes est directement reliée au Directeur et produit des sondages et études pour identifier des relations entre facteurs afin d'asseoir les décisions en partie sur des faits et des données ;
- toutes les décisions intègrent des données et leur analyse, le but étant d'atteindre le même niveau de rigueur que pour des décisions techniques ;
- l'équipe d'analystes est également en charge de présenter ses études et conclusions aux dirigeants et managers afin d'induire leur changement de compréhension, d'attitudes et de style de management.

Les études incluent des modèles de conservation permettant de prévoir quelles personnes sont les plus susceptibles de quitter la firme et personnaliser les outils de rétention, des modèles évaluant le succès futur des personnes en recrutement...

Le projet Oxygen, qui visait à répondre à la question « le management est-il important ? » (la réponse fut positive) se transforma en programme de mesure et de déploiement d'attitudes-clé de management, au moyen d'entretiens avec les employés quittant Google, puis utilisant les résultats des entretiens annuels, et enfin des méthodes plus sophistiquées comme des analyses multi-variées, des entretiens en double aveugle... [3].

Ainsi l'utilisation de modèles issus des données irrigue quantités de process RH (recrutement, appréciation, rétention...) et d'opérations de décision (embauche, augmentation, promotion...), ceux-ci constituant le cœur de compétences, non seulement des RH, mais

également de quiconque en charge d'une équipe. La jeunesse de Google (18 ans) et sa profonde culture quantitative aident certainement à des décisions « en connaissance des données ». Cependant, davantage que sur l'analyse, les efforts ont porté sur l'influence et la persuasion des managers et dirigeants, à intégrer, dans leurs décisions, les modèles issus des données.

Ce cas rapidement décrit illustre la méthode que nous employons depuis plusieurs années, tout d'abord le plan global Big Data, puis la conduite des initiatives d'amélioration des performances.

2.3 Difficultés de mise en œuvre des démarches Big Data

Bien peu de firmes sont aussi avancées que Google en utilisation des données. Maintenant si le « Quoi », c'est-à-dire l'intérêt du Big Data et de décisions basées en partie sur des données n'est plus remis en question, le « Comment » pose problème. Les acteurs du conseil et des technologies constatent que les projets Big Data ne dépassent que rarement le stade de prototype ou POC (*Proof of Concept*). S'il est relativement facile de trouver des modèles plus performants que l'existant (très souvent un mix d'expérience, d'intuition et de quelques modèles inscrits dans les fichiers de tableur, les applications informatiques...), la difficulté est d'insérer ces modèles dans les bons mécanismes de décision ou d'opération. Dès lors, la transformation reste souvent difficile à réaliser.

Trois phénomènes principaux y participent : la définition peu précise du Big Data pour quantités d'acteurs des TI, sa transversalité, et la spécificité des connaissances utiles aux projets.

Ceci se traduit par :

- une faible compréhension commune, partagée et rationnelle des apports potentiels du Big Data ainsi que de sa claire position dans des projets d'amélioration de performance ou de transformation ;
- le manque, peut-être particulièrement en France, de compréhension par les décideurs d'entreprise, de l'induction (mathématiques appliquées, ceci incluant l'économétrie, les statistiques inductives...), de ses intérêts mais également de ses limites ;
- une absence de méthode face à des projets très singuliers, ce qui freine également la création d'un marché entre clients (DSI) et fournisseurs des différents services utiles aux projets : cabinets de conseil, intégrateurs, spécialistes des data tels les statisticiens, administrateurs des données...

Nous allons reprendre dans l'ordre ces facteurs limitants, les caractériser et aborder des éléments permettant de les réduire.

3. Démarches & Méthodes

3.1 Compréhension commune : le mythe et ses déclinaisons

Les technologies, après les religions, la nature et les idéologies, sont aujourd'hui porteuses de mythe. Le battage médiatique entourant le Big Data brouille sa compréhension et de ce fait limite les ambitions des porteurs possibles de projets.

« Like any new technology, there is a lot of confusion surrounding Big Data. There are endless debates about what is and isn't Big Data, armies of consultants who are eager to muddy the waters in return for a hefty retainer fee and the usual amount of hype and alphabet soup of acronyms and buzzwords » [4].

Comme pour toute nouvelle technologie, il y a beaucoup de confusion autour du Big Data. Des débats sans fin à propos de ce qui est ou n'est pas le Big Data, des armées de consultants désireux de brouiller les pistes pour des honoraires conséquents et la quantité habituelle de battage médiatique avec les acronymes et les mots à la mode.

