

# Thamous

proposé par

Alain Herreman

avril-mai 2004

Solution informatique pour la gestion d'une base de données bibliographiques, pluridisciplinaire, scientifique et collective pourvue de liens entre les références.

Principales caractéristiques de la base de données :

- données : bibliographiques, biographiques, institutionnelles, revues, éditeurs, ...
- multi-projets
- liens typés entre références
- usage public et privé
- validation scientifique
- signature scientifique et individuelle

## Plan

- [Introduction - présentation générale](#)
- [Caractéristiques](#)
  - [pluridisciplinarité : projets](#)
  - [multi-types](#)
  - [liens typés](#)
  - [usage public et privé](#)
  - [Entrée d'une nouvelle référence](#)
  - [Suppression et modification d'une référence](#)
  - [Recherche de données](#)
  - [Contraintes de publicité des liens](#)
  - [Contrainte de liaison](#)
  - [Validation scientifique](#)
  - [Signature](#)
- [Représentation des données](#)
- [Statuts](#)
- [Projets scientifiques](#)
  - [Histoire et épistémologie des sciences mathématiques grecques anciennes . Transmission des textes \(Bernard Vitrac\)](#)
  - [Géométrie algébrique réelle, de Descartes à nos jours \(Michel Coste, Alain Herreman, Marie-Françoise Roy\)](#)
  - [Les limites du vivant du XVIIIe au XXe siècle \(Stéphane Tirard\)](#)
  - [Le Répertoire Bibliographique des Sciences Mathématiques \(Philippe Nabonnand, Laurent Rollet, Laurent Guillopé\)](#)
  - [Histoire du calcul graphique et graphomécanique \(Dominique Tournès, Joachim Fischer, Konstantinos Chatzis\)](#)
  - [Sources mathématiques numérisées sur Gallica \(Frédéric Métin\)](#)
  - [Sciences et milieux scientifiques en France dans l'entre-deux-guerres \(Hélène Gispert\)](#)
- [Spécifications](#)

- [Cahier des charges](#)
- [Faisabilité](#)

## Introduction - Présentation générale

Ce programme se propose de tirer parti des possibilités offertes par internet en développant une base de données bibliographique, pluridisciplinaire, scientifique et collective pourvue de liens entre les références.

Un moteur de recherche permet aujourd'hui d'obtenir de nombreuses informations sur une référence bibliographique. Néanmoins, cette sélection ne répond à aucun critère scientifique et aucune garantie n'est donnée quant à la fiabilité et à l'intérêt des résultats obtenus. Il importe de mettre en place une base de données bibliographique dont la pertinence des sélections soit garantie par des experts. Par ailleurs, internet permet de nouvelles formes de collaborations collectives. Il est en effet fastidieux de rentrer des données bibliographiques et il est inutile que chacun entre dans sa base de données personnelle des références que d'autres ont déjà certainement entrées : idéalement, **une même information ne doit être saisie qu'une seule fois** et corollaire : **une information valide qui peut être rendue publique ne doit pas rester privée**. Chacun pourrait ainsi soumettre une référence, avec les informations utiles la concernant, et celle-ci serait effectivement enregistrée dans la base de données publique après avoir été validée.

La validation se fait par le comité scientifique d'un **projet**. Un projet délimite un thème : l'histoire des mathématiques grecques, la géométrie algébrique réelle, la génétique au 20<sup>e</sup> siècle, Darwin et le darwinisme, La réception de *la critique de la raison pure* de Kant, etc. Quand l'accès à la base de données se fera à partir d'un projet déterminé, il sera alors possible à un utilisateur de soumettre des données qui seront ou non validées par le comité scientifique de ce projet. Les experts pourront ainsi valider les données sans avoir à les entrer eux-mêmes. Il ne s'agit pas de développer un système de recensement bibliographique systématique (automatique). Cette approche très utile dans certains domaines l'est moins dès que l'on ne s'intéresse pas seulement aux publications récentes. De plus, un tel système n'a pas besoin d'experts ou plutôt : il n'en tire parti. Thamous se propose au contraire de favoriser l'acquisition et la diffusion de la connaissance de ces experts. Les recherches automatiques, aussi utiles soient-elles, ne sauraient les remplacer.

Afin de permettre aux projets de se développer suivant leurs spécificités, leurs responsables doivent pouvoir définir eux-mêmes les informations (notamment les mots-clés) pertinentes et leur niveau de publicité. Cela conduit à distinguer des **données publiques** et des **données spécifiques à un projet**. Les données comme le titre, le nom de l'auteur, l'éditeur, url, etc. seront nécessairement publiques (« contrainte de publicité »). La base contenant aussi des noms de personnes, d'institutions, de revues, d'éditeurs, url, etc. une part de ces informations seront aussi publiques. Ainsi, les données saisies au sein d'un projet constitueraient toujours une contribution à une diffusion publique commune et accessible à tous. Ayant été validée par des experts, leur pertinence et leur intérêt seraient ainsi garantis. De la sorte, toutes les données avant d'être publiques auront été validées par un comité scientifique. C'est la distinction et la conciliation d'un niveau public et d'un niveau rapporté à un projet scientifique spécifique qui permet de développer un système aussi attrayant pour les experts que pour le grand public, chacun bénéficiant des contributions de l'autre. Thamous permettant à tous les usagers d'entrer des données, ceux-ci contribueront à la collecte des url des documents disponibles sur internet, cette donnée étant bien sûr publique.

Il convient aussi de s'assurer de l'intérêt de chacun à alimenter la base de données. Offrir des fonctionnalités puissantes est nécessaire mais ne suffit pas. Suivant le principe qu'une même information ne doit être saisie qu'une seule fois, il importe que les usagers n'aient pas à saisir dans une base de données personnelle les données qu'ils saisissent dans la base publique. Pour cela il convient de leur offrir la possibilité d'avoir des **données privées**. Ces données privées seront stockées sur une base de données personnelle coordonnée avec la base de données publique. De cette

manière, tous les usagers bénéficieront de l'apport de tous les autres, mais chacun conservera la possibilité de saisir des données qu'il ne veut pas rendre publiques. Ces bases de données privées pourront être sur un ordinateur personnel ou sur un serveur distant. Afin néanmoins de toujours garantir une contribution à la base de données collective, les données comme le titre, le nom de l'auteur, l'url, etc. seront obligatoirement publiques. Enfin, afin que les contributions de chacun soient reconnues, les données saisies seront signées par celui qui les aura entrées. Cette signature permet aussi à chacun de retrouver plus facilement les données qui sont pour lui particulièrement pertinentes.

La base de données permet la **création de liens entre références**. Cela permet notamment de saisir les références citées par une référence donnée. De nombreux autres types de liens peuvent être définis. De tels liens complètent les recherches par mots-clefs. Les liens entre références sont aussi des données objectives qui peuvent être saisies par une seule personne et être profitables à toutes. De plus, il est facile quand on lit un livre ou un article de savoir qu'il comporte des références, des remarques, des commentaires, etc. sur d'autres livres ou articles. Il est impossible en revanche de le savoir à partir de ces autres livres ou articles... Créer un lien permet en introduisant une information facilement accessible de récupérer une information difficilement accessible. Une partie de l'information récupérée par chacun lors de ses lectures est ainsi préservée pour lui-même et aussi mise à la disposition de tous. Les références pour lesquelles les textes cités auront été intégralement saisies seront ainsi distinguées et progressivement de plus en plus nombreuses. D'autres types de liens peuvent être créés suivant les trois niveaux : public, projet, privé. Il appartient à chaque comité scientifique de définir les types de liens adaptés à son projet et d'établir une liste de références dont il convient de saisir en priorité intégralement la bibliographie. Ce système de liens permet ainsi de connaître facilement (dans la mesure où les informations ont été saisies...) quels textes, quels auteurs, quelles institutions sont cités ou citent le texte considéré et ainsi de naviguer entre les références, en remontant ou en progressant dans le temps. Ce simple parcours donne à lui seul un riche aperçu historique sur un texte et le réseau de références dans lequel il s'inscrit. La contribution de tous les utilisateurs assortie d'une validation permet d'obtenir un nombre important d'informations, les liens et leur validation au sein de projets en garanti la cohérence et la validité. Ces données objectives sont utiles aux chercheurs, aux étudiants, aux enseignants et à toute personne intéressée par un texte ou un ensemble de textes considéré.

La diversité des projets, le système de publicité et de liaison des données favorisent la **pluridisciplinarité et l'interdisciplinarité**. En effet, les données et les liens saisis à partir de projets scientifiques contiennent toujours une partie publique. Les liens créés à partir de différents projets sont donc accessibles à tous. Il est ainsi possible de rendre visible l'insertion d'une référence dans les disciplines et les périodes couvertes par l'ensemble des projets scientifiques.

## Caractéristiques

### Pluridisciplinarité : projets

L'accès à la base se fait à partir d'un **projet**. Tout projet a un **nom de projet** et un **comité scientifique**.

A chaque projet correspond un environnement propre : mots-clefs, données accessibles, système de validation.

Les projets peuvent être hébergés sur divers serveurs à la convenance de leurs responsables.

### Multi-types

La base de données comprend différents **types de références** : bibliographiques, biographiques, revues, institutions, liens (voir section suivante).

Chaque référence contient différents **types de données**. Une référence bibliographique contient les types de données habituels (y compris des urls pour les documents disponibles sur internet) : intitulé, nom d'auteur, éditeur ou nom revue, mots-clefs, etc.

Idem pour référence biographique : nom, date naissance, date décès, pays, mots-clefs, etc.

Chaque type de donnée a un **type d'accessibilité** unique: public, privé, nom d'un projet.

## Liens typés

Liens typés entre deux références.

Lien quel que soit le type des références, le type des deux références peut être différent.

Exemples de types de liens: « réf. citée » (pour indiquer qu'une référence biblio. ou une personne est citée par une référence biblio.) « correspondance » (pour indiquer que deux références bio. ont correspondu), « élève » (pour indiquer qu'une référence bio. a été l'élève d'une autre) etc.

Les types de liens sont définis par chaque projet.

Pour chaque type de référence, une sous-liste de ses types de données sera distinguée pour définir les liens. Au moins l'un d'entre eux doit être de type public (voir section suivante).

Navigation et recherche sur les liens. Par exemple : liste de toutes les références qui citent livre donné.

## Usage public et privé

La base de données comprend plusieurs bases de données coordonnées. Toutes sont accessibles par internet.

La **base publique** contient toutes les données de type public.

Une base pour chaque projet qui contient les données propre au projet, i.e. les données non publiques. Ces bases peuvent être situées sur des serveurs différents.

Différents **droits d'accessibilité** sont définis : public, privé, nom de projet.

NB : public (resp. privé) n'est qu'un cas particulier de projet. La distinction est conservée par commodité.

## Entrée d'une nouvelle référence

L'entrée des données se fait toujours à partir d'un projet. Leur validation, y compris celle des données publiques, appartient au comité scientifique du projet. Il lui appartient notamment de veiller à ne pas introduire de doublons. Les données publiques sont sauvegardées sur la base publique, les données propres au projet sur la base du projet. Les données privées sont directement sauvegardées sur la base de données privée sans passer par le comité scientifique du projet.

Idem pour un lien (l'exclusion des doublons est cette fois automatique).

## **Suppression et modification d'une référence**

Dès lors qu'une référence contient une donnée publique, sa suppression ou sa modification doit être validée par le comité scientifique de Thamous (afin de garantir la cohérence et éviter les conflits entre deux projets).

Le comité scientifique d'un projet peut supprimer ou modifier toutes les données relevant de son projet.

Idem pour un lien.

## **Recherche de données**

Les recherches portent sur les données publiques et les données du projet à l'intérieur duquel se fait la recherche. Les données publiques et celle du projet sont visibles, avec les données privées de la personne à laquelle elles appartiennent.

## **Contraintes de publicité des liens**

- les types de lien « réf. citée », « réf. non citée » (liste à compléter) sont nécessairement publics.
- Un lien contient nécessairement au moins un type public (« réf. citée », « correspondance », etc.).
- un lien entre deux références bibliographiques contient nécessairement « réf. citée » ou « réf. non citée » comme type public.

En conséquence, la navigation avec les liens à partir d'une référence relevant d'un projet peut faire apparaître des références d'autres projets (seules les données publiques seront visibles).

## **Contrainte de liaison**

- Une référence bibliographique doit toujours être liée.

Avec la contrainte de publicité, assure la liaison de toute référence à la base publique.

## **Validation scientifique**

La validation se fait par  $n$  membres du comité scientifique.  $n$  et la composition du comité scientifique sont fixés par les responsables du projet. Ces paramètres sont des informations accessibles aux utilisateurs.

Des autorisations complètes peuvent être attribuées à des individus : leurs modifications des données du projet seront automatiquement validées.

Afin de coordonner l'ensemble des projets et permettre la modification des données publiques, un comité scientifique Thamous doit être créé. Il est composé de représentants des principaux projets.

## **Signature**

Les personnes proposant des modifications doivent s'identifier.

L'identification comprend : un nom, une adresse courriel, un mot de passe, un identifiant numérique.

Chaque modification, une fois validée, est signée par le numéro d'identification. Cela afin de reconnaître et d'inciter les contributions personnelles. La modification est aussi signée par le projet qui l'a validée.

La liste des contributeurs avec le nombre de contributions (nombre de références entrées et modifiées, plusieurs modifications sur une même référence ne comptant qu'une fois...).

## Représentation des données

### **Recherche conditionnelle dans plusieurs champs**

(...)

### **Editions des liens**

(...)

### **Edition de chronologies**

(...)

## Statuts

La base publique n'est la propriété d'aucune personne ou institution. Elle peut être diffusée et copiée librement. (...)

Les données des projets ont leur statut propre.

## Projets scientifiques

### Définition

Les projets doivent offrir les garanties suivantes :

- Compétence scientifique des responsables pour le domaine couvert ;
- Amorçage : capacité raisonnable à alimenter la base de données et à atteindre le niveau de complétude défini par le projet ;
- Mise à jour: maintenir le niveau de complétude de la base défini par le projet.

La définition d'un projet doit permettre aux utilisateurs d'apprécier la valeur du résultat d'une recherche.

Histoire et épistémologie des sciences mathématiques grecques anciennes.

## Histoire et épistémologie des sciences mathématiques grecques anciennes. Transmission des textes

Responsable : Bernard Vitrac (CNRS - Centre Louis Gernet)

Ce projet recense les références bibliographiques portant sur les disciplines que les Anciens eux-mêmes qualifient de mathématiques (Arithmétique, géométrie, astronomie, optique, mécanique,

canonique (théorie mathématique de la musique), géographie mathématique ...). Pour cet aspect, la période historique concernée est : VI<sup>e</sup> siècle avant JC - VI<sup>e</sup> siècle après.

Il suit la transmission des textes conservés (pas la transmission des "idées"), pour ces mêmes disciplines, sur une période plus longue : de l'Antiquité jusqu'au début du XVI<sup>e</sup> siècle.

### **Géométrie algébrique réelle, de Descartes à nos jours**

Responsables : Michel Coste, Alain Herreman, Marie-Françoise Roy dans le cadre de l'IRMAR (UMR CNRS-Université de Rennes I)

Ce projet recense les principales publications de la Géométrie algébrique réelle. Il propose une liste de quelques dizaines de mots-clefs pertinents et définit différents liens entre les références recensées. Il est lié au projet européen RAAG (The European Research Training Network : [www.ihp-raag.org](http://www.ihp-raag.org)).

### **Les limites du vivant du XVIII<sup>e</sup> au XXI<sup>e</sup> siècle**

Responsable : Stéphane Tirard (Université de Nantes)

Ce projet recense les références traitant de l'histoire et de l'épistémologie de notions telles que la vie ralentie, la vie suspendue, l'anhydrobiose, la cryptobiose, la diapause, la dormance, la reviviscence, la quiescence... ainsi que des références se rapportant à la vie dans les milieux extrêmes.

### **Le Répertoire Bibliographique des Sciences Mathématiques**

Responsables : Philippe Nabonnand (Archives Poincaré), Laurent Rollet (Archives Poincaré), Laurent Guillopé (Cellule Math-Doc).

Le Laboratoire de Philosophie et d'Histoire des Sciences - Archives Henri Poincaré (Université Nancy 2) travaille à la publication de la correspondance scientifique, administrative et privée d'Henri Poincaré. L'exploration de la correspondance mathématique de Poincaré a permis d'apporter un éclairage nouveau sur un épisode important de l'organisation internationale de la communauté mathématique du 19<sup>e</sup>-20<sup>e</sup> siècle. En effet, durant toute sa vie, le mathématicien nancéen dirigea une entreprise internationale de constitution d'une bibliographie exhaustive des sciences mathématiques. Le Répertoire Bibliographique des Sciences Mathématiques est devenu un objet très rare et notre projet vise à le saisir dans son intégralité dans une base de données informatique. Ce projet de recherche s'inscrit dans le cadre d'une campagne de numérisation de revues mathématiques anciennes et actuelles et il bénéficie de soutiens de la Bibliothèque Nationale de France et de la cellule Math-Doc. Voir [présentation détaillée](#) et <http://math-sahel.ujf-grenoble.fr/RBSM/>.

### **Histoire du calcul graphique et graphomécanique**

Responsables : Dominique Tournès (REHSEIS, CNRS et université Paris 7), Joachim Fischer (Technische Universität Berlin), Konstantinos Chatzis (LATTS, École nationale des ponts et chaussées)

L'objectif est de recenser, principalement pendant la période 1750-1950, les références bibliographiques en rapport avec les divers aspects du calcul graphique (calcul par le trait, statique graphique, nomographie, intégration graphique, instruments mathématiques de type graphomécanique).

Ce projet, qui bénéficie du soutien du programme "Histoire des savoirs" du CNRS, veut contribuer à

l'inventaire et à l'étude des savoirs pratiques de calcul en usage dans les différents corps de métiers, notamment chez les ingénieurs.

### Sources mathématiques numérisées sur Gallica

Responsable : Frédéric Métin (APMEP)

Ce projet recense les sources relatives aux mathématiques publiées par la bibliothèque Gallica (<http://gallica.bnf.fr/>). L'accès à une source se fera directement grâce à son url.

### Sciences et milieux scientifiques en France dans l'entre-deux-guerres

Responsable : Hélène Gispert (GHDSO - Université de Paris-Sud XI)

Ce projet a pour objectif de contribuer à recenser les sources primaires et secondaires concernant les sciences et les milieux scientifiques les plus larges en France dans l'entre-deux-guerres. Il s'intéressera à la production des savoirs, à leur circulation et leur promotion dans les milieux scientifiques ainsi qu'aux différents cercles d'acteurs concernés par ces multiples formes de l'activité scientifique. Un intérêt particulier sera porté aux enjeux institutionnels, sociaux, culturels, politiques de la science dans la société française de cette période.

## **Cahier des charges**

- Programme libre conformément aux normes de la FSF (Free Software Foundation)
- Le programme doit être bien écrit afin d'en permettre la compréhension et la modification.
- Il doit offrir les fonctionnalités décrites dans les spécifications.
- Le développeur est invité à prendre des initiatives pour la présentation et les fonctionnalités offertes.
- Transfert des données de la base actuelle dans la nouvelle base.

(...)

**Document réalisé sous TeXmacs.**