Česká společnost uživatelů otevřených systémů EurOpen.CZ
Czech Open System Users' Group
www.europen.cz

XXIX. konference – XXIXth conference

Sborník příspěvků
Conference proceedings

Hotel Eliška Mikulov
1.–4. října 2006

Sborník příspěvků z XXIX. konference EurOpen.CZ, 1.–4. října 2006

Plzeň 2006. První vydání.
Editor: Vladimír Rudolf, Jiří Felbáb
Sazba a grafická úprava: Ing. Miloš Brejcha – Vydavatelský servis
Vytiskl: IMPROMAT CZ, spol. s r. o., U Hellady 697/4, Praha 4
provozovna Copyshop, Smetanovy sady 6, 301 37 Plzeň

Příspěvky neprošly redakční ani jazykovou úpravou.

ISBN 80-86583-11-2

Obsah

Programový výbor

Bechynský Štěpán, Microsoft Praha
Dostálek Libor, Siemens Business Services, Praha
Felbáb Jiří, mgm technology partners, Praha
Rudolf Vladimír, Západočeská univerzita v Plzni

Electronic passports

Zdeněk Říha

E-mail: zriha@fi.muni.cz

Klíčová slova: elektronic passports, biometric passports, basic access control, active authentication

Abstrakt

Elektronický pas vypadá na pohled stejně jako pas klasický – neelektronický, rozdíl je však v tom, že v elektronickém pase je umístěn bezkontaktní čip, který obsahuje údaje o držiteli pasu a vydávající instituci.

Technickou specifikaci pasů má na starost Organizace pro civilní letectví (ICAO), která vydala standardy v této oblasti. ICAO zatím nestanovila povinnost začít vydávat elektronické pasy a nechává toto rozhodnutí na vydávajících státech. Situaci v Evropské unii však upravila Evropská komise, která rozhodla, že nejpozději od 28. srpna 2006 musí státy EU začít vydávat elektronické pasy. Elektronické pasy jsou tedy realitou i v České republice.

Hlavním důvodem zavedení elektronických pasů je jejich vyšší bezpečnost. Integrita dat uložených v čipu je chráněna digitálním podpisem, proti kopiím dat se lze bránit aktivní autentizací a pro ověření, zda pas opravdu patří jeho současnému držiteli, jsou v pase uložena biometrická data (nejprve pouze fotografie, což není nijak revoluční, ale později to budou také otisky prstů).

1 Introduction

A passport is a government issued identification document proving that the holder is a citizen of a particular country; belongs under its protection and is authorized to enter foreign countries. The passport is also connected with legal protection abroad and the right of the holder to return to the state in which he holds citizenship.[10] Even if some 100 years ago no documents were required to travel to many countries, already during WW I, travel documents were widespread. Initially passports took the form of a sheet of paper, but since

the 1920s passports have looked like as know them today. Standardization in the field of passports has been the task of the League of Nations, the United Nations and the International Civil Aviation Organization – (ICAO), a UN agency founded in 1944 to facilitate safe, secure and effective civil aviation. It is clear that due to the important role of passports in modern society, standardization in this field is very important.

For many centuries international travel was not monitored. However, as a result of unwanted migration and lately also due to the threat of terrorism the inspection of people crossing borders draws more attention. The passport inspection uses a combination of 2 authentication techniques:

- Something you have (token) – the passport is a physical object that was issued by an authority and this authority specifies some properties of the holder (name, surname, nationality, diplomatic status etc.). During the border control it is important to verify that the document is an original issued by a relevant authority and that it has not been modified in an unauthorized way. For this reason the passport is resistant to easy counterfeiting, i. e. contains several protection factors, which are difficult to imitate by using commonly available technology. Protective technologies include special paper, watermark, protective iron stripe, special ink, microprint, items visible only under ultraviolet light etc. Similar techniques are used to protect banknotes. Techniques which make unauthorized modification more difficult include sealing the personal data page with a special foil/wrap and recently also a so called digital photo: the photo is instead of being glued, first scanned and then printed onto the personal data page (often twice – in a larger and smaller variant). This makes the exchange of the photo significantly more difficult. All this security techniques are becoming increasingly available to general public. Therefore passports have to employ newer and newer features. That's nothing new and relates also to banknotes. The increasing availability of security printing techniques are however one of the reasons why electronic passports are being introduced.

- Something you are (biometrics) – to avoid misuse of lost and stolen passports it is necessary to verify that the passport really belongs to the holder. Therefore the personal data page includes the photo of the holder. The match of the holder with the photo is done manually by the frontiersman who is trained and experienced in face recognition. Even so the possibility of misuse of a passport by a similar-looking person is substantial. It is well known that people are not good in recognizing races other than their own (e. g. Asians have problems recognizing Europeans, Europeans have problems recognizing Africans).

After the authentication phase during which the identity of the traveler is recognized and verified may follow an authorization phase during which the holder's right to travel is investigated. Checking the expiration date may already be part of the previous stage. Further it is checked whether the passport validity was not revoked for other reasons (e. g. accusation/charge against the holder). To verify the passport validity we can use an off-line list of revoked passports or on-line national or international databases (e.g. Interpol operates a database of lost and stolen passports).

To verify whether a passport is listed in a computer database (local or networked) of invalidated passports it is necessary to enter the passport number, holder name or other information usable for finding the record in the database. Manual transcription of data in the passport is slow and prone to errors. For this reason already in the 1980s the ICAO standardized (standard number 9303) how to store some passport data in two machine processible lines. This zone includes basic information about the passport and its holder (name, surname, passport number, expiration date etc.) and is printed in a standardized font so that a computer can read it by OCR (optical character recognition) and process it.

Because the amount of data storable in the MRZ is only very small (88 characters) and the only security factor is the check digit, new ways of storing data for automated processing were investigated. The new version of the ICAO standard 9303 from 2003 uses the technology of contactless smartcards, asymmetric cryptography and optionally also biometrics.[5] New passports equipped with RFID chips are called electronic passports. The RFID chip is usually integrated in the paper cover page of the passport (but other places are not excluded) and an electronic passport is not visibly different from an old-style passport (except for the logo on the cover page).

2 Adding the chip

The main reason to introduce electronic passports is the increased security of passports, faster reading speed and larger storage space. Several storage technologies were discussed. 2D optical/bar codes dot not provide sufficient storage capacity and are not writeable (ICAO would like to use electronic passports to store electronic visas in the future), contact smartcards only have a limited lifetime due to friction caused by contact (moreover dirt or high humidity can cause problems in some regions).

Contactless smartcards do not require contact with the reading device, they use fast communication protocols and when equipped with modern chips, have a large memory (tens of kB) as well as cryptographic coprocessors. ICAO chose the protocol ISO 14443 [3] for use in electronic passports. Such devices are able to communicate within the anticipated range of 0–10 cm from the reader. There

are two subtypes of ISO 14443 devices, so called type A or B. These two types differ in number of parameters including the communication protocol, but both of them are usable in electronic passports.

Electronic passports bring several advantages. Firstly that is higher security, which is connected with digital signatures of the stored data (and with their storage capacity that enables storage of biometric data). In the future writeable electronic passports will be able to store electronic visas and possibly other data. Another advantage can also be fast reading of passport data. This is however complicated by some protection mechanisms.

Data stored in the electronic passport must be digitally signed by the issuing institution. This is an important security factor, because even in cases when a counterfeiter has the newest technical equipment for printing and personalization of the passport at his disposal, he will not be able to create the correct digital signature of the fake data without access to the proper private key. This way of protecting the data is called passive authentication and is an obligatory part of every electronic passport. Passive authentication cannot prevent production of exact copies of data (also called cloning). To avoid such copying, additional techniques are necessary (biometrics and active authentication, see below). The PKI hierarchy is in this case reduced to a single level.[2] Every state creates its own national CA which signs the document signing authority keys; these authorities then sign data in electronic passports. To distribute certificates of signing authorities the ICAO introduces a special infrastructure. Also the CRLs have to be dealt with; they are issued by member states at maximum every 90 days. In the case of an incident (i. e. a private key leak) the CRL has to be redistributed within 48 hours (CRLs are distributed primarily bilaterally but as a secondary channel, the ICAO infrastructure can be used). It is interesting to note that a private key leak does not automatically mean invalidity of ALL documents signed by that key, but warrants increased care when inspecting such documents.

Electronic passports also bring some disadvantages.[1] These mostly relate to the use of contactless technologies for data transmission. Firstly it is possible to detect the presence of a passive RFID chip even if we do not communicate with the chip. So for example a thief can find out that there is an RFID chip in a bag and focus on that bag. Secondly even without a successful reading of the data from the chip it is possible to find out some information about the chip used. For example anticollision algorithms need to know the chip serial number before communication with the chip to e. g. for authentication. Similarly some error codes may leak the manufacturer of the chip or the chip model and therefore possibly also the issuing state. Such information can be misused by e. g. terrorists in the construction of a bomb which will be activated as soon as there is a person with a passport issued by a particular country or group of countries near the bomb. Or (based on the chip number) it is possible to track a

particular person (even if other information from the passport does not have to be available to the attacker). Both these disadvantages (i. e. determination of the existence of RFID and the chip serial number or other indirect information) can be eliminated by the use of a Faraday cage i. e. placing the chip into a metal package (e. g. aluminum cover). This way it will not be possible to detect the chip and communicate with the passport until the passport is removed from the cover (in the case of an aluminum cover until we open the booklet). Such a packaging cannot eliminate the possibility of eavesdropping once the communication is in progress (after the passport is open). If the electronic passport does not use additional protection mechanisms (and their use is not obligatory according to the ICAO standard), it is possible to read data from the passport without any authentication and the communication between the reader and the chip is not encrypted.

Due to strong fears of unobservable unauthorized reading of the data from the passport chip it was necessary to implement a way to control the access to the data. Because the basic data must be readable by border control staff of any country (including enemies), it would be really difficult to implement secret (encryption or authentication) key management in a way that the border control staff (and other authorized parties) can read the data while everybody else cannot. Therefore it was decided to implement a system, which will allow access to anybody who is able to read some data from the personal data page. Because the authentication requires the knowledge of such passport data and these data items can be only obtained after the passport is open, it is possible to expect that successful authentication will mean the reader has the passport in their physical possession (i. e. this happens with the consent of the passport holder). The passport number, birthday of the holder and the passport expiry date is SHA-1 hashed and such a hash is used to obtain two 3DES keys to authenticate and establish a common encryption key (protocol is based on ISO 11770-2), which is then used to secure subsequent communication (Secure Messaging). This way the whole communication is secured against eavesdropping. Such a mechanism for restricting access to the chip data is called Basic Access Control (BAC).

The disadvantage of the basic access control is the small entropy in the data used to authenticate. Although the theoretical maximum is about 56 bits (birthday from the range of max. 100 years i. e. about 15 bits, expiration date max. 10 years, i. e. about 11 bits, 9-digit passport number about 30 bits) or even 73 bits if the document number is alphanumeric, not all the values are equally probable and thanks to additional knowledge it is possible to execute an attack in a more efficient way than just by trying all possible values. In particular the knowledge of the passport number range can make the attack much faster. This has been done and published for Dutch passports.[8] With the knowledge of the numbering scheme and by restricting the ranges of other values the entropy drops to about 35 bits. This is still too much to execute

an on-line attack against the passport (i. e. trying all the possible values in a real communication with the passport), but if we can eavesdrop a successful communication then we can do an off-line attack when we get the encryption key used to secure the communication and so decrypt the transmitted data. To make such attacks more difficult it would be possible to introduce a more random passport numbering scheme.

Another disadvantage of basic access control is the need to access the authentication data (i. e. scanning and recognizing the MRZ or typing the data manually) before we can start reading the data from the chip. This significantly reduces the overall speed of obtaining the data from the passport. It is also necessary to realize that by using basic access control anyone who has physical access to the passport (e. g. a hotel receptionist) can read the data from the passport. In the case where the chip includes only data visible on the personal data page it does not change the situation, but once some other data (e. g. a fingerprint) is stored in the passport without additional protection, the subject could get access to that new information.

The use of digital signatures for data integrity does not imply that the attacker cannot read all the data including the digital signatures from the chip and create another chip into which this data is loaded. Securing the passport by traditional means (printing technologies) and the check that the data stored on the chip correspond with data in the MRZ still plays an important role. Even this check does not prevent making copies of whole passports including all printed data. Against such attacks the electronic passports can be equipped with a technology that is called active authentication. In the passport chip an asymmetric private key is stored. Such a key never leaves the chip (there is no command to read the key), the reader can only verify whether the chip has access to the private key. The public key of the chip (including the digital signature made by the issuing authority) is part of the readable data. The reader reads the key and then by using a challenge-response protocol (concretely the reader sends a random number, the chip adds another random number and digitally signs it) verifies if the chip has got access to the private key corresponding with the public key. The counterfeiter cannot therefore make a complete copy of the chip, because the private key cannot be read from the original chip. He also cannot create another pair of keys because the public key must be digitally signed by the issuing authority (verification of the public key digital signature is therefore an important part of active authentication). The impossibility of reading the private key from the chip is based on the assumption of the chip's tamper resistance. If the chip is not sufficiently tamper resistant and the private key can be obtained from the chip then it will be possible to create exact copies of the chip and even active authentication cannot reveal that the chip is a copy.

Active authentication can be misused in an interesting way against the privacy of the holder. The random numbers which the reader sends to the passport

to sign (the passport chip adds another random part) do not have to be completely random. In cases where the random value has some semantics (e. g. day‖time‖place) it may be possible to misuse the returned signed number as a proof that the holder was at a particular place at a specific time. This certainly cannot be used as evidence in court as the passport signs anything, anytime Even so it is considered a serious disadvantage, thanks to which some countries (e. g. Germany) do not implement active authentication at all.

The ICAO organization has not specified any obligation of member countries to introduce electronic passports. The EU has regulatory power over member states and requires the introduction of electronic passports by 28. August 2006. Commission decision C(2005) 409 allows the use of both variants of the ISO 14443 standard (A and B) in accordance with the ICAO standard. Also active authentication is left as optional. The decision further specifies that Basic Access Control is obligatory for electronic passports of member states.

It is also necessary to handle the case when the chip in the electronic passport is not functioning at all. Some attempts to intentionally destroy the passport chip have already appeared (e. g. the RFID Zapper [7] or the use of a classical microwave oven).

3 Adding biometrics

Electronic passports enable the storage of several kilobytes of data. Such storage space is sufficient to store also biometric data. Storing biometric data in identification documents is not completely new; even current passports include the signature and the picture of the holder. Before the widespread use of photographic techniques (i. e. a long time ago) passports included a textual description of the holder and in some periods certain identification documents also included fingerprints. What is new is the possibility of automated authentication by using biometric technologies and thanks to digital signatures the linkage of biometric data to other data in the passport.

In accordance with the ICAO standard the chip has to include a facial image of the holder. Other biometric characteristics (possible alternatives are fingerprints or iris scans) are only optional and the decision whether to store them on the chip or not is left up to the issuing state. Member states of the EU have to start issuing biometric passports with the face of the holder by 28 August 2006 and with fingerprints by 28 June 2009.

The possibility of biometric verification is an important security factor for electronic passports accompanied by biometric data (so called biometric passports). Even if the picture of the holder can also be verified manually, automated biometric verification is more accurate and can be done without the presence of border control staff. The major disadvantage of biometric verification based on

the face matching is its high error rate (even so, automated verification is usually more accurate manual verification). In the case of controlled light conditions, the error rate (in terms of rejecting authorized holders – FRR) can reach about 10 % (and the probability of an unauthorized acceptance – FAR of 1 %), in cases where the light conditions cannot be optimized for the biometric system the error rate can even reach 50 %.[4] It is clear that with such error rates not every rejected person can be investigated thoroughly. The use of biometrics is most beneficial in the case of fingerprints or iris scans. The accuracy achievable with such biometric systems is higher by an order of magnitude (FRR around 0,5 % at FAR of 0,1 % for fingerprints).[4]

4 Data structure

The data structure on the chip uses a filesystem defined by ISO 7816-4, where the directories are called dedicated files (DF) and files are so called elementary files (EF). The data is stored in several files in a common directory.[6] One file (EF.COM) is reserved for metadata (data format version and the list of stored data groups), one file (EF.SOD) contains information about security (digital signatures of hashes of all the files) and other files include the data, which is grouped into several data groups (DG). The list of data groups is shown in the following table.

Data group	**Stored data**
DG1	Machine readable zone (MRZ)
DG2	Biometric data: face
DG3	Biometric data: fingerprints
DG4	Biometric data: iris
DG5	Picture of the holder as printed in the passport
DG6	Reserved for future use
DG7	Signature of the holder as printed in the passport
DG8	Encoded security features – data features
DG9	Encoded security features – structure features
DG10	Encoded security features – substance features
DG11	Additional personal details (address, phone)
DG12	Additional document details (issue date, issued by)
DG13	Optional data (anything)
DG14	Reserved for future use (public key for the EAC)
DG15	Active Authentication public key info
DG16	Address of relatives to be notified

The ICAO standard supports storage of biometric data in the form of facial image including encoded facial geometry (data group DG2), encoded fingerprints (DG3), encoded iris scans (DG4) and further supports storage of the picture of the holder as printed in the passport (DG5) and the signature of the holder (DG7). Biometric data (DG2-4) are bundled with metadata as in the case of CBEFF (Common Biometric Exchange Format Framework, see ISO 19785), whilst to store biometric template it is possible to use ISO 19794-5 for the face (the picture must be a frontal facial image with neutral expression), ISO 19794-2, 19794-3, 19794-4, 19794-8 for fingerprints and ISO 19794-6 for iris scans. Digitized data printed in the passport (DG5 a DG7) are in the case of face and signature data stored in accordance with ISO 10918.

Commission decision C(2005) 409 further specifies that the facial image must be stored as an image file in the JPEG or JPEG2000 (recommended) format. For fingerprints, the WSQ (lossy compression algorithm optimized for fingerprints) compression algorithm must be used and the fingerprints must be stored as compressed images in accordance with ISO 19794-4. For accessing the fingerprints Extended Access Control must be used (details of key management are not specified by the decision, they are still under discussion).

In the future the DG17 will be used for automated border clearance, DG18 for electronic visas and DG19 for travel records. Currently the format of these data groups is not standardized. Also DGs 8–10 specifying the security features of the passport booklet (secure printing etc.) are to be defined yet.

5 Conclusions

The main potential advantage of electronic (or biometric) passports is increased security. This will only be effective in longer term as it will take some time to equip borders with the necessary verification technology. Moreover it is expected that counterfeiters will focus on older passports whose validity has not been restricted by the introduction of electronic passports and on passports of countries which have not yet introduced electronic passports. It is also probable that it will not be possible to verify every person crossing the border biometrically, but only in random (or otherwise selected) cases/flights etc. In any case electronic passports make counterfeiting more difficult and biometric passports also make the misuse of lost/stolen passports more difficult, which is good news not only for governmental institutions, but also for every passport holder.

The disadvantages of electronic passports include above all the high cost of introducing relevant technologies (of issuing passports as well as to verifying passports and their holders) and problems related to possible privacy violations (remote readability of data, storage of biometric data).

6 Notes

This paper is based on [9].

„The views expressed are purely those of the writer and may not in any circumstances be regarded as stating an official position of the European Commission."

References

[1] Juels, A., Molnar, D., Wagner, D. *Security and Privacy Issues in E-passports.*

[2] ICAO, MRTD *PKI for Machine Readable Travel Documents offering ICC Read-Only Access.*

[3] ICAO *ICAO 9303 specification; including Supplement 9303, 2005-4 V3.0.*

[4] Ezovski, G. M. *Biometric Passports: Policy for International and Domestic Deployment.* Journal of Engineering and Public Policy. vol. 9, 2005.

[5] ICAO TAG MRTD/NTWG *Biometrics Deployment of Machine Readable Travel Documents, version 2.0, including annexes A-J.* http://www.icao.int/mrtd/download/documents/

[6] ICAO *Development of a Logical Data Structure LDS for Optional Capacity Expansion Technologies, V 1.7.* http://www.icao.int/mrtd/download/documents/

[7] MiniMe (pseudonym), Mahajivana (pseudonym) *RFID-Zapper.* http://events.ccc.de/congress/2005/wiki/RFID-Zapper(EN)

[8] Robroch, H. *Privacy issues with new digital passport, riscure.com.* http://www.riscure.com/news/passport.html

[9] Říha, Z., Vakalis, I. *„Elektronické pasy", Data Security Management.* volume X, number 3, 2006. ISSN 1211-8737.

[10] Wikipedia *Passport.* http://en.wikipedia.org/wiki/Passport

Elektronické pasy – jak fungují

Luděk Rašek

E-mail: ludek.rasek@r73.info

1 Důvody zavedení

1.1 EU

V reakci na rostoucí požadavky na identifikaci osob cestujících po světě a zejména z důvodu, že tato opatření požadují Spojené státy americké pro zachování bezvízového styku rozhodly orgány Evropské unie o vydávání pasů s biometrií ve členských státech. V níže uvedených rozhodnutích jsou odkazy na dokumenty Mezinárodní organizaci pro civilní letectví (ICAO), které definuje technologické vlastnosti cestovních dokladů. Více viz odstavec 1.3. Zejména následující dokumenty se věnují zavádění biometrických pasů: [EU Bio] a [EU Bio Tech]

1.2 Česká legislativa

Do české legislativy níže uvedené zákony a vyhlášky zavádějí pojem „cestovní doklad s biometrickými prvky“. Tyto zákonné úpravy byly zavedeny z důvodu harmonizace českého práva s právem EU (viz 1.1. V české republice se vydávají následující druhy cestovních dokladů a pro všechny platí, že od 1. 9. 2006 budou vydávány nové druhy dokladů již s čipem jako nosičem biometrických údajů:

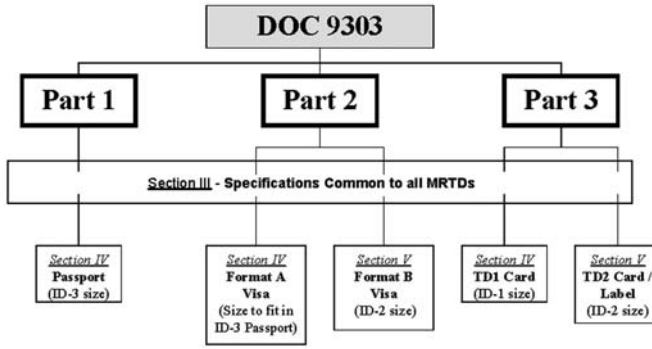
- cestovní doklad občana české republiky – vydávány obcemi s rozšířenou působností
  - pro občany 5 až 15 let věku jsou vydávány doklady s platností 5 let
  - pro občany starší 15 let jsou vydávány doklady s platností 10 let
- uprchlické pasy – jsou vydávány cizineckou a pohraniční policií osobám, které v ČR získali statut uprchlíka
- cizinecké pasy – jsou vydávány cizineckou a pohraniční policií osobám, které mají v ČR povolení k trvalému pobytu a doloží, že si nemohou pořídit pas dle svého občanství nezávisle na své vůli

- diplomatické pasy – jsou vydávány MZV představitelům českého státu (prezident, člen vlády, poslanec, senátor, soudce Ústavního soudu, předseda Nejvyššího soudu, předseda Nejvyššího správního soudu, prezident Nejvyššího kontrolního úřadu, diplomat, manžel dříve uvedených a další dle rozhodnutí MZV)

- služební pasy – jsou vydávány MZV osobám vyjmenovaným v zákoně, které cestují do zahraničí ve věcech České republiky a to na dobu pobytu v zahraničí

Pro úplnost je třeba dodat, že i nadále budou vydávány cestovní doklady bez biometrických údajů v případě rychlého vydání pasu (tzv. blesk) a pro děti mladší 5 let.

Vydávání cestovních dokladů se řídí zejména následujícími zákony a vyhláškami 325/1999 Sb, 326/1999 Sb, 329/1999 Sb a 642/2004 Sb.



Obr. 1 Struktura dokumentů ICAO Doc9303 od cestovních dokladech

1.3 ICAO

Mezinárodní organizace pro civilní letectví je organizací při OSN, která koordinuje způsob, jakým je provozována civilní letecká doprava na světě. Jedna z částí této organizace je vyčleněna pro definování jednotných vzorů pro doklady prokazující totožnost osob. V celé řadě specifikací vydávaných touto skupinou v rámci ICAO byly postupně definovány vlastnosti pasů, ID karet (rozměry, obsah, formát strojově čitelné zóny apod.) V současné době se skupina zabývá také specifikací požadavků na pasy nesoucí biometrické údaje. Pokud je to možné, nejsou touto skupinou vytvářeny zcela nové technologické specifikace, ale jsou využívány a referencovány již existující zdroje a specifikace, které prošly některým z mezinárodních normalizačních úřadů (nejčastěji ISO). Jen tam, kde existující normy nestačí, jsou navržena nová řešení (struktury dat ale s využitím ASN.1

apod.). Jako nosič biometrických údajů v pasu byl zvolen bezkontaktní čip odpovídající specifikacím ISO 7816 a ISO 14443. Některé dokumenty jsou publikovány na stránkách ICAO a valná většina tohoto článku bude výkladem jednotlivých technických specifikací. V některých případech jsou referencovány dokument tav. Doc. 9303, což je sada předpisů a specifikací popisujících veškeré vlastnosti ID dokladů (vč. pasů). Sada Doc9303 je dostupná jako placená dokumentace.

2 Využité technologie

2.1 Pasová knížka

Pasová knížka pro e-pasy je tvořena stejně jako pasová knížka dosud používaných pasů (tedy pasů se strojově čitelnou zónou – MRZ) ve formátu ID3 (ISO7810) o velikosti 125×88 mm (B7). Ve světě je používáno několik různých konstrukcí pasů s biometrickým údaji. Konstrukcí pasu rozumíme umístění bezkontaktního čipu v pasové knížce. Bezkontaktní čip vyžaduje speciální konstrukci některé ze stran pasu, která musí zakrýt a ochránit anténu, s jejímž využitím pas komunikuje. Používány jsou zejména následující konstrukce:

- čip v deskách pasu
- čip ve speciální papírové stránce uvnitř pasu
- čip v polykarbonátové stránce, na které je rovněž vytištěna datová stránka (osobní údaje, fotografie, MRZ)

Způsob vložení čipu do pasové knížky je jedním z důležitých ochranných prvků pasu, protože by měl znemožnit např. výměnu čipu mezi pasy nebo alespoň usnadnit detekci takovéhoto pokusu.

V českých pasech je využito technologie polykarbonátové stránky obsahující čip s laserovou personalizací osobních údajů do vnitřní světlocitlivé vrstvy s celou řadou dalších dodatečných ochranných prvků vytvořených s využitím technologií ceninového tisku a dalších moderních technologií.

2.2 Datová stránka a MRZ

V pasu je umístěna tzv. datová strana, na které jsou vytištěny veškeré základní údaje o držiteli pasu, včetně jeho fotografie a volitelně vzoru podpisu.

Údaje na datové stránce jsou definovány v Doc9303. Je přesně specifikováno pořadí a význam jednotlivých prvků na stránce. Ve spodní části strany je vytištěna strojově čitelná zóna vhodná pro snímání pomocí OCR.

Jednotlivé položky uvedené v MRZ jsou povinné, mají různou délku a definovaný formát. Jsou uvedeny v tabulce 1.