De plus, si les DSI connaissent bien les données des différents composants de l'entreprise, elles ne connaissent pas toujours les potentiels d'amélioration des décisions et opérations, et ne sont pas toujours légitimes pour les identifier et les évaluer. D'autre part, ces composants métier, lorsqu'ils connaissent bien leurs données, connaissent peu leurs complémentaires dans le corpus interne, alors que, pour le Big Data, la valeur des données réside souvent dans cette mise en correspondance, dérivant de leur partage effectif.

Il est ainsi souvent nécessaire de susciter une compréhension commune, partagée et... rationnelle de ce que recouvre le Big Data pour les acteurs économiques, quelles sont les perspectives business ou opératoires autorisées, quelles sont les technologies nécessaires, les organisations pertinentes, les savoirs utiles... Plusieurs méthodes et moyens, dont la pédagogie inversée, les séminaires et les *learning expeditions*, peuvent être employés.

3.2 Insérer l'induction

Les démarches pragmatiques de statistiques inductives permettant la création de modèle à partir de données sont peu pratiquées en France. L'excellence Française en mathématiques, entre autres en médailles Fields, se focalise sur les mathématiques théoriques.

Les mathématiques appliquées, aujourd'hui intimement liées à l'utilisation des Technologies de l'Information, sont moins prisées et pratiquées ; l'OCDE, dans son étude sur l'innovation et les données a comparé, dans 28 pays, le taux d'emploi des *data-specialists* (mathématiciens, actuaires, statisticiens et spécialistes des bases de données...). La France se positionne avant-dernière, entre le Portugal et la Turquie.

Le sujet ne se pose pas qu'en termes de compétences mathématiques, utiles en ce qu'elles peuvent améliorer les décisions et opérations, mais dans leur insertion dans ces mécanismes de décisions, pour lesquelles le facteur limitant est plutôt RH et managérial que technique. Des réticences existent, dès lors direction et exemplarité sont fondamentales. Ceci implique d'accepter que l'expérience n'est pas la seule source de rationnel, tolérer les questions, voire les désaccords, sous conditions d'arguments factuels et basés sur des « données », etc.

"Becoming data-driven is very difficult for many executives. They are reluctant to turn over decision-making to people who make decisions on the basis of data rather than expertise" [5].

Décider en influence forte des données constitue une réelle difficulté pour de nombreux dirigeants. Ils sont réticents à déléguer les décisions à des personnes qui utilisent pour cela les données plutôt que l'expertise.

La pédagogie est ici tout à fait essentielle, en permettant aux décideurs de comprendre les *data-specialists* et de conjuguer leurs arguments reposant sur des données, donc issus du passé, avec leur vision adressant l'avenir. Cette pédagogie est nécessaire à l'initiation de la démarche comme nous l'avons vu, mais également en accompagnement, à tout niveau de décision et à toute étape des projets de changement, afin que l'apprentissage soit collectif et efficace.

3.3 Méthodes de conduite des initiatives

De notre expérience en conduite de projet ou en conseil sur le Big Data, il nous semble essentiel d'adopter des outils d'identification des données pertinentes (figure 1) ainsi qu'une méthode lisible visant à concevoir un plan Big Data. L'objectif de ce plan est de poser et rationaliser les moyens (technologies, profils nouveaux de *data specialists*, sources d'innovation par exemple pour le *Machine Learning*) et lançant un nombre limité d'initiatives.

Cette matrice classe les données en fonction de leur volume et de leur densité en information, les deux axes sont logarithmiques. Une base de données est composée de tables, comme celles d'un tableau, elles-mêmes constituées de lignes et de colonnes, à l'intersection desquelles se trouvent des cellules. Le volume des données est mesuré comme étant le nombre de cellules, et la densité des données comme étant la proportion de cellules « signifiantes », c'est-à-dire non vides ou non nulles. Ceci nous permet de définir quatre cadrans :

- les données à forte densité en information et faible volume (cadrant en bas à droite) sont des données de SI classique d'entreprise. Typiquement les bases de données d'entreprise somment entre quelques mégaoctets et gigaoctets, et entre 5 et 10 % des informations y sont utiles, c'est-à-dire remplies et/ou non nulles ;
- le quadrant en haut et à droite est un peu particulier. Les deux logos présents sont ceux de Axiom et ChoicePoint, courtiers de données ;
- le quadrant en bas à gauche est indéterminé. Le temps dira dans quel sens se fera l'évolution, par exemple vers le Big Data par accumulation ;
- enfin, le quadrant, en haut à gauche est typique du Big Data ; un fort volume de données, elles-mêmes à faible densité en information.

Les seuils entre cadrans sont indicatifs. Celui du volume, donné à mille milliards de cellules, et celui de la densité, donné à une cellule renseignée sur mille, nous semblent aujourd'hui corrects, ou tout au moins plausibles. Ils sont évolutifs suivant les progrès, principalement des performances du stockage et du traitement des données.

Étant donnée la focalisation « data » de ces initiatives, elles doivent faire l'objet d'une attention particulière : examen des distributions pour identifier les outliers, tests des données sur leur validité...

Ensuite, ces initiatives sont à objectifs ouverts, et ce dans une bien plus grande ampleur que dans les projets Agiles.

De la même manière, la caractéristique exploratoire de ces projets ne doit pas conduire à ne pas manager les équipes, même si la méthode employée doit être différente. Parmi les points d'attention figurent la coordination de l'équipe, sa communication, la rigueur dans l'analyse... De plus, en phase d'initialisation, le nombre d'embauches ou d'appel à des interventions externes peut être conséquent. Ainsi, une définition précise des rôles est cruciale. S'appuyer sur des fiches standards des différents métiers intervenants dans les initiatives, comme nous le faisons, s'avère très utile.

Enfin l'assurance qualité doit être présente (et visible) pour assurer l'absence de bug dans le code produit comme une erreur dans le *parsing* des *data* (= analyse des données), une mauvaise utilisation des statistiques (confusion dans les coefficients de corrélation, de détermination...), une communication induisant des erreurs de compréhension des résultats (causalité, corrélation, causalités de Granger...). Là encore, l'utilisation d'une méthode sécurisée et accélère le process.

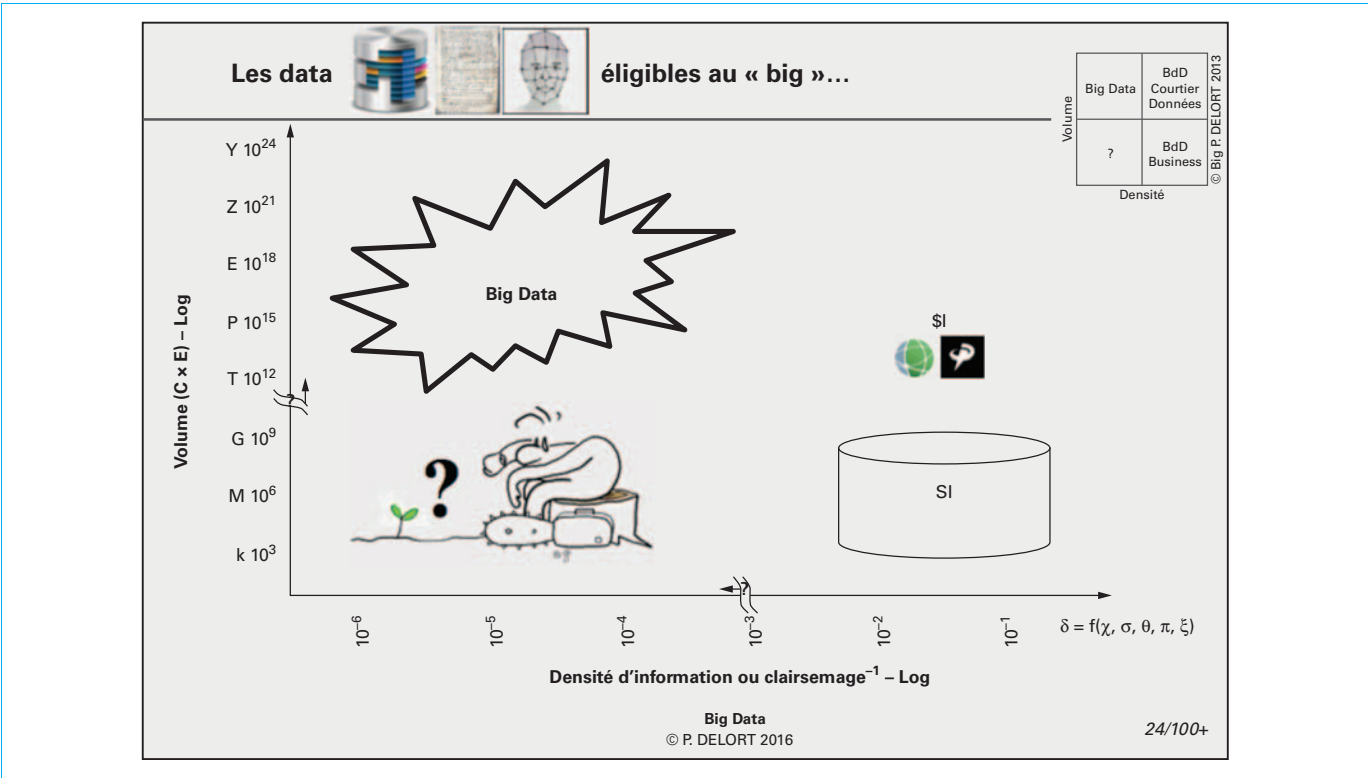


Figure 1 – Matrice volume x densité, quelles sont les data éligibles au « Big » ?

