

# Attribution de coordonnées standardisées aux cantons et localités suisses pour une saisie homogène des étiquettes lacunaires

Autor(en): **Vust, Matthias**

Objektyp: **Article**

Zeitschrift: **Bulletin de la Société Vaudoise des Sciences Naturelles**

Band (Jahr): **102 (2023)**

PDF erstellt am: **10.08.2024**

Persistenter Link: <https://doi.org/10.5169/seals-1046410>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# Attribution de coordonnées standardisées aux cantons et localités suisses pour une saisie homogène des étiquettes lacunaires

Mathias VUST<sup>1</sup>

VUST M., 2023. Attribution de coordonnées standardisées aux cantons et localités suisses pour une saisie homogène des étiquettes lacunaires. *Bulletin de la Société vaudoise des Sciences naturelles* 102 : 67-73.

## Résumé

La numérisation des herbiers anciens révèle des cas où seule la ville, voire le canton, est indiquée comme localité. D'autres échantillons n'ont pas d'indication de lieu, mais la biographie du récolteur laisse penser qu'ils proviennent de Suisse. Faute d'indications plus précises, la question se pose de savoir quelles coordonnées et quelle incertitude attribuer à ces échantillons, afin de les transmettre aux bases de données nationales et internationales, qui les représenteront sur des cartes de répartition. Une standardisation est proposée et discutée.

**Mots-clés :** cantons, géolocalisation, herbiers, numérisation, standardisation.

VUST M., 2023. Allocation of standardized coordinates to Swiss cantons and localities for consistent entry of incomplete labels. *Bulletin de la Société vaudoise des Sciences naturelles* 102 : 67-73.

## Abstract

Digitization of old herbariums reveals cases where only the town, or even the canton, is indicated as the locality. Other samples have no indication of localization, but the biography of the collector suggests that they came from Switzerland. In the absence of more precise information, the question arises as to what coordinates and uncertainty should we assign to these samples in a digitization's goal targeting national and international databases in order to represent them on distribution maps. Standardization was proposed and discussed.

**Keywords :** counties, digitization, geolocalization, herbaria, standardization.

VUST M., 2023. Zuweisung von standardisierten Koordinaten an Schweizer Kantone und Orte für eine einheitliche Erfassung von lückenhaften Etiketten. *Bulletin de la Société vaudoise des Sciences naturelles* 102 : 67-73.

## Zusammenfassung

Bei der Digitalisierung alter Herbarien tauchen Fälle auf, in denen nur die Stadt oder sogar der Kanton als Ort angegeben ist. Andere Proben haben keine Ortsangabe, aber alles deutet aufgrund der Biografie des Sammlers darauf hin, dass seine Herkunft in der Schweiz liegt. Mangels genauerer Angaben stellt sich die Frage, welche Koordinaten diesen Proben zuzuordnen sind und wie unsicher sie sind, damit sie in nationale und internationale Datenbanken übertragen werden können, die sie auf Verbreitungskarten darstellen. Eine Standardisierung wird vorgeschlagen und diskutiert.

**Schlüsselwörter :** Digitalisierung, Geolokalisierung, Herbarien, Kantone, Standardisierung.

<sup>1</sup> Institut de biologie, Université de Neuchâtel, mathias.vust@unine.ch

## INTRODUCTION

Un organisme, déterminé par son nom latin, observé en un endroit, à une date et par un naturaliste donné constitue une *donnée scientifique*. Il s'agit du minimum requis pour être publié, compilé dans une base de données ou représenté sur une carte. Cela répond aux questions quoi? où? quand? et par qui? Il existe ensuite deux cas de figure, soit un témoin de l'espèce est récolté pour être conservé, il y a alors un *échantillon*, soit il n'y en a pas et la donnée correspond alors à une *observation*. L'échantillon constitue une collection matérielle alors que l'observation est une collection immatérielle. Les échantillons sont conservés dans des collections personnelles, des musées ou des conservatoires, tandis que les observations sont rassemblées sous forme de texte, de tableaux ou, à terme, rentrées dans des bases de données. La compilation de telles données permet ensuite de documenter la répartition des espèces dans l'espace et dans le temps (GROOM *et al.* 2019). À l'heure de la centralisation des données, au niveau national ou international, notamment avec le réseau Global Biodiversity Information Facility (GBIF), il apparaît essentiel de standardiser les types d'information recueillies ainsi que les informations elles-mêmes. L'objectif étant de rendre compatibles les différentes bases de données de sorte qu'un flux de données soit possible entre elles. Les standards ABCD (ABCD TASK GROUP 2007) et Darwin Core (WIECZOREK *et al.* 2012) répondent à la question de la standardisation des types d'information, fournissant une référence standard stable pour le partage des informations sur la diversité biologique (<https://dwc.tdwg.org>), utile aussi bien pour la gestion des collections naturalistes que pour celle des bases de données. Le défi est double : documenter l'état actuel de la répartition des espèces sur le terrain, et numériser les données anciennes contenues dans les publications ou conservées sous forme d'échantillons dans les musées d'histoire naturelles. Le but de l'initiative SwissCollNet est justement d'encourager la numérisation des collections en favorisant la collaboration des institutions suisses de sciences naturelles. C'est dans ce cadre que s'est fait sentir le besoin de répondre à la question « Quelles coordonnées attribuer à des échantillons qui n'en ont pas? Et particulièrement lorsque seule la ville ou le canton est indiqué? ».

Dans le cas de la numérisation de la collection des lichens de l'herbier de Neuchâtel (NEU), il est apparu que les échantillons ayant uniquement un canton comme lieu de récolte ont été récoltés durant une période limitée du XIX<sup>e</sup> siècle. En effet, dans la première flore de Suisse, HALLER (1768) signale des localités par le nom de lieux habités, de lieux-dits ou de sommets. À cette époque, tous les cantons ne font pas encore partie de la Confédération suisse. Il faut attendre 1803 (avec l'indépendance du canton de Vaud, entre autres), puis 1815 (avec l'entrée des cantons de Genève, Neuchâtel et du Valais), pour voir apparaître une image de la Suisse moderne composée de 25 cantons. Ce n'est donc qu'à partir de ces dates que les naturalistes ont pu être tenté·e·s, du moins en Suisse romande, de ne signaler la localisation de leurs échantillons que par le canton de récolte. Il faut préciser que le but des naturalistes suisses du début du XIX<sup>e</sup> siècle était souvent davantage la (re)connaissance des espèces et l'édification de flore, à une période de foisonnement des nouvelles espèces décrites, plutôt que la documentation de la répartition de ces espèces. Il n'est donc pas rare de trouver dans les herbiers des échantillons de cette époque sans récolteur·ice, sans date ou sans localité. Cependant, le ou la récolteur·ice peut parfois être reconnu·e à son écriture (VUST *et al.* submit.). Dans ces cas, la date peut être extrapolée par un intervalle correspon-

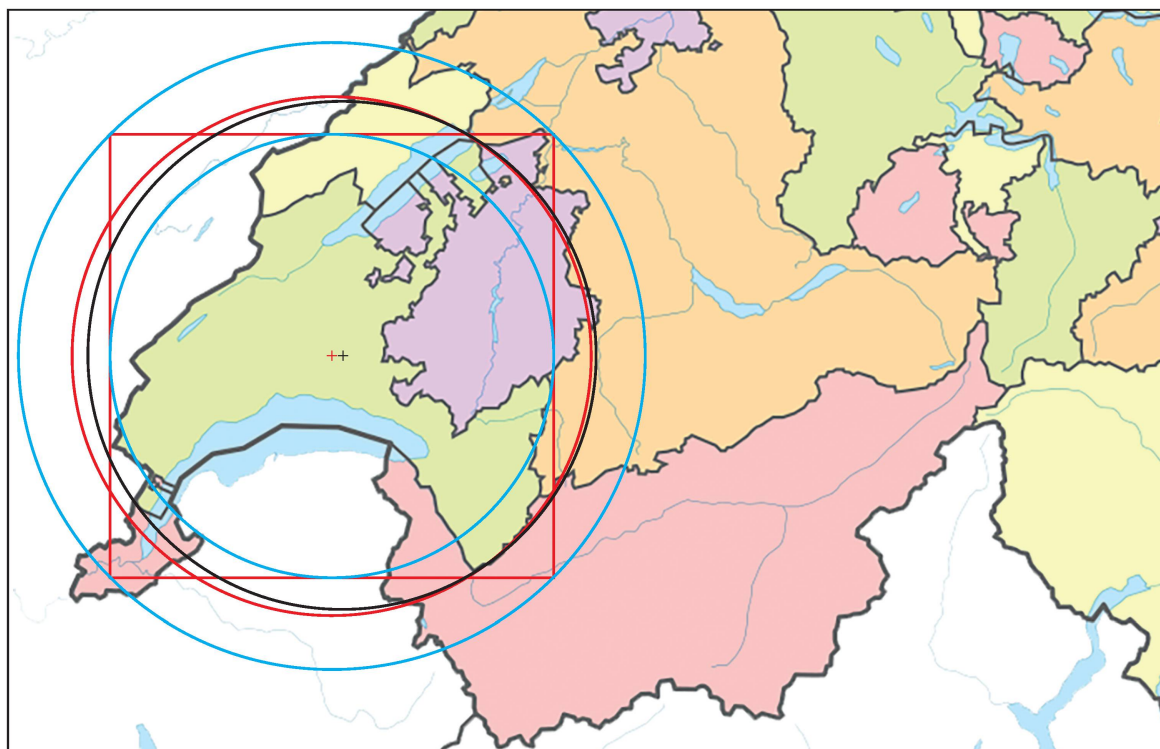
dant à la période d'activité du ou de la naturaliste. Ce court article se concentre donc sur la problématique des échantillons sans localité précise, dont on soupçonne fortement une récolte en Suisse d'après la biographie du récolteur. La problématique est la même lors d'une localité réduite à un canton suisse, ou une ville, ou une localité indiquée sans coordonnée. Comme une uniformisation est souhaitable au sein de chacun des champs d'information, une standardisation des coordonnées à attribuer dans de tels cas est recherchée. L'enjeu est non seulement une standardisation, mais surtout l'espoir de gain de temps énorme en cas d'automatisation de l'attribution de telles coordonnées.

## PROPOSITION DE STANDARDISATION - LA SUISSE ET SES CANTONS

L'enjeu est ici de trouver un *point central*, dont les coordonnées seront représentatives du territoire considéré, et une *incertitude* permettant d'intégrer l'ensemble de ce territoire, que ce soit un canton ou la Suisse entière. Il paraît en effet légitime qu'au lieu d'attribuer une incertitude unique par défaut, un canton de petite surface ait une incertitude plus réduite qu'un canton de plus grande taille.

Une première méthode consiste à déterminer les coordonnées limites  $X_{\min}$  et  $X_{\max}$ , de même que  $Y_{\min}$  et  $Y_{\max}$  de chaque canton. Les coordonnées du centre des parallélogrammes obtenus sont données par  $X_{\min} + [(X_{\max} - X_{\min})/2]$ ;  $Y_{\min} + [(Y_{\max} - Y_{\min})/2]$ . Ces parallélogrammes représentent les rectangles les plus petits contenant entièrement la surface de chaque canton. Comme il n'y a en principe qu'une seule incertitude, et non une incertitude différenciée pour les X et les Y, il n'est pas envisageable de garder le parallélogramme comme représentatif de l'incertitude. Un moyen d'avoir une unique incertitude est donc de considérer le plus grand côté de ce parallélogramme comme le côté du carré représentant l'incertitude à partir du point central. L'incertitude équivaut alors au demi-côté du carré. Seulement, il apparaît que cette représentation de l'incertitude correspond en fait au cercle inclus dans le carré et non au carré lui-même, or il peut se trouver des territoires cantonaux inclus dans le carré, mais situés hors du cercle de l'incertitude (figure 1). Il s'agit alors de déterminer le rayon minimal du cercle intégrant l'ensemble du territoire cantonal. Il est clair que la forme du périmètre du canton joue un rôle. Plus le périmètre du canton est proche d'un cercle, plus le demi-côté du carré sera représentatif. Au contraire, plus le périmètre du canton sera proche d'un carré, ou avec des avancées vers les coins du carré, plus ce sera la demi-diagonale qui sera représentative (figure 1). Mathématiquement, pour un carré de 1 km de côté, la surface est de 1 km<sup>2</sup>. Le disque inclus dans le carré a une surface de  $\pi r^2$ , soit 0,78 km<sup>2</sup>. La diagonale est égale à la racine carrée de la somme des carrés des deux côtés et vaut 1,41 km. Le disque incluant le carré, d'un rayon équivalant à la demi-diagonale, soit 0,7 km, a quant à lui une surface de 1,57 km<sup>2</sup>. Dans le cas du disque inclus dans le carré, 22 % du carré ne sont pas intégrés dans le disque. Dans le cas du disque incluant le carré, le disque intègre 57 % de surface en plus. Graphiquement, un disque de rayon intermédiaire paraît idéal (figure 1). Quel serait le rayon d'un disque de même surface que le carré? Il équivaut à la racine de  $1/\pi$ , soit 0,56 km, soit un peu moins que le 0,6 km correspondant à l'intermédiaire exact entre le rayon du disque inclus et celui du disque incluant le carré. L'incertitude est alors le rayon du disque équivalant au demi-côté du carré multiplié par 1,12, (0,5 multiplié par 1,12 équivalant à 0,56), afin que la surface du disque prévu équivale celle du carré intégrant la surface du canton.





**Figure 1.** Représentation du plus petit carré incluant un territoire et des disques susceptibles de représenter l'incertitude: le disque de rayon équivalent au demi-côté du carré et celui de rayon équivalent à la demi-diagonale (cercles bleus), le disque de rayon intermédiaire à la surface équivalente à celle du carré (cercle rouge) et le plus petit disque incluant l'entier du canton et centré sur le centroïde déterminé grâce à la fonction *centroïdes de polygone* de QGIS (cercle noir). Exemple du canton de Vaud, Suisse, en vert. Source: Office fédéral de topographie.

Une deuxième méthode consiste à faire calculer à un système d'information géographique (SIG en français, GIS en anglais) le centroïde du périmètre du canton, défini comme étant le centre du plus petit disque incluant totalement le périmètre. L'incertitude recherchée équivaut alors à ce rayon. La fonction *centroïdes de polygone* existe dans les outils de géométrie mis à disposition par QGIS par exemple. Les cantons suisses présentent toutefois quelques difficultés, plusieurs cantons comportant des enclaves d'autres cantons voisins!

Ces deux méthodes s'appliquent également aux communes, comme à la Suisse entière.

## PROPOSITION DE STANDARDISATION - LES AGGLOMÉRATIONS ET LIEUX-DITS

L'Office fédéral de topographie Swisstopo met à disposition des coordonnées dans les documents *swissnames3d*, de même que certains sites internationaux comme *geonames.org*. Trois difficultés se présentent, 1) l'utilisation de différents systèmes de coordonnées, 2) le manque d'explications quant à la méthode utilisée pour définir les coordonnées centrales proposées et 3) l'absence d'incertitude. Deux méthodes différentes et complémentaires sont proposées. La première consiste à utiliser les polygones des agglomérations Suisses de *swissnames3d\_ply*. Le document fournit des coordonnées et une altitude. L'incertitude serait attribuée en fonction du nombre d'habitants fourni dans la colonne « *Einwohnerkategorie* » (tableau 1). Il apparaît que les agglomérations les plus grandes, de plus de 50 000 habitants, peuvent présenter une

concentration d'habitants différentes les unes des autres, si bien que les incertitudes ont été attribuées au cas par cas, selon leur extension géographique. Les agglomérations plus petites peuvent aussi présenter de telles différences de concentration, notamment quand il s'agit de quartiers de villes comparés à des villages de campagne. La surface occupée par les agglomérations diminue néanmoins globalement en fonction de la diminution du nombre d'habitants, mais conserve un rayon minimal de 100 m pour les hameaux et lieux-dits de moins de 50 habitants (tableau 1) en raison de la dispersion de l'habitat dans de tels cas. Au total, 54 931 agglomérations correspondent à la classe d'objet « TLM\_Siedlungsname ».

La deuxième méthode, complémentaire, consiste à utiliser les noms figurants sur les cartes topographiques suisses, considérés comme des points dans le document `swissnames3d_pkt`. Des coordonnées sont fournies, ainsi qu'une altitude. Une incertitude standardisée de 100 m y a été attribuée. Une sélection a été faite : les noms de localité, les sommets et les cols ont été conservés tandis que les arrêts de bus, car et bateaux, les bornes, les lieux sacrés, les sources ou les établissements scolaires ont été retranchés. Le total s'élève à quelques 287 000 noms.

Le principal problème persistant est que de nombreux noms ne sont pas uniques ! Il reste donc à ajouter le canton pour séparer les agglomérations homonymes. Cependant, cela ne suffit pas toujours car des homonymes se trouvent parfois dans le même canton. Il existe par exemple 43 lieux-dits « Zopf » en Suisse.

**Tableau 1.** Incertitudes attribuées aux agglomérations selon leur nombre d'habitants.

Nombre d'habitant · e · s	Incertitude [m]	Nombre d'habitant · e · s	Incertitude [m]
< 20	100	2 000 à 9 999	500
20 à 49	100	10 000 à 49 999	1 000
50 à 99	100	50 000 à 100 000	500-2 000
100 à 999	150	> 100 000	1 500-2 500
1 000 à 1 999	250		

## DISCUSSION ET CONCLUSION

L'attribution de coordonnées uniques à un territoire, c'est-à-dire à une surface et non à un point, est un problème universel qui se pose lors de la numérisation d'échantillons naturalistes dont la localisation sur l'étiquette se réduit à un pays ou une subdivision de ce dernier. L'important dans ce cas n'est pas la précision des coordonnées, mais l'incertitude à donner pour qu'elles soient représentatives du territoire dans lequel l'échantillon a pu être récolté. Le *point central* et l'*incertitude* permettront une visualisation cartographique représentative de la réalité, avec une petite incertitude pour un petit territoire et une plus grande pour un territoire plus étendu. Les deux notions sont géométriquement différentes. Le territoire considéré étant représenté dans un système de coordonnées orthogonales, il s'insère dans un parallélogramme. L'incertitude unique sur la localisation de l'échantillon se représente mathématiquement de la manière la plus simple comme un rayon à l'intérieur duquel l'échantillon a pu être récolté. L'idée intuitive que l'incertitude puisse correspondre au demi-côté du plus petit carré englobant complètement le territoire considéré s'avère juste, mais peu satisfaisante puisque 22 %

de la surface du carré ne se trouve pas dans le disque du rayon de l'incertitude. Une solution mathématiquement simple consiste alors à choisir comme incertitude le rayon d'un disque de surface équivalente à celle du carré incluant le périmètre, soit 1,12 fois plus grand que le rayon du disque inclus dans ce carré. Cette méthode a les avantages d'être réalisable à la main et en peu de temps; elle a le désavantage de ne pas inclure dans l'incertitude obtenue l'absolue totalité du territoire. Néanmoins, devant l'impossibilité de faire correspondre le carré incluant le territoire et le disque représentant l'incertitude, il est rassurant de considérer deux formes partageant le même centre et ayant la même surface. C'est donc pour son élégance intellectuelle que cette solution est proposée.

L'alternative des outils fournis par les SIG est performante et efficace, mais étonnamment les résultats sont contrastés pour les cantons suisses. Si l'obtention du centroïde est une fonction classique, celle du rayon du disque incluant le polygone l'est moins. Le fait est que l'algorithme fournit une solution censée être parfaite, si bien que les imperfections obtenues laissent perplexes. Si le calcul des centroïdes et imprécisions des communes n'a pu être mis en défaut, celui des cantons a été corrigé à la main dans 21 cas sur 26 pour obtenir une intégration vraiment complète du territoire considéré. Autant la première méthode intégrait par définition des territoires potentiels dans les angles du carré qui n'étaient pas pris en compte dans le disque de l'incertitude, autant la méthode du centroïde est définie pour intégrer la totalité du territoire. La limite de cette méthode se trouve alors peut-être dans la taille des polygones et surtout dans les milliers de points qui les constituent dans le cas des cantons suisses. L'avantage considérable de l'utilisation des SIG est l'efficacité de la méthode pour obtenir les centroïdes des 2149 communes de Suisse.

Faute de consensus en Suisse, chaque conservateur·ice ou préposé à la saisie fait de son mieux selon sa propre méthode pour attribuer des coordonnées aux échantillons qui n'en n'ont pas. Les documents mis à disposition dans les annexes représentent donc des propositions de standards pour la Suisse, ses cantons, ses communes et ses lieux-dits. À l'inverse, une donnée naturaliste ayant l'une de ces coordonnées et incertitude pourra être reconnue comme une donnée sans localité précise, à qui il a été attribué les coordonnées par défaut. Dans le cas des agglomérations et lieux-dits, les coordonnées proposées sont à considérer comme des aides permettant de gagner du temps, mais dont il conviendra de vérifier l'adéquation. Cette liste de coordonnées standards n'est pas une solution directe et absolument valable dans tous les cas, du moins pas pour l'instant, en raison des nombreuses localités homonymes!

## REMERCIEMENTS

Cet article a été réalisé dans le cadre des projets Sinergia « Héritages botaniques des Lumières : exploration de sources et d'herbiers historiques à l'intersection des lettres et des sciences / Botanical Legacies from the Enlightenment : unexplored collections and texts at the crossroad between humanities and sciences » (Subside n° 186227) financé par le Fond National Suisse, et SwissCollNet « Lichens of the Enlightenment : Reconditioning, digitization, databasing and revision of the lichen collections in Neuchâtel, Lausanne, and Geneva » (Project number SCN206-NE) financé par la Confédération suisse. Nos plus vifs remerciements s'adressent à InfoFlora, et à Monique Vilpert en particulier, pour les échanges et la mise à disposition des données des centroïdes des cantons et communes de Suisse.



## BIBLIOGRAPHIE

- ABCD TASK GROUP. 2007. Access to Biological Collection Data (ABCD), Version 2.06. Biodiversity Information Standards (TDWG). <http://www.tdwg.org/standards/115>. (6 November 2019, date last accessed).
- GROOM, Q., DILLEN, M., HARDY, H., PHILLIPS, S. WILLEMSE, L. & WU, Z. 2019. Improved standardization of transcribed digital specimen data. *Database* Vol. 2019: article ID baz129; doi: 10.1093/database/baz129.
- HALLER A. VON, 1768. *Historia stirpium indigenarum Helvetiae inchoata*. Sumptibus Societas Typographica, Bernae.
- QGIS. [www.qgis.org](http://www.qgis.org).
- SWISSTOPO. <https://www.swisstopo.admin.ch/fr/geodata/landscape/names3d.html>, consulté en août 2023.
- WIECZOREK, J., BLOOM, D., GURALNICK, R. *et al.* 2012. Darwin core: an evolving community-developed biodiversity data standard. *PLoS One*, 7, e29715.
- VUST M., DI MAIO E. & GRANT J. Un cahier inédit d'étiquettes de botanistes. *Bulletin de la Société neuchâteloise des sciences naturelles*. Submitted.

## ANNEXES

**Annexe 1.** Coordonnées et incertitudes standardisées de la Suisse et de ses 26 cantons dans les systèmes de coordonnées WGS84 et CH1903+.

**Annexe 2.** Coordonnées et incertitudes standardisées des communes suisses dans les systèmes de coordonnées WGS84 et CH1903+.

**Annexe 3.** Coordonnées et incertitudes standardisées des agglomérations et des lieux-dits de Suisse dans les systèmes de coordonnées WGS84 et CH1903+.

À consulter sur <https://wp.unil.ch/svs/publications/bulletins/>