Tabulka 1 Příklad výpočtu MRZ

Číslo položky	Význam	Hodnota z příkladu
01	typ dokumentu	P
02	vydavatelský stát	CZE (Česká republika)
03	jméno držitele	EMILIE NOVAKOVA (Emílie Nováková)
04	číslo pasu – 9 nejvýznamnějších číslic; pokud je číslo pasu delší, než 9 číslic, je zbytek čísla pasu uložen v datové položce č. 12, kratší číslo je doplněno <	0000000017< (00000017)
05	kontrolní číslice k číslu dokladu; pokud je číslo dokladu delší než 9 číslic, bude zde výplňový znak < a kontrolní číslice bude v datové položce č. 12	8
06	občanství	CZE (Česká republika)
07	datum narození	680512 (12. 5. 1968)
08	kontrolní číslice data narození	6
09	pohlaví	F (žena)
10	datum platnosti	160613 (13. 06. 2016)
11	kontrolní číslice data platnosti	3
12	volitelné pole – pokud je číslo pasu delší než 9 číslic, jsou zde uvedeny nejméně významné číslice čísla pasu a kontrolní číslice čísla pasu oddělená výplňovým znakem <	6855121515<<<<
13	kontrolní číslice pole 12	9
14	celková kontrolní číslice (viz níže)	4

Strojově čitelná zóna je umístěna relativně k levému dolnímu rohu datové strany dokladu. Zóna je vytištěna fontem OCR-B o velikosti Size1 (cca 14 bodů). Více informací o fontu lze nalézt ve specifikaci ISO Standard 1073-2:1976.

MRZ na pasu se skládá ze dvou řádků po 44 znacích. Neobsahuje mezery, jsou nahrazeny tzv. výplňovým znakem (filler character) ‚<‘. Pokud je potřeba, jsou položky doplněny znakem ‚<‘ do požadované délky. MRZ obsahuje pouze ASCII znaky a je definován způsob transkripce znaků v ASCII nepřítomných (např. Ä→AE). Dle následujícího schématu popíšeme kódování:

Obr. 2 Příklad kompletní datové stránky

Ptiiinnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn ######### CbbbYYMMDDCsyymmddCpppppppppppppCX

- P – pro pasy je vyhrazeno písmeno P
- t – typ pasu přiřazený vydavatelem o délce jednoho znaku – např: S – služební, D – diplomatický apod.; pro běžný pas nebo pokud není využito nahrazuje se ‚<‘
- iii – kód země vydávající pas – pro ČR je to CZE (kódy vycházející z tabulky ISO 3166 3-písmenných kódů zemí rozšířené např. o příznak uprchlíka bez příslušnosti: XXA a další z Doc9303)
- nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn – jméno držitele v pořadí – příjmení <<jméno<další<jména – vše velkými písmeny, doplněné zprava ‚<‘ do délky 39 znaků; pokud jméno obsahuje doplňky typu Jr, II apod, kodujíé se do příjmení a oddělují se ‚<‘ ######### – číslo pasu, pokud číslo pasu obsahuje jiné znaky než číslice a písmena, jsou nahrazeny ‚<‘
- C – kontrolní číslice, nebo ‚<‘ pokud je číslo pasu delší než 9 číslic, zbytek čísla je pak v poli pppppppppppppp
- bbb – státní příslušnost – pro ČR – CZE
- YYMMDD – datum narození
- s – pohlaví
- yymmdd – datum ukončení platnosti pasu

- C – kontrolní číslice data ukončení platnosti pasu
- ppppppppppppppp – pomocné pole složené zbytku čísla pasu, kontrolní číslice tohoto čísla a ukončeného ‚<‘ zbytek do délky 14 znaků je možno vyplnit čímkoli; je zde např. rodné číslo
- C – kontrolní číslo z ppppppppppppppp
- X – složená kontrolní číslice vypočtená ze znaků: 1–10, 14–20, 22–43 druhé řádky MRZ

Nástroj pro generování MRZ pro pas lze najít na [MRZ Calc].

2.2.1 Postup výpočtu kontrolní číslice

Každý znak se nahradí číselnou hodnotou – číslice mají svou hodnotu, písmena se převedou dle tabulky 2 (která mimochodem obsahuje výčet povolených znaků MRZ).

Tabulka 2 Tabulka pro výpočet kontrolní číslice

<	A	B	C	D	E	F	G	H	I	J	K	L	M
0	10	11	12	13	14	15	16	17	18	19	20	21	22

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
23	24	25	26	27	28	29	30	31	32	33	34	35

MRZ se tedy převede na řadu číslic a poté vypočteme vážený součet této řady. Váhy, které se používají se cyklicky berou z této řady (7, 3, 1, 7, 3, 1, 7, 3, 1 atd).

Vypočtený součet vydělíme 10 a zbytek po tomto dělení je kontrolní číslice. V tabulce 3 je uveden krátký příklad.

Tabulka 3 Příklad výpočtu kontrolního čísla (viz výše uvedený příklad MRZ)

Vstup	1	6	0	9	0	5	–	
Hodnota	1	6	0	9	0	5	–	
	*	*	*	*	*	*	–	
Váha	7	3	1	7	3	1	–	
	7+	18+	0+	63+	0+	5=	93÷ 10=	3

2.3 RFID

Pracovní skupina pro cestovní doklady nové generace (NTWG) rozhodla, že nosičem biometrických údajů bude čip a jeho rozhraní bude bezdrátové radiové. Vybrána byla existující specifikace bezkontaktního rozhraní ISO/IEC 14443. Samotný čip pak musí vyhovovat specifikacím pro čipové karty z řady ISO/IEC 7816 s využitím komunikačního protokolu T=CL (contactless). Další varianty, které přicházely v úvahu skupina vyhodnotila jako nevhodné (např. 2D čárový kód pro malou kapacitu, holorafickou paměť pro nezralost technologie apod.). Popis využití bezkontaktních čipů v e-pasech jsou popsány v [ICAO Bio Annex I].

Vybraná technologie je v současné době známa také pod názvem RFID. RFID zahrnuje celou řadu nejrůznějších zařízení, která dokáží komunikovat bezdrátově na bázi technologie popsané v ISO/IEC 14443. Čipy zpřístupňované tímto protokolem zahrnují celou škálu od jednoduchých paměťových čipů až po plnohodnotné čipy mikroprocesorové vybavené koprocesory pro kryptografii (symetrickou i asymetrickou).

Právě díky využití technologie RFID vzniká celá řada nedorozumění týkajících se tzv. čipové totality a velkého bratra, kdy je možno využívat bezkontaktní čipy ke sledování pohybu označeného předmětu a v případě pasu konkrétního člověka. Mezi RFID čipem použitým jako značkovače zboží v Tesku a mezi čipem použitým v e-pasu je rozdíl zhruba odpovídající rozdílu mezi flashdiskem a PDA.

2.3.1 Hardware

Je až s podivem, co v všechno dokáží výrobci vměstnat na jeden mikroprocesor, který je nutno nejen ovládat, ale hlavně napájet modulovaný elektromagnetickým polem. Existuje několik výrobců, kteří dokáží vyrobit čipy s požadovanými vlastnostmi. Čipy do e-pasu vyrábějí například následující výrobci: Philips (SmartMX P5CT072), Infineon (SLE66CLX641P), Toshiba (TLCS900), Sharp, Atmel (AT90SC12872RCFT), Samsung (S3CC9GCX), STMicroelectronics.

Pro využití v e-pasu specifikuje ICAO následující požadavky:

- kompatibilita s ISO 14443 s modulací typ A nebo B
- kompatibilita s ISO 7816-4 a vyšší pro práci s daty na kartě
- dostatečná paměť pro uložení požadovaných dat (obličej, otisk)
- podpora kryptografie na čipu pro zabezpečení dat (RSA, EC)
- volitelně implementace BAC

2.3.2 Operační systém

Operační systémy využívané pro implementaci požadavků ICAO v e-pasech pokrývají širokou škálu od jednoúčelových OS specializovaných na e-pasy po JavaCard applety. OS pro e-pasy dodávají např. následující dodavatelé: SC2 (Apollo OS), IBM(JCOP), Axalto (Axseal), T-Systems (TCOS), G&D (Starcos 3.1PE), Oberthur (ID One e-pass), GemPlus (GemBorder) a další.

2.3.3 Vlastnosti pro ochranu soukromí – náhodné UID

Jednou z největších obav odpůrců e-pasů je potenciální možnost, že držitel e-pasu bude bez svého vědomí identifikovatelný. Tedy že největší výhoda tzv. RFID tagů (náhrada čárového kódu schopná identifikovat konkrétní kus zboží) se promění ve slabinu ohrožující soukromí držitele e-pasu.

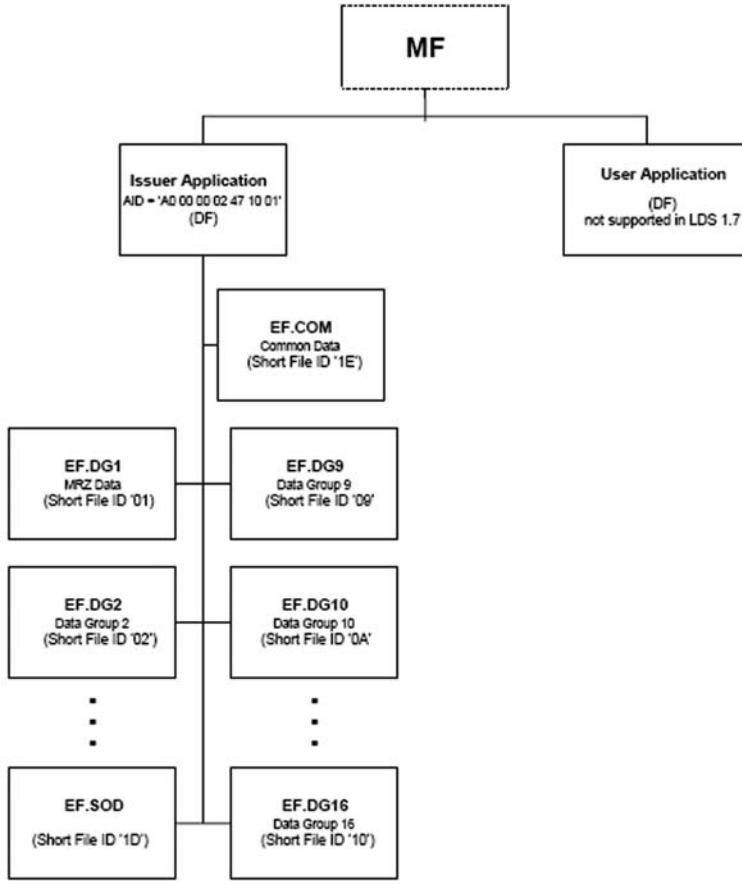
Identifikovatelnost čipu umístěného v e-pasu vyplývá ze specifikací normy ISO 14443, kde je definováno tzv. UID (universal identifier), které se používá v rámci tzv. antikolizního mechanismu pro výběr čipu pro případ, kdy je v elektromagnetickém poli čtecího zařízení více čipů schopných komunikovat. Když čtecí zařízení zahajuje komunikaci s čipy ve svém okolí, vyšle příkaz REQ. V případě, že čip zaznamená tento požadavek, v definovaných časových okamžicích dle hodnoty jednotlivých bitů svého UID vysílá či mlčí a zároveň naslouchá vysílání ostatních. Takto je možno mezi „naslouchajícími“ čipy zvolit jeden s nejvyšším UID (z hlediska mechanizmu volby). Pokud to není ten správný, čtečka mu jako další odešle příkaz HALT, který způsobí, že čip již v rámci „volby“ neodpovídá a je možné zvolit čip s dalším UID dle velikosti.

Jak je vidět, je využití UID z hlediska funkcionality čipu nutné a toto UID musí být konstantní po dobu pobytu čipu v el. mag. poli. Tohoto požadavku využívají dodavatelé čipů pro elektronické pasy a implementují tzv. náhodné UID, kdy čip v pasu drží UID právě jen po dobu, kdy je přítomen v elektromagnetickém poli jednoho zařízení a při každém vstupu do pole (= zapnutí napájení) generuje nové UID pomocí pseudo random generátoru.

2.4 LDS

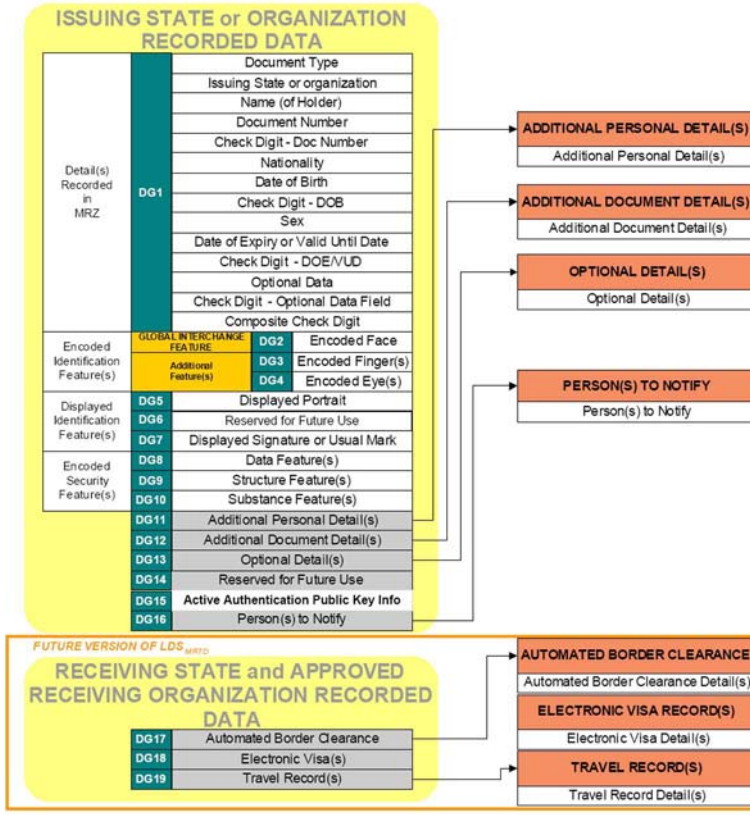
Pro ukládání dat do pasu (na čip) se děje dle standardů z rodiny ISO/IEC 7816, která společně s ISO 14443 a ICAO dokumenty (viz výše) detailně specifikuje rozhraní elektronické části pasu a datové struktury pro e-pas používané.

LDS definuje identifikaci aplikace e-pasu na čipu pomocí RID (Registered Application Identifier) a strukturu souborů v rámci této aplikace. Data jsou obsažena v tzv. Data Groups Definovány jsou datové skupiny (viz obr. 4:



Obr. 3

Skupiny DG1(kopie MRZ) a DG2(obraz obličeje) jsou povinné. Od 28. 6. 2009 budou v pasech vydávaných členskými státy EU (s výjimkou GB, IRL) povinné rovněž otisky prstů (DG3). Pro pasy nesoucí daktyloskopické údaje pak rozhodnutí EU požaduje využití tzv. Extended Access Control (EAC). V rámci EAC je definován mechanismus tzv. chip authentication (viz níže). Pro účely tohoto mechanismu byla ICOA vyhrazena skupina DG14. Datové skupiny jsou mapovány na elementární soubory (EF) v rámci struktury karty. Uvnitř souborů jsou pak data uložena s využitím kódování TLV (tag–length–value), které je definováno v ISO 7816-4 a -6 a vychází z ASN.1 DER kódování. Hodnoty jednotlivých tagů, jak za sebou následují a v jaké struktuře definuje tzv. šablona. Například data, která nesou biometrickou fotografii držitele pasu jsou v EF.DG2 uložena takto:



Obr. 4

```
‘75‘ ‘82319C‘
   ‘7F61‘ ‘823197‘
       ‘02‘ ‘01‘ ‘01‘
       ‘7F60‘ ‘82318F‘
          ‘A1‘ ‘26‘
              ‘80‘ ‘02‘ ‘0100‘
              ‘81‘ ‘01‘ ‘02‘
       ‘83‘ ‘07‘ ‘20020315133000‘
              ‘84‘ ‘08‘ ‘2002040120070331‘
              ‘86‘ ‘02‘ ‘0001‘
              ‘87‘ ‘02‘ ‘000A‘
              ‘88‘ ‘02‘ ‘0004‘
          ‘5F2E‘ ‘823162‘ ‘... 12642 bajtů biometrických dat
                                    ve formátu CBEFF
```

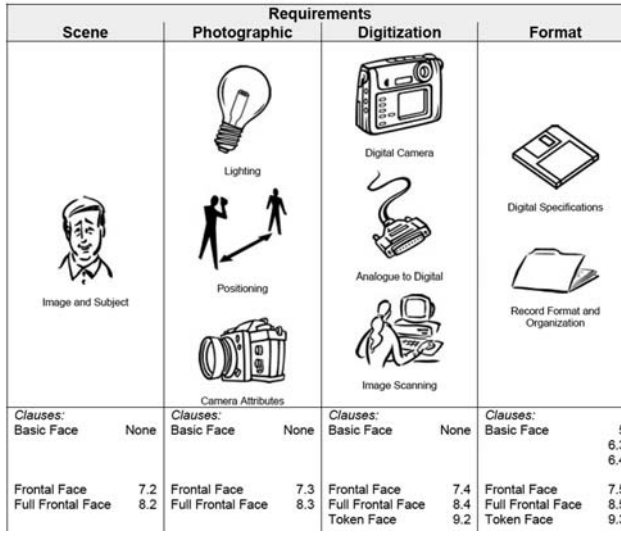
Pro všechny skupiny jsou v dokumentu [ICAO Bio LDS] definovány šablony. Dle těchto šablon jsou pad data formátována a jediným stupněm volnosti je možnost vynechat volitelné položky.

2.5 Biometrie

V elektronických cestovních pasech je v současné době využíváno biometrického obrazu obličeje a od 28. 6. 2009 budou pasy členských zemí EU vybaveny navíc ještě otisky prstů. Specifikace ICAO se v případě biometrie odkazuje na mezinárodní normy z rodiny ISO/IEC JTC 1/SC 37.

2.5.1 Obraz obličeje

ICAO pro biometrický obraz obličeje přebírá specifikaci [ICAO Bio Annex D]. V této normě jsou definovány požadavky na obraz obličeje, který je vhodným vstupem pro algoritmy rozpoznávání obličeje. Norma specifikuje několik druhů obrazů obličeje od základního obrazu obličeje (basic face image type), přes frontální obraz obličeje (frontal face image type), plný frontální obraz obličeje (full frontal image type) až po tzv. token image (token face image type) určený pro uložení do biometrického tokenu. Norma postupně zpřesňuje požadavky na jednotlivé typy obrazů. Každý následující vychází z předchozího.

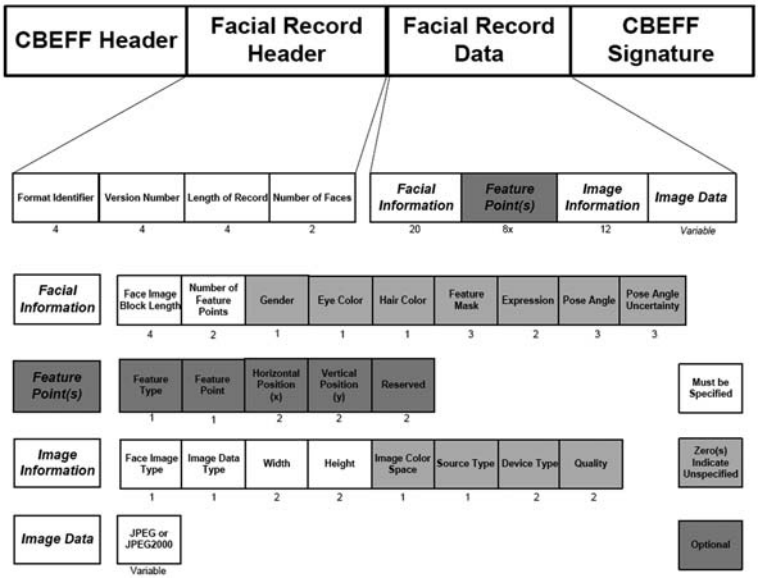


Obr. 5

Na obrázku 5 jsou uvedeny jednotlivé požadavky na výsledný obraz – požadavky na scénu, požadavky fotografické, požadavky na digitalizace a požadavky

na formát dat. Např. na základní typ obrazu nejsou kladeny žádné požadavky krom výsledného datového formátu. Token image naopak musím splňovat omezení konkrétního nastavení scény, pozice fotografovaného, nastavení přístroje a osvětlení, způsob digitalizace (umístění očí) a také výsledného datového formátu.

Výsledné obrazy jsou ukládány do obálky CBEFF, která pro obraz obličeje stanovuje dodatečné typy metadat. Kompletní výčet a uspořádání metadat v CBEFF je uveden na obrázku níže. Metadata obsahují např. informaci o pohlaví, barvě očí, barvě vlasů, výrazu obličeje a v neposlední řadě pak již nalezené tzv. feature points, což jsou pozice konkrétních bodů obličeje (oči, nos, ústa, brada apod.)



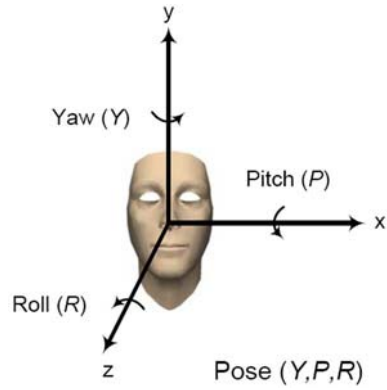
Obr. 6

Pozice obličeje na fotografii definuje pozice (pose) pomocí rotací ve třech osách dle obrázku 7:

Příklad nalezených MPEG4 feature point je uveden na obrázku 8 (jde pouze o některé body).

Požadavky na frontální obraz (frontal image) jsou následující (příkladem je obrázek s vyznačenými MPG4 body:

- požadavky na scénu
  - pozice – žádný z měřených úhlů nesmí být od střední polohy odchýlen více jak o 5 stupňů v absolutní hodnotě



Obr. 7

Obr. 8

– na fotografii nesmí být další obličej
– fotografovaný musí stát čelem ke kameře bez natáčení hlavy
– osvětlení by mělo být rovnoměrné – žádně směrové světlo
– na obličeji by neměly být žádné stíny
– oční důlky by neměly být zastíněné

  - je třeba zamezit vytváření tzv. hot-spot – přesvětlených míst
  - brýle nesmí mít silné obroučky a obroučky nesmí zakrývat oči, na brýlích by neměly vznikat odlesky, jiné než čiré brýle jsou povoleny jen ze zdravotních důvodů
  - páska přes oko smí být použita pouze ze zdravotních důvodů a musí být popsána v metadatech

- fotografické požadavky
  - správná saturace (ani pře- ani podexponováno)
  - fotografie by měla být ostrá od nosu po uši a od brady po temeno hlavy
  - barevné vyvážení obrazu by mělo být neutrální (důležité je nastavit dobře bílou), červené oči nejsou akceptovatelné
  - nemělo by docházet ke zkreslení vlivem objektivu pozorovatelnému lidským okem (zkreslení objektivu typu rybí oko)

- požadavky na digitalizaci obrazu
  - poměr velikosti pixelů musí být 1 : 1
  - souřadnice v obrázku musí mít počátek v levém horním rohu a rostou směrem doprava a dolů do kladných hodnot

- barevný profil
  - hustota šedé – v oblasti obličeje musí být dynamický rozsah minimálně 128 stupňů šedé bez přepálení
  - barevný prostor – RGS, YUV, grayscale, pomocí barevného profilu zařízení musí být generován normalizovaný obrázek např. sRGB

- obrázek nesmí být prokládaný, ani nesmí vzniknout z prokládaného obrázku pomocí filtrování

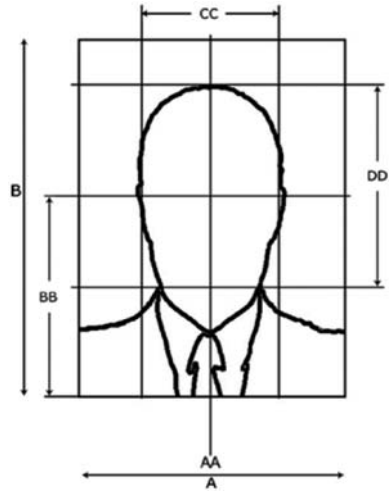
Požadavky na plný frontální obraz (full frontal) zahrnují požadavky na frontalní obraz a doplňují

- fotografické požadavky
  - obličej musí být na fotografii horizontálně vycentrovaný (linka AA musí být vertikální osou fotografie)
  - pozice očí (délka BB) musí být vzdálena od spodního okraje na vzdálenost mezi 50 % do 70 % z výšky obrazu

  - šířka hlavy (délka CC) busí být taková (vůči šíři celého obrazu), aby platilo A : CC = 7 : 5
  - výška hlavy (BB) by neměla přesáhnout 80 % výšky obrázku

- požadavky na digitalizaci – obrázek má mít minimální šířku min. 180 pixelů a vzdálenost očí musí být minimálně 60 pixlů

Uvedené délky jsou definovány podle obrázku 9:



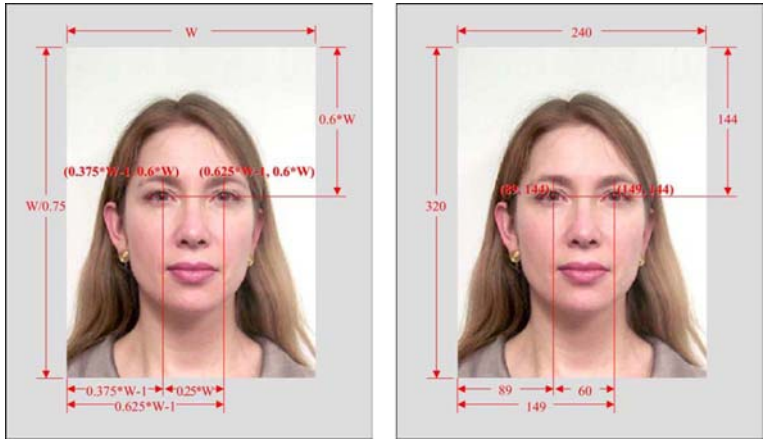
Obr. 9

A jako poslední uvedeme požadavky na tzv. token image (obraz pro umístění do čipu)

- poměr stran obrazu je 3 : 4
- pozice očí od horního okraje je 0,6× šířka obrázku
- vzdálenost očí je 0,25× šířka obrázku
- pozice pravého oka od levého okraje snímku je (0,625 × šířka obrázku) − 1
- minimální šířka obrázku je 240 pixlů (60 pixlů mezi očima)

Další požadavky na obraz obličeje jsou: uniformní pozadí 18% šedé.

Doporučeným formátem pro ukládání fotografií do čipu je JPEG2000 a povolen je rovněž JPEG.



Obr. 10

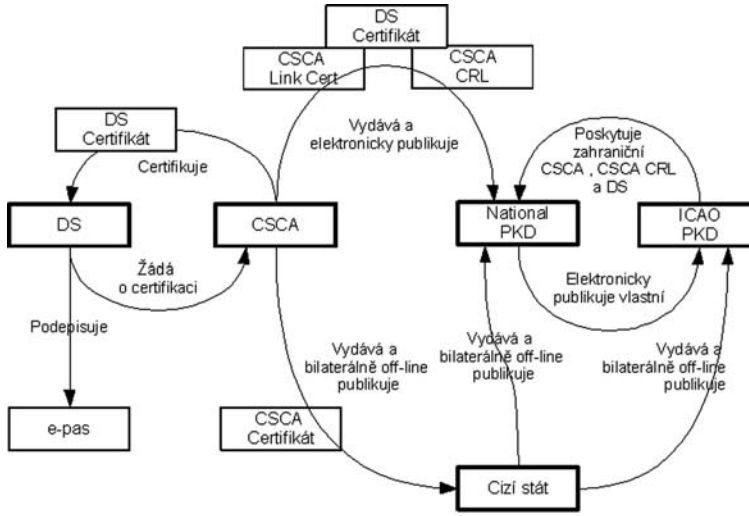
2.6 PKI

Pro ověřování pravosti elektronických dat v pasech byla navrženo v dokumentu [ICAO Bio PKI] relativně jednoduchá infrastruktura veřejných klíčů. Tato infrastruktura je jednoúrovňová. Každý stát je povinen pověřit nějaký subjekt provozováním národní CA (CountrySigning CA), která certifikuje klíče výrobce dokladů tzv. document signera(DS). Využití PKI a role jednotlivých komponent je znázorněo na obrázku 11 (tučně jsou označeny organizace, jinak jde o přenášená data).

2.6.1 Národní certifikační autorita (NCA)

Country Signing Certification Authority (CSCA) je zřízena státem, nebo státem pověřenou organizací. Jde o off-line certifikační autoritu. Vydává kořenový self-signed certifikát ke svému privátnímu klíči. Privátní klíč se zpravidla obnovuje každých 5 let. Při obnovení privátního klíče se vydává nový kořenový certifikát a tzv. link certifikát (speciální certifikát využívaný při obnově klíče CA, který slouží k provázání starého a nového klíče; nový klíč je certifikován s využitím starého a při včasné obnově umožňuje distribuovat nový certifikát CA s využitím důvěry ve starý – odpadá nutnost protokolárního předání a nový certifikát lze šířit elektronickou cestou). Doba platnosti certifikátu CSCA se vypočte jako Doba používání klíče + Délka platnosti certifikátu DS (typicky 5 let +(10 let + 3 měsíce)).

CSCA pravidelně vydává CRL s periodou maximálně 90 dnů.



Obr. 11

2.6.2 Document signer (DS)

Document signer je subjekt, který elektronicky podepisuje elektronický data v pasu – obvykle výrobce dokladů. Pro podepisování používá privátní klíč ocertifikovaný CSCA. Privátní klíč DS se zpravidla používá po dobu cca 3 měsíců a každý klíč DS je ocertifikován CSCA. Platnost certifikátu DS musí pokrývat celou dobu, kdy jsou platné dokumenty podepsané s využitím tohoto certifikátu. Při platnosti dokumentu 10 let je délka platnosti certifikátu DS 10 let a 3 měsíce.

2.6.3 Distribuce certifikátů a ICAO Public key directory (PKD)

Důvěryhodné předání národního kořenového certifikátu je základem pro bezpečně pracující systém pro inspekci e-pasů. Spoléhající se stát musí být přesvědčen o původu tohoto kořenového certifikátu. Proto se kořenové certifikáty jednotlivých národních certifikačních autorit se primárně distribuují bilaterálně offline diplomatickou poštou na datových nosičích. Kořenové certifikáty se takto distribuují mezi státy a také do centrálního systému nazvaného ICAO PKD. Rovněž bilaterálně se distribuují CRL.

ICAO PKD je adresářový server, který slouží k publikaci certifikátů DS a CRL vydávaných národními autoritami jednotlivých států, které vydávají e-pasy. Tento server je primárním zdrojem certifikátů DS a sekundárním zdrojem CRL. CSCA certifikáty se pomocí ICAO PKD nedsitribuují, ale jsou využívány v rámci ICAO PKD k ověřování certifikátu DS a CRL, kteréjsou do ICAO PKD importovány.

2.6.4 Národní Public key directory (PKD)

Pro komunikaci s cizími státy a pro distribuci certifikátů cizích států do systémů uvnitř státu se doporučuje využívat jednoho centrálního systému – Národního PKD.

2.6.5 Ověřování platnosti podpisu v pasu

ICAO stanovuje, že i v případě selhání ověření elektronického podpisu nebo certifikátu, kterým je ověřen klíč použitý k vytvoření podpisu může být nadále pas považován za platný ale podléhá podrobnějšímu zkoumání při průchodu přes inspekční systémy. Dojde-li tedy v průběhu období platnosti pasu (a tedy i certifikátů použitých pro zabezpečení) ke kompromitaci některého z použitých privátních klíčů a jk jeho uvedení na CRL, pas mlže být nadále používán, ale držitel pasu je vystaven důkladnějším prohlídkám při překračování hranic.

2.6.6 Kryptografické algoritmy a jejich síla

Pro vytváření a ověřování elektronického podpisu v systému e-pasů je možno využívat asymetrické algoritmy RSA, DSA a Eliptické křivky DSA (ECDSA). Vzhledem k relativně dlouhé platnosti certifikátů a podpisů je nutno volit silnější algoritmy, než pro běžné krátkodobé použití. Klíč CSCA by měl být délky nejméně 3 072 bitů pro RSA a DSA a 256 bitů pro EC; pro DS pak 2048 bitů pro RSA a DSA a 224 pro ECDSA; pro Aktivní autentizaci pak 1024 pro RSA a DSA a 160 bitů pro ECDSA. Jako hash funkci je doporučeno používat funkce ze třídy SHA-2 (SHA-256, 512 apod). Pro vytváření elektronického podpisu s využitím RSA je doporučeno používat podpisové schéma RSA-PSS.

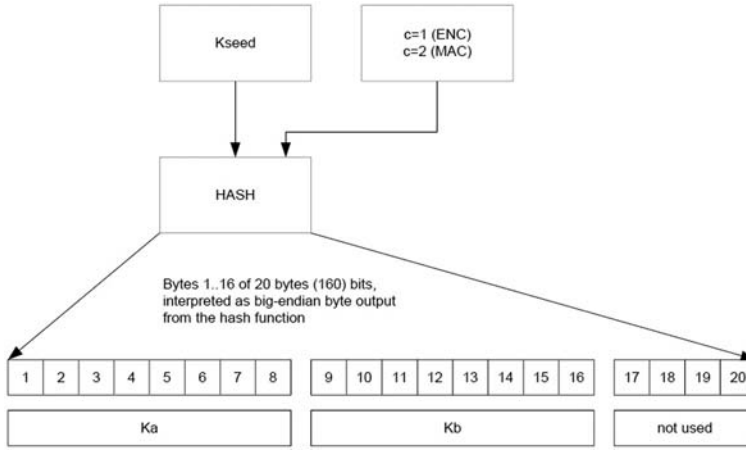
2.7 Zabezpečení elektronické části

Celý pas je vybaven bezpečnostními prvky, které mají zamezit vyrobení falešného pasu. Rovněž elektronická část pasu je chráněna před falšováním. Kromě ochrany před kopírováním bylo nutno z důvodu využití bezkontaktního rozhraní (a možnosti čtení na dálku bez vědomí držitele) zavést ochranu před neoprávněným čtením dat z čipu.

2.7.1 Basic access control

Mechanismus Basic access control je navržen s cílem zabránit neoprávněnému čtení pasu bez vědomí držitele. Je založen na vlastnosti čipových karet tzv. secure messagingu. V rámci secure messagingu se používají jeden klíč pro autentizaci zpráv pomocí MAC a druhý pro šifrování zpráv algoritmem 3DES. Klíče se odvozují jednoduchým mechanismem z údajů v MRZ datové stránky – číslo pasu

(9 znaků), datum narození držitele (6 znaků data YYMMDD), datum ukončení platnosti pasu (6 znaků datumu YYMMDD). Klíče se odvozují dle obrázku 12:



Obr. 12

Pro vygenerování klíčů je tedy nutno přečíst MRZ. Díky této vlastnosti je možno data z čipu číst pouze pokud držitel pasu umožnil přečíst MRZ, což chrání pas před neoprávněným čtením bez vědomí držitele. Díky šifrování jsou data rovněž chráněna před odposlechem.

Nevýhodou mechanismu BAC je jeho slabost. Zdrojem pro generování klíče nejsou dostatečně náhodná data a v mnoha případech (zejména při sekvenčním číslování pasů). Síla ochranného mechanismu je dostatečná pro on-line ochranu přístupu k pasu, takže útočník nedokáže přečíst data z pasu, který má např. Člověk procházející kolem čtecího zařízení. Při odposlechnutí oprávněného čtení pasu (možné až na vzdálenost metrů) a zaznamenání komunikace je možné zachycená data dešifrovat s běžně dostupným HW v krátkém čase (hodiny).

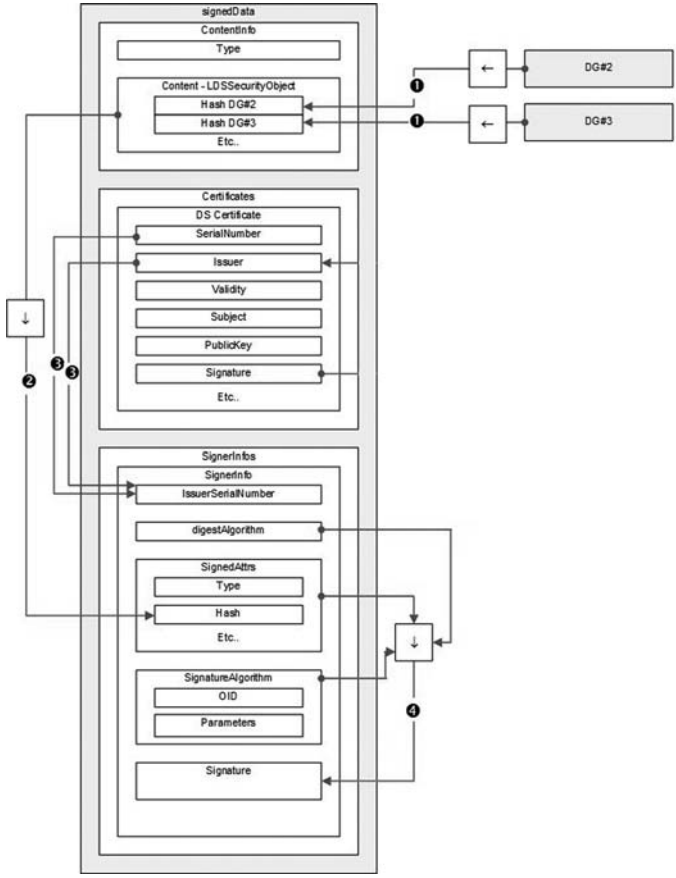
Implementace mechanismu BAC není ve specifikacích ICAO povinná. EU pro členské státy určuje BAC jako povinný.

2.7.2 Elektronický podpis data v čipu

Data uložená v čipu jsou elektronicky podepsána vydavatelem pasu DS s využitím privátního klíče certifikovaného národní certifikační autoritou (CSCA).

Při generování resp. verifikaci podpisu jsou prováděny následující kroky

1. Z všech datových skupin DGn jsou vypočteny hodnotu otisku a jsou shromážděny do ASN.1 struktury LDSSecurityObject



Obr. 13 Schéma závislostí dat při vytváření podpisu (autor: M. Hrdlička)

2. DER kódování LDSSecurityObject je datovým obsahem CMS zprávy typu SignedData – je vypočten otisk a je uložen mezi podepisované atributy CMS v CMS zprávě

3. Do CMS zprávy je vložen certifikát DS a je provázán s obsahem SignerInfo struktury pomocí IssuerAndSerial hodnoty

4. S použitím odpovídajícího algoritmu pro otisk a elektronický podpis se vytvoří samotná hodnota elektronického podpisu z DER kódování podepsaných atributů CMS zprávy

Takto vytvořený elektronický podpis zabraňuje tomu, aby si při falšování pasu mohl padělatel naplnit elektronickou část dle svého.

2.7.3 Aktivní autentizace

Výše popsaný elektronický podpis dat v čipu e-pasu nebrání tato data kopírovat. Doporučení ICAO proto specifikuje ještě volitelný mechanizmus tzv. aktivní autentizace (AA), který s využitím vlastností čipu bezpečně uchovat privátní klíč zajistí, že čip nebyl zkopírován. V čipu e-pasu je uložen privátní klíč (generovaný přímo v čipu nebo v rámci bezpečného prostředí personalizace). Veřejný klíč příslušející k tomuto privátnímu je zapsán do DG15 a je zahrnut do dat vybavených elektronickým podpisem (viz předchozí kapitola). Tímto elektronickým podpisem je zajištěn správný původ tohoto klíče. Pokud by se padělatel pokusil pas zkopírovat, nutně musí zkopírovat i DG15 beze změny a tedy i hodnotu veřejného klíče. Odpovídající hodnota privátního klíče je však bezpečně uložena v čipu bez možnosti ji přečíst. Při kontrole takto falšovaného pasu pak mechanizmus AA selže (nebude v souladu veřejná část klíče v DG15 a privátní část v čipu).

Aktivní autentizace probíhá takto:

1. Čtečka provede čtení dat z pasu včetně DG15 a ověří jejich integritu a původ pomocí elektronického podpisu
2. Čtečka vygeneruje náhodná data a pošle je do čipu
3. Čip v pasu tato data podepíše s využitím privátního klíče pro AA a odešle podepsaná data do čtečky
4. Čtečka ověří elektronický podpis dat pomocí klíče získaného z DG15 – pokud se podpis podaří ověřit, je pas v pořádku, jinak se může jednat o falzifikát

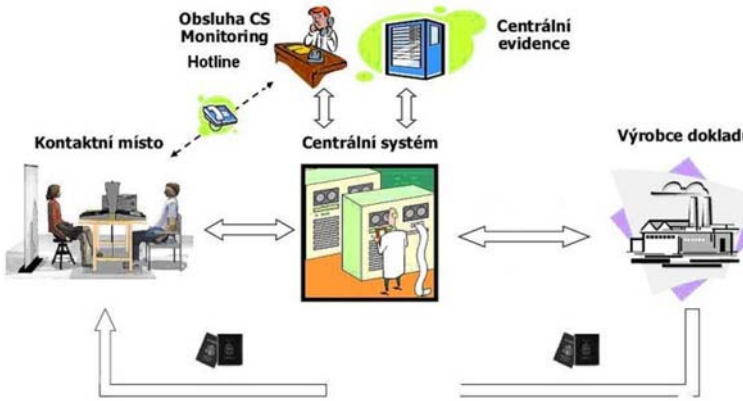
Aktivní autentizace je náchylná k útoku na soukromí s využitím tzv. challenge sematic (viz bod 2 postupu AA). Pokud je výzva místo náhodného čísla generována tak, aby nesla hodnotu, existuje možnost sledování pohybu držitele pasu, protože v odpovědi na výzvu se pas identifikuje. Z tohoto důvodu AA není implementována v pasech USA a Spolkové republiky Německo a dalších zemí. V Německu byl v nedávné době zaznamenán úspěšný případ kopírování elektronické německého e-pasu, což vzhledem k blízkému zavedení výroby e-pasů ve většině zemí Evropy, vyvolalo silný mediální ohlas.

3 Procesy vydávání pasu v ČR

V České republice se elektronické pasy vydávají od 1. 9. 2006, kdy došlo ke spuštění systému CDBP (Cestovní doklady s elektronickými prvky). Žádosti o vydání e-pasu jsou sbírány na úřadech v obcích s rozšířenou působností. Žadatel se dostaví s doklady prokazujícími nárok na vydání cestovního pasu (občanský průkaz,

rodný list, starý platný pas případně potvrzení o občanství). Veškerá data pro zápis do pasu jsou získávána v elektronické podobě z centrálních evidence obyvatel. Žadatel nevyplňuje žádná data. Z biometrických dat se v současné snímá době pouze fotografie a digitalizovaný podpis.

Fotografie je pořizována přímo na místě ve speciální fotokabině, která slouží k zajištění konstantních podmínek pro fotografování bez ohledu na pozice oken, osvětlení místnosti na úřadu apod. Po pořízení fotografie jsou přímo na sběrném místě provedeny její úpravy do tvaru požadovaného ICAO specifikací (full frontal image pro personalizaci na datovou stránku, token image pro uložení do čipu). Zároveň jsou na obraze specializovaným modulem provedeny kontroly, zda je v souladu s požadavky odpovídajících norem, aby mohl úředník v případě nevyhovující fotografie pořídit novou.



Obr. 14 Zdroj: MV

4 Testy interoperability

Pro úspěšné nasazení e-pasů v celosvětovém měřítku je zásadní schopnost plné interoperability e-pasů nejrůznějších výrobců HW i SW se čtecími zařízeními a inspekčními systémy. Z tohoto důvodu se v Berlíně uskutečnil celosvětový test interoperability. Zde uvedeme stručnou sumarizaci některých výsledků:

- bylo testováno 47 typů čteček od 37 výrobců
- celkem bylo zkoumáno 443 e-pasů, z toho 423 s BAC a 161 s AA
- algoritmus podpisu je u 37 pasů založen na EC a u 404 pasů na RSA
- algoritmus otisku je 289× SHA1, 142× SHA256, 10× SHA512
- formát fotografií byl 302× JPEG a 132× JPEG2000

- doba čtení pasu byla průměrně 5,8 s; u pasů s BAC 5,7 s; u pasů s AA 6,3 s.

Podrobné výsledky testů lze najít na webových stránkách věnovaných této události [Berlin Interop]

5 Budoucnost – otisky prstů v roce 2009

Dle rozhodnutí komise EU ([EU Otisky]) se od 28. 6. 2009 musí země EU vydávat pasy vybavené otisky prstů. Vložení otisků prstů vnáší do pasu daleko citlivější osobní údaj, než jakým je obličej. Úspěšnost rozpoznávání obličeje a jednoznačnost sejmuté fotografie je v porovnání s otisky prstů nižší. Otisky prstů identifikují jednoznačně jejich nositele, nelze je vyměnit a proto je jejich utajení velice citlivé. Před uvedením otisků prstů nejsou v e-pasu v podstatě žádné nové údaje proti dosavadním zvyklostem (obraz obličeje je v pasech od jejich samého počátku).

Zavedení otiskl prstů proto vyžaduje zvýšení ochrany dat v pasu uložených. Mechanismu vyššího zabezpečení se ve specifikacích ICAO nazývá Extended Access Control (EAC) a není v těchto specifikacích popsán. Velkou aktivitu na tomto poli vyvíjí německý BSI a jejím výsledkem je návrh pro realizaci EAC v dokumentu [EAC TR].

V následujících odstavcích se nebudeme zabývat oblastí biometrie. Zaměříme se pouze na způsob zabezpečení čipu před neoprávněným čtením.

5.1 Bezpečnost

Pro zavádění otisků prstů bylo vytyčeno několik požadavků

- ochrana soukromí držitelů pasů (bezpečný kanál pro přenos dat mezi čtečkou a pasem) – lepší, než BAC – řešením je DH výměna klíčů v rámci CA
- vylepšení ochrany před kopírováním a odstranění challenge semantic – řešením je funkčnost bezpečné komunikace navázané dle prvního bodu v rámci CA
- ochrana soukromí držitelů pasů nastavením přístupových oprávnění pro čtení daktyloskopických údajů – řešeno s pomocí TA

aby vydávající stát měl možnost určovat, kdo si může otisky prstů jeho občanů, číst. Dodatečným požadavkem, ze kterého EAC vychází je vylepšení ochrany čipu před kopírováním tentokrát bez nebezpečí challenge semantics (viz výše).

Ochrana proti kopírování pasu se nazývá Chip authentication (CA). Pro zjištění, zda terminál může či nemůže číst (či dokonce zapisovat) citlivá data určuje

Terminal authentication (TA). Kombinace těchto metod se nazývá Extended Access Control (EAC) a je specifikována v [EAC TR].

5.1.1 Chip authentication (CA)

CA je mechanismus založený na algoritmu výměny klíčů Diffie-Hellman. Čip využije pro výměnu klíčů speciální statický asymetrický klíčový pár určený pro tuto operaci. Čtečka použije dočasný klíčový pár. Na základě těchto klíčů proběhne DH protokol pro výměnu klíčů a po jeho skončení budou obě strany (čip i čtečka) mít k dispozici sdílené tajemství, na jehož základě odvodí klíče pro kryptografickou ochranu komunikace následující po CA.

Klíč, který čip použil pro autentizaci resp. jeho veřejná část, se ověří vůči DG14, kde je tato veřejná část uložena (obdobně jako veřejná část klíče pro AA je uložena v DG15) a je tak podepsána v rámci struktury LDSSecurityObject. Pokud veřejná část klíče použitá v rámci CA je shodná s obsahem DG14 a zároveň čip rozumí komunikaci chráněné na základě sdíleného tajemství, prokazuje tak držení privátního klíče příslušného ke klíči veřejnému z DG15 a tedy prokazuje, že čip nebyl zkopírován.

Důkaz původu čipu (pasu) je v podstatě vedlejším efektem vytvoření bezpečného komunikačního kanálu. Po CA již čip a čtečka komunikují s využitím secure messagingu s využitím kryptograficky silného klíče (v porovnání s BAC).

5.1.2 Terminal authentication (TA)

Vydávající státy v rámci ochrany osobních údajů svých občanů (držitelů pasu) musí mít možnost řídit, která organizace či stát má právo citlivá biometrická data číst a která ne. Za tímto účelem umožní čip v pasu přečtení citlivých dat pouze zařízením, která prokáží, že jsou k tomu oprávněna (obdobný posup je využit u platebních karet, kdy karta umožní transakci pouze autorizovaným platebním terminálům). Zařízení prokazuje svá oprávnění opět s využitím asymetrické kryptografie. Každému zařízení, které chce číst citlivá data z čipu je vybaveno asymetrickým klíčovým párem a v rámci komunikace prokáže, že vlastní privátní klíč odpovídající klíči veřejnému, který byl zaslán do čipu. Jakmile jednou terminál prokáže, že vlastní odpovídající privátní klíč k předloženému privátnímu, musí čip nejak poznat, že zařízení vybavené tímto tento klíčovým párem, je oprávněn číst citlivá data. K přiřazení konkrétní identity k subjektu reprezentovanému párem klíčů se využívá X.509 certifikátů. Někdy jsou v těchto certifikátech umístěny rovněž informace o přístupových oprávněních (certifikáty vydávané MicrosoftCA mohou obsahovat seznam rolí uživatele v doméně).

X.509 certifikáty jsou pro využití v omezeném výpočetním prostředí čipu příliš komplikované. Pro potřeby čipových karet byla proto definována zjednodušená specifikace tzv. card verifiable certificate. V podstatě se jedná o datovou strukturu nesoucí podobné informace, které nese X.509 certifikát:

- CV Certificate
  - Certificate body
    - Certificate profile identifier – verze (= 0)
    - Certificate authority reference – max. 16 znakový (ISO 8859-1) identifikátor autority sestavený z kódu země, jména autority a ID klíče (např. CZ-CVCA-01)
    - Public key – hodnota veřejného klíče dle algoritmu
    - Certificate holder reference – max. 16 znakový (ISO 8859-1) identifikátor držitele sestavený z kódu země, jména držitele a ID klíče (např. CZ-CPP-01)
    - Certificate holder authorization – zakódovaná role (oprávnění) držitele certifikátu (viz níže)
    - Certificate effective date – začátek platnosti certifikátu ve formátu YYMMDD kódované v BCD v zóně GMT
    - Certificate exspiration date – konec platnosti certifikátu ve formátu YYMMDD kódované v BCD v zóně GMT
  - Signature – vypočtena z Certificate body

Takto zjednodušený certifikát odešle inspekční terminál do čipu a ten ho ověří vůči klíčům, které má uloženy. Pokud ověření uspěje, jsou terminálu přidělena oprávnění dle nastavení čipu a hodnoty v položce holder authorization (viz níže).

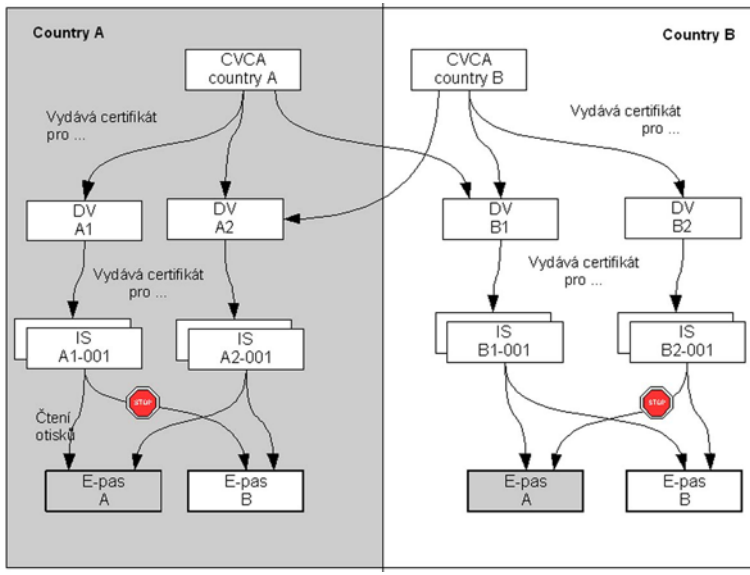
Pro vysvětlení autorizací je nejprve potřeba zmínit, jak vypadají hierarchie certifikátů.

5.2 PKI pro kontrolu e-pasů

Celé PKI pro ověřování dokladů je postaveno na tzv. card verifiable certifikátech (nikoli na X.509 certifikátech). Hierarchie certifikátů je navržena tak, aby byla co nejvíce flexibilní (pro úkoly, které má plnit). Celá hierarchie vychází z toho, že čip v pasu důvěřuje pouze zařízením, které prokáží, že jejich klíč (nepřímo s využitím DV) certifikovala CVCA, jejíž certifikát (klíč) má čip uložen uvnitř své bezpečné paměti.

5.2.1 Country Verification Certifiation Authority (CVCA)

Každý stát provozuje (přímo či prostřednictvím pověřené organizace) alespoň jeden subjekt ověřující pravost dokladů (CVCA), který spravuje množinu povolených kontrolních systému (document verifier – DV). Certifikát CVCA se vydává na dobu 6 měsíců (minimum) až 3 roků(maximum). CVCA je povinno zpracovat žádost o vydání certifikátu pro DV do 72 hodin (ověří správnost žádosti) a následně vydat certifikát do 24 hodin. CVCA vypracuje a zveřejní pravidla



Obr. 15

vydávání certifikátů DV, aby cizí státy požadující certifikaci svých inspekčních systému věděli, jak postupovat.

Jak je vidět, maximální platnost CVCA certifikátu je kratší, než typická doba platnosti e-pasu (10 let). Z tohoto důvodu při obnově klíče CVCA je nutno vygenerovat tzv. link certifikát. Tento certifikát je pak distribuován všude tam, kde je třeba používat TA a do čipu se pak při autentizaci terminálu posílá celý řetěz: CVCA Link1 – CVCA Link2 – . . . CVCALinkN – DVCert – ISCert. V čipu je uložen prvotní CVCA klíč a klíče link certifikátů a je možno celý řetěz ověřit přímo v čipu. Certifikát DV a IS se posílá do čipu při každé autentizaci, klíče link certifikátů CVCA by mely být v čipu uloženy.

Doba překryvu platnosti starého a nového certifikátu by měla být minimalilzována. Délka doby překryvu vychází z času potřebného pro distribuci nového certifikátu na všechna inspekční místa.

Primárním komunikačním kanálem pro komunikaci s CVCA je e-mail (mohou být i jiné). Pro výměnu dat v e-mail rámci zpráv musí být použito formátu MIME. Odesílatel by měl ve zprávě formulovat požadavek na doručenku (dle odpovídajících standardů). Pokud doručenka nedorazí v předpokládané době, může odesílatel komunikaci opakovat. Komunikační rozhraní CVCA podporuje tyto zprávy (u nichž je definováno mapování na obsah emailu):

- Register (Subject: Register; Body: URI,které říká, z jaké internetové adresy lze se státem komunikovat)

- CVCA certificate (Subject: CVCA Certificate, přílohy: jeden nebo více CVCA Link certifikátů)
- DV certification request (Subject: DV Certification Request; příloha: žádost(i) ve formě self-signed certifikátu)
- Odpověď na žádost (Subject: [Reply to] DV Certification Request; Body: případné důvody nevydání; příloha: vydané certifikáty DV)

Ve standardu není explicitně uvedeno, zda a jak má být komunikace chráněna, ale vzheldem k její citlivosti by autor doporučoval ochranu S/MIME s využitím elektornického podpisu (kritická je integrita a původ zpráv, nikoli utajení jejich obsahu).

Document Verifier (DV) Každý stát určí alespoň jeden subjekt pověřený ověřováním dokladů (DV), který bude spravovat množinu povolených inspekčních systémů (IS). DV žádá o certifikaci svého klíče u CVCA. Formátem žádosti je tzv. self-signed certifikát. CVCA vydává card verifiable certifikát a naplní jeho položku Certificate holder authorization podle toho, kdo o certifikát žádá. CVCA tedy busí mít implementován systém, kde bude definováno jak se prokazuje skutečnost, že daná žádost pochází od konkrétního subjektu, který je oprávněn číst citlivá data z pasu. Dále CVCA rozhoduje o hodnotách položky validity.

DV je certifikační autoritou, která certifikuje klíče konkrétních inspekčních systémů. V České republice by v roli DV mohla vystupovat Cizinecká a pohraniční policie, která bude spravovat inspekční systémy pro ekektronické pasy, kterým i budou vybavení příslušníci na letištích a dalších místech, kde se pasy prověřují.

Pokud organizace provozující DV potřebuje verifikovat pasy té které země, musí zažádat u příslušné CSCA provozované zvolenou zemí (např. v případě, že CPP ČR bude chtít číst otisky prstůz pasů vydaných občanům Polska, musí se obrátit s žádostí o certifikaci k CVCA provozované Polskem).

Délka platnosti certifikátu DV je od 2 týdnů do 2 měsíců a je určena vydávajícím CVCA. DV je povinen zpracovat žádost o vydání certifikátu do 24 hodin a vydat pak certifikát do 48 hodin.

Obnova DV V případě, ze DV žádá o obnovení certifikátu, autentizuje tuto žádost tím, že ji navíc podepíše pomocí starého klíče. K uložení podpisu slouží položka Certificate holder authorization.

Postup generování žádosti o obnovu DV probíhá takto (v tomto bodě je specifikace nejasná a jde o interpretaci autora):

1. generuji nový klíčový pár KP_NEW

2. generuji certifikát obsahující údaje nového certifikátu, ale podepsaný starým klíčem KP_OLD

3. generuji self-signed certifikát pomocí KP_NEW, přičemž do položky Certificate holder authorization zanořím certifikát generovaný v kroku 2

4. certifikát získaný v kroku 3 se odešle e-mailem k zpracování

Inspection system (IS) Inspekční systém je technologická jednotka, která již přichází do kontaktu s e-pasem a která pro čtení z pasu používá svůj klíč. Typicky je IS např. příruční počítač vybavený čtečkou pasů a snímačem otisku prstů a vybavený modulem pro uchování klíče pro autentizaci vůči pasu (tzv. SAM). SAM může mít např. formu čtečky kontaktních čipových karet a klíče jsou pak uloženy na čipové kartě.

Délka platnosti certifikátu IS je od 1 dne do 1 měsíce. Délka platnosti musí být vždy nejdéle stejná jako platnost odpovídajícího DV typicky však kratší. Autorizace uvedené v certifikátu IS jsou podmnožinou autorizací, které má přiděleny vydávající DV.

5.3 Autorizace

Ve všech certifikátech je naplněna položka Certificate holder authorization, která obsahuje informace, jaká je role toho kterého systéimy a jaké operace jsou mu povoleny. Autorizace jsou uloženu podle následující tabulky a je nazvána *relativní autorizací*:

7	6	5	4	3	2	1	0	**Description**
x	x	-	-	-	-	-	-	**Role**
1	1	-	-	-	-	-	-	CVCA
1	0	-	-	-	-	-	-	DV (domestic)
0	1	-	-	-	-	-	-	DV (foreign)
0	0	-	-	-	-	-	-	IS
-	-	x	x	x	x	x	x	**Access Rights**
-	-	0	0	0	0	-	-	RFU
-	-	-	-	-	-	1	-	Read access to DG 4 (Iris)
-	-	-	-	-	-	-	1	Read access to DG 3 (Fingerprint)

Obr. 16

Pro určení výsledných oprávnění držitele konkrétního certifikátu je potřeba cestu od certifikátu až ke kořenové autoritě a posbírané hodnoty pole autorizace

bitově logicky vynásobit (AND). Výsledná hodnota je tzv. *efektivní autorizací*, a podle ní čip řídí přidělené oprávnění.

5.4 Správa klíčů a času v čipu

Po vyrobení pasu (a personalizaci) obsahuje čip aktuálně platný klíč CVCA a využívá ho jako důvěryhodnou kotvu pro autentizaci terminálu. Postupem času je klíč CVCA vyměněn (za život pasu i několikrát) a aby čip mohl i nadále ověřovat klíče IS, potřebuje celý řetězec vedoucí až ke klíči CVCA, se kterým byl personalizován. Čip umožňuje autorizovaným terminálům importovat následné klíče CVCA, provede jejich ověření a uloží si je do úložiště důvěryhodných klíčů. Díky tomu, že si čip klíče po ověření uloží, postačuje při běžných autentizacích zasílat do čipu pouze certifikáty DV a IS.

Při ověřování platnosti předložených certifikátů je kromě kryptografického ověření také zjistit, zda je certifikát ještě platný. Čip nemá k dispozici zdroj času (když je mimo elektromagnetické pole, je zcela vypnutý). Proto je v čipu implementován tzv. aktuální čas čipu. Jde o hodnotu, která reprezentuje poslední časový údaj, který je prokazatelně již minulostí. Při personalizaci je tato hodnota inicializována na čas personalizace. Během doby života čipu je pak aproximována na základě certifikátů, které čip zpracovává v rámci importu a autentizací. Aktuální čas čipu je nastaven po úspěšné autentizaci na největší hodnotu atributu Effective date ze zpracovaných CVCALink certifikátů, DV certifikátů a „domácích“ IS certifikátů. Pokud navíc při importu a nastavení interního času čip identifikuje, že v oblasti důvěryhodných klíčů jsou vypršené certifikáty, musí je označit jako neaktivní a je možno je vymazat pro ušetření paměti.

Díky doporučenému způsobu překrývání platnosti při obnově CVCA by oblast důvěryhodných klíčů neměla nikdy obsahovat více jak 2 certifikáty CVCA.

Literatura

[ICAO BioMRTD] *Biometrics deployment of Machine Readable Travel Documents 2004.*
http://www.icao.int/mrtd/download/documents/Biometrics%20deployment%20of%20Machine%20Readable%20Travel%20Documents%202004.pdf

[ICAO Bio Annex A] *Annex A – Photograph Guidelines.*
http://www.icao.int/mrtd/download/documents/Annex%20A%20-%20Photograph%20Guidelines.pdf

[ICAO Bio Annex B] *Annex B – Facial Image Size Study #1.*
http://www.icao.int/mrtd/download/documents/Annex%20B%20-%20Facial%20Image%20Size%20Study_1.pdf

[ICAO Bio Annex C] *Annex C – Facial Image Size Study #2.*
http://www.icao.int/mrtd/download/documents/Annex%20C%20-%20Facial%20Image%20Size%20Study_2.pdf

[ICAO Bio Annex D] *Annex D – Biometric Data Interchange Formats – Part 5: Face Image Data (ISO./IEC JTC 1/SC 37 N 506).*
http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/2262372/2263034/2300191/JTC001-SC37-N-506.pdf?nodeid=3924597&vernum=0

[ICAO Bio Annex E] *Annex E – Biometric Data Interchange Formats – Part 6: Iris Image Data (ISO./IEC JTC 1/SC 37 N 504).*
http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/2262372/2263034/2300191/JTC001-SC37-N-504.pdf?nodeid=3924512&vernum=0

[ICAO Bio Annex F] *Annex F – Biometric Data Interchange Formats – Part 4: Finger Image Data (ISO/IEC JTC 1/SC 37 N 466).*
http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/2262372/2263034/2300191/JTC001-SC37-N-466.pdf?nodeid=3924168&vernum=0

[ICAO Bio Annex G] *Annex G – Biometrics – Biometric Data Interchange Formats – Part 2: Finger Minutiae Data (ISO/IEC JTC 1/SC 37 N 464).*
http://www.icao.int/mrtd/download/documents/Annex%20G%20-%20Fingerprint%20Minutiae.pdf

[ICAO Bio Annex H] *Annex H – Biometrics Data Interchange Formats – Part 3: Finger Pattern Spectral Data (ISO./IEC JTC 1/SC 37 N470).*
http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/2262372/2263034/2300191/JTC001-SC37-N-470.pdf?nodeid=3924237&vernum=0

[ICAO Bio Annex I] *Annex I – Use of Contactless Integrated Circuits.*
http://www.icao.int/mrtd/download/documents/Annex%20I%20-%20Contactless%20ICs.pdf

[ICAO Bio Annex J] *Annex J – ICAO. May 2003 Press Release.*
http://www.icao.int/mrtd/download/documents/Annex%20J%20-%20ICAO%20May%202003%20Press%20Release.pdf

[ICAO Bio Annex K] *Annex K – ICAO. Supplementary Requirements to ISO14443 – v2.*
http://www.icao.int/mrtd/download/documents/Annex%20K%20-%20ICAO%20Supplementary%20Requirements%20to%20ISO14443%20-v2.pdf

[ICAO Bio Annex L] *Annex L – ePassports Data Retrieval Test Protocol.*
http://www.icao.int/mrtd/download/documents/Annex%20L%20-%20ePassports%20Data%20Retrieval%20Test%20Protocol.pdf

[ICAO Bio PKD] *Issues of the ICAO. Public Key Directory (PKD).*
http://www.icao.int/mrtd/download/documents/Issues%20of%20the%20ICAO%20Public%20Key%20Directory%20PKD.pdf

[ICAO Bio LDS] *Logical Data Structure (LDS) version 1.7.*
http://www.icao.int/mrtd/download/documents/LDS-technical%20report%202004.pdf

[ICAO Bio PKI] *PKI for Machine Readable Travel Documents offering ICC read-only access v1.1.*
http://www.icao.int/mrtd/download/documents/TR-PKI%20mrtds%20ICC%20read-only%20access%20v1_1.pdf

[ICAO Bio 9303 Sup] *Supplement to Doc 9303 – ePassports.*
http://www.icao.int/mrtd/download/documents/9303%20Supplement%20-%20December%202005.pdf

[CDBP MV] *Cestovní doklad s biometrickými prvky – stránky Ministerstva vnitra.*
http://www.mvcr.cz/sprava/informat/biometrika/

[Obsah Datapge] *Popis obsahu datové strany.*
http://www.highprogrammer.com/alan/numbers/mrp.html

[Datapage Example] *Příklad datové strany.*
http://travel.state.gov/visa/temp/without/without_1990.html

[MRZ Calc] *Kalkulátor obsahu MRZ.*
http://www.highprogrammer.com/cgi-bin/uniqueid/mrzp

[EAC TR] *Extended Access Control TR 03110.*
http://www.befreite-dokumente.de/eingereichte-akten/tr-03110-eac-1.0/

[EU Bio] *Nařízení Rady (ES) č. 2252/2004 ze dne 13. prosince 2004 o normách pro bezpečnostní a biometrické prvky v cestovních pasech a cestovních dokladech vydávaných členskými státy (publikováno dne 29. prosince 2004 v Official Journal of the European Union číslo L385/1).*
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32004R2252:CS:HTML

[EU Bio Tech] *Rozhodnutí Komise ze dne 28. února 2005, kterým se stanoví technické specifikace norem pro bezpečnostní a biometrické prvky v cestovních pasech a cestovních dokladech vydávaných členskými státy – K(2005) 409.*
http://ec.europa.eu/justice_home/doc_centre/freetravel/documents/doc/c_2005_409_cs.pdf

[EU Otisky] *Rozhodnutí komise ze dne 28. 6. 2006 kterým se stanoví technické specifikace norem pro bezpečnostní a biometrické prvky v cestovních pasech a cestovních dokladech vydávaných členskými státy.*
http://ec.europa.eu/justice_home/doc_centre/freetravel/documents/doc/c_2006_2909_cs.pdf

[329/1999 Sb.] *Zákon č. 329/1999 Sb., o cestovních dokladech a o změně zákona č. 283/1991 Sb., o Policii České republiky, ve znění pozdějších předpisů, (zákon o cestovních dokladech), ve znění zákona č. 217/2002 Sb., zákona č. 320/2002 Sb., zákona č. 539/2004 Sb., zákona č. 559/2004 Sb. a zákona č. 136/2006 Sb.*
http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_number1=329/1999&PC_8411_l=329/1999&PC_8411_ps=10#10821

[326/1999 Sb.] *Zákon č. 326/1999 Sb., o pobytu cizinců na území České republiky a o změně některých zákonů, ve znění zákona č. 140/2001 Sb., zákona č. 151/2002 Sb., zákona č. 217/2002 Sb., zákona č. 222/2003 Sb., zákona č. 436/2004 Sb., zákona č. 501/2004 Sb., zákona č. 539/2004 Sb., zákona č. 559/2004 Sb. a zákona č. 136/2006 Sb.*
http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_number1=326/1999&PC_8411_l=326/1999&PC_8411_ps=10#10821

[325/1999 Sb.] *Zákon č. 325/1999 Sb., o azylu a o změně zákona č. 283/1991 Sb., o Policii České republiky, ve znění pozdějších předpisů, (zákon o azylu), ve znění zákona č. 2/2002 Sb., zákona č. 217/2002 Sb., zákona č. 320/2002 Sb., zákona č. 519/2002 Sb., zákona č. 222/2003 Sb., zákona č. 539/2004 Sb., zákona č. 57/2005 Sb. a zákona č. 501/2004 Sb.*
http://portal.gov.cz/wps/portal/_s.155/701?number1=325%2F1999&number2=&name=&text=

[642/2004 Sb.] *Vyhláška č. 642/2004 Sb., kterou se provádí zákon o občanských průkazech a zákon o cestovních dokladech.*
http://portal.gov.cz/wps/portal/_s.155/701?number1=642%2F2004&number2=&name=&text=

[ISO 1073-2:1976] *ISO norma o alfanumerickém fontu pro OCR část 2.*
http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=5568

[Berlin Interop] *Testování interoperability v Berlíně.*
http://www.interoptest-berlin.de/

Zdokonalení autentizace použitím jednorázových hesel

Daniel Kouřil, Luděk Šulák

E-mail: kouril@ics.muni.cz, ludek.sulak@seznam.cz

1 Úvod

Spolehlivá autentizace neznamená jen volbu spolehlivého protokolu, ale i ochranu údajů používaných pro identifikaci. V řadě systémů je nejpoužívanější uživatelskou autentizační metodou kombinace uživatelského jména a hesla. Protokoly založené na této metodě nevyžadují složitou implementaci, ani nejsou náročné na výpočetní kapacitu. Uživatelé jsou na tento přístup zvyklí a není je nutné školit na jeho používání. Je také jednoduché připravit aplikace a jejich uživatelské rozhraní tak, aby umožňovaly zadávání a zpracování hesla. Pro jednoduchost používání je autentizace uživatelským heslem základem řady sofistikovaným protokolů, jako je např. Kerberos.

Na druhou stranu vykazují hesla jistá omezení, která mohou výrazně ovlivnit bezpečnost celého systému. Pokusíme se demonstrovat, že systém založený na heslech může být poměrně hodně zranitelný. Uživatelské heslo musí být snadno zapamatovatelné pro člověka, proto je omezená míra entropie, kterou heslo obsahuje. Na systém založený na uživatelských heslech lze tedy nasadit útoky, které se snaží uhodnout správné heslo, resp. kombinaci uživatelského jména a hesla. Útok může být veden buď tzv. hrubou silou, kdy jsou postupně zkoušeny všechny možné kombinace znaků ze zvolené abecedy, nebo tzv. slovníkový útok, který je založen na skutečnosti, že lidé často volí hesla odvozená od existujících slov v přirozeném jazyce. Zkušenost ukazuje, že tyto útoky mohou být velmi účinné. Obranou proti nim může být kontrola kvality hesla při jeho nastavování, resp. změně, nebo nějaká omezení na straně služby, která hesla ověřuje (např. omezený počet pokusů). Vždy je ale potřeba pečlivě volit mezi úrovní zabezpečení systému a jeho použitelností, aby se pravidla nestala příliš omezující pro uživatele, kteří potom hledají (a zpravidla nacházejí) způsoby, jak pravidla obejít, např. vylepením silného hesla na monitor. Na tomto místě je však vhodné poznamenat, že je vhodnější zvolit silné heslo i obtížně zapamatovatelné, které si uživatel zapíše na papír, který je bezpečně uložen, např v peněžence. Další možností útoku na systém založený na heslech je zaměřit se na místo, odkud se heslo zadává, tj.

na uživatelský počítač nebo přímo na samotného uživatele. Obliba a rozšířenost hesel samozřejmě nemohla zůstat nepovšimnuta počítačovým pirátům, kteří produkují viry, trójské koně a různý tzv. spyware, který se nepozorovaně nainstaluje do uživatelova počítače, snaží se zachytávat citlivé údaje zadávané uživatelem a následně je dopravuje zpět útočníkovi. Vedle těchto technických metod lze také nasadit útoky založené na technikách sociálního inženýrství, příkladem je známé rhybaření, kdy např. pomocí podvodných webových stránek nebo e-mailů jsou z uživatelů lákány citlivé údaje. Heslo lze také překvapivě snadno odpozorovat během zadávání na klávesnici. Bezpečnosti nepřispívá, že odcizení hesla nelze detekovat a rovněž je těžké rozlišit neoprávněné použití hesla útočníkem od korektního provozu.

Vedle hesel jsou k dispozici samozřejmě i jiné autentizační mechanismy, které umožňují uživatelskou autentizaci. K nejpoužívanějším patří metody založené na digitálních certifikátech a PKI, stále více se také prosazují biometrické techniky. Tyto metody však nejsou pro uživatele tak pohodlné a zejména v případě PKI jsou citlivé na správné používání a ochranu citlivých dat. Navíc jsou obtížnější na implementaci, což snižuje mj. i možnosti integrace se stávajícími aplikacemi. V tomto příspěvku popisujeme přístup založený na jednorázových heslech, která nabízí alternativní přístup k autentizaci založené na heslech. Jednorázová hesla umožňují řešit řadu problémů, které mají standardní hesla při zachování rozumné úrovně uživatelské přívětivosti a minimalizuje problémy s implementací a nasazením.

2 Jednorázová hesla

Jednorázové heslo (*one time password – OTP*) lze použít pouze pro ustavení jednoho autentizovaného spojení, po použití se už tedy nedá použít k další autentizaci. Potenciální útočník tak odposlechem nic nezíská. Vzhledem k tomu, že jsou generována náhodně, tak jsou i odolná proti standardním útokům prováděným proti běžným heslům. Z uživatelského pohledu fungují podobně jako běžná hesla, na která jsou uživatelé zvyklí. Řadu protokolů a aplikací, které podporují autentizaci heslem lze snadno používat i s jednorázovými heslem, často bez jakéhokoliv dodatečné úpravy. Úpravou samozřejmě musí projít komponenta, která heslo ověřuje, ale to je z pohledu nasazení zpravidla akceptovatelná změna.

Z pohledu uživatele je největší změnou zadávání vždy jiného hesla při přihlašování. Obecně lze říci, že uživatel spravuje seznam hesel, která postupně používá pro autentizaci. Existuje řada mechanismů pro správu jednorázových hesel, které se liší způsobem generování hesel, jejich uložením, použitím apod. Pro správu OTP se často používají tokeny, což mohou být buď specializovaná hardwarová zařízení nebo aplikace (tzv. soft-tokeny) zpravidla určené pro instalaci do mobilních zařízení (PDA, mobilní telefony). Tyto tokeny umožňují zavést

tzv. dvoufaktorovou autentizaci, kdy uživatel musí prokázat něco co má (tj. token) a něco co zná (PIN nebo heslo k tokenu). Vlastnictví tokenu je prokazováno tak, že token pro výpočet OTP použije tajný klíč uložený na tokenu, který nemůže být vyexportován mimo token. Tento tajný klíč je během zavádění tokenu bezpečně dopraven na místo, kde se budou OTP ověřovat. V případě soft-tokenů je samozřejmě tato ochrana slabší, protože klíč zadává uživatel při inicializaci soft-tokenu a je uložen v přístupné paměti zařízení.

2.1 Pevná posloupnost hesel vytvořená předem

V tomto modelu jsou OTP vygenerována před prvním použitím a jsou předána uživateli, který je postupně používá pro autentizaci. Často se tisknou na papír nebo se pro jejich správu používá soft-token. Server zpravidla během autentizačního dialogu oznámí index očekávaného hesla tak, aby uživatel věděl, které má zadat. Tento systém je jednoduchý, nevýhodou však je nutnost nové inicializace seznamu po vyčerpání všech hesel, což vyžaduje dodatečnou interakci s uživatelem. Na tomto principu jsou postaveny systémy S\Key[1] a OPIE[2], které jsou oba založeny na tzv. Lamportově schématu. L. Lamport navrhl systém pro generování a správu OTP, který elegantně využívá jednocestné funkce a umožňuje snadnou implementaci ověřování na straně serveru. V tomto schématu si uživatel zvolí nějaké počáteční heslo (pass-phrase), ke kterému server vygeneruje tzv. seed, což je náhodný řetězec znaků. Na uživatelovo heslo a inicializační hodnotu serveru je aplikována hašovací funkce, jejíž výsledkem je tzv. nulté heslo. Na nulté heslo je N-krát aplikována hašovací funkce, což vyústí v N jednorázových hesel. Seznam se používá odzadu, tj. od posledního vygenerovaného hesla, a na serveru je uloženo jen naposledy použité heslo a jeho index v seznamu. Autentizace uživatele probíhá na základě principu výzva-odpověď, kdy server předá uživateli výzvu obsahující seed a pořadové číslo požadovaného hesla. Při ověření OTP se na toto ověřované heslo aplikuje hašovací funkce a zkontroluje se, že výsledek je totožný s uloženým heslem. Po úspěšné autentizaci se uložené heslo nahradí právě použitým heslem pro ověření další transakce.

Hesla vygenerovaná jednocestnou funkcí jsou 64bitová čísla, která nejsou příliš pohodlná na zadávání. Proto je ve specifikaci definován slovník, pomocí kterého je heslo převedeno do šesti anglických slov (např. JIM, NOB, EARN, WIFE, RAIN, HATH). S tímto je i spojeno samotné uchovávání hesel resp. jejich generování na straně uživatele. Uživatel si může hesla vygenerovat dopředu a vytisknout si je na papír. Tento způsob má zjevnou nevýhodu při ztrátě či odcizení seznamu, ovšem na druhou stranu je to řešení nejtriviálnější a nejlevnější. Existují ale i implementace soft-tokenů[1], které lze zprovoznit v mobilních zařízeních podporujících Javové aplikace. Tyto aplikace lze snadno instalovat do

[1] http://marcin.studio4plus.com/en/otpgen/, http://www.ii.uib.no/~janfrode/jotp/

mobilního zařízení, ať už přímo stažením z internetu nebo nahráním přes datové rozhraní zařízení. Aplikace umožňují vygenerovat pro zadané heslo, seed a index příslušné OTP, které uživatel přepíše do autentizačního dialogu.

2.2 Generování hesel modelem výzva–odpověď

Tato hesla jsou generována pomocí uživatelova PINu a výzvy od autentizačního serveru. Typickým představitelem této třídy je token CryptoCard RB-1[3]. Tento token velikosti platební karty obsahuje malý displej a klávesnici, pomocí které uživatel zadává svůj PIN k tokenu a výzvu od serveru. Na displej je následně zobrazeno výsledné OTP, které se použije pro autentizaci. Generování OTP je založeno na hašování pomocí algoritmu DES, který jako vstup používá výzvu od klienta a tajný klíč uložený na kartě, který je sdílen s ověřovacím serverem. Token nemá omezenou dobu používání. Tokeny jsou programovatelné a umožňují nastavení různé úrovně ochrany (délka PINu, počet zadání). Podobně jako např. u čipových karet je token zablokován, pokud uživatel nezadá správný PIN v několika po sobě jdoucích pokusech.

2.3 Posloupnost hesel generovaná v závislosti na čase

Tato hesla jsou generována na základě aktuálního času a nevyžadují tedy zvláštní interakci s autentizačním serverem. Na druhou stranu ale vyžadují synchronizaci hodin tokenu a serveru, což může být složitý problém. Tento způsob může být také náchylný na tzv. replay útoky, protože heslo je pro dané časové okno pořád platné i po použití. Na modelu časových OTP jsou založeny tokeny RSA SecurID[4], které obsahují pouze displej, kam se jednou za minutu vygeneruje nové heslo. Uživatel pro autentizaci použije heslo, které je zrovna zobrazeno a k němu připojí svůj osobní PIN, který zabraňuje zneužití tokenu v případě jeho ztráty nebo odcizení. Jako u většiny tokenů je pro vygenerování OTP použit tajný klíč, který je do tokenu nahrán při výrobě a nelze jej změnit. Tokeny SecurID mají omezenou životnost (jeden až pět let, v závislosti na typu), po jejímž uplynutí je nutné pořídit nový token. Tento systém značně zvyšuje provozní náklady ve srovnání s jinými tokeny, přesto tokeny SecurID patří již řadu let k nejrozšířenějším zařízením svého druhu.

Pro generování časově závislých OTP je k dispozici i open-source implementace soft-tokenu MOTP[2], který lze nahrát do mobilního telefonu nebo PDA. Jeho součástí je i serverová část, která demonstruje možnost použití. Za zmínku stojí pečlivé uložení inicializačního klíče, který nelze zobrazit.

[2] http://motp.sourceforge.net/

2.4 Podpora OTP v aplikacích

Pro podporu OTP existuje relativně bohatá podpora. K dispozici jsou open-source knihovny pro mechanismy S\Key a OPIE. Podpora systému S\Key byla integrována s programem OpenSSH, který tak podporuje autentizaci jednorázovými hesly. Podobně existuje i řada PAM modulů, které umožňují integraci OTP do aplikací, které podporují PAM. Spolu s hardwarovými tokeny je možné také pořídit specializované aplikace, které zajistí ověření OTP vygenerované tokenem. Řada dalších aplikací může profitovat z podpory OTP v protokolu SASL, která umožňuje použít OTP v libovolné aplikaci, která používá SASL (např. různé SMTP servery). V prostředí WWW lze použít specializované moduly pro server Apache, které buď využijí dostupný PAM modul nebo externí aplikaci instalovanou na serveru. Lze tedy konstatovat, že řada aplikací je připravena na použití OTP. Společným nedostatkem ale je minimální možnosti pro efektivní nasazení OTP v rozsáhlejším prostředí. Většina zmíněných řešení používá lokální databázi OTP, která je uložena přímo vedle aplikace, což znemožňuje efektivní správu a využití různými službami z různých míst v síti. Vyjímkou jsou samozřejmě komerční servery dodávané k hardwarovým tokenům, ale ty zase bývají uzavřené a nepokrývají všechny požadavky. V následující kapitole představíme rozsáhlé distribuované prostředí a popíšeme návrh zavedení OTP do jeho infrastruktury.

3 Integrace OTP do prostředí *META Centra*

Projekt *META Centrum* je aktivita sdružení CESNET, které je provozovatelem české akademické vysokorychlostní sítě CESNET2 a provádí výzkum a vývoj v oblasti aplikací využívajících toto prostředí. Jednou z klíčových aplikací je *META Centrum*, které nad sítí CESNET2 buduje a produkčně provozuje infrastrukturu pro realizaci náročných výpočtů, tzv. *grid*. Cílem projektu je vybudovat národní gridové prostředí, které skrývá rozdíly v jednotlivých zapojených systémech a vytváří uživatelům iluzi jediného výpočetního uzlu. *META Centrum* se také podílí na převážné většině všech národních i mezinárodních gridových projektů, které se v ČR řeší.

Zabezpečení *META Centra* je založeno na systému Kerberos[5] který poskytuje centrální správu identit. Pro přihlašování do systému Kerberos se standardně používá kombinace uživatelského jména a hesla, v rámci *META Centra* jsme také realizovali podporu pro použití prostředků PKI a podporu čipových karet. PKI přihlašování řeší řadu bezpečnostních problémů, které přináší statická hesla, na druhou stranu ale vytváří některé další problémy a není také obecně použitelné z libovolného počítače, jelikož potřebuje instalaci základního programového vybavení na klientskou stanici (což není vždy možné nebo bez-

pečné, např. z internetové kavárny). Použití OTP proto vidíme jako vhodný komplement ke stávajícím metodám. Cílem této kapitoly je popsat návrh zavedení podpory OTP do stávajícího prostředí.

Primárním požadavkem je zachování současného prostředí tak, aby zůstaly beze změny používané mechanismy pro přihlašování a také koncové služby. OTP musí vytvořit pouze další metodu pro přihlášení, aniž by byly ovlivněny stávající procedury. Dalším požadavkem je také maximální flexibilita pro používané metody OTP. Zavedený systém by měl umožňovat použít více typů OTP a také podporovat více typů tokenů. V budoucnu se dá předpokládat, že někteří naši uživatelé budou vybaveni OTP tokeny pro přístup k jiným projektům a bylo by vhodné, aby tento token byl použitelný i pro přístup k *META Centru*.

3.1 Kerberos a OTP

Kerberos je postaven na důvěryhodné službě (KDC), která obsluhuje celou organizační doménu. KDC vydává lístky, což je forma digitálního certifikátu určená adresně pro klienta i server. Platnost lístku je několik hodin. Lístky jsou podepsány symetrického klíčem, který má KDC uloženo ve své databázi. V případě uživatelů jsou tyto klíče odvozeny od jejich hesla. Z hlediska zavádění dalšího autentizačního mechanismu je důležité, že autentizace se uživatele se provádí pouze v jednom místě, během získávání tzv. TGT lístku. Zbytek použití systému Kerberos je nezávislý na použitém přihlašovacím mechanismu a je výhradně založen na použití lístků. Nejsou tedy nutné žádné úpravy na straně koncových služeb, což výrazně usnadňuje zavedení nové autentizační metody.

Podpora OTP v systému Kerberos zahrnuje několik kroků. Nejprve je nutné rozšířit samotný autentizační protokol tak, aby podporoval použití OTP, dále rozšířit KDC server o podporu ověřování OTP a v neposlední řadě nachystat klientské nástroje pro přihlašování pomocí OTP. Pro změnu protokolu jsou možné dva přístupy, buď lze použít OTP na místě běžného hesla, které se používá pro standardní autentizaci nebo použít tzv. pre-autentizaci, což je mechanismus, který se používá pro dodatečné zabezpečení autentizace. KDC může například vyžadovat, aby uživatel nejprve prokázal znalost správného hesla před tím, než je mu vydán samotný lístek (který je tímto heslem zašifrován). KDC se tak chrání před útočníky, kteří si vyžádají lístek a poté zkusí jeho dešifrování např. hrubou silou. Metody pre-autentizace jsou rozšiřitelné o libovolnou další metodu a používají se např. pro rozšíření založené na PKI. První přístup by byl výrazně jednodušší a nevyžadoval žádné změny v protokolu a umožňoval by i použití stávajících klientů. Na druhé straně by bylo možné jej použít výhradně pro OTP, která nevyžadují odezvu od serveru (KDC) a byly by tak vyřazeny implementace OTP založené na systému výzva-odpověď. Perspektivněji proto vidíme použití pre-autentizace, která umožní implementaci i složitých metod. Zejména umožní realizovat flexibilní podporu více typů OTP, kdy KDC může

nabídnout klientovi více metod v jedné pre-autentizační zprávě. Uživatel si vybere preferovanou metodu, jejíž výsledek pošle zpět KDC k ověření. V žádném případě se nesmí posílat OTP v otevřené formě, protože zprávu nelze chránit proti odposlechu a útočník by heslo mohl odchytit a zneužít. Je proto nutné posílat nějakou šifru, která je pomocí OTP vytvořena a kterou může ověřit také KDC (protože OTP zná). Tento mechanismus je ostatně totožný se způsobem ověřování běžného hesla. Podobný přístup zvolila pracovní skupina pro Kerberos standardizačního sdružení IETF, která se OTP začala v posledních měsících intenzivně věnovat[6]. Jedinou, ale poměrně významnou výhradou k tomuto přístupu je nemožnost použití současných klientů, kteří jsou velmi rozšíření, např. jsou standardní součástí všech moderních linuxových distribucí. *META Centrum* však distribuuje základní klientskou sadu pro nejrozšířenější Linuxové distribuce a MS Windows, která obsahuje některé nestandardní nástroje (např. podporu PKI) a je tedy snadné přidat do těchto balíčků i podporu OTP. Není to sice univerzální řešení, ale věříme že je plně dostačující pro uživatele *META Centra* a jejich způsob používání.

Pokud jde o podporu ověřování OTP na straně KDC, plánujeme zavést obecné rozhraní pro práci s OTP, které usnadní použití různých typů OTP a jejich zapojení do KDC. Toto rozhraní by mělo oddělit logiku KDC od vlastního ověřování OTP, což umožní snadnější správu všech komponent i vývoj samotného kódu. Vlastnostmi se toto rozhraní bude blížit např. známému standardu PKCS11, ale bude určeno výhradně pro OTP, tudíž výrazně jednodušší. Vzhledem k tomu, že pro ověřování existuje řada PAM modulů, chceme se zaměřit na podporu PAM rozhraní, např. tak, že naše obecná knihovna bude vytvářet jednoduchou vrstvu nad PAM rozhraním. Tento způsob by výrazně eliminoval objem kódu pro vlastní ověřování OTP a zároveň by umožňoval využít již hotové a otestované implementace OTP. Elegantně tak také řeší i podporu pro hardwarové tokeny, jejichž dodavatelé zpravidla dodávají PAM modul schopný obsloužit jejich token.

V neposlední řadě bude potřeba pozměnit klientské aplikace, které provádějí autentizaci a komunikují s uživatelem tak, aby byly schopné uživateli zobrazit výzvu od KDC, která je posílána v pre-autentizačních zprávách a je potřebná pro vygenerování resp. zvolení OTP. V případě OTP, která nepotřebují vstupní informace je postup jednoduchý a v podstatě se neliší od standardního použití. Pro podporu OTP generovaných na základě výzvy od serveru bude situace složitější. V případě příkazu `kinit`, který provádí autentizaci bude situace poměrně jednoduchá. Stačí přidat zvláštní přepínač, který způsobí, že kinit nebude čekat na zadání hesla po zadání uživatelského jména a kontaktuje KDC, které vrátí vstup pro generování OTP pro daného uživatele. V případě GUI, která nemáme pod kontrolou však bude situace obtížnější. OTP je nutné podporovat také při autentizaci proti SSH serverům a pro autentizaci v prostředí WWW. Pro SSH server je v *META Centru* použit PAM modul, který je schopen heslo ověřit.

Bude tedy nezbytné upravit tento modul tak, aby uživateli zobrazil příslušnou zprávu potřebnou pro výběr OTP. V instalacích *META Centrum* jsou použity PAM moduly s lokálními úpravami, nebude proto příliš složité přidat rozšíření i pro OTP. V principu se bude jednat o podobný zásah jako pro příkaz `kinit`. V oblasti WWW používáme pro ověření Kerberovského hesla modul pro Apache mod_auth_kerb, který implementuje obsluhu metody Basic. Podpora OTP, které vyžadují dodatečný ustup generovaný serverem by nebyla jednoduchá, protože by vyžadovala dvojitou interakci s uživatelem (jednou zadání jména, podruhé OTP). Pro použití v prostředí WWW tedy budeme spíše uvažovat o podpoře pouze OTP, která lze vygenerovat nezávisle, resp. o jiných přístupech (např. PKI).

3.2 PKI a OTP

META Centrum je primárně založeno na mechanismu Kerberos, ale jeho uživatelé také silně využívají PKI, zejména pro přístup do zahraničních projektů. PKI má ovšem uplatnění i v rámci *META Centra*, v případech, kde poskytuje lepší vlastnosti než jméno a heslo, resp. nativní autentizace lístkem. Příkladem je např. prostředí WWW, kde podpora pro systém Kerberos v prohlížečích není tak dobrá jako podpora PKI. Pro účely *META Centra* používáme tzv. on-line certifikační autoritu, která vydává automaticky certifikáty uživatelům, kteří se autentizují pomocí syst mu Kerberos. Zdánlivě jde myšlenka on-line CA proti standardnímu pohledu na PKI, kdy je CA zpravidla držena zcela odděleně od sítě a všechny operace jsou prováděny off-line. Na tuto on-line CA se ale lze dívat jako na transformační mechanismus, který převádí některé typy autentizačních údajů na X.509 certifikát a klíč. Z pohledu *META Centra* je PKI zajímavé jako prostředek principu single sign-on, kdy se uživatel přihlásí pouze jednou a má kompletní sadu přihlašovacích údajů k různým službám a není nutné aby se v určitém časovém intervalu (zpravidla deseti hodin) autentizoval znovu.

Pro implementaci on-line CA používáme aplikaci MyProxy[7], která má podporu pro SASL. Na základě tohoto mechanismu a příslušného kódu na straně serveru lze realizovat podporu i pro OTP, vývojáři MyProxy například úspěšně demonstrovali použití MyProxy s tokeny Cryptocard. Z pohledu *META Centra* je ovšem důležité mít službu MyProxy zakomponovanou do zbytku prostředí tak, aby se OTP dala použít i pro přístup k jiným službám (zejména získání lístku systému Kerberos). Pro integraci podpory OTP se službou Myproxy proto plánujeme použít Kerberos jako prostředek pro ověřování hesel, kdy se všechna hesla budou přeposílat KDC a případná komunikace od KDC zase klientovi. Tímto způsobem bude možné zavést podporu OTP, která jsou centrálně spravována pro celý projekt. Zároveň uživatelé budou používat autentizační údaje jednotně, bez ohledu na to, zda žádají o kerberovský lístek nebo o X.509 certifikát. Technická realizace bude založena na modulu PAM, podobně jako v případě KDC.

Podpora OTP ve službě MyProxy umožní snadný přístup uživatelů k digitálním certifikátům, které bude možné používat např. pro snadný přístup k WWW zdrojům, kde je PKI přirozeným prostředkem a je podporováno všemi standardními prohlížeči.

4 Závěr

V příspěvku jsme představili systém jednorázových hesel, která nabízí řešení některých problémů se standardními hesly. Výhodou OTP je to, že nabízejí obranu proti odposlechu a lze je tedy použít i v nedůvěryhodném prostředí, především na cizím počítači, kde nevíme, jestli není podstrčený upravený klient, či v internetové kavárně, kde nevíme zda nejsou logovány stisknuté klávesy. V příspěvku jsme také představili návrh integrace OTP do stávajícího distribuovaného systému, který je založen na mechanismu Kerberos. Očekáváme, že zavedení OTP do tohoto prostředí zvýší bezpečnost celé infrastruktury.

Literatura

[1] HALLER, N. *The S/KEY One-Time Password System.* IETF RFC 1760. 1995.

[2] HALLER, N., METZ, C., NESSER, P. STRAW, M. *A One-Time Password System.* IETF RFC 2289. 1998.

[3] Domovská stránka CRYPTOCard Inc. http://www.cryptocard.com/

[4] Domovská stránka RSA Security Inc. http://www.rsasecurity.com/

[5] NEUMAN, B. C., TS' O, T. Kerberos: An Authentication Service for Computer Networks. *IEEE Communications*, roč. 32, s. 33–38, září 1994.

[6] RICHARDS, G. *OTP Kerberos.* IETF Internet-Draft draft-richards-otp-kerberos-00. Expires on December 4, 2006. Work in progress.

[7] NOVOTNY, J., TUECKE, S., WELCH, V. An Online Credential Repository for the Grid: MyProxy. *Proceedings of the Tenth IEEE Symposium on High Performance Distributed Computing (HPDC10).* August 2001.

AJAX

Štěpán Bechynský

E-mail: Stepan.Bechynsky@microsoft.com

V současné době se na Internetu objevují on-line aplikace, které se svou funkčností snaží přiblížit aplikacím klasickým. Klasickou ukázkou je Outlook Web Access z dílny Microsoftu a některé aplikace z dílny Googlu. Myšlenka on-line aplikací je velmi jednoduchá. Uživateli by měl stačit jen počítač s prohlížečkou a připojení do Internetu. Ostatní aplikace si bude spouštět on-line v prohlížeči.

Co je největším omezením pro tento typ aplikací? V prvé řadě se jedná o bezpečnost. Cokoliv spuštěného v rámci prohlížeče nemá, resp. by nemělo mít, přístup k lokálním zdrojům počítače. Další omezení vyplývá z protokolu http a původní myšlenky webových aplikací. Pokud uživatel potřebuje data ze serveru, je jeho požadavek na server odeslán prohlížečem a celá stránka se překreslí podle odpovědi ze serveru. Tím se nesmírně stěžuje vývoj, protože uchovat stav složitých aplikací v prostředí protokolu http je poměrně těžké a náchylné k chybám.

Další typ aplikací, kterých je podstatně více, jsou tzv. Mashups. Jde o aplikace, kde se kombinují informace z různých zdrojů. Typickým zástupcem jsou mapové aplikace. Mapu dodá např. Virtual Earth nebo Google Maps, autor ji vloží do své stránky a dodá vlastní vrstvu. Data pro novou vrstvu mohou opět pocházet z externího zdroje. Takovýmto zdrojem bývají často RSS, rozšířený o uživatelské informace. Zde je největší problém stahování dat z různých zdrojů, tak aby se vše nemuselo nejdříve stáhnout na jedno místo, zde zkombinovat a výsledek odeslat klientovi. Snahou je, co nejvíce věcí kombinovat na klientovi tedy v prohlížeči.

Řešením nejen pro on-line aplikace a Mashups je technologie AJAX vyvinutá firmou Microsoft koncem devadesátých let minulého století. Zjednodušeně řečeno, AJAX umožňuje stahovat data v rámci webové stránky nezávisle na prohlížeči. Typické použití jsou tzv. našeptávače. Do textového políčka píšete text, který je průběžně odesílán na server a zpět dostáváte nápovědná slova, která byste mohli chtít napsat. Tato slova se zobrazují jako seznam pod textovým polem.

AJAX – Asynchronous JavaScript and XML – vyžívá objekt XMLHttpRequest instalovaný na straně klienta, který je součástí prohlížeče. JavaScript ve stránce může pomocí XMLHttpRequest zavolat server a získat od něj potřebná data v reakci na akci uživatele. Získaná data opět zpracuje JavaScript

na straně klienta bez nutnosti překreslovat celou stránku. Prohlížeč se o tom v podstatě nedozví. Volání je asynchronní, takže webová stránka dál reaguje na vstup uživatele, i když stažení dat trvá delší dobu. Zde narážíme na první problém technologie AJAX, která vyžaduje rychlou odezvu ze strany serveru, aby uživatel zbytečně nečekal a dojem z aplikace se co nejvíce blížil aplikaci klasické.

Ukázka kódu:

```
var request = null;

if (window.ActiveXObject)
{
        request = new ActiveXObject("Microsoft.XMLHTTP");
}
else
{
        request = new XMLHttpRequest();
}

request.onreadystatechange = readData;

request.open(‘ GET‘ , ‘ http://localhost/ServerTime.ashx‘ , true);
request.send(null);

function readData()
{
        if (request.readyState != 4)
        {
                return;
        }

        var serverTime = request.responseText;
        document.getElementById(‘ time2‘ ).innerHTML = serverTime;
}
```

Data získaná ze serveru lze interpretovat jako text nebo jako XML. Objekt XMLHttpRequest metodou `responseXML` vrací přímo objekt DOM document.

Ukázka kódu:

```
...
var xml = request.responseXML;

var items = xml.getElementsByTagName(‘ item‘ );

for (i = 0; i < items.length; i++ )
{
        var item = items.item(i);
```

```
        ...
}
...
```

Výhody

Asi jedinou prokazatelnou výhodou, kvůli které se AJAX používá, je možnost stahovat a zobrazovat data bez překreslení celé stránky. Na to navazují další výhody, které ovšem nemusí vždy platit. Záleží, co aplikace dělá a jak je napsána. Jedná se zejména o menší přenos dat a menší počet dotazů do databáze. Jednou stažená data zůstávají na klientovi a stahují se jen data nová.

Nevýhody

Protože prohlížeč neví, že si stránka stahuje nová data, zůstává adresa stránky pořád stejná. Tím pádem tlačítko zpět prohlížeče a záložky nefungují, jak uživatel očekává. Při neuváženém použití může dojít k významnému nárůstu komunikace mezi klientem a serverem.

S pohledu vývojáře

Největším problémem, z mého pohledu, je nedostatečná podpora JavaScriptu ve vývojářských nástrojích. Pokud chci vyvíjet, potřebuji mít možnost debugovat kód, nastavit breakpoints a vypisovat obsah proměnných. Další problém, který je z části řešitelný, je nekompatibilita prohlížečů. Vytvoření objektu XMLHttpResponse je pro jednotlivé verze prohlížečů různé a práce s objekty stránky také. Tento problém se snaží řešit projekty, které generují AJAX kód na straně serveru podle prohlížeče na straně klienta.

Quo vadis, SI?
aneb Lesk a bída softwarového inženýrství

Karel Richta

E-mail: richta@fel.cvut.cz

Abstrakt

Příspěvek je podkladem pro diskuzi na téma „softwarové inženýrství“. Pokouší se vymezit obsah termínu „softwarové inženýrství“ a porovnat jej s ostatními inženýrskými disciplinami, jako je např. stavební inženýrství. Lze si položit otázku, zda existuji specifické postupy, které odlišují softwarové inženýrství od ostatních inženýrských disciplin. Renomovaní autoři se k těmto otázkám občas vrací, neboť řada principů je stále platných, jiné naopak zmizely v propadlišti dějin. Softwarové inženýrství bude brzy slavit čtyřicátiny. Zajímavá otázka je, jak se softwarové inženýrství za tuto dobu vyvinulo? Jaké trendy či změny jsou nejpodstatnější? Jiným zajímavým hlediskem je otázka, proč se některé trendy prosazuji s určitým zpožděním, pokud je softwarové inženýrství srovnatelné s ostatními inženýrskými disciplinami. Co je v současnosti základní problém, se kterým se softwarové inženýrství potýká? Je možné se na vývoj software dívat jako na uměni? Poučili jsme se již ze zkušeností, podobně jako se to stalo v jiných inženýrských disciplinách? To jsou témata, která by chtěl tento příspěvek navodit.

1 Úvodem trochu historie

Za okamžik zrození termínu „softwarové inženýrství“ (SI) se obvykle považuje rok 1968, kdy NATO sponzoruje první konferenci na toto téma [7]. Konala se ve Spolkové republice Německo ve známém středisku Garmisch-Partenkirchen a řídil ji profesor Bauer z Mnichovské techniky. Účastnilo se jí asi 50 odborníků z různých oblastí, z praxe i ze škol. Termín „softwarové inženýrství“ byl vybrán úmyslně jako provokativní – naznačující, že produkce software musí přejít na jiné postupy a být podložena teoretickými disciplinami, podobně, jako je tomu u inženýrského přístupu v jiných oborech. Její účastníci formulovali směry, kterými by se výzkum v oboru SI měl ubírat. Byly ustaveny pracovní komise pro

zkoumání těchto směrů a účastníci se snažili definovat jejich náplň. Technologický obsah se pak snažili vymezit na následující konferenci, která se konala o rok později v Římě [11]. Význam termínu softwarové inženýrství formulovali následovně [1]:

> *„Softwarové inženýrství je disciplina, která se zabývá zavedením a používáním řádných inženýrských principů do tvorby software tak, abychom dosáhli ekonomické tvorby software, který je spolehlivý a pracuje účinně na dostupných výpočetních prostředcích.“*

Hlavní hnací silou pro uspořádání této konference a vůbec pro vznik disciplíny softwarového inženýrství byl vzrůstající počet neúspěšných projektů. Řada projektů nebyla dokončena vůbec, jiné nebyly dokončeny včas, příp. přesáhly původní odhad rozpočtu.

Některé nedostatky v projektech byly dokonce příčinou známých katastrof, např. sonda Mariner I., která měla letět k Venuši, ale nedoletěla. Příčinou byla chyba při ručním přepisu specifikace do kódu. Drobné přehlédnutí kodéra způsobilo, že se místo derivace rychlosti uvažovala rychlost, což způsobilo chybu v navigaci a sonda musela být destruována. Podobný známý případ se stal v projektu Mercury, kde šlo o syntaktickou chybu – v následujícím příkazu v jazyce FORTRAN:

```
DO 17 I = 1,10
```

byla zaměněna čárka za tečku:

```
DO17I = 1.10
```

a příkaz byl interpretován jako přiřazení do proměnné `DO17I` (mezery zde nejsou podstatné). Odpovídající větev programu nebyla nikdy otestována, což mohl být zdroj katastrofy, ale problém se podařilo objevit dříve, než došlo k neštěstí.

Srovnáme-li to se situací, kdy byla spuštěna na vodu loď Vasa a potopila se, je zde drobný rozdíl – v době jejího spuštění neexistovaly ještě formulace zákonů, podle kterých by neštěstí bylo možno zabránit. Dnes můžeme přesně vypočítat, jak by měla být loď konstruována, aby byla schopna plavby. V 17. století ale příslušné zákony známy nebyly a byly používány tabulky s rozměry, které se osvědčily v minulosti. Pikantní je, že ze současných dokumentů víme, že plány na stavbu lodi Vasa byly změněny poté, co se na lodi začalo pracovat (změna specifikace za chodu). Král chtěl mít na palubě více děl než obvykle, loď byla postavena s vysokou nástavbou, se dvěma palubami pro děla. Pro vyvážení bylo dno lodi naplněno velkými kameny, které sloužily jako zátěž udržující loď stabilně na vodě. Ale Vasa byla příliš vratká a v prvém větším závanu větru se potopila. Přestože dnes řadu zákonů známe, chyby děláme dále, viz např. jedna z prvních verzí rakety Ariane-5, která nedoletěla kvůli numerické chybě.

Nedokončené projekty a neschopnost tvůrců zlepšit efektivitu tvorby způsobily velké zklamání. Tomuto jevu se obvykle říká „softwarová krize“. V řadě případů za problémy stál nedostatek profesionální discipliny, své ale sehrála také neznalost důležitých principů. Tvorba software je kreativní proces, který samozřejmě velmi závisí na kvalitě zúčastněných. Přesto nezáleží jen na nich, bez patřičné discipliny a dokumentace nelze očekávat úspěšné řešení zejména větších projektů, ani uvažovat např. o opakovaném využívání komponent, apod.

Snaha o zavedení discipliny do tvorby software znamená mimo jiné zavedení a využívání modelů, nových jazyků, metodik a nástrojů, tak aby se odstranily nedostatky v komunikaci, řízení apod. V letech 1968–1980 se formulovaly definice životního cyklu, vodopádový model a techniky pro jednotlivé fáze tohoto cyklu. Vzniká strukturovaná analýza, strukturovaný návrh a strukturované programování. Principem je využívání omezené sady strukturovaných konstrukcí, které mají jeden vstup a jeden výstup a o nichž je možno něco dokázat. Nahrazují tzv „špagetový kód“, kde se poskakuje libovolně z místa na místo. Ovlivňuje to programovací jazyky jako Pascal, či C. Později vznikají jejich objektová rozšíření. Vznikají techniky pro popis uživatelských požadavků. Vznikají modely pro odhad nákladů, rizikové faktory, metriky. Vznikají databázové technologie.

Léta 1980 až 1990 se zabývají formalizací a snahou využívat co nejvíce automatizaci zejména při zpracování analýzy. Zkouší se prototypování, spirálový model vývoje, neboť vodopádový model byl opět zavrhnut. Ukázalo se, že jeho použití na celý projekt příliš prodlužuje čas mezi zadáním a odevzdáním produktu. Vznikají modely pro specifikaci dynamického chování. Začínají vznikat CASE nástroje, v těchto letech zejména systémy pro strukturované modely. Vzniká internet jako médium.

V létech 1990 až 2000 se formuluje tzv. „softwarový proces“, modely CMM (Capability Maturity Model), opět se na scénu vracejí objektově-orientované technologie, začínají se využívat distribuované výpočty. Jako technologická novinka vznikají např. návrhové vzory, aplikační stavebnice (frameworks), definují se technologie používající softwarové komponenty, CORBA, RPC, RMI. Nové IT technologie v řadě případů předbíhají vývoj softwarového inženýrství, které někdy zaostává a snaží se tyto technologie dohonit. V CERNu vzniká projekt WWW, vzniká jazyk HTML, protokol HTTP a první webová stránka.

Léta současná lze charakterizovat tzv. webovým inženýrstvím. Lze nalézt určitou podobnost ve vývoji technologie webových aplikací, která připomíná 60-léta a softwarovou krizi [5]. Vývojáři dnešních webových aplikací obecně patří do mladší generace, jsou obeznámeni s novými technologiemi a prožívají svou suverenitu a nadřazenost podobně, jak to prožívali jejich předchůdci v šedesátých letech. Nutně se jim musí jevit současné systémy a aplikace jako obecně těžkopádné, zatížené dědictvím různých zastaralých technologií. Podobnost těchto scénářů varuje před nebezpečím, že vývojáři webových aplikací mohou opět opakovat chyby svých předchůdců. Programátoři šedesátých let obecně podceňovali

starší systémy a respektovali pouze svou technologickou nadřazenost. Nevěnovali dostatečnou pozornost potřebám uživatelů, nevěnovali se vývoji formálních metod pro budování a testování aplikací.

Další novinky současnosti jsou servisně-orientované architektury (SOA), velký rozvoj prožívá specifikací řízený styl vývoje, označovaný jako MDD (Model Driven Development), či MDA (Model Driven Architecture). Podstatou je, že se snažíme pracovat s modelem, kód z něj odvozujeme automatizovaně, příp. pomocí reverzního inženýrství synchronizujeme model a kód. Pro datové modely se již tato technologie často využívá, pro kód se rozvíjí. Tvůrci jednoduchých webových aplikací ji obvykle považují za nereálné monstrum.

2 Softwarová krize

Dokud výkon počítačů nepřesáhl určitý rozměr, bylo možno se spolehnout na programátorské „hvězdy". Často se počítače se využívaly pro vědecko-technické výpočty, kde záleželo spíše na preciznosti řešení, než na efektivitě tvorby programů. Podle tzv. Moorova zákona ale vzrůstá výkon hardwaru zhruba dvakrát za dva roky a přestože sám autor prohlásil svou extrapolaci jako „pěkně divokou", zákon zhruba platí dodnes a firma Intel nedávno zveřejnila výsledky výzkumné zprávy uvádějící, že Moorův zákon pravděpodobně přestane platit až kolem roku 2021 (křemík se dostane na hranici svých možností).

S nástupem počítačů třetí generace v polovině 60. let rostou možnosti počítačů a rovněž jejich spolehlivost. Přichází tak možnost uvažovat o zcela nových aplikacích, např. v bankách, pojišťovnách, rezervačních systémech. Vzniká potřeba řešit mnoho nových projektů, často mnohem většího rozsahu, než dříve. Metody tvorby software se ale oproti vývoji hardwaru nemění zdaleka tak rychle, tvorba software narůstá zhruba lineárně, jak přibývá programátorů. Jak říká Dijkstra [2]:

> *„Hlavní příčinou softwarové krize byl nárůst výkonu hardware. Neomaleně řešeno, programování nemělo problémy, dokud neexistovaly počítače. Dokud jsme měli slabé počítače, mělo programování jen snesitelné těžké problémy. Nyní máme gigantické počítače a k nim gigantické problémy se softwarem."*

Příčinou softwarové krize vždy byl nesoulad mezi složitostí vytvářeného produktu a relativní nedostatečností a nezkušeností softwarové profese. Tento rozdíl způsobuje rozevírání nůžek a důsledkem pak jsou softwarové krize. Krize se projevují v následujících bodech:

- Projekty překračují rozpočet.
- Projekty překračují čas.

- Software nemá dostatečnou kvalitu.
- Software neodpovídá požadavkům.
- Projekt není dobře řiditelný a software je obtížně udržovatelný.

Při pohledu na tyto body a současné projekty se zdá, že softwarová krize trvá stále. Svůj vliv zde má též výjimečné postavení softwarových expertů, kteří jakkoli jsou geniální, mohou přímo mentálně obsáhnout jen určitý rozsah projektů. Řešení velkých projektů nelze proto nechat pouze na nich. Je nutno použít standardní techniku řešení obtížných problémů – „rozděl a panuj“ – velký problém je třeba rozdrobit na problémy menší.

Známý příklad je projekt vývoje operačního systému OS/360, který se začal zpožďovat v době, kdy řešitelský tým čítal cca 200 osob. Přestože se ve snaze urychlit vývoj později se zvýšil až na 1 000 osob, urychlit se jej nepodařilo. Šéf týmu Fred Brooks z toho dokonce odvodil zákon [11], který vyjadřuje komunikační potřebu v nestrukturovaném řešitelském týmu. Z něj vyplývá dokonce možnost záporného nárůstu výkonu při najmutí nových sil, neboť se tím zvětší potřeba komunikace. Hrubá interpretace tohoto zákona by mohla znít:

„Přidání nových kapacit na zpožděný projekt prodlouží jeho řešení.“

Všechny tyto problémy související se softwarovou krizí vedly tedy nakonec k pokusu udělat z vývoje produkovaného nadšenci inženýrskou disciplinu. V 70. letech dochází k formulaci základních principů tohoto oboru. Vzniká první generace nástrojů pro podporu této discipliny, které jsou označovány jako CASE (Computer Aided Software Engineering).

CASE nástroje prvé generace podporují zpravidla jeden krok ve vývoji, např. modelování dat. Druhá generace těchto nástrojů vzniká v 80. letech – jsou to nástroje, které se snaží podporovat více kroků ve vývoji, až po současné integrované nástroje, které v řadě případů podporují celý vývojový cyklus.

Velkým přínosem pro tuto oblast byl vznik unifikovaného vyjadřovacího jazyka UML (Unified Modeling Language). Ten představuje jakési esperanto oboru a tvůrci nástrojů se mohou opřít o notaci, která by měla být obecně srozumitelná. Součástí definice UML jsou i výměnné formáty, umožňující přenos a sdílení modelů.

Institucionálně se obor softwarové inženýrství ustavuje v roce 1993 [8], kdy nejprve IEEE a později i ACM vytvářejí komise pro ustavení SI jako nové profese. Obě komise se nakonec v roce 1994 spojily a zabývaly se definicí náplně profese softwarového inženýrství. Teprve v roce 1997 je v USA softwarové inženýrství certifikováno jako obor [1].

Existuje velká sada norem pro softwarové inženýrství – viz např. [12, 2]. Základní terminologie je definována standardem IEEE 610.12. Pak následuje celá

dlouhá sada norem, popisujících jak jednotlivé kroky ve vývoji, tak i např. primitivní data a jejich význam. Jako příklad lze uvést normu na reprezentaci pohlaví lidí – ISO/IEC 5218: Information technology – Codes for the representation of human sexes.

3 Je vývoj softwaru umění, věda nebo rutina?

Softwarové inženýrství má blízko k různým disciplínám. Na jedné straně je možné jej považovat za inženýrství, neboť se jedná o disciplinované využívání pragmatických zkušeností, tj. postupy, které se očekávají od inženýra. Srovnáme-li tento pohled s inženýrem stavebním, pak stavební inženýr realizuje stavbu podle modelu, programátor programuje podle modelu. Chemický inženýr navrhuje postup výroby nějaké látky z ingrediencí, softwarový návrhář navrhuje skladbu celku z rozmanitých komponent.

Na druhé straně softwarový architekt podobně jako inženýr architekt komponuje celkovou architekturu softwarového díla. Zde lze pohlížet např. na konceptuální modelování jako na určitou formu umělecké činnosti. Návrh grafického rozhranní připomíná práci návrháře technických výrobků a má svou estetickou a uměleckou hodnotu. Pojem „dobrý program“ je velmi subjektivní, a je často hodnocen více citem, než měřením. Konec konců jedno ze základních děl v oboru programování od Donalda Knuta se jmenuje „Umění programovat“.

Do značné míry má softwarové inženýrství společné rysy s matematikou a logikou. Konceptuální modelování stejně jako matematika používá exaktní metody a formální postupy, účelem ale není model sám, spíše jeho pragmatické vlastnosti. Programy mají matematicky odvoditelné vlastnosti, např. konečnost použitého algoritmu. Praktické ověřování těchto vlastností je ale často velmi obtížné. Programování je matematická disciplína v tom smyslu, že program je vlastně konstruktivní důkaz že vstupy a výstupy jsou v jisté relaci, dané specifikací problému, který program řeší.

Tvorba programů má také řadu rysů společných s tovární výrobou. Firmy produkují jeden typ aplikace (např. webové stránky) jako na běžícím pásu. Jen charakter spolupráce u běžícího pásu pohybuje není stejný, neboť nefunguje tak, že jeden programátor přidá do aplikace tlačítko, další programátor vytvoří pro tlačítko popisek, atd. některé metodiky se snaží k tovární výrobě přiblížit, neboť je to efektivní.

V neposlední ředě do softwarového inženýrství patří také řízení projektů. Řada činností souvisejících s vytvářením softwarového produktu je obdobná činnostem při řízení projektů v jiných oborech. Realizace stavby má mnoho společného s realizací softwarového díla, neboť je třeba naplánovat kapacity, zajistit koordinaci subdodavatelů apod.

Pokusíme-li se srovnat práci softwarového inženýra např. s inženýrem chemickým, zjistíme některé podobnosti. Chemičtí inženýři již dávno pochopili, že procesy, které byly použity v laboratoři nemusí fungovat v továrně. Navíc, přechod z laboratoře do továrny obvykle nelze učinit jedním skokem – je třeba provádět pokusy v jakýchsi pilotních verzích továrny, aby se získalo dosti zkušeností s většími objemy a nechráněným prostředím. Tvůrci programových systémů jsou postaveni před podobný problém, zdá se však, že jej zatím neuchopili.

4 Softwarové týmy a softwarové profese

Jedním z projevů přechodu od ruční výroby k manufaktuře je definice softwarových profesí. Řešení velkých projektů vyžaduje spolupráci mnoha řešitelů a práci je nutno rozdělit. Dělba práce vyžaduje organizaci týmů řešících větší softwarové projekty. Týmy lze organizovat jako strukturované nebo nestrukturované. Nestrukturované týmy dělí práci podle objemu:

- „Osamělí vlci“ – geniální programátor není schopen s nikým kolaborovat, ani komunikovat, vše řeší sám. „Ohmův zákon“ jim netřeba sdělovat, vymyslí si jej sami.
- „Horda“ – organizace, která je založena na předpokladu, že pokud jeden parník pluje z Londýna do New Yorku čtyři dny, pak dva parníky tam plují dny dva.
- „Demokratická skupina“ – efektivní organizace, pokud jsou všichni ochotni se dohodnout a podřídí se celkovému cíli.

Strukturované týmy využívají dělby práce podle profese. Mohou být organizovány jako:

- „Chirurgický tým“ – vše se točí kolem chirurga, všichni jsou připraveni mu podat skalpel. Velmi výkonná organizace, ale poněkud drahá.
- „Tým hlavního programátora“ – jedná se o variantu chirurgického týmu, kdy hlavní řešitelé představují tandem (aby si byli schopni oponovat náměty a byli zastupitelní) a ostatní profese si najímají dle potřeby.
- Více-týmová organizace – pro projekty většího rozsahu je nutno rozdělit řešení na více částí. Známé psychologické pravidlo 7 ± 2 určuje přibližné rozsahy týmů.

Volba organizace je dána rozsahem projektu. Pro větší projekty se samozřejmě hodí strukturované týmy, které využívají různé profese. Ne každý v týmu dělá všechno, určité činnosti vykonávají specialisté. Na větší projekty je třeba více-týmová organizace, máme-li dost prostředků, je nejvýkonnější chirurgický tým.

5 Oddělení programů od dat – databáze

Jedním z důležitých důsledků snahy řešit softwarovou krizi a formulace požadavků na softwarové inženýrství je princip oddělení správy dat od programů, které s nimi pracují. Opět se jedná o využití techniky „rozděl a panuj“ – správa dat má na starosti bezpečné uložení dat s minimálním rizikem jejich ztráty. Programy operující nad databází se naopak zaměřují na aplikační logiku. Databázové systémy představují odvětví softwarového inženýrství, které vzniká opět konce 60. let, doby vzniku prvých hierarchických databází. Jejich nevýhody se později pokoušejí odstranit databáze síťové. Prostou záměnou ukazatelů (adres) klíčovými položkami dospíváme k relačním databázím. Článek pana Codda pochází opět z roku 1970, byť prvé produkční implementace jsou k dispozici až kolem roku 1980.

V letech devadesátých vznikají databáze objektové, umožňující persistentní uložení nejen dat, ale i metod. Jejich využívání však zaostává za databázemi relačními. Důvodem možná je právě čas uvedení na trh, také ale jednodušší práce s daty v relačním modelu, pokud vlastnosti objektové databáze nevyužijeme.

Kompromisem jsou databáze objektově-relační, které se snaží umožnit využívat výhod obou přístupů. V souvislosti s rozšířeným využíváním technologie XML vznikají nativní XML databáze, do jisté míry varianta na téma hierarchické databáze.

6 Opakované použití a vzory

Opakované použití řešení jednou vyřešeného problému přináší na jedné straně nesporně mnohem vyšší efektivitu. Má ale svá úskalí v tom, že opakované užití předpokládá nejen správné vyřešení vzorového případu, ale i jeho precizní dokumentaci – vytváření vzorů.

Vzory lze vytvářet na mnoha různých úrovních. Na úrovni programovacího jazyka existují různé opakovaně použitelné techniky – např. funkce, která vrací hodnotu typu T bude velmi pravděpodobně potřebovat proměnnou typu T, ve které se bude kumulovat výsledek.

Na úrovni návrhu lze využívat obecnějších vzorů – např. pro sekvenční průchod kolekcí položek lze vytvořit vzor „iterátor“, který řešiteli napoví, jak se to obvykle dělá. Na úrovni architektury systému lze připomenout známé vzory klient/server, třívrstvá architektura, virtuální stroj apod. Možná nejúčinnější jsou vzory analytické, které vlastně představují ucelené řešení skupiny problémů. Za příklad poslouží třeba vzorové řešení jednoduchého účetnictví. Informační komunita by měla požadovat od tvůrců zákonů přesnou specifikaci takových vzorů. Jejich širší využití je věcí blízké budoucnosti.

Nelze se domnívat, že využívání vzorů vyřeší všechny problémy, ale na druhé straně to jistě přispívá k projasnění řešení. Opakovaně využívaný vzor zvyšuje robustnost řešení, snižuje pravděpodobnost chyby.

7 Výuka SI

Jak jsme již konstatovali, institucionálně se obor softwarové inženýrství ustavuje v roce 1993 [8], kdy nejprve IEEE a později i ACM vytvářejí komise pro ustavení SI jako nové profese. Obě komise se nakonec v roce 1994 spojily a zabývaly se definicí náplně profese softwarového inženýrství (The Body of Knowledge and Recommended Practices), etickým kodexem a standardy (The Code of Ethics and Professional Standards) a pravděpodobně nejznámějším výsledkem jejich činnosti je definice přepokládané sady znalostí softwarového inženýra, tzv. SEEK (Software Engineering Education Knowledge), která představuje jakousi šablonu pro vytváření osnov pro výuku tohoto oboru [14], viz též. (Guide to the Software Engineering Body of Knowledge – SWEBOK).

Jedná se pokračování aktivit, které začaly nejprve společným programem pro informatiku a počítačové inženýrství (Computing Curricula 1991 – CC2001, Computer Science and Computer Engineering). Později pokračovala inovovaným programem CC2001 (později označovaným jako informatika – Computer Science, CS2001), na které navázaly snahy o přesnější definici dalších podoborů informatiky – informačních systémů (IS2002), softwarového inženýrství (SE2004), počítačového inženýrství (Cumputer Engineering CE2004) a informačních technologií (IT2006). Všechno by měl zastřešovat společný dokument CC2005. Počítačové inženýrství se chápe jako disciplina zahrnující tvorbu softwarových i hardwarových komponent – jako kombinaci informatiky (CS – Computer Science) a elektrického inženýrství (EE – Electrical Engineering). Struktura informatiky se trochu mění a bývá zobrazována následovně:

	Před rokem 1990	Po roce 1990
Hardware	EE + CE	EE + CE
Software	CS	CE + CS + SE
Organizace	IS	IS + IT

Prvý bakalářský program v obosu SI vypsalo v roce 1996 vysoké učení technické v Rochesteru. Zpočátku byl odmítnut, ale později akreditaci získalo v roce 2003 spolu s inženýrskou školou v Milwaukee. Prvý inženýrský program byl vypsán již v roce 1979 na universitě v Seattlu, kde v roce 1982 udělili prvý titul v tomto oboru.

8 Čekají nás další softwarové krize?

Jak jsme se již zmínili dříve, byla softwarová krize konce 60. let způsobena několika faktory – výkonnost hardwaru přerostla schopnosti programátorů té doby. Techniky používané pro komunikaci a specifikaci cílů, odhady požadavků na zdroje, techniky plánování, používané technologie tvorby, vše nestačilo novým potřebám. Pravděpodobně lze na situaci aplikovat Parkinsonův zákon – „Každý člověk zastává místo, na které nestačí. Člověk nastoupí do organizace a zastává určité místo. Pokud se osvědčí, znamená to, že na toto místo stačí a je povýšen na místo vyšší. Takto se postupně dostane na místo, na které nestačí a již není dále povyšován." Podobně se každá metodika tvorby softwaru postupně s vývojem možností hardwaru dostane do situace, kdy se již nehodí.

V současné době internetových aplikací existuje mnoho nástrojů pro přípravu kódu, usnadňující tvůrcům aplikací práci. Programování, nebo možná lépe vytváření kódu, se tak dostává do rukou mnohem více lidem s různými znalostmi. Navíc platí zásadní princip:

„Kdo je na trhu první, obvykle vyhrává."

Tento princip aplikuje řada firem, včetně těch největších. Používá se i na situace, kdy se její použití tak úplně nehodí, uživatelé pak fungují jako betatestéři.

Do jisté míry to má u internetových aplikací opodstatnění – budoucí uživatele systému ještě úplně neznáme, někdy ani neznáme jejich požadavky. Snažíme se ale ovládnout trh, tak raději předhodíme uživatelům cosi jednoduššího, co ale už funguje. Postupně pak budeme aplikaci doplňovat o další možnosti. Problém začne, když uživatelé začnou od nových aplikací vyžadovat složitější a komplexnější služby. Pak pravděpodobně opět přijdou na řadu specialisté a potřeba sofistikovanějších aplikací.

Zopakujme si jaké problémy stály za softwarovou krizí. Byla to jednak špatná komunikace mezi zúčastněnými na všech úrovních (zákazník, analytik, návrhář, programátor, testér, manažer). Dnes je nutnost a význam komunikace znám, většina metodik komunikaci podporuje a některé dokonce pro ni rezervují speciální roli.

Jiný zdroj potíží byl nesprávný přístup k vývoji, kdy se vývojář snažil vytvořit „umělecké" dílo, nezáleželo mu tolik na spokojenosti zákazníků. Dnes jsou v řadě případů zákazníci zatahováni do procesu vývoje a jejich spokojenost je primární (někdy možná až příliš). Také dělba práce je mnohem lépe propracovaná, členové týmu zastávají různé role a každý by měl v danou chvíli provádět to, co přináší z hlediska souhry týmu největší užitek.

Dalším problémem bylo nesprávné plánování a špatné odhady. Tyto nedostatky se v současnosti snažíme odstranit propracovanými postupy a metodami.

Přesto jsou asi správné odhady stále trochu magické a hodně záleží na zkušenosti a umu. Důležité je včasné docenění hrozeb a rizik, neboť jejich podcenění se později nevyplatí. Některé metodiky se proto nechávají rizikem řídit – snažíme se nejprve odstranit a vyřešit nejvíce rizikové části.

Nové technologie většinou přinášejí nové možnosti, je proto třeba je sledovat a studovat. Nelze si ale představovat, že za nás vyřeší všechny problémy a potíže. Vhodná volba technologie je pouze součástí řešení, které může případně usnadnit.

9 Quo vadis, SI?

Další vývoj v oboru softwarového inženýrství závisí na více faktorech. Jedním z nich je zatím stálý růst výkonu hardware, se kterým lze počítat ještě pro příští dekádu. Pak se může objevit jiná technologie, než je křemík (např. biologické procesory), nebo jiné paradigma (např. kvantové počítače). Takový obrat přinese do produkce software jistě potřebu nových konceptů. Pokud se nová technologie neobjeví, bude opět nutno přejít od expanzivních metod k metodám intenzivním – zvyšování schopností bude zapotřebí dosáhnout zlepšením struktury, kódu apod. Současný software v řadě případů plýtvá pamětí, či časem, neboť tyto potřeby zaštítí výkonnější hardware. Pokud se vývoj hardware pozastaví, bude třeba sáhnout do rezerv. Kdyby se výkon hardware naopak dramaticky zvýšil, bude možno opět uvažovat o nových aplikacích a metodách, což by opět rozevíralo nůžky softwarové krize.

Další faktor pro směřování softwarového inženýrství jsou metodiky. Zde si můžeme pro další vývoj softwarového inženýrství lze stanovit některé cíle [4]:

- Musíme být schopni vytvářet softwarové komponenty, jejichž chování jsme schopni specifikovat a pochopit, a které tedy budeme schopni využívat v jiných komponentách a aplikacích. Jako příklad lze uvést např. komponentu zajišťující provedení platby.

- Musíme být schopni prostřednictvím software poskytovat služby, kde samozřejmě důležitá je nejen poskytovaná funkčnost, ale i další nefunkční charakteristiky – rychlost provedení, doba odezvy apod. Celkového chování služby dosahujeme kooperací použitých komponent, služba tedy může být zprostředkovaná. Takovou službou může být provedení platby prostřednictvím komponenty zajišťující provedení platby.

- Musíme být schopni tyto služby vytvářet takovým způsobem, aby byly připravené na přicházející změny a byly schopny se těmto změnám přizpůsobit, přičemž adaptaci bude možno jednoduše spravovat. Např. pokud přibude nový způsob platby, komponenta pro zajištění plateb by měla být schopna integrovat tento nový způsob na základě jednoduchého pokynu.

- Musíme být schopni podporovat nové struktury a schémata, nové metody pro rozmanité nové aspekty ve vývoji softwaru.

- Musíme být schopni přizpůsobit dobré vlastnosti existujících metod a nástrojů pro nová prostředí, nové agilní, příp. i extrémní metodiky a vůbec neklasické použití, neboť je o čas.

Z hlediska produkce softwaru přicházejí nová paradigmata ovlivňující vývoj softwaru:

- Vytvářej software tak obratně, aby se prodával v co největších kvantech.

- Vysoká kvalita a spolehlivost software vedou k nižší ceně, de-facto standardům, zvyšují přidanou hodnotu.

- Vyráběj z prefabrikátů formou „Vyber a kombinuj".

- Vytvářej produkty kolaborativně, využívej globální kooperace a konkurence.

Jaké zbraně máme k dispozici:

- Porozumění současnému stavu přináší lepší znalost. Kvalita služby je vždy ovlivněna znalostí, která za službou stojí. Je třeba využívat služeb, za kterými stojí zkušenost.

- Životní cyklus softwarových systémů se zkracuje. Potřebujeme technologie, které podporují postupný vývoj, přidávání dalších komponent a „zásuvných modulů" (plugins). Měli bychom vytvářet technologie a metodiky, které jsou schopny pracovat s „černými skříňkami", případně s nespolehlivými komponentami.

- Zařídit administrativní prostředí, které podporuje inovace obchodní i technické.

- Podporujme typy lidí, kteří jsou schopni absorbovat tyto aktivity. Podporujme výukové systémy, ve kterých se propaguje kooperativnost, distribuované kognitivní myšlení.

- Vývoj softwarových metodik pro vývoj softwaru evolucí. Metodiky orientované na určitou doménu, orientované na architekturu. Využívání softwarových vzorů, zejména analytických vzorů, návrhových vzorů a idiomů. Prohlubujme znalosti objektově-orientovaného a funkcionálního programování. Vytvářejme prostředí pro kooperativní distribuovaný vývoj v počítačové síti.

- Podpora školení pro lidí z průmyslu tak, aby pochopili, že vytvářený software je má podporovat v jejich běžné práci, že není účelem, ale prostředkem.

Podle analytiků z Gartner Group [9] lze v následujícím desetiletí předpokládat změny v profesi IT. V roce 2010 to postihne šest z deseti IT profesionálů, velké firmy budou redukovat IT. Bude vznikat potřeba nového druhu IT profesionálů, kteří budou ovládat nejen technologie, ale zejména budou rozumět obchodním procesům. Nové cíle IT profesionálů lze formulovat asi takto: „Strávil jsem dva roky tím, že jsem pomáhal navrhnout a realizovat proces prodeje přes internet, což přineslo zvýšení obratu o 20%“.

Gartner radí IT profesionálům, aby si probrali své znalosti a soustředili se na jejich vylepšování směrem ke zkušenosti v určité oblasti. Předpokládá vzestup požadavků na všestrannost – přednost budou mít ti, kteří jsou schopni syntetizovat znalosti a přinášet novou hodnotu v širším okruhu. Primární hnací silou pro tyto změny jsou

- Globální zdroje – díky vysokorychlostním globálním sítím s možností vyhledávání znalostí a služeb, stanou se globální zdroje standardní součástí portfolia a vznikne tím pro mnoho IT profesionálů konkurence.
- Automatizace v oblasti produkce a nasazování software rovněž změní produkci a vývoj softwaru a ovlivní související profese. Např. automatizace testování, monitorování vzdálených systémů, zajištění technické podpory apod.
- Mnohem větší význam bude mít konzumní využívání služeb. Prostřednictvím technologií jako jsou osobní počítače, dostupné služby, mobilní telefony apod. budou služby požadovány větším okruhem méně znalých lidí.
- Restrukturalizace obchodu – slučování firem, akvizice, odhalování rezerv, konsolidace, řešení nezaměstnanosti, využívání vnějších zdrojů, bankroty firem – to vše by mělo přivést IT profesionály k otázce, zda mají zůstat u „čisté technologie“, nebo se pokoušet rozšířit znalosti do oblasti průmyslu, obchodu, porozumět základním procesům, což jim přinese výhodu všestrannosti v této doméně.

Za primární oblasti využívání softwarových technologií považuje Gartner [9]:

- Rozvoj infrastruktury a služeb (hardware, sítě, služby s tím související).
- Vzrůstající požadavky na „obchodní inteligenci“ (business intelligence), poskytované on-line služby.
- Návrh procesního zpracování, vylepšování obchodních a technických procesů, případně jejich automatizace.
- Správa distribuovaných vztahů a zdrojů.

10 Závěr

Softwarové inženýrství jako disciplina vzniká v 60. letech, kdy se poprvé projevilo zpoždění vývoje softwaru za vývojem možností hardwaru. Důsledkem byla softwarová krize a reakcí na ni snaha zavést do tvorby software inženýrské postupy. Inženýrským přístupem se obecně myslí opodstatněné využívání nových vědeckovýzkumných poznatků, disciplinovaný postup při řešení konstruktérských problémů.

Softwarové inženýrství přináší v letech 70. nové postupy, jazyky a nástroje, někdy úspěšné, někdy méně. To bylo asi příčinou toho, že nástup osobních počítačů v letech 80. nezpůsobil další velkou krizi, přestože se software v řadě případů vrací o několik desítek let zpátky – např. operační systémy začínají v podstatě znovu.

V 90. letech přichází internet a opět se tvorba software vrací o něco zpět, neboť se mění styl využívání a roste okruh uživatelů. Navíc nové jazyky a vizuální nástroje umožňují vytvářet programy mnohem většímu okruhu uživatelů.

Co přichází na počátku tisíciletí? Asi lze za takový nový prvek považovat architektury orientované na služby – aplikace řeší problémy distribuovaně, ve spolupráci s poskytovateli různých služeb. A co v následující dekádě? Odhady se různí, zdá se ale, že jedním z hlavních směrů bude konzum – ať se nám to líbí, či nikoliv, nové prostředky pro přístup k informacím a službám jako jsou mobilní zařízení apod. budou využívána skupinou lidí, kteří nemají žádný zájem na softwarovém vývoji, chtějí pouze konzumovat poskytované služby – viz zpráva Gartner group.

Druhým rysem, asi mnohem méně akcentovaným, možná bude modelem řízený vývoj, který by opět vynesl do popředí teoretické disciplíny, vrátil by tvorbě software neobvyklost a preciznost. Rozhodně je třeba nezapomenout na sítě a infrastrukturu, na nichž ještě dlouho bude pokrok v tomto oboru závislý.

Z toho, co zde bylo uvedeno plyne, že softwarové krize asi již v rozsahu katastrof roku 1968 nepřijdou. Softwarový svět se trochu poučil, snaží se vyvíjet nástroje na obranu před takovými jevy. Pokud by se měla použít paralela, kapitalistický svět se také poučil z velké krize v roce 1933, ale to neznamená, že v současnosti se žádné krize vyskytnout nemohou a také se vyskytují. Zdá se, že je tato možnost přímo zakotvena v principu věci. Buďme tedy připraveni na nejhorší, a zkoumejme, co nám může pomoci. Jiný pohled na věc říká, že softwarová krize ještě vůbec neskončila, že probíhá neustále od 60. let, katastrofy jsou na běžné pořádku, jen se elegantněji vysvětlují.

Softwarové inženýrství znamená zavedení disciplíny do volné tvorby software. Žádný opravdový programátor ho proto nemůže mít příliš v lásce, neboť jej nutí vytvářet „nesmyslnou“ dokumentaci a další podobné artefakty. Rozumný programátor si ale uvědomuje, že bez disciplíny to nejde. Asi bychom se měli držet Einsteinova výroku:

„Snažte se věci udělat tak jednoduché, jak to jde, ale ne jednodušší.“

Na závěr parafrázujme výrok z [1]:

„Geniální návrh vytvářejí geniální návrháři. Jakákoliv geniální metodika může podpořit a osvobodit tvořivou mysl, nemůže osvítit nebo inspirovat nádeníka. Rozdíl lze přirovnat k rozdílu mezi Mozartem a Salierim.“

Literatura

[1] BROOKS, F. *The Mythical Man Month 2Oth, Anniversary Edition.* Addison-Wesley, 1995.

[2] *ČNI, přehled norem pro softwarové inženýrství.* http://domino.cni.cz/NP/NotesPortalCNI.nsf/key/technicka_normalizace~informace_o_normach~e_byznys~ceska_normalizace~softwarove_inzenyrstvi?Open

[3] DIJKSTRA, E. The Humble Programmer. *Communication on ACM.* 15 (1972), 10, str. 859–866.

[4] FINKELSTEIN, A., KRAMER, J. *The Future of Software Engineering, Software Engineering Roadmap.* http://www.softwaresystems.org/future.html

[5] JELÍNEK, I. Perspektivy webového inženýrství. *Sdělovací technika.* 11 (2004). http://www.stech.cz/articles.asp?ida=435&idk=325

[6] KADLEC, V. *Agilní programování.* Computer Press, 2004. ISBN 80-251-0342-0.

[7] KNUTH, D. E. *The Art of Computer Programming.* Addison-Wesley, 1997.

[8] MEAD, N. R. *Issues in Licensing and Certification of Software Engineers.* CMU SEI, 2002. http://www.sei.cmu.edu/staff/nrm/license.html

[9] MORELLO, D. *The IT Professional Outlook: Where Will We Go From Here?* Gartner report G00130462; 2005.

[10] *NATO Software Engineering Conference 1968*, Garmisch, Germany. Report: Naur. P, Randell, B. eds, NATO, Brusel 1969.

[11] *NATO Software Engineering Techniques Conference 1969*, Rome, Italy. Report: Buxton, J. N., Randell, B. eds, NATO, Brusel 1970.

[12] *Popis norem IEEE pro softwarové inženýrství.*
http://standards.ieee.org/reading/ieee/std_public/description/se

[13] RICHTA, K., SOCHOR, J. *Softwarové inženýrství I.* Praha : Vydavatelství ČVUT, 1996 (dotisk 1998). ISBN 80-01-01428-2.

[14] *Software Engineering 2004: Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering. The Joint Task Force on Computing Curricula of IEEE Computer Society and ACM.*
http://sites.computer.org/ccse/

Softwarový inženýr včera, dnes a zítra

Václav Pergl

E-mail: vpergl@kerio.com

MOTTO [Švandrlík 1991]
Keď bude vojna,
nastúpite v prilbách do reky
a bojové jednotky
pojdu cez vás
jako cez most!
Poručík Čaliga

Abstrakt

Softwarový inženýr je člověk, který se úporně snaží aplikovat inženýrské metody při tvorbě softwaru a nedbá na neúspěchy. Podobně, jako jeho kolega hardwarový inženýr se neustále a překotně vyvíjí. Příspěvek pojednává zejména o požadavcích kladených na znalosti a dovednosti softwarového inženýra, od „Franto, udělej nám program na řešení kvadratické rovnice“ až po řízení týmu několika set navzájem spolupracujících softwarových webových AJAXových inženýrů. Zmíněn bude i způsob a vývoj vzdělávání softwarového inženýra.

Co by měl znát dnešní softwarový inženýr?

Guide to the Software Engineering Body of Knowledge [SWEBOK]

Publikace nazvaná „Guide to the Software Engineering Body of Knowledge“ (http://www.swebok.org/ironman/pdf/SWEBOK_Guide_2004.pdf) představuje ve své aktuální verzi poměrně úspěšný pokus autorů o výčet oblastí znalostí softwarového inženýra:
SWEBOK Knowledge Areas (KAs):

- Software requirements
- Software design

- Software construction
- Software testing
- Software maintenance
- Software configuration management
- Software engineering management
- Software engineering process
- Software engineering tools and methods
- Software quality

ISO/IEC TR 15504 – Information technology – Software process assessment

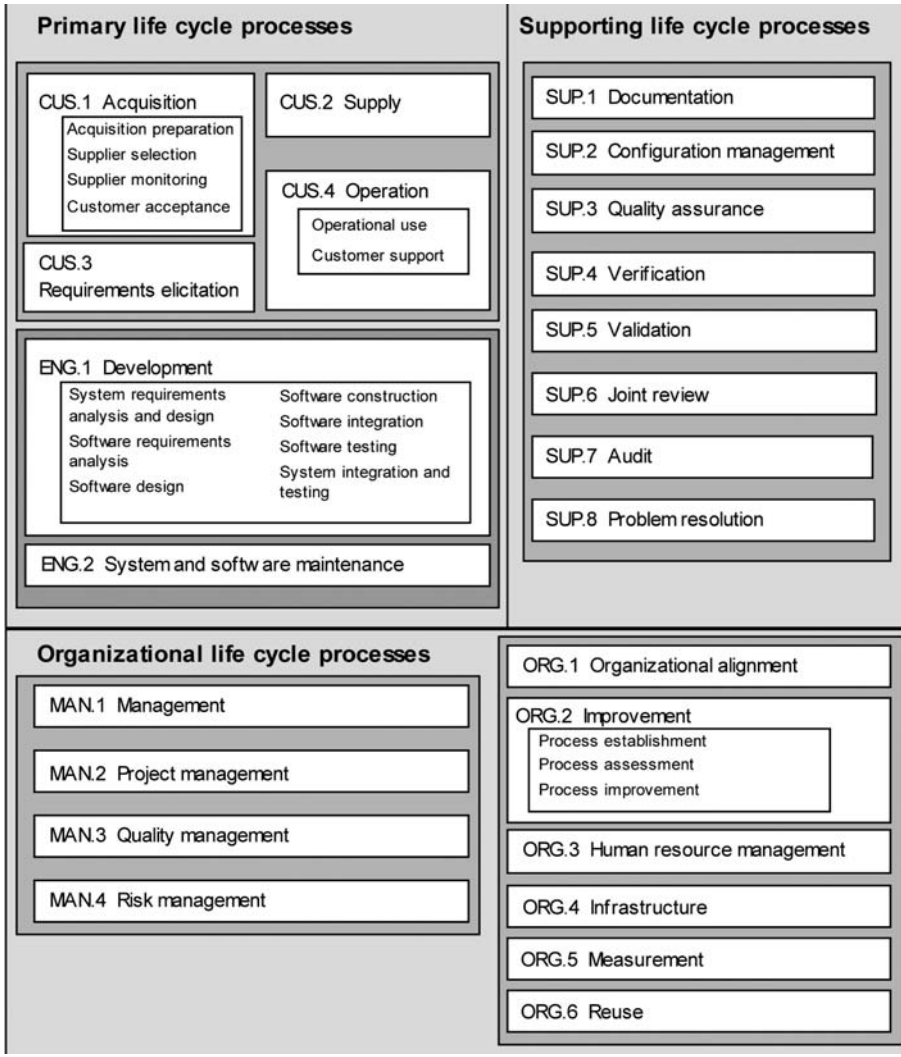
Nepoměrně širší pohled na znalosti, které pokrývají oblast softwarových procesů, představuje materiál ISO/IEC TR 15504 – Information technology – Software process assessment. Jedná se o sadu dokumentů tvořící rámec pro plánování, řízení, monitorování, kontrolu a zlepšování procesů v oblastech akvizice, dodávky, vývoje, provozu, rozvoje a podpory software.

Software Engineering 2004 – Computing Curricula Series

Představuje třetí pohled na oblast znalostí současného softwarového inženýra. Cílem dokumentu je zejména poskytnout informaci o tématech, ve kterých by měl být vzděláván a posléze i zkoušen člověk, který chce získat titul „softwarový inženýr“.

Co musel znát včerejší softwarový inženýr!

Na tomto místě autor opustí rovinu zahraničních teorií a vychrlí ze sebe tuzemské zkušenosti. Pokusíme se na konkrétním exempláři nastínit rozvoj softwarových znalostí českého „softwarového inženýra“ v období od podzimu roku 1969 až do včera. Období nebylo samozřejmě zvoleno náhodně, neboť na podzim roku 1969 autor příspěvku vytvořil svůj první počítačový program vyděrováním strojového kódu do děrné pásky pro východoněmecký počítač Celatron SER-2D s bubnovou operační pamětí. A včera jsem upravoval část PHP kódu pro dolování dat z interních systémů naší společnosti.



Obr. 1

Programátor

Období 1969 až 1973 bylo vyplněno nesmělými pokusy o přechod od transformace absolutního kódu programu, zapsaného ručně na papír v primitivním assembleru k něčemu luxusnějšímu. V roce 1973 se začalo blýskat na lepší softwarové časy. VŠSE v Plzni vlastnila polskou kopií ICL, počítač ODRA 1204

Table 1: SEEK Knowledge Areas and Knowledge Units

KA/KU	Title	hrs	KA/KU	Title	hrs
CMP	**Computing Essentials**	**172**	**VAV**	**Software V & V**	**42**
CMP.cf	Computer Science foundations	140	VAV.fnd	V&V terminology and foundations	5
CMP.ct	Construction technologies	20	VAV.rev	Reviews	6
CMP.tl	Construction tools	4	VAV.tst	Testing	21
CMP.fm	Formal construction methods	8	VAV.het	Human computer UI testing and evaluation	6
			VAV.par	Problem analysis and reporting	4
FND	**Mathematical & Engineering Fundamentals**	**89**	**EVL**	**Software Evolution**	**10**
FND.mf	Mathematical foundations	56	EVO.pro	Evolution processes	6
FND.ef	Engineering foundations for software	23	EVO.ac	Evolution activities	4
FND.ee	Engineering economics for software	10			
PRF	**Professional Practice**	**35**	**PRO**	**Software Process**	**13**
PRF.psy	Group dynamics / psychology	5	PRO.con	Process concepts	3
PRF.com	Communications skills (specific to SE)	10	PRO.imp	Process implementation	10
PRF.pr	Professionalism	20			
MAA	**Software Modeling & Analysis**	**53**	**QUA**	**Software Quality**	**16**
MAA.md	Modeling foundations	19	QUA.cc	Software quality concepts and culture	2
MAA.tm	Types of models	12	QUA.std	Software quality standards	2
MAA.af	Analysis fundamentals	6	QUA.pro	Software quality processes	4
MAA.rfd	Requirements fundamentals	3	QUA.pca	Process assurance	4
MAA.er	Eliciting requirements	4	QUA.pda	Product assurance	4
MAA.rsd	Requirements specification & documentation	6			
MAA.rv	Requirements validation	3			
DES	**Software Design**	**45**	**MGT**	**Software Management**	**19**
DES.con	Design concepts	3	MGT.con	Management concepts	2
DES.str	Design strategies	6	MGT.pp	Project planning	6
DES.ar	Architectural design	9	MGT.per	Project personnel and organization	2
DES.hci	Human computer interface design	12	MGT.ctl	Project control	4
DES.dd	Detailed design	12	MGT.cm	Software configuration management	5
DES.ste	Design support tools and evaluation	3			

s opravdovým assemblerem a slušným překladačem jazyka ALGOL 60. Časem byl ALGOL vyměněn za FORTRAN (zejména pro oblast výrobního a finančního účetnictví státního podniku STAVOPROJEKT Plzeň), ale asembler si svoji nepostradatelnost udržel až do roku 1986. Po jmenování do funkce „samostatný programátor analytik“ byl tento honosný titul s assemblerem neslučitelný.

Programátor analytik

Přechod na skutečného programátora analytika, byl kromě opuštění assembleru, charakterizován zejména zrozením analytika. Člověka, který neřeší pomocí počítače problémy své (svých stejně praštěných kolegů programátorů), ale problémy normálních lidí. Ti ale považují vtrhnutí počítačů do svého odborného revíru za nutné zlo, kterému bohužel nešlo zabránit.

Hlavní námi tehdy používanou metodou bylo nejprve Jacksonovo strukturované programování a později metodologie Jackson System Development. Vývoj programátora lze v tomto období (do konce roku 1989) charakterizovat poznáváním nových jazyků (Pascal, Lisp, C), vtrhnutím velkých i malých databází (ORACLE a dBase klony).

Tvůrce nabídky, analytik, návrhář, programátor, tester a vedoucí projektu

V období kolem revolučního roku 1989 přišli další změny. Objevili se první prakticky použitelné CASE systémy doprovázené lépe či hůře použitelnými metodologiemi vývoje. Používali jsme zejména variace na SSADM (Structured System Analysis and Design Metod), kterou původně vyvinula firma LBMS pro CCTA. Dobrých výsledků jsme s touto metodologií dosahovali zejména ve spojení kvalitním CASE nástrojem firmy LBMS. Hlavním úkolem tehdejšího vedoucího projektu bylo dodržet termín a smluvně závaznou metodologii. Současný lze stav charakterizovat změtí procesů, metod, metodologií a teorií velmi podobný diagramu ISO/IEC TR 15504. Zkrátka a dobře

Ferda Brablenec práce všeho druhu!

A co zítřejší softwarový inženýr?

Tady je každá rada drahá. Pomoci nám může kvalitní křišťálová koule a nakouknutí pod pokličku normálních řemesel. Pokud vezmeme lidské činnosti, které lidstvo provozuje dostatečně dlouho a zdá se nám, že mají ledacos společného s tvorbou softwaru, pak se snad příliš nezmýlíme. Návrh obsahuje následující kandidáty:

- stavaře,
- felčary,
- hudebníky.

„To be continued“

Literatura

[Švandrlík 1991] ŠVANDRLÍK, M. *Černí baroni.* Praha : Mladá fronta, 1991.

[SWEBOK] http://www.swebok.org/ironman/pdf/SWEBOK_Guide_2004.pdf

[ISO 15504] ISO/IEC TR 15504 – Information technology – Software process assessment

[SEEK 2004] *Software Engineering 2004 – Curriculum Guidelines for Undergraduate – Degree Programs in Software Engineering – A Volume of the Computing Curricula Series August 23, 2004.*

Peopleware revisited in the age of off- and near shoring

Till Gartner

E-mail: till@mgm-tp.com

In 1987 Tom DeMarco and Timothy Lister published their book Peopleware. It quickly became one of the most discussed books in the IT industry – an industry that wasn't too well-defined and was still building up its history at that time. What made the book such a success in terms of public attention – and sales figures – was the fact that DeMarco and Lister were the first to recognize and articulate that workers in the software industry, mainly software developers, are just different to workers in other industries.

Until then it was a common practice to apply approaches and management concepts from other industries also to the software industry. This led to expert advices and books that transferred techniques from the car industry for example into the field of software engineering. And they often transferred the production parts and the lessons learned in mass production to software production. As we know today that doesn't really work.

Peopleware is a book that addresses management issues. It is therefore meant to be read by managers in the IT industry. The standard way how those managers usually get a copy of this book is that people from their teams give them one. That is because software developers always hope to work for managers that adopt the ideas expressed in Peopleware. But also the press and book reviews acclaimed it to a noticeable level. Some describe the book as an anti-Dilbert manifesto. Successful IT companies almost always try to adopt the concepts – or at least try to maintain that illusion.

A comment on Peopleware from a Microsoft employee that I found on the Internet reads as follows: *Ever wonder why everybody at Microsoft gets their own office, with walls and a door that shuts? It's in there [Peopleware]. Why do managers give so much leeway to their teams to get things done? That's in there too. Why are there so many jelled SWAT teams at Microsoft that are remarkably productive? Mainly because Bill Gates has built a company full of managers who read Peopleware. I can't recommend this book highly enough. It is the one thing every software manager needs to read... not just once, but once a year.*

Another famous company that builds its reputation in the technical arena very much on the values presented in Peopleware is Google. Generally this has become an important weapon in the arena of IT recruiting.

In Peopleware DeMarco and Lister set up a list of lessons for managers of software development teams that corrected the former view on those topics. Having engineer type minds themselves the book is structured in a concise way: Every chapter holds and discusses one management principle. A strong tool the authors often use is inversion to explain soft facts: Rather than explaining how to achieve something they explain how to make sure not to achieve something. This technique is very powerful when trying to nail down soft facts, human feelings and reactions. And it proves very helpful: The reader quickly understands the point and it is easy for him to check on a given situation whether one of those killer-factors is present. An example that is often cited is the *Teamicide technique*: Techniques to make sure a team kills itself and never reaches the glorious and productive jelled state.

Most of DeMarco and Lister's findings and principles are based on solid arguments, experiences and statistics. They often analyze pathologies – projects or teams where things went badly wrong. This also proves to be a strong tool for isolating typical management errors. Nevertheless the book illustrates the situations with real life project situations and anecdotes which gives it an easy to read and easy to understand, lightweight style. Since I don't have that solid statistical foundation of my arguments and comments they have a far more personal and subjective character.

The goal of this little mental stroll is to revisit some of their lessons – from today's perspective. And more specifically with the background of an industry that has evolved a lot during those past 20 years. Nowadays we work with distributed teams, we span software development around the globe, we implement follow-the-sun-work, best-of-breed-practices and tune projects and the software development process trying to meet car-production-type ideals: cheap mass production by assembly lines like processes.

About the author

I was told to give you some background information about my own experiences so you can better estimate the value of my comments.

I am 38 years old and currently hold the position of Managing Director of mgm technology partners, s. r. o., a Prague based 100% daughter of the German company mgm technology partners GmbH. We do IT projects for enterprise customers. We cover the entire life cycle: Analysis, design, code, test, deploy, train, support and project management. We have $\sim$ 160 people working for mgm us based in Munich (HQ), Prague, Hamburg and Grenoble (France). We are a team of about 40 people working for in the Prague office.

I studied mathematics and financed my studies and hobbies by programming – in whatever was asked: Pascal, C/C++, Clipper, SQL, PowerBuilder, Java, ...

During my studies I founded a software company called TPS Labs together with some friends. We developed and sold a CRM product – at times before CRM was really known. We were serving mainly large customers: Compaq, Volkswagen, DASA (now a part of EADS) and others.

Some interesting situations I was involved in and in which I gathered experiences that could be relevant for the following comments:

- Product development, all phases, development team $\sim$ 40–50 people, support team $\sim$ 20 people. All in one office, open space, in Germany, during 5 years.
- Product development between Germany and India. Dev team in Germany $\sim$ 15, in India $\sim$ 20, during 2 years.
- Project development ranging from 5–50 people, mainly technical or consultancy, in different roles: developer, project manager, account manager. Across different regions and countries: Germany, France, Italy, UK, US, ...
- Development aid in Mauritania – Africa. Almost 1 year developing software for a bank. Working with software developers that had an MBA equivalent in Computer Science but had never touched a keyboard.

The principles of Peopleware

In the following I will go through some selected concepts and principles of Peopleware and will roughly explain them and then have a look at what happened to them. Are they still valid? Have they become an industry standard? Have they may be proven to be plain wrong by now...

When going through those topics please bare in mind that I will look at them in regards of companies that do off- or near shore development. That means large, multinational companies rather than cozy little start ups.

I will mainly follow the order that DeMarco and Lister used in Peopleware since they used a pretty straightforward one: The parts I will extract from are:

- Managing the human resource
- The office environment
- Growing productive teams

The books 1st edition contains one other part, the 2nd edition contains 2. They are not less interesting to read, but it would go beyond the time and space assigned to this article to discuss all of them. I tried to select the ones that are most interesting in the context of near- or off shore development locations.

Managing the human resource

Somewhere today, a project is failing

On page 4 of their book, DeMarco and Lister come up with the main message of the entire book:

The major problems of our work are not so much *technological* as *sociological* in nature.

Almost all the management concepts, errors, suggestions and best practices of the book are just different aspects of this core message. And that was the revolutionary message that made the book such a success.

Software developers are human people within their social environment – in a first place. And the standard error made by software development managers is to think of them as machines or robots. May be they often do so because software engineers usually are ratio driven people. Or because they read about Ford's way of building cars in a mass production and hope to be able to do so with software too.

But software developers are very different to robots and to chain workers. They are sensible, socially demanding individuals. Actually I personally compare the job of managing software developers often to *Prima Donna management* (A Prima Donna is the leading female singer in the opera – and often regarded as egotistical, unreasonable and irritable).

So what do we think of this statement nowadays? I guess it is as true as it used to be 20 years ago. And unfortunately it still has not become a common knowledge amongst IT managers. And it certainly still is common knowledge amongst software developers.

Since many managers still don't accepted this principle the working conditions of software developers often lack of fundamental characteristics required to get hold of social problems. When looking at off shore development it strikes us: how should I handle sociological problems of a team that is spread across continents? It is obvious that distributing teams across different sites and offices, quite often across different time zones makes the handling of sociological issues even more difficult. Paired with denying that sociological problems are often the core problems that need to be tackled this situation is a clear worsening compared to the teams usually located in one site as we had those 20 years ago. And this will reflect in many of the following principles we will look at...

There ain't no such thing as overtime

DeMarco and Lister suggest a push-me pull-me type of effect. After an extended period of uncompensated overtime, a person will either slack off greatly or, if continuing to work, will start working backwards from fatigue.

Another picture that stands for our industry and profession: The half ate pizza together with its bottle of Coke standing next to the keyboard. And a camp bed behind the programmer's chair.

While this picture probably still holds true for start up companies it has become a rare one within large, professional IT organizations. With older people that are parts of software development teams (often in management positions) they tend to have understood that there is another aspect to life than just work. And they often give an example to their younger colleagues that follow the style. Especially in large companies where reward is anonymous, money driven rather than personal approval and group recognition developers tend to live a more balanced life. This is what I have met and seen in the recent past. I could very well imagine though, that the situation may be a different one in more research type work environments where the true reward becomes esteem of the clique.

Quality – if time permits

Since at least 20 years quality, the time invested on it and the methods used are amongst the top 10 most discussed topics in the IT world. DeMarco and Lister look at it from the motivating or demotivating perspective: Producing high quality software is an inherent motivation for software developers – and having to deliver low quality software due to time pressure and management decisions à la *„that's enough quality, let's leave the product as it is, the market wouldn't notice and reward the difference anyway"* are heavily demotivating.

So the question rises whether quality has a better advocate nowadays in near- or off-shore development setups. One thing is for sure: Quality is a major concern of every manager that thinks about giving a software project or a bigger development task to an off shore team. But does this imply that he then gives it the needed time and effort? Unfortunately my experience still shows that quality improving measures are the first tasks that are crossed out of the to do list if projects run out of time or budget – and distributed projects run out of time even more often than on shore ones. Therefore in average off shore developed software is of lower quality. This hurts even more due to the extra effort needed to fix the bugs: The users sit far away from the development team, quick and flexible solutions to late problems and post rollout bugs are less feasible.

Together with the extra effort and invest required needed for organization and communication the quality of the delivered software is the biggest issue in multi-site development – and the one its image suffers of most.

Laetrile

Laetrile is a colorless liquid pressed from the soft bitter insides of apricot pits. In Mexico, you can buy it for fifty dollars a drop to „cure" your fatal cancer.

Of course, it doesn't cure anything. People who are desperate enough don't look very hard at the evidence. So do desperate project managers...

Amongst the cures offered to those desperate project managers DeMarco and Lister identified the following „Seven False Hopes of Software Management":

1. There is some new trick you've missed that could send productivity soaring.
2. Other managers are getting gains of one hundred percent or two hundred percent or more.
3. Technology is moving so swiftly that you're being passed by.
4. Changing languages will give you huge gains.
5. Because of the backlog, you need to double productivity immediately.
6. You automate everything else; isn't it about time you automated away your software development staff?
7. Your people will work better if you put them under a lot of pressure.

Again: How desperate and how blind are managers that are managing or are part of distributed off shore setups? Well what is probably the best cure to fake remedies? I would say: People that sit together, that build up a trust relationship in which they feel free to speak out their thoughts – ideally an after work situation of 2–3 managers sitting together at a beer. Picture this: Manager A: *„Does one of you understand how this new technology should solve our problem... ?"* Manager B: *„Well almost, not really actually... "* Manager C: *„Now that you say it, I also didn't get it – Why don't we have a look at it again?"* Manager A: *„Yes, it doesn't sound too realistic to me either."* – Do we expect those managers to have an after-work-beer-cross-continent-conference-call?

The office environment

The furniture police and open space offices

In 1987 DeMarco and Lister encountered many offices where the ones in charge of the office space had a police mentality and were dreaming of well kept, clean offices without any sign of personalization. Needless to say that this is in direct conflict with the Prima Donna characters of software developers.

Luckily as far as I have experienced it this problem is solved nowadays – in most organizations, even large ones. Even the slickest manager that works with software developers has understood by now that they want personal, cozy work environments. That they like their posters, that they like their geek T-Shirts –

and that discussing these topics can only diminish productivity. And this is probably the same all around the globe.

Unfortunately we're far less evolved when it comes to the office space: offering office space to software developers that is adapted to their work and needs is still a true USP. Most large companies still have open space or cubicles – some even change back from offices for 3–4 people to open space setups.

Brain time versus body time and the telephone

DeMarco and Lister introduce the concept of *flow*: *Flow is a condition of deep, nearly meditative involvement. In this state, there is a gentle state of euphoria, and one is largely unaware of the passage of time.* This is the most productive state, but it comes infrequently at work. And when being disturbed it takes at least 15 minutes to get back into the state of flow. And leaving and reentering this state can only be performed a limited number of times during a day since it requires a considerable mental effort.

DeMarco and Lister call the time in flow *brain time* whereas the time in which the developer is distracted and interrupted is so called *body time* – the worker is physically present without producing anything of value.

One of the main reasons for disturbance that they identified is the telephone. Please bear in mind that telephones were different 20 years ago. They had one ring tone – and usually a loud one. An answering machine or voice mail was far from being a standard.

So where are we today with our environment in terms of flow and disturbances? I think generally we progressed – in the right direction. At least the fact that working without permanent disturbance is far more efficient is well recognized by all managers. And generally the same applies to the sites of large enterprises in different countries and continents. And the technical equipment is better adapted to it: Not only are we all used to high-tech-monsters as telephones (my phone is so complex I can barely manage my voice mails...) and all organizations heavily use email. Emails have the great advantage of being non-real-time-disturbing since you can read them once you have finished your session of *flow*.

Then again the new technical equipment came along with some new diseases in our communication patterns. I would call the 2 most serious ones *EmailOverflow* and *ConCallMania*.

In large organizations – and even worse in the IT departments – many employees get over 150 or even over 200 emails per day. And that is with the obvious spam already filtered out by more or less smart algorithms on the mail server. That means that the poor person has to go through all of those emails and decide what to do. I we assume that somebody takes 2–3 minutes per email to read and answer to it that sums up to 300 to 600 minutes per day – that's

between 5 and 10 hours. Since this is obviously not feasible people adapt their reflexes in order to deal with these quantities: They ignore them („In case it's urgent the person will send it again or call me"). Furthermore Emails tend to become semi-legal mini documents that ranges somewhere between a phone call and an official letter. They are sent out almost as quickly as a phone call and they are almost physical – thus have a company document character. Therefore many workers and managers of large organizations put many people on CC in order to be able to state that everybody was informed. This „legal backup practice" dramatically increases the number of emails. Needless to say that the fact that teams are spread across different locations increases the email flow even more – and so does the fact that colleagues sit in different time zones and have only few or no working hours in common.

The *ConCallMania* suffers of similar problems: Since setting up conference calls is easy once the infrastructure is in place people tend to abuse it. When setting up physical meetings people were aware of the fact they need to prepare them properly: Who do I want to invite, what is the point or decision I want to make, what is the agenda, etc. Telephone conferences are often just set up in a more sloppy way. And since nobody has to leave its desk to attend we invite those other 10 people on top – just in case they would like to be informed... And again: The fact that we sit in different locations increases the frequency as well as the duration of conference calls.

In general information management is a main skill in today's large IT organizations. And it's one that is rather a social issue than a technical one... And tackling social issues in teams that consist of complex social creatures via Email and telephone is definitely more difficult than it used to be. So generally we're in a worse situation than we were 20 years ago. This time though it's not the managers fault...

Growing productive teams

Jelled teams – The whole is greater than the sum of the parts

One of the two most repeated topics and terms from Peopleware is the one of the *jelled team*. DeMarco and Lister use it to designate a team that works together in a smooth, productive, creative fun way.

Once a team begins to jell, the probability of success goes up dramatically. The team can become almost unstoppable, a juggernaut for success. (...)

A few very characteristic signs indicate that a jelled team has occurred. The most important of these is low turnover during projects and in the middle of well-defined tasks. The team members aren't going anywhere till the work is done. Things that matter enormously prior to jell (money, status, position for advancement) matter less or not at all after jell. People certainly

aren't about to leave their team for a rinky-dink consideration like a little more salary.

Since Peopleware jelling teams is every good manager's ultimate goal – well it should be... When reading about the team jelling process, about actions a manager can take to help or stop a team from jelling one thing becomes crystal clear: Jelling a team takes personal engagement of the manager. And personal relations. Between the team members, the manager and the team, the manager with its peers and the entire company. I think there is **one** thing that is obviously more difficult to set up across distributed teams: personal relations. At least it takes far more time, it is expensive and it is very delicate. Therefore jelling teams takes so much more in distributed off shore situations. Furthermore it is even more difficult for jelled teams to survive over long periods – but there we already go for the other most important term of Peopleware...

Teamicide

Teamicide is the idea that it is impossible to define what it is that causes a team to form or jell but very possible to define what prevents teams from forming or achieving success.

Definition of teamicide: *„... to inhibit the formation of teams and disrupt project sociology.“*

„... a short list of teamicide techniques...“

- *defensive management*
- *Bureaucracy*
- *Physical separation*
- *Fragmentation of people's time*
- *Quality reduction of the product*
- *Phony deadlines*
- *Clique control*

Believe me, these are the words of DeMarco and Lister from 20 years ago! It is obvious that it is almost impossible to avoid these teamicide situations in a true distributed team.

Aspects beyond Peopleware

Although DeMarco and Lister presented a very large and complete list of techniques and topics for software development management I could think of two aspects that I would add to their list. May be they just became more important in the last 20 years...

Email etiquette

I described earlier how Email flooding hinders people from productive work in large organizations. I would therefore hope that our industry and its players – by that I mean all of us – get used to a more polite and meaningful email etiquette. Probably this is something that should be discussed in each company since this discussion itself would already sensitize the people.

Predictability of management

Another aspect that I always found very important – especially when managing software developers – is predictability. By predictability in this context I mean that the managed ones should understand why their manager is taking certain actions. This should be true for almost all actions – and even more for actions that have a direct impact on the people themselves. I wouldn't want to work for a boss that takes decisions that I don't understand and that doesn't bother explaining me why he took them...

mgm technology partners way

mgm technology partners is involved in many international projects. We have offices in Munich and Hamburg (Germany), Prague (Czech Republic) and Grenoble (France). I truly think that mgm is a great company because we handle the issues highlighted in Peopleware far better than any other company I personally know. We are not the perfect company – but we are very good. And I think that the reason for this is the fact that we are a company for technical people that is run by technical people. All of us worked as software engineers. Some of us (me included) still spend many evenings programming – just for the fun of it. The people that have management type jobs at mgm technology partners therefore feel like the software developer characters rather than the managers when reading Peopleware. When selecting a company to work for the values promoted in Peopleware were key for my decision – and the same applies for the large majority of my colleagues.

In order to maintain good personal relationships amongst us we invest a lot of time and money in formal and informal communication. Distributed teams try to meet face to face at least every 2nd month, we have joint events that are of informal character: for example working for a Munich based company typically implies that you are invited to the big beer festival Oktoberfest – no matter whether you are based in Munich, Prague or Grenoble. We also try to create a general work environment that is adequate to our colleagues work: offices for 2–6 people, areas and facilities for coffee and lunch breaks...

Conclusion

I would like to put forward the following bottom line statement:

DeMarco and Lister's findings from 1987 still hold true. Almost all of them. Some even more than when they were originally written.

I think the Peopleware is till a book that every manager in our industry should read – and I agree that we should read this book repeatedly.

Another book that I would like to mention here for its similar value within our industry is Scott Adams's *The Dilbert Principle: A Cubicle's – Eye View of Bosses, Meetings, Management Fads, & other Workplace Afflictions.*

Thank you for your attention.

Geografické informační systémy v informační společnosti

Václav Čada

E-mail: cada@kma.zcu.cz

Abstrakt

Geografické informační systémy (GIS), jejich definice ve vazbě na historický vývoj. Budování národní geoinformační infrastruktury a interakce s návrhem infrastruktury prostorových dat v Evropě (INSPIRE), publikovaným ve formě směrnice Evropského parlamentu (Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community), a možnosti ovlivnění obsahu a naplňování datových bází na národní úrovni.

Prostorově lokalizovaná data a jejich základní vlastnosti s ohledem na funkčnost GIS. Základní datové zdroje a informační systémy prostorově lokalizovaných dat v ČR. Problematika lokalizace geoprostorových dat, volba souřadnicových systémů, úroveň podrobnosti dat a jejich přesnost. Metadatové informační systémy – data o datech. Informační systém katastru nemovitostí, jeho budování a správa.

Webové služby jako moderní způsob poskytování aktuálních geoprostorových informací pro projekty a uživatele GIS.

Geografické informační systémy a jejich vývoj

Pod pojmem **geografický informační systém** (Geographic Information System – GIS) v současnosti rozumíme informační systém zabývající se informacemi, které se týkají jevů přidružených k místu vztaženému k Zemi. Jedná se o funkční celek vytvořený integrací technických a programových prostředků, dat, pracovních postupů, obsluhy, uživatelů a organizačního kontextu, zaměřený na sběr, ukládání, správu, analýzu, syntézu a prezentaci prostorových dat pro potřeby popisu, analýzy, modelování a simulace okolního světa s cílem získat nové informace potřebné pro racionální správu a využívání tohoto světa.

Jedna z prvních definic pocházející od R. F. Tomlinsena z roku 1963, ve které byl GIS definován *jako systém určený pro práci s daty nesoucími informace o terénu, využívající nové technologie (tj. počítače)*, je z pohledu současnosti nekomplexní a neúplná. Dříve často frekventované definice akcentovaly skutečnost, že prostředí GIS je tvořeno počítačovým hardware (HW) a software (SW). Pod pojem geografický informační systém byla zahrnována *výpočetní technika*

i programové vybavení pro sběr a kontrolu dat, jejich uskladnění, výběr, analýzu, manipulaci a prezentaci. S nástroji, které takovýto komplex poskytuje, lze získávat informace o druhu a kvalitě krajinných prvků a o jejich vzájemných vztazích. Za klíčovou charakteristiku, která GIS definuje, byla považována schopnost GIS souhrnně analyzovat atributy dat spolu s jejich prostorovým umístěním. Je zřejmé, že tyto definice opomíjely nejen veledůležitou roli metod a technologických postupů, personální obsluhy systémů, ale především funkce uživatelské.

Některé definice GIS naopak technické a programové prostředky opomíjí. GIS je definován jako *informační systém, který je určen k práci s daty vztaženými k prostorovým nebo geografickým souřadnicím. Jinými slovy, GIS je jak databázový systém se specifickými schopnostmi práce s prostorově vztaženými daty, tak soubor operací a metod pro práci s těmito daty. V tomto smyslu může být GIS chápán jako „mapa vyšší kvality" (higher-order map)* (viz např. Jeffrey, Estes 1990). Existuje také řada abstraktních definic jako např. *GIS je systém, jehož cílem je snižovat entropii informací o území.*

Z uvedených příkladů definice GIS vyplývá, že se obecně jedná o informační systém, jehož nástroji jsou zpracovávána **geodata**[1], a tím jsou zprostředkovány a poskytovány **geoinformace**[2] o jevech a objektech na Zemi. Geografické informační systémy jsou databázové aplikace, jejichž data jsou prostorově lokalizována a tato lokalizace je využívána jako primární vlastnost pro uložení, správu, analýzu a poskytování dat.

Veškeré tyto základní funkce GIS jsou limitovány lidským faktorem jak na úrovni obsluhy vlastního informačního systému, tak v širším organizačním kontextu okolí systému jako je vzájemná interakce dílčích IS vůči sobě. Zcela zásadní jsou především potřeby koncových uživatelů GIS, kterým mají poskytovat kvalitativně vyšší informace pro jejich řídící a rozhodovací procesy. Šíře potřebných znalostí tvůrců, správců a uživatelů GIS zapříčinila vznik a rozvoj nových mezioborových vědních disciplín jako **geomatika** a **geoinformatika**[3].

Geomatika je podle dokumentu ISO/TR 19122 (Geografická informace/Geomatika – kvalifikace a certifikace personálu) „vědním oborem, zabývajícím se

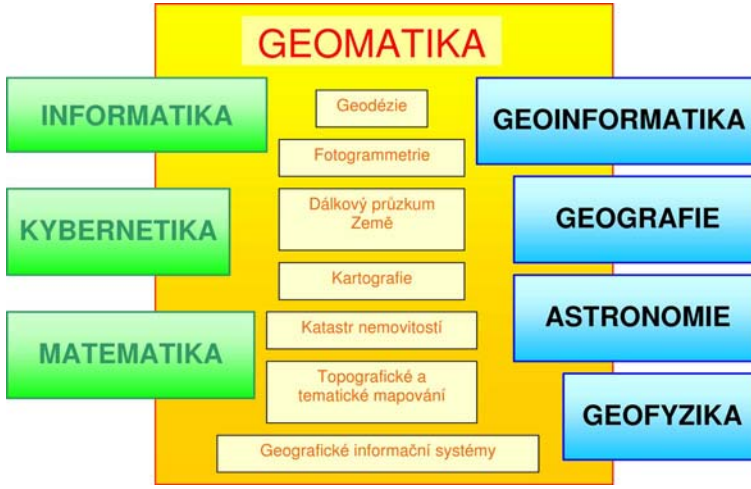
[1] **Geodata** – data vztahující se k územním prvkům; jedná se o geometrická, prostorová a popisná data. – data identifikující geografickou polohu a charakteristiky přírodních a antropogenních jevů a hranic mezi nimi. Reprezentují abstrakce entit reálného světa ve formalizované, počítačově zpracované podobě. Jiná používaná synonyma – geoprostorová data, geografická data

[2] **Geoinformace** – význam, který je přisouzen geodatům.

[3] **Geoinformatika** nemá podobnou mezinárodní definici. Pouze z Lexikonu geoinformatiky Univerzity v Rostocku (SRN) vyplývá, že ekvivalentem pro termín geoinformatika je geografická informační (geoinformační) věda, což je podle ISO/TR 19122 „multidisciplinární vědní obor zajišťující rozvoj geoinformačních technologií" (tedy zejména teoretické a technologické zajištění tvorby a užití geografických informačních systémů – GIS). Česká definice ve Standardu Úřadu pro veřejné informační systémy (2001) to potvrzuje, když geoinformatiku považuje za „specifickou část informatiky, zabývající se geodaty, geoinformacemi a geografickými informačními systémy".

sběrem, distribucí, ukládáním, analýzou, zpracováním a presentací geografických dat nebo geografických informací“. Podle standardu NATO AAP-6 je geomatika „vědou a technologií vedení geoprostorových informací včetně získávání, ukládání, analýzy a zpracování, zobrazování a šíření georeferencovaných informací“. Geomatika vychází z integrovaného přístupu ke sběru, ukládání, přenosu (distribuci), analýze, poskytování a prezentaci dat vztažených k Zemi, označovaných jako základní prostorová (geoprostorová) data. Data pocházející z mnoha zdrojů jsou pořizována rozličnými metodami geodézie, kosmické geodézie, mapování, kartografie, dálkového průzkumu Země a fotogrammetrie. Pro jejich zpracování, správu a analýzu se využívá informační technologie, zejména geografické informační systémy (GIS).

Z úvodní definice vyplývá, že je tato vědní disciplína široce mezioborová, patří do oblasti aplikovaných věd a takovéto prostředí je jen těžko prosaditelné na fakultách zaměřených na tradiční obory (stavební fakulty, fakulty geografie, fakulty životního prostředí apod.). V roce 1995 byl na Západočeské univerzitě v Plzni, Fakultě aplikovaných věd akreditován nový studijní program Geomatika. Tento program byl koncipován jako jednooborový s těmito specializacemi (specializace je pojem, který na ZČU označuje tzv. segment studijního programu, tedy strukturu předmětů a modulů v dané skladbě; základem popisu specializace je struktura státní zkoušky):
01 Geodézie a geoinformační systémy
02 Kartografie
03 Katastr nemovitostí a občanské právo
04 Geodézie a katastr nemovitostí (je nabízena v kombinované formě studia a využívána zejména pracovníky z technické praxe a pracovníky katastrálních úřadů pro doplnění požadované odborné kvalifikace).



V současné době je na ZČU v Plzni akreditován studijní program Geomatika pro všechny stupně strukturovaného studia (bakalářský, magisterský, doktorský) jak v denní, tak i kombinované formě. Absolventi magisterského programu Geomatika získávají po složení státních závěrečných zkoušek a obhajobě diplomové práce titul Ing. Protože se podařilo prosadit vyjmenování tohoto oboru ve vyhlášce č. 31/1995 Sb., kterou se provádí zákon č. 200/1994 Sb., pro udělení úředního oprávnění pro výkon a ověřování výsledků zeměměřických činností podle § 13 odst. 1 písm. a) b) a c) zákona č. 200/1994 Sb., mohou naši absolventi toto oprávnění po splnění dalších požadovaných náležitostí získat. Další informace jsou k dispozici na http://gis.zcu.cz/.

Budování národní geoinformační infrastruktury v ČR

Problematika infrastruktury prostorových dat celosvětově vede k hledání nástrojů, prostředků a vytváření podmínek k maximálnímu využívání geodat. Jedná se o prostředky a postupy určené ke sběru, zpracování, záznamu a uchování geodat a jejich distribuci za účelem uspokojení širokého spektra potřeb uživatelů. Funkční geoinformační infrastruktura se tak stává dalším přirozeným požadavkem budované informační společnosti. Racionální správa a smysluplné využívání zdrojů je prioritním požadavkem trvale udržitelného rozvoje společnosti. Pro řízení těchto procesů musí společnost vytvořit a garantovat jednotné a závazné mechanismy a postupy tak, aby prostředky vynakládané do oblasti informačních systémů (IS) byly investovány efektivně s garancí návratnosti. K tomuto účelu je vytvářena infrastruktura prostorových dat (SDI – spatial data infrastructure, GDI – Geographic (geospatial) data infrastructure, GII – Geographic (geospatial) information infrastructure).

V českém prostředí se v současnosti nejčastěji hovoří o problematice „geoinformační infrastruktury (GII)“ nebo též „národní geoinformační infrastruktury (NGII)“. K této problematice zpracovalo sdružení Nemoforum[4] ve spolupráci se zástupci soukromého i veřejného sektoru programový dokument „Národní geoinformační infrastruktura České republiky, program rozvoje v letech 2001–2005“[1], který projednala a podpořila Rada vlády pro státní informační

[4]**Nemoforum** je účelové sdružení založené podle § 829–§ 841 Občanského zákoníku. Účastníky Nemofora mohou být orgány veřejné správy, profesní svazy, komory či sdružení a vysoké školy, přičemž každý účastník jmenuje jednoho svého zástupce jako člena pléna Nemofora. Nemoforum se chce na základě spolupráce institucí z veřejné, profesní (soukromé) i akademické sféry podílet na vybudování funkčního e-governmentu v České republice, který jejím občanům umožní – v oboru informací o nemovitostech a území – rychlou a jednoduchou komunikaci s veřejnou správou. Záměrem Nemofora je přispívat k řešení nejrůznějších problémů s tím spojených, v souladu s vývojem v Evropě i ve světě, za účelem uspokojení potřeb uživatelů, pořizovatelů i správců těchto informací.

politiku (SIP). NGII je v tomto dokumentu popsána jako soubor vzájemně provázaných podmínek, které v prostředí ČR umožňují zajistit a zpřístupnit co největšímu okruhu uživatelů širokou škálu geoinformací uživatelsky vhodnou formou při plném využití potenciálu moderních (geo)informačních a komunikačních technologií.

Samostatnými programy, které jsou v souladu s projektem NGII a jejichž nositelem byl Český úřad zeměměřický a katastrální (ČÚZK), byla v uplynulém období řešena problematika:

- Základní báze geografických dat (ZABAGED),
- Ortofotografického zobrazení území ČR,
- státního mapového díla,
- prostorového referenčního rámce,
- normalizace.

V některých oblastech státních mapových děl závazných na území státu jako je katastrální mapa nedošlo a dosud nedochází ke koncepčním a racionálním řešením v uspokojování potřeb a požadavků na geodata katastru nemovitostí. Z mnoha objektivních i subjektivních důvodů stagnuje projekt tvorby digitální katastrální mapy (DKM), dlouhodobě se nedaří sjednotit obsah a naplňování základního registru územní identifikace a nemovitostí. Opomíjí se především problematika vymezení základních geoprostorových dat.

V této souvislosti jsou legislativní opatření uváděná v platnost v poslední době velice problematická. Schváleným zákonem č. 319/2004 Sb., který nabývá účinnosti dnem 1. ledna 2004 a novelizuje zákon 200/1994 Sb., o zeměměřictví, je v § 4a „obsah, správa, užití a rozšiřování dat databáze“ stanoveno, že pro tvorbu informačních systémů veřejné správy, obsahujících geografická, topografická a geodetická data z celého území ČR, jsou závazná data základní báze geografických dat České republiky (ZABAGED).

Nařízením vlády č. 430/2006 Sb., o stanovení geodetických referenčních systémů a státních mapových děl závazných na území státu a o zásadách jejich používání, jsou nově zaváděny historické souřadnicové systémy gusterbergský a svatoštěpánský na celém území státu pro zeměměřické činnosti ve veřejném zájmu[5] a pro výsledky zeměměřických činností využívaných ve veřejném zájmu.

[5]podle § 4 odst. 1 zákona č. 200/1994 Sb., ve znění pozdějších předpisů **Zeměměřickými činnostmi ve veřejném zájmu** jsou:
a) budování, obnova a údržba bodových polí,
b) vyhotovení nového souboru geodetických informací katastru nemovitostí,
c) vyhotovení geometrického plánu a dokumentace o vytyčení hranice pozemku,
d) plnění úkolů pro potřeby obrany státu včetně k tomu nezbytné mezinárodní spolupráce

Základní geoprostorová data, charakteristické vlastnosti

Problematika vymezení základních geoprostorových dat (geodat) má v oblasti geografických informačních systémů zásadní význam. Ekonomické rozvahy budovaných projektů GIS kalkulují s náklady na geodata, programové (SW) a technické prostředky (HW) přibližně v poměru 80 : 15 : 5. Vysoké vstupní náklady prvotního pořízení, následné správy a vedení včetně aktualizace geodat jsou často nejen základním limitujícím faktorem rozšíření technologií GIS v řadě oblastí řízení společnosti, ale zároveň nejperspektivnější oblastí pro optimalizaci a racionalizaci tohoto procesu s maximálním efektem.

Geoprostorovými daty rozumíme lokalizační data (přímé i nepřímé lokalizace a výškopisu) zabezpečující potřebnou integraci tematických[6] a aplikačních datových bází. Základní geoprostorová data mají integrační charakter a tvoří společný základní obsah většiny tematických či aplikačních datových bází využívaných v prostorově orientovaných rozhodovacích procesech. Základní geoprostorová data zajišťují lokalizaci základních geoprvků, jsou aplikačně nezávislá a umožňují integraci dat z více zdrojů. Společnou charakteristikou těchto dat je, že se v jejich rámci předpokládá přímá lokalizace na úrovni přesnosti a podrobnosti dosažitelné pro pozemkovou úroveň (nikoli topografickou).

Požadované kvalitativní atributy geoprostorových dat

Generování komplexní geoprostorové informace je podmíněno polohově lokalizovanými a časově definovanými daty, která jsou produkována různými zdroji, institucemi nebo dokonce odlišnými geovědami. Systemové nástroje GIS zajišťují správu těchto dat a generování požadovaných komplexních geoprostorových informací. Tyto informace jsou zásadní pro rozhodovací procesy řízení trvale udržitelného rozvoje společnosti a jejich kvalitu. Kvalita rozhodovacích procesů je závislá na kvalitě geoprostorových dat a geoprostorových informací. Pro hodnocení kvality geoprostorových dat je možné stanovit jednoznačná kritéria: **přesnost**, **rozlišovací schopnost** (podrobnost), **konzistentnost** a **komplexnost** dat.

---

a vědecko-technického rozvoje,
e) tvorba, obnova a vydávání základních a tematických státních mapových děl,
f) vyhotovení zeměměřických podkladů a dokumentace pro výkon státní správy,
g) vyměřování státních hranic,
h) standardizace jmen nesídelních geografických objektů z území České republiky a jmen sídelních a nesídelních geografických objektů z území mimo Českou republiku,
i) vedení informačních systémů v zeměměřictví podle písmen a) až h),
j) dokumentace a archivace výsledků zeměměřických činností.

[6] **Tematické datové báze** – datové báze vytvářené pro uživatelské aplikace nebo těmito uživatelskými aplikacemi generované.

Přesnost

Přesnost je vždy vázána na základní prvky datového modelu a je relevantní vlastností daného prvku. Nesoulad mezi kódovanou a skutečnou hodnotou atributu daného prvku je nazýván **chybou**. Přesnost je vždy pouze relativní vlastnost prvku, protože je vázána na specifický popis prvku a úroveň generalizace.

Polohová přesnost geoprvku[7] (feature) je charakteristika lokalizace a geometrických parametrů prvku. Pro prvky je polohová přesnost definována jako prostorová vzdálenost mezi kódovanou (modelovou) polohou prvku a jeho skutečnou polohou v reálném světě = polohovou chybou. Polohová chyba je popisována klasickými statistickými veličinami, jako je *směrodatná odchylka, střední souřadnicová chyba* nebo *střední polohová chyba* charakteristických bodů prvku.

Časová přesnost vyjadřuje časové limity, ve kterých je daný prvek platný. Jedná se o stanovení času nebo časového intervalu, ve kterém se určují, nebo byly určeny atributy vybraného prvku.

Tematická přesnost je kritérium spolehlivosti tematických atributů daného prvku a je analyzována v závislosti na typu hodnoceného objektu. Tematická přesnost analyzuje i vhodnost zařazení objektů do kategorií (bodová, polygonální, areálová apod.)

Rozlišovací schopnost

Rozlišovací schopnost geoprvků je závislá na zvoleném datovém modelu a vyjadřuje množství a hodnoty detailů, které rozlišujeme v prostoru, čase i tématu. Rozlišovací schopnost je limitována především úrovní generalizace (zjednodušení) daného geoprvku nebo jevu. Je zřejmé, že dvě báze geoprostorových dat stejné přesnosti s rozdílnou rozlišovací schopností nejsou stejně kvalitní.

Prostorová rozlišovací schopnost vyjadřuje hodnotu ještě uvažovaných prostorových detailů nebo minimální velikosti geoprvku. V případě obrazové interpretace prvku je prostorová rozlišovací schopnost většinou vyjádřena velikostí pixelu rastru.

Časová rozlišovací schopnost je podmíněna minimálním časem trvání události (jevu). Je závislá na typu a charakteru zkoumaného jevu.

Tematická rozlišovací schopnost odpovídá přesnosti zařazení do dané kategorie geoprvku. Pro stanovení kategorií představuje tematická rozlišovací schopnost vymezení dvou příbuzných kategorií.

Konzistentnost dat

Konzistentnost dat představuje úroveň rozporných nebo vícečetných dat pro popis daného geoprvku nebo jevu. Konzistentnost dat podmiňuje výslednou spolehlivost a kvalitu generovaných informací.

[7] **Geoprvek** – modelový obraz geografické entity reálného světa, který je dále nedělitelný na jednotky stejné třídy, nebo sada takových entit se společnou hodnotou atributu

Prostorová konzistentnost dat je testována topologickými funkcemi GIS a možností z těchto dat generovat skladebné geoprvky.

Časová konzistentnost dat je dána podmínkou, že pouze jediná událost se může vyskytnout v konkrétním čase na stanoveném místě.

Na **tematické konzistentnosti** atributů jsou závislé interakce časových a prostorových vazeb generovaných informací.

Úplnost dat

Nesoulad mezi daty uloženými v datových bázích a daty potřebnými k popisu generalizovaného vyjádření objektu definujeme jako parametr **úplnosti dat** a je možné jej hodnotit z hlediska prostoru, času i tematického obsