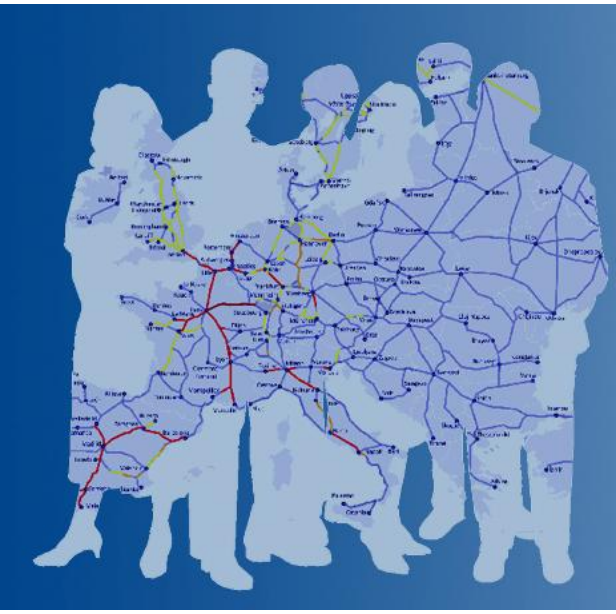
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

EuroSDR



1st EuroSDR Workshop

Crowd Sourcing for Updating National Databases

Identifying common interests and future research issues in the field of crowd sourcing for the updating of national databases

Workshop report

Workshop organized by
EuroSDR
and
Swiss Federal Office of Topography

Thursday / Friday August 20-21, 2009
Federal Office of Topography, swisstopo
Wabern, Switzerland

Workshop Report: Crowd Sourcing for the Updating of National Databases

This document is the report of the 1st EuroSDR Workshop on Crowd Sourcing for the Updating of National Databases, held at the Federal Office of Topography (swisstopo), Wabern, Switzerland on 20-21 August 2009.

Workshop Information

Title: 1st EuroSDR Workshop on Crowd Sourcing for Updating National Databases
Date: Thursday 20th and Friday 21st August 2009
Venue: Swiss Federal Office of Topography (swisstopo), Wabern, Switzerland

Goals

The workshop, organized by EuroSDR and the Swiss Federal Office of Topography swisstopo, dealt with the use of Web 2.0 technologies and communities for the mutual use of user generated content in the framework of national databases. Current applications and data management processes were reviewed and experiences and expectations, as well as current limitations and challenges, discussed.

Topics

- Web 2.0 technologies and communities
- User acceptance
- Data management
- Quality issues
- Legal tasks
- Advantages and Challenges for National Mapping and Cadastral Agencies

On day one of the workshop participants from the private sector, universities and government gave presentations on current projects and the state of the art. On day two common themes and future research needs were explored.

Workshop homepage

Information and further details can be obtained from the workshop homepage at http://www.swisstopo.ch/ and
http://www.eurosdr.net/workshops/crowdsourcing_2009/index.htm

Report

The Workshop was held at the Swiss Federal Office of Topography (swisstopo), Wabern, Switzerland, Thursday 20th and Friday 21st August 2009.

- Welcome and Introduction
- Presentations I
- Presentations II
- Summary of day one
- Break-out sessions and presentations
- Conclusions
- Next steps

Introduction

39 delegates from the United Kingdom, Sweden, Belgium, Germany, The Netherlands, Spain, France and Switzerland attended the Workshop "Crowd Sourcing for Updating National Databases" on 20-21 August 2009 in Wabern, Switzerland. The Workshop was organized by EuroSDR Commission III and the Swiss Federal Office of Topography (swisstopo).

The Workshop was opened by ***André Streilein***, Chair of EuroSDR Commission III, who welcomed the delegates to the Swiss Federal Office of Topography (swisstopo). He high-lighted the motivation and goals of this first workshop by identifying common interests and future research issues in the field of crowd sourcing for updating national databases. The use of Web 2.0 technologies and communities for the mutual use of user generated content in the framework of national databases is definitively an issue of importance for all national mapping agencies and map providers.