4. Conclusion

Aujourd'hui, les DSI découvrent avec les fournisseurs de consulting, de service ou de technologies, un apport nouveau aux performances de leur entreprise par l'utilisation de modèles plus performants issus des données.

Trouver ces modèles demande, comme nous l'avons vu, des démarches outillées par des méthodes, en premier lieu pour identifier les domaines favorables, et ensuite pour découvrir dans les données massives des modèles performants.

Maintenant, si des modèles performants sont relativement faciles à trouver, la difficulté réside en leur insertion dans les bons mécanismes de décision ou opération pour des améliorations de performances (au premier ordre, s'articulant autour des Coûts, de la Qualité, de la Flexibilité, des Délais ou des Services).

Nous constatons que, plus tôt dans la démarche générale sont pris en compte ces objectifs, meilleure sera cette insertion et donc les résultats, ceci allant de pair avec des changements multiples placés au cœur de la Transformation Numérique des entreprises : changements Technologiques (Applications/Outils/Composants), Organisationnels (Structures/Process/Opérations), RH (Management/Postes/Compétences) et Business (Modèle/Offre/Produits).

Systemes d'Information (SI) d'Entreprise et conduite des projets Big Data

par **Pierre DELORT**

DSI & Président de l'Association Nationale des Directeurs des Systèmes d'Information (ANDSI), Consultant
Professeur Invité Telecom ParisTech, France

Sources bibliographiques

- [1] DELORT (P.). – *Information contre Institution*. Mines ParisTech (1997).
- [2] SULLIVAN (J.). – *Don't Look Now, But Google Is a Big Talent Competitor to Your Company*. <http://www.tlnt.com> (27 février 2013).
- [3] GARVIN (D.). – *How Google Sold Its Engineers on Management*. Harvard Business Review (Dec 2013).
- [4] SATELL (G.). – Forbes (2013).
- [5] SCHMARZO (W.). – *Fostering a data-driven culture*. Economist Intelligence Unit (2013).
- [6] CARR (N.). – *It doesn't matter*. Harvard Business Review (2003).
- [7] DELORT (P.). – *Le Big Data*. Presses Universitaires de France, collection *Que Sais-je ?*.

Événements

Big Data Paris, a lieu tous les ans à Paris
<http://www.bigdataparis.com>

Brevets

Matrice volume x densité déposée sous enveloppe SOLEAU 517545, 2014.

Gagnez du temps et sécurisez vos projets en utilisant une source actualisée et fiable !

15 DOMAINES D'EXPERTISE

- ✓ Automatique - Robotique
- ✓ Biomédical - Pharma
- ✓ Construction et travaux publics
- ✓ Électronique - Photonique
- ✓ Énergies
- ✓ Environnement - Sécurité
- ✓ Génie industriel
- ✓ Ingénierie des transports
- ✓ Innovation
- ✓ Matériaux
- ✓ Mécanique
- ✓ Mesures - Analyses
- ✓ Procédés chimie - bio - agro
- ✓ Sciences fondamentales
- ✓ Technologies de l'information



Articles de référence
disponibles en français
et en anglais

Détails des offres et sommaires
à retrouver sur le site

www.techniques-ingenieur.fr

Les offres Techniques de l'Ingénieur permettent d'accéder à une **base complète et actualisée d'articles** rédigés par les meilleurs experts et validés par des comités scientifiques, avec :

+ de 10 000 articles de référence et 1 000 fiches pratiques opérationnelles.

3 000 quiz dans + de 1 000 articles interactifs.

+ de 550 bases documentaires, 8 Parcours Pratiques répartis dans plus de 90 offres.

1 280 auteurs contribuent chaque année à enrichir cette ressource.

Service de Questions aux experts.

Les Archives, technologies anciennes et versions antérieures des articles.

+ de 300 000 utilisateurs de techniques-ingenieur.fr chaque mois !

NOS ÉQUIPES SONT
À VOTRE DISPOSITION

Par téléphone
 33 (0)1 53 35 20 20

Par email
 infos.clients@teching.com