

PatentToNet : l'exploration libre des brevets par les réseaux

L'utilisation des brevets comme source d'information dans le processus de veille des organisations est connu mais contraint à l'utilisation de logiciels coûteux pour sa mise en œuvre. Par ailleurs, l'*Office Européen des Brevets* a libéré l'accès à sa base de plus de 80 millions de brevets via une API. Dans cet article nous décrivons un nouvel outil de collecte de données brevets pour visualisation, manipulation puis exploration d'un 'univers brevet'. Pour un ensemble de brevets répondant à une requête donnée, nous construisons un ensemble de réseaux dynamiques d'associations des données bibliographiques qui seront visualisés et manipulés pour fournir une présentation globale et claire, plus facilement compréhensible, de l'ensemble exploré et dans une perspective temporelle. Nous illustrerons au travers de quelques exemples sous l'angle d'un processus de veille intelligente au service du processus de recherche et développement des organisations (innovation frugale, veille technologique, veille concurrentielle).

Dans la propriété industrielle, le Brevet joue un rôle clé, de par son rôle stratégique puisqu'il représente un droit de propriété pour une invention, un produit ou un procédé qui offre une nouvelle solution technique à la résolution d'un problème. L'*Intelligence Compétitive* (IC) est un processus de veille intégrant la recherche d'information pertinente et utile à l'ensemble des prises de décision. Ce processus est opéré en général à l'externe des organisations et l'Internet s'est montré, depuis son avènement, comme une excellente source informationnelle. L'ouverture des données et le "Big-data" se déversent depuis peu participant ainsi à la croissance exponentielle des données du réseau ce qui contribue à enrichir encore cette source informationnelle d'intérêt pour l'IC. La volumétrie et la dynamique du réseau imposent alors la mise en œuvre d'une nécessaire instrumentation adéquate (Chen *et al.*, 02) d'intérêt fondamental pour l'IC. Ainsi, l'*Open Patent Services* (OEB 14) ouvre l'accès à la base mondiale de brevets de l'*Office Européen des Brevets*, construite par agrégation de diverses bases nationales (90 pays). L'interface web¹ offre aujourd'hui l'accès aux quelques 80 millions de brevets déposés. Ces brevets, rédigés dans langue du pays de dépôt, sont décrits et classés par des données dites « bibliographiques » qui ont pour obligation d'être traduites, à minima, en anglais. Cette base fait partie du web invisible car seul le formulaire web ou l'API ouvrent l'accès aux données des brevets. Les fonctionnalités ouvertes en mode API (*Application Programming Interface*) libèrent la totalité de la base à des requêtes automatiques.

L'instrumentation des processus d'IC est aujourd'hui incontournable (Quoniam, 13). Il convient cependant de séparer dans la chaîne de traitement précédent l'analyse différentes phases spécifiques :

- Une phase de collecte de données automatisée (Young *et al.*, 08), tri puis sélection de données (Liu *et al.*, 11) ;
- Une phase de visualisation (Yang *et al.*, 08) et navigation pour exploration (Salampasis & Hanbury, 13) .

Des solutions commerciales telle Intellixir², MathéoPatent³, provenant de la communauté des sciences de l'Information et de la Communication, ou encore Patent Integration⁴ existent. Mais sur celles-ci les fonctionnalités d'exploration par cartographies réseau sont très limitées. Par ailleurs, la faible pénétration de la bibliographie brevet dans la littérature académique,

¹<http://www.epo.org/>

²<http://www.intellixir.com>

³<http://www.matheo-software.com/fr/produits/matheo-patent-veille-brevet.html>

⁴<https://patent-i.com/>

pourtant source informationnelle technologique de première importance, suggère la nécessité d'offrir l'accès à cette ressource potentielle en appui complémentaire à la démarche scientifique ou encore à la vulgarisation scientifique.

En suivant ce découpage fonctionnel, nous présentons ci-après *PatentToNet*⁵ après avoir rappelé quelques notions élémentaires de réseaux et de visualisation/spatialisation, utilisées par la suite.

Quelques notions élémentaires de réseau

Depuis l'inscription de la théorie des graphes comme une branche des mathématiques par Euler (1736). Depuis, beaucoup de développements ont été faits notamment grâce à l'utilisation de graphes en sociologie avec le développement de la sociométrie (Parlebas, 1992): Pour ce dernier (Degenne et Forsé, 99) la théorie des graphes est « [...] un arsenal de concepts , théorèmes , algorithmes et raisonnement propice à l'analyse de réseau , et à travers lequel la représentation graphique va au-delà de la simple représentation pour permettre un traitement mathématique et générer de nouvelles connaissances ». Les informaticiens ont développé des algorithmes puissants qui permettent de décrire ces structures complexes dans un délai raisonnable. Mais ils ont également mis au point des logiciels de visualisation et permettre ainsi la représentation, la manipulation et l'exploration de grandes structures en réseau.

Il ya plusieurs affirmations de sens de la notion de réseau. (Albornoz *et al.*, 05) en énumèrent quatre :

- Le réseau systémique qui représente au plan fonctionnel les échanges et interactions des acteurs individuels et/ou institutionnels ayant des objectifs communs ou complémentaires. Cette représentation dénote à la fois un sens descriptif et sens normatif.
- une version socio-technologique par allusion aux nouvelles formes d'organisation qui surgissent et évoluent, associées au développement de certaines technologies ;
- une version économique, dans laquelle le concept de réseau est partie de la théorie de l'innovation ou, plus précisément, de la théorie des systèmes sociaux de l'innovation (Amable, 02);
- une version d'augmentation de capacités, comme une forme organisationnelle qui permet d'obtenir une masse critique en forme distribuée.

Ces considérations sont si variées que nous pouvons affirmer de la nécessité d'en déterminer les contenus objets et les relations *a priori* entre les objets qui seront représentés par un graphe. La sémantique de cette représentation, c'est à dire la combinaison d'arcs ou relations est aussi à spécifier (citation, lien économique, relation privilégiée, etc.) aura une influence sur les possibles interprétations.

La visualisation, élément indispensable pour l'exploration

L'utilisation d'images visuelles est courante dans de nombreuses branches de la science. Alfred Crosby (1997) a porté la Visualisation, en complément de la mesure, comme responsable de l'explosion de l'ensemble de la science moderne. Dans notre cas, la visualisation est utilisée comme un instrument transdisciplinaire pour comprendre la complexité de réseaux (Freeman, 00). L'utilisation des représentations spatiales sera nécessaire (Tufte & Graves-Morris, 83) en nous projetant parfois où la complexité défie l'art de la représentation (Díaz *et al.*, 02).

⁵<https://github.com/Patent2net/Patent2Net>

La littérature sur la visualisation des réseaux est très fournie (Di Battista *et al.*, 99) (Eades, 84) (Herman *et al.*, 00). Parmi les nombreuses spatialisations existantes (Eldridge, 06), les suivantes effectivement applicables dans notre approche favorisent les représentations suivantes :

- Affichage des divisions : Openord (Martin *et al.*, 11)
- Expression des complémentarités : Forceatlas (Bastian *et al.*, 09), Frushterman - Reingold (Fruchterman Reingold, 91), Yifan Hu (Hu, 05)
- Affichage des classifications : circulaire (Breitkreutz *et al.*, 03), radiale (Huffaker *et al.*, 99),
- Disposition géographique (Becker *et al.*, 95) (Misue *et al.*, 95),
- Décrire les connexions (graphiques non orientés) : Kamada - Kawai (Kamada & Kawai, 89).

Le lecteur intéressé peut se reporter à (Díaz *et al.*, 02) (Landesberge *et al.*, 11) pour une description plus complète et de nombreuses méthodes de visualisation de données plus sophistiquées.

Un outil libre de collecte de données brevet

Patent2Net est une suite de script python écrit sous licence Cecil B incluant :

- *OPSGather-PatentList* : envoie une requête à l'API OPS et récupère la liste de brevets. La demande suivra le format de CQL qui permet la construction de requêtes complexes. Un nom de fichier placé comme premier paramètre du script est utilisé pour écrire la liste de brevets résultants.
- *OPSGather-BiblioPatents* : utilise le fichier résultat précédent pour collecter l'ensemble des données bibliométriques (inventeurs, mandataires, date de dépôt, pays, classification et statut).
- *PatentToNet* qui produit le réseau temporel des entrées bibliographiques associées aux brevets, au format *gexf* avec des attributs utiles à la manipulation et l'exploration.

Ainsi, dans le réseau produit un brevet sera lié à tous ses attributs de données bibliographiques, et chacun d'eux s'est appuyé également à l'autre produisant un graphe complet pour chaque brevet.

Ainsi PatentToNet produit un graphe complet de données bibliographiques sur les brevets. Le graphique complexe créé sera manipulé avec Gephi⁶, dont les fonctions de filtrage, de regroupement, de sous-réseaux et de visualisation seront utilisées pour l'explorer.

L'analyse des brevets avec PatentToNet

L'analyse traditionnelle des brevets réponds à de nombreuses questions au-delà de la simple veille technologique (Hidalgo-Nuchera *et al.*, 09) par exemple : l'impact de la structure sociale sur la créativité individuelle (Flemming *et al.*, 07), la diffusion des connaissances (Singh, 03), la dynamique régionale (Breshi *et al.*, 09), un cartographie des technologie (Engelsman *et al.*, 94) ou encore les prévisions de l'emploi (Dou *et al.*, 12). Certains logiciels d'analyse de brevet fournissent déjà l'utilisation de la représentation en réseau pour explorer le contenu des brevets tels que Matheo-Patent (Dou *et al.*, 09), mais nous donnons ci-après une solution plus souple qui ouvre la voie à de nouvelles forme d'analyse de brevets.

Ainsi réseaux co-inventeurs et des réseaux co-mandataires qui sont fréquemment étudiés sont inclus et seront traités en utilisant le processus de filtrage en Gephi. D'autres sous-réseaux produits présentent plusieurs connaissances issues des données brevets recueillies comme par exemple :

- Les croisements technologiques (réseau IPCR-IPCR, cf. Figure 1),
- L'intérêt national pour la technologie protégée (à travers le réseau IPCR-Pays);

⁶ <http://www.gephi.org>

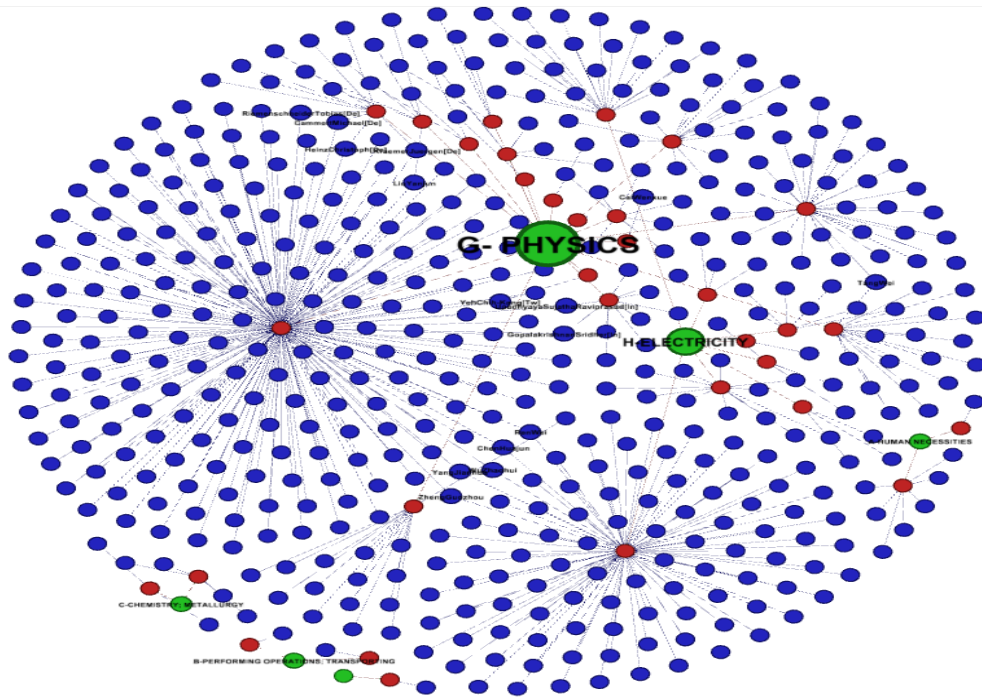


Figure 4 : Les inventeurs (bleu) et classifications IPCR 1 (vert), IPCR4 (rouge) des brevets "Big Data"

- Inventeur et technologies (IPCR-inventeur). En analysant ce réseau on peut identifier-qui « est spécialiste de » (cf. figure 1) ;
- Inventeur et entreprises (Inventeur-Mandataire) montrant le qui « travaille pour qui », ou, sous un autre aspect, l'interrelation de données tangibles et intangibles (respectivement investisseur et la connaissance de l'inventeur) ;
- Similairement étudier les aspects technologiques des sociétés via les réseaux mandataires-technologie (cf. Figure 2, Figure 3. En perspective temporelle entre 1995 et 2005).

Figure 1 : Croisements technologiques et applications autour de l'optique adaptative en représentation circulaire

Figure 2 : L'impression 3D en 1995, quelques brevets US et européens. Une société dominante (*Image Technology Int*). Deux brevets mondiaux. Applications classées en G03B (appareils photo) et H04N (transmission image).

Figure 3 : L'impression 3D en 2005 : CorpZ est l'entreprise qui détient le plus de brevets. Les types d'applications majoritaires ont changé (B29C - façonnage matière plastique) également. Apparitions de brevets dans de nombreux pays (Corée, Mexique, Chine, etc.)

La spatialisation du réseau par l'algorithme de Fruchtermanet Reingold positionne les inventeurs (en bleu) et les sous catégories de niveau 4 (IPCR4) autour des grandes catégories de la classification internationale des brevets (IPCR) (nœuds verts). Par cette représentation (cf. Figure 1) les réseaux complémentaires (attachés principalement aux grandes catégories comme physique ou électricité) forment des réseaux quasiment séparés et leur interrelation se fait via les inventeurs « multidomains » dont les nœuds sont légèrement plus importants et leur nom apparaît.

Conclusion

En conclusion nous ouvrirons sur les évolutions potentielles de tels traitements informationnels devenant indispensables (Masiakowski *et al.*, 13) par la création complémentaire de connecteurs à d'autres bases complémentaires (par exemple la base complémentaire *Patent Scope* gérée par l'Organisme Mondial de la Propriété Intellectuelle - OMPI, (WIPO, 14), le traitement de curation en réponse (Carayol & Cassi, 09) aux limites engendrées par les nombreuses erreurs que comporte la base (typographie, orthographe, dénominations, inconsistances, etc.), ou encore la 3D pour la navigation. L'ensemble des résultats s'inscrit dès lors comme de l'*open innovation* la base contenant aussi les demandes de brevet (même non abouties). Les outils de collecte et de navigation deviennent alors des outils à outils (Trippe, 03); les résultats d'analyse des inventions, des inventions elles-mêmes pour "orienter des recherches autour du brevet depuis l'écriture de nouveaux brevets, l'invalidation de brevet existant, la création d'information à valeur ajoutée, des liens avec d'autres systèmes d'information" (Quoniam, 13). Sous un autre aspect, l'outil PatentToNet permet l'enseignement de cette approche information et communication dans les masters InfoCom et Gestion qui n'auraient pas accès aux ressources payantes comme certaines actuellement en France, au Brésil et au Sénégal.

Références

- (Albornoz *et al.*, 05) Albornoz M. and Alfaraz, C. (2005), *Redes de conocimiento: construcción, dinámica y gestión*, RICYT, Buenos Aires, Argentina, august, <http://www2.ricyt.org/novedades/74-nueva-publicacion-qredes-de-conocimiento-construccion-dinamica-y-gestionq>, link followed in Feb. 2014.
- (Amable, 02) Amable B. 2002. « Les systèmes d'innovation ». In *L'Encyclopédie de l'innovation*. Paris Economica.
- (Bastian *et al.*, 09) Bastian, M., Heymann, S. et Jacomy, M. (2009) Gephi: an open source software for exploring and manipulating networks. In : *ICWSM*. p. 361-362.
- (Becker *et al.*, 95) Becker, R. A., Stephen G. Eick, and Allan R. Wilks (1995). "Visualizing Network Data." *IEEE Transactions on Visualization and Computer Graphics* 1(1): 16-28.
- (Breitkreutz *et al.*, 03) Breitkreutz, B.-J., Stark, C. and Tyers M. (2003). "Osprey: a network visualization system." *Genome Biology* 4(3): R22.
- (Breschi *et al.*, 09) Breschi, S. and Lissoni F., (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9(4):439-468.
- (Carayol & Cassi, 09) Carayol N., and Cassi L., (2009). Who's Who in Patents. A Bayesian approach. *Cahiers du GREThA* 2009--07.
- (Chen *et al.*, 02) Chen, H. Chau, M. Zeng, D. (2002). "CI Spider: a tool for competitive intelligence on the Web." *Decision Support Systems* vol. 34, n. 1 pp. 1-17.
- (Díaz *et al.*, 02) Díaz, J., Petit, J., et Serna, M. (2002) A survey of graph layout problems. *ACM Computing Surveys (CSUR)*, vol. 34, no 3, p. 313-356.
- (Degenne et Forsé, 99) Degenne, A. et Forsé, M (1999). *Introducing social networks*. Sage.
- (Dou *et al.*, 12) Dou, H., Mohellebi, D., & Kister, J. (2012). , L'importance du traitement bibliométrique des brevets pour développer l'activité industrielle. Exemple des bitumes en Algérie. *RIST (Revue Scientifique et Technique)*, p. vol 19 n°1.
- (Dou *et al.*, 09) Dou, H., Mannina B. et Massari G. (2009) « Análise de patentes para melhorar a competitividade tecnológica e o pensamento inovador », *International Journal of Competitive Intelligence, Strategic, Scientific and Technology Watch*.
- (Di Battista *et al.*, 99) Di Battista, G., P. Eades, R. Tamassia, and I.G. Tollis (1999). *Graph Drawing: Algorithms for visualization of graphs*, Prentice Hall.
- (Eades, 84) Eades, P. (1984). "A Heuristic for Graph Drawing." *Congressus Numerantium* 42: 149-160.
- (Eldridge, 06) Eldridge, J. (2006) "Data visualisation tools--a perspective from the pharmaceutical industry." *World Patent Information* vol. 28, n. 1. pp. 43-49.
- (Engelsman *et al.*, 94) Engelsman, E. C. & van Raan, A. F. J., (1994) "A patent-based cartography of technology", *Research Policy*, Elsevier, vol. 23(1), pages 1-26, January.
- (Flemming *et al.*, 07) Fleming, L. and S. Mingo, D. Chen (2007). Collaborative Brokerage, Generative Creativity, and Creative Success. *Administration Science Quarterly*, 52: 443-475
- (Freeman, 00) Freeman, L. C. (2000) Visualizing social networks. *Journal of social structure*, vol. 1, no 1, p. 4.
- (Freeman, 79) Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215-239.
- (Fruchterman Reingold, 91) Fruchterman, T. M. and Reingold, E. M. (1991) Graph drawing by force-directed placement. *Softw. Pract. Exper.* 21 (11), pp 1129-1164
- (Herman *et al.*, 00) Herman, I., Melançon, G. and Marshall M. S. (2000). "Graph Visualization and Navigation in Information Visualization: A Survey." *IEEE Transactions on Visualization and Computer Graphics* 6 (1): 24-43
- (Hidalgo-Nuchera *et al.*, 09) Hidalgo-Nuchera, A., Iglesias-Pradas, S., Hernández-García, Á., (2009) Utilización de las bases de datos de patentes como instrumento de vigilancia tecnológica, *El Profesional de la Información*, Volume 18, Number 5, September - October, pp. 511 - 520.
- (Hu, 05) Hu, Y. (2005) « Efficient, high-quality force-directed graph drawing ». *Mathematica Journal* 10, n° 1: 37-71.

- (Huffaker *et al.*, 99) Huffaker, B., Evi Nemeth, and K. Claffy (1999). "Otter: A general-purpose network visualization tool." *Internet Society INET'99 Conference*.
- (Kamada & Kawai, 89) Kamada, T., and S. Kawai (1989). "An algorithm for drawing general undirected graphs." *Information Processing Letters*, 31 (1): 7-15.
- (Landesberge *et al.*, 11) von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J.J., Fekete, J.-D. and Fellner, D.W. (2011), Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. *Computer Graphics Forum*, 30: 1719–1749. doi: 10.1111/j.1467-8659.2011.01898.x
- (Liu *et al.*, 11) Su-Houn Liu, Hsiu-Li Liao, Shih-Ming Pi, and Jing-Wen Hu. 2011. Development of a Patent Retrieval and Analysis Platform - A hybrid approach. *Expert Syst. Appl.* 38, 6 (June).
- (Martin *et al.*, 11) Martin, S.; Brown, W. M. Klavans, R., *et al.*(2011)OpenOrd: an open-source toolbox for large graph layout. In : *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics. p. 786806-786806-11.
- (Masiakowski *et al.*, 13) P Masiakowski, S Wang. (2013) "Integration of software tools in patent analysis" - World Patent Information, Volume 35, Issue 2, June, Pages 97–104, Elsevier
- (Misue *et al.*, 95) Misue, K., P. Eades, W. Lai, and K. Sugiyama (1995). "Layout Adjustment and the Mental Map." *Journal of Visual Languages and Computing* 6(2): 183-210.
- (OEB 14) Open Patent Services (OPS) Office Européen des Brevets [en ligne: <http://www.epo.org/searching/free/ops.html>] - consulté le 13/01/2014.
- (Porter *et al.*, 05) Porter, A. L. et S. W. Cunningham(2005),*Tech mining : Exploiting new technologies for competitive advantage*. NewYork : Wiley InterScience,.
- (Quoniam, 13) Quoniam, L., (2013) "Le brevet comme objet de recherche", dans *Recherches ouvertes sur le numérique : approches pratiques en information-communication* sous la dir. de Fabrice Papy. - Paris : Hermès ; Lavoisier. - 319 p. - Traité des sciences et techniques de l'information, série Environnements et services numériques d'information, p 95-114.
- (Ravaschio *et al.*, 10) Ravaschio, Juliana de Paula, L. Innocentini Lopes de Faria, et Quoniam L.(2010) "A informação patentária em trabalhos acadêmicos da engenharia química: considerações sobre o seu uso no Brasil e EUA." Dans *Seminário Nacional de Bibliotecas Universitárias 16, 2010*. Rio de Janeiro: UFRJ (Brésil).
- (Salampasis & Hanbury, 13) M. Salampasis and A. Hanbury (2013), "A Generalized Framework for Integrated Professional Search Systems" dans *Multidisciplinary Information Retrieval, Lecture Notes in Computer Science* volume 8201, p. 99-100.
- (Singh, 03) Singh, J. (2003), Social networks as drivers of knowledge diffusion, SSRN, Harvard University.
- (Spangler *et al.*, 11) S. Spangler, C. Ying, J. Kreulen, S. Boyer, T. Griffin, A. Alba et al. (2011), "Exploratory analytics on patent data sets using the SIMPLE platform". *World Patent Information*, N°33 pp. 328–339.
- (Tufte & Graves-Morris, 83) Tufte, E. R. et Graves-Morris, P. R. (1983) *The visual display of quantitative information*. Cheshire, CT : Graphics press.
- (Trippe, 03) A.J. Trippe, (2003)Patinformatics: tasks to tools, *World Patent Information*, 25, pp. 211–222.
- (Wang & Cheung, 11) W.M. Wang, C.F. Cheung, (2011)"A Semantic-based Intellectual Property Management System (SIPMS) for supporting patent analysis", *Engineering Application and Artificial Intelligence*, N°24, pp. 1510–1520.
- (Wasserman et Faust, 94) Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- (WIPO, 14) Organisation mondiale de la propriété intellectuelle, OMPI <http://www.wipo.int/patentscope/fr/> consulté le 14/01/14.
- (Yang *et al.*, 08) Y. Yang, L. Akers, T. Klose, C. Barcelon (2008) Text mining and visualization tools – impressions of emerging capabilities, *World Patent Information*, 30, pp. 280–293.
- (Young *et al.*, 08) Young G. K., Jong H. S., and S. C. Park. (2008), Visualization of patent analysis for emerging technology. *Expert Syst. Appl.* 34, 3 (April), 1804-1812.