Presentations I

The first session of presentation was opened by ***Jantien Stoter*** from ITC Enschede, TU Delft and Kadaster, The Netherlands, with a presentation entitled "Crowd sourcing for updating national databases in The Netherlands". She illustrated the present crowd sourcing activities in The Netherlands by three examples. The introduction of Web 2.0 technologies by the Netherland's Kadaster and the Municipality of Rotterdam is currently under discussion but is hampered mainly by fears regarding quality, reliability and licensing issues. However in the provinces some Web 2.0 like applications help to gather specific geo-located information, for instance, for bird counting or damage reporting. The main problems are identified as reliability and commercial licensing.

Malcom Havercroft, Head of Operations from Ordnance Survey Great Britain, presented in his talk "Crowd-sourcing – some experiences and thoughts", recent research and projects in Great Britain. First, he described crowd sourcing applications in the broadest sense, where vernacular geographical place names are harvested in the web. He then focused on the Ordnance Survey "Explore" portal, where the outdoor community can contribute with, and ask for, hiking routes in the UK (http://explore.ordnancesurvey.co.uk/). The presentation was summarized with a graph, showing the main interacting/dependant variables of a crowdsourcing process: data volume, sourced data value, solution technology, economic aspects and level of accreditation system. Malcom Havercroft identified the main tasks to proceed as identifying the crowd, specification of the data looked for, accreditation aspects, technology and business model.

Geotagged photos in Web 2.0 sharing sites was the topic of the talk of ***Vyron Antoniou***, University College of London, UK, "Do photo sharing websites represent a sufficient database to aid in national map updating or change detection?". In his PhD work he assessed the spatial dimension of the geotagged photos phenomenon in the UK by examining the popular Web 2.0, photo-sharing websites of Flickr, Panoramio, Picasa Web and Geograph. He suggested that such Web sources can be categorized into spatially implicit and spatially explicit ones. Spatially explicit applications urge their users to interact directly with spatial features while at the same time encourage the content to be distributed. In contrast spatially implicit Web applications are more social-oriented and do not focus on the spatial element. He showed, that the covered area is as expected dominated by tourist sites and has, on a countrywide perspective, gaps. Spatially implicit applications such as Flickr, Picasa Web and Panoramio end up with a huge amount of data but poor spatial coverage whereas spatially explicit sources such as Geograph provide acceptable coverage with moderate data volumes.

Jean-Christophe Guélat, swisstopo, Switzerland, presented in his talk "Integration of user generated content into national databases" the actual and planned revision workflow web tool of the Swiss Federal Office of Topography. The present revision workflow is a one way approach from revision notifications of users, analysed by a centralised revision dispatcher and distributed to the different production processes of swisstopo. The future developments will include vector and GPS attachments and an exchange with, and accreditation of, the data user.

Presentations II

The second session started with the presentation of the OpenAdresses project by ***Hans-Jörg Stark***, University of Applied Sciences Northwestern Switzerland "An OpenGeodata Project: OpenAddresses". OpenAdresses (www.openaddresses.ch) is a project to collect geo-located addresses and linked attributes within a web interface by crowdsourcing. The project started in north-western Switzerland but has already expanded to some sites in Austria. Official geo-located address datasets already exist in Switzerland, but it is expensive and difficult to access. Crowdsourced or user generated datasets are free for the community and may be tailored to specific users needs.

Ross Purves, University of Zürich, Switzerland tackled in his talk about "Using crowd sourcing to explore the semantics of place" the paradigm between the crisp, well defined location of places in official datasets (e.g. addresses, districts, borders, etc.) and the sometimes fuzzy or vague human perspectives of "places". His research on exploring how people describe place is methodologically based on the analysis of tagged, geo-located information in databases. This approach is probably not "crowdsourcing" in the narrow sense, but does rely on user generated data.

François Golay, EPFL, Switzerland, pointed out some aspects of "The Cognitive and Social Stakes of Crowd Sourcing". First he presented a short typology of data collaboration dynamics from coexistence to fusion. A typology of data production approaches may be separated into "traditional production", "autoproduction", "crowdsourcing" and "coproduction". Each of the approaches has his own difficulties and advantages. With a focus on crowdsourcing, Golay further discussed aspects of the socio-cognitive appropriation mechanisms of collaborative data production and pointed out the consequences for practice.

Jean-Marc Viglino, IGN, France "Handling partner's feedbacks through the web" explained the needs of IGN France for a partnership with institutional partners and the public for updating geo-datasets in time and with low costs. The Ripart ("site de *r*emontée d'*i*nformation *p*artagée") feedback website of IGN (ripart.ign.fr) is one element of their programme for collecting user generated content. Viglino outlined some further aspects of the motivation and expectations of contributing crowds.

Nicolas Schmidt, a PhD student of HSR, Switzerland "QA of Crowdsourced Geographic Content" presented working ideas and had some questions to the audience in his talk. He highlighted the trust and quality aspects of some crowdsourcing methods.

Summary of day one

In his summary , André Streilein expressed his thanks to the contributing participants for giving such an insight and overview of the aspects of crowdsourcing. The presentations had shown, that national mapping agencies have a distinct need to include user generated information into their update process of their GIS data sets and maps. An active scientific community exists in this field. Some research is in progress but the coordination and focus needs to be strengthened.
A consultative poll revealed the interest of the audience in two main topics for the next day break-out sessions. A first group wanted to discuss interaction aspects with the crowd, whereas a second group was interested more in quality aspects.

Break-out sessions and presentations

After the summary of day one, the participants had discussions in break-out groups on the selected topics: "Interacting with the crowd", "Stakes of user generated contents (UGC) and quality management" and "Quality aspects of crowdsourcing". The moderated discussions in the groups were intensive and gave the participants an opportunity to bring in personal experiences, questions and conclusions. Afterwards the outcomes of the break-out sessions were presented by the moderators and discussed in the plenum.

Group 1

Break-out group one (moderator Christian Heipke, University of Hannover Germany), discussed different aspects of "**Interacting with the crowd**". The group agreed that it is a fact that there is a lot of information out in the crowds, which is worth looking at and potentially harvesting. How to identify fruitful crowds for mapping agencies and how to interact raise open questions which need more clarification and even definition. First the term "crowd sourcing" is understood as interaction with the community and not as a data mining approach on the web. The Web 2.0 is, in addition, not equal to crowd sourcing. Break-out group one identified several types of (mapping) crowds according to the effort they spend for collecting data and to the potential value (see Fig. 1).

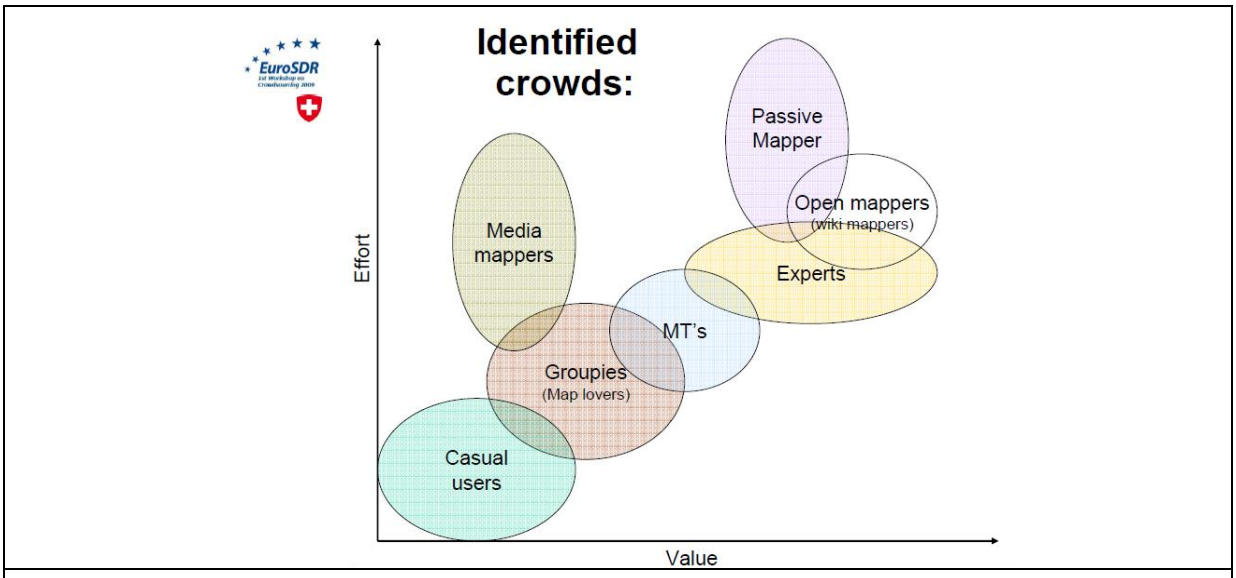


Fig. 1. Identified crowds as proposed by break-out group one (Moderator C. Heipke)

The identified crowds for

- **The "groupies"**: These are small groups of map (or any geodata) lovers which produce trustable and very valuable data of great value. The motivation is given by the public or the community specific acknowledgment and recognition. It is a key issue to get the groupies organized in community forums.
- **The "casual users"**: Casual users are often identified as hikers, bikers, mountaineers etc. This group partly overlaps with the groupies but are

distinguished by a lesser effort and therefore less valuable data production. They contribute occasionally. Moderated community platforms are necessary to group the casual users in communities of critical mass.

- **The "experts"**: Experts are active people and leading map experts in organisations like mountain rescue, fire brigades, civil protection, traffic guides, etc. They are motivated by the feeling that they may make their own life easier, they can contribute with very valuable and trustable information. But the experts have to be identified by, and committed to, the mapping agencies with (financial) agreements.
- **The "media mappers"**: Media mappers are potentially large groups, activated sporadically by regional up to international media campaigns (television, radio, internet forums, print media, etc.). These are mostly once-off mappers specially motivated by competitions, mapping parties, etc. The contributions are limited in time and extent and a big initial effort for initiation of campaigns is needed. Over the long-term it would be advisable for the mapping agencies to work to migrate the "media mappers" to "casual mappers" or even "groupies".
- **The "open mappers"**: The open mappers are small groups which spend a lot of time to contribute very valuable and large information to open source data sets or data systems (OpenStreetMap, MapShare, Google MapMaker, etc.). They are motivated by contributing (and using) good public data. The enhancement and simplification of the associated Open Source mapping tools will help to grow the "open mappers" groups.
- **The "passive mappers"**: Mobile phones which incorporate GPS positioning are a prime example of the new technologies that enable the passive collection of information about position, time and speed of individuals. The use of this information is restricted by law and / or acceptance of the individuals. The utilisation of such resources for mapping agencies is unclear and needs interpretation and intersection with existing data sets. The potential crowd is certainly very large. In specific cases an overlap to open mappers exists (for example mapped waypoints of GPS devices).
- **The "mechanical turks"**: The Amazon Mechanical Turk is a crowd sourcing marketplace and service where human beings can contribute to posed tasks for a monetary payment. The motivation of this medium sized group is money. They can be used to gather needed data for mapping agencies, but the delivered data would need careful checking.

Identification of the ideal crowds is one matter. However, to achieve a fruitful interaction with the crowds opens up further research questions on a technological and social level. Further research on quality issues such as completeness, trust, interpretability etc. of crowd sourced data as well as on data integration is needed.

Group 2

The second break-out group, moderated by F. Golay, EPFL Switzerland, was discussing the main "**Stakes of user generated contents (UGC) and quality management**". They came up with a flowchart of the mapping agency to crowd interactions (Fig. 2). The need to offer an optimised data model to the crowd, a better internal workflow adaptation to included crowd sourced data and shorter consolidation and updating time intervals were identified as elements of high priority on the side of the mapping agencies. Involvement of excellent crowd members (super-users), effective tools and interfaces and finding the trade-off between prescribed (top-down) and emerging (bottom-up) data models improves the quality and added value of user generated contents. Fostering synergies between the mapping agencies and the crowd or other communities like OpenStreetMap help to build trust in the data in the crowds and in the interactions. Open communication and discussions of legal aspects helps to find acceptable agreements between the crowds and the mapping agencies.

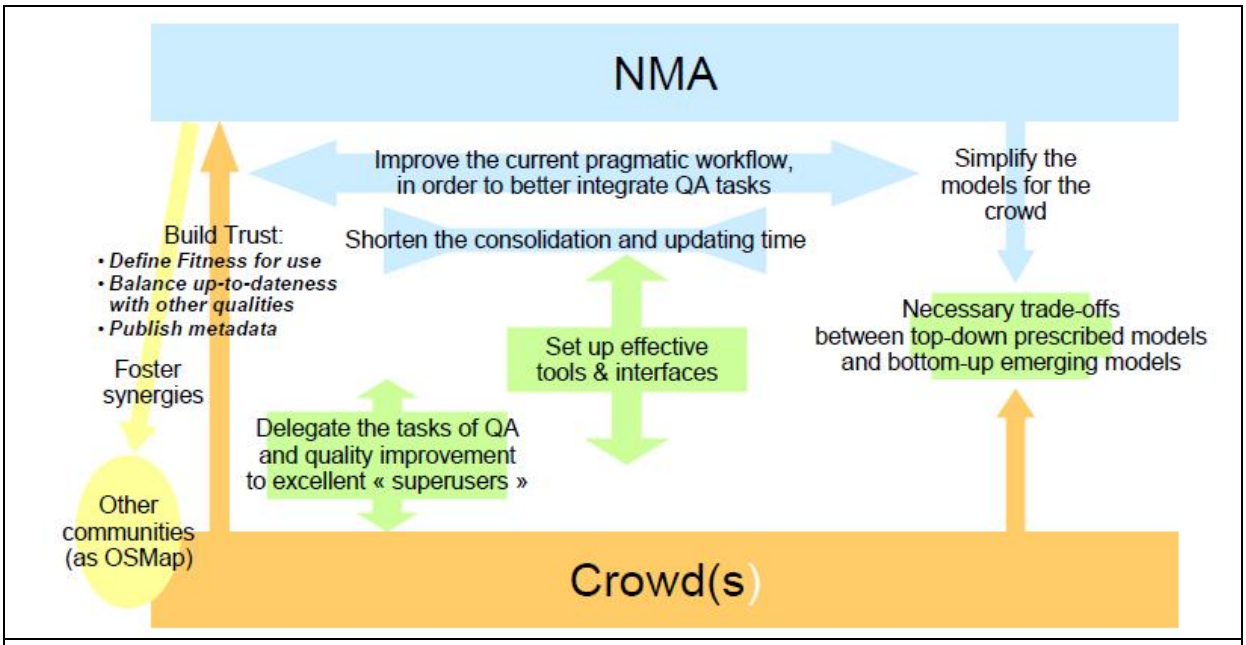


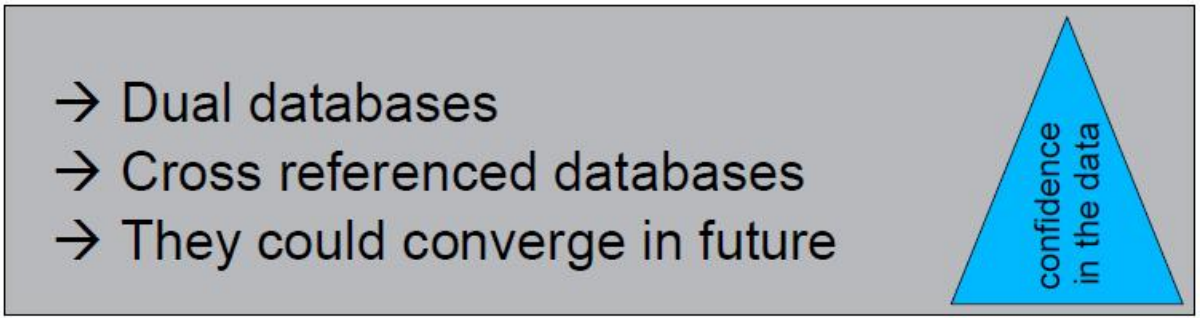
Fig. 2. Sketched summary of main stakes of user generated contents (UGC) and quality management

Group 3

The third break-out group, moderated by Mike Jackson, Centre for Geospatial Science, University of Nottingham, UK, focussed on the "**quality aspects of crowd sourcing**". Quality is seen as "fitness for purpose". Quality of crowd sourced data has multiple dimensions such as absolute and relative geometric accuracy, up-to-dateness or currency, topological correctness, the accuracy of the metadata, the legal aspects and the validation stamp by the officials (in case of national mapping agencies). The assessment of quality also needs to address the more fuzzy spatial concepts such as expressions like "the midlands", "the north", etc. The perceived quality of crowd sourced data is not solely a matter of the measurement of physical parameters.. Trusting in data is an additional dimension. Trusting in data is also

influenced by marketing elements, brand, origin ("national mapping agencies are reliable"), type of information ('we always trust in photos').
Measuring the quality of crowd sourced data is context dependent – it ideally requires a knowledge of the intended purpose that it is to be used for and the market needs. A changing market may drive the need for different quality levels:

- Low cost crowd sourced location data for specific "high-usage" areas where relative rather than absolute position is important but which also incorporates rich complementary information (metadata) which is frequently (even constantly) up-dated and validated (by officials such as mapping agencies). Expensive data for a seamless coverage of an entire country.



The Break-out group suggested that national mapping agencies should follow such a direction of dual databases, well linked, which may converge in the future. The development and growth will certainly raise awareness of the usability of the data and therefore enable trust to be a developed in the context of usage.

Conclusions

The findings of the three break-out groups were discussed in plenary in a spirited and constructive discussion. André Streilein summarised the expressed needs for a strong collaboration of the mapping agencies and the research communities in the field of user generated content, which is identified as an important and valuable input in map or geodata updating procedures, content generation and interaction with the users. There is a lack on experiences, knowledge on interaction schemes, on techniques, on legal aspects, on production process integration aspects. etc. The dialogue between crowds and mapping agencies is in most cases not established or on a very basic level. How to establish, support and motivate a community to achieve tailored user generated content is seen as one main field of investigation.
The audience agreed, that research, exchange of facts and figures as well as further workshops on crowd sourcing for national mapping agencies would be welcome and even necessary.

Next steps

In order to pave the way for further activities on the topics of crowd sourcing, the organising committee has subsequently asked the attendees of the workshop the following questions by email:

1. Does your organisation have research and development projects, which are finished, ongoing or planned, with respect to the issues of the workshop?
    a. If yes, what are the topics of these projects?

2. Is your organisation willing to actively collaborate in a small group, which investigates into the following topics:
    a. How to collaborate with the crowd?
    b. How to assure quality and reliability of the data collected?

Programme

Thursday, 20th August 2009	Friday, 21st August 2009
13:30 Welcome **13:40 Presentations (Part I)** **15:00 Coffee Break** **15:30 Presentations (Part II)** **16.50 Summary of the day** **17:20 End** **19:45 Dinner**	**8:30 Introduction to the break-out sessions** **8:40 Break-out session I, II and III** **10:00 Coffee Break** **10:30 Plenary presentations of break-out sessions I - III & discussion** **11:30 Conclusion from the workshop sessions and future steps** **12:00 End of Meeting**

Presentations (Part I)		**Thursday, 20th August 2009**
13:40	Web 2.0 technologies and Dutch Cadastre	J. Stoter, ITC Enschede
14:00	Some OS experiences of crowd-sourcing and some questions?	M. Havercroft, OS UK
14:20	Do photo sharing websites represent a sufficient database to aid in national map updating or change detection?	V. Antoniou, University College London
14:40	Integration of user generated content into national databases	J.-Chr. Guélat, swisstopo, CH
Presentations (Part II)		**Thursday, 20th August 2009**
15:30	OpenAddresses.ch - An OpenGeodata project	H.-J. Stark, University of Applied Sciences NW Switzerland
15:50	Using crowd sourcing to explore the semantics of place	R. Purves, University of Zurich, CH
16:10	The cognitive and social stakes of crowd sourcing	F. Golay / M. Noucher, EPFL Lausanne, CH
16:30	Project "Echange" and the mapshare portal RIPart of IGN France	J.-M. Viglino, IGN France
16:50	QA of Crowdsourced Geographic Content	N. Schmidt, Univ. of appl. Techn. Sciences, Rapperswil, CH

Break-out sessions		Friday, 21st August 2009
Session 1	Interacting with the crowd	Moderated by: Ch. Heipke, University of Hannover, D
Session 2	Quality aspects of CS (group 1)	Moderated by: F. Golay, EPFL, CH
Session 3	Quality aspects of CS (group 2)	Moderated by: M. Jackson, University of Nottingham, UK

Hosting group

Federal Office of Topography - swisstopo
Seftigenstrasse 264
P.O. Box
3084 Wabern
info@swisstopo.ch

André Streilein, Jesko Schaper, Tobias Kellenberger, Thomas Koenig, Jean-Christophe Guélat

The Federal Office of Topography (official name) is the Swiss National Mapping Agency since 1838. Swisstopo has about 300 employees. In the function of the Federal Geo-Information centre swisstopo produces high-quality spatial reference data and products derived thereof. It coordinates the activities for federal basis geodata and promotes their broad application. Swisstopo guarantees to provide data and fundamental bases for a monitoring of the spatial development in Switzerland.

List of participants

Last name	First Name	Affiliation	Country
Antoniou	Vyron	Dept. Civil, Environmental & Geomatic Engineering, University College London	UK
Baumgartner	Michael	MFB-Geoconsulting	CH
Buogo	Alain	Federal Office of Topography, swisstopo	CH
Dräyer Relling	Susanne	Federal Office of Topography, swisstopo	CH
Ertz	Olivier	Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud (HEIG-VD)	CH
Feurer	Christine	Federal Office of Topography, swisstopo	CH
Flandroit	Dominique	Coordination Web, Institut géographique national de la Belgique	B
Freyre	Jérôme	Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud (HEIG-VD)	CH
Gleerup	Andrea	Lantmäteriet, division Informationsförsörjning	S
Golay	Francois	GIS Research Laboratory, Ecole polytechnique fédérale de Lausanne (EPFL)	CH
Gschwend	Christian	Department of Geography, University of Zurich - Irchel	CH
Guélat	Jean-Christophe	Federal Office of Topography, swisstopo	CH
Havercroft	Malcolm	Data Collection & Management, Ordnance Survey UK	UK
Heilig	Markus	Federal Office of Topography, swisstopo	CH
Heipke	Christian	Institute of Photogrammetry and GeoInformation, Leibniz University of Hannover	D
Jackson	Mike	Centre for Geospatial Science, The University of Nottingham	UK
Käuferle	Dominik	Federal Office of Topography, swisstopo	CH

Last name	First Name	Affiliation	Country
Kellenberger	Tobias	Federal Office of Topography, swisstopo	CH
Keller	Stefan F.	Institut for Software, HSR Technical University of Applied Sciences	CH
Kjellström	Bengt	Lantmäteriet, division Informationsförsörjning	S
Koenig	Thomas	Federal Office of Topography, swisstopo	CH
Lanini	Michael	MFB-Geoconsulting	CH
Lechner	Marco	Department of Physical Geography, University of Freiburg	D
Mouton	Jean-Pierre	Coordination Web, Institut géographique national de la Belgique	B
Nackaerts	Kris	VITO - Flemish Institute for Technological Research, TAP - Centre for Remote Sensing and Earth Observation Processes	NL
Neun	Moritz	ESRI Switzerland AG	CH
Oesterlin	Nils	Federal Office of Topography, swisstopo	CH
O'Sullivan	Liam	Federal Office of Topography, swisstopo	CH
Philipona	Claude	Camptocamp SA	CH
Pla	Maria	Department of Topographic Databases, Institut Cartogràfic de Catalunya	E
Purves	Ross	Department of Geography, University of Zurich - Irchel	CH
Rappo	Daniel	Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud (HEIG-VD)	CH
Schaper	Jesko	Federal Office of Topography, swisstopo	CH
Schmidt	Nicolas	Institut for Software, HSR Technical University of Applied Sciences	CH
Schrotter	Gerhard	GIS Competence Centre, City of Zürich	CH

Last name	First Name	Affiliation	Country
Stark	Hans-Jörg	Institute of Survey and Geoinformation, University of Applied Sciences Northwestern Switzerland	CH
Stoter	Jantien	Geo-information processing / ITC Enschede and Kadaster / TU Delft	NL
Streilein	André	Federal Office of Topography, swisstopo	CH
Tomko	Martin	Department of Geography, University of Zurich - Irchel	CH
Viglino	Jean-Marc	Project "Echanges", IGN-France	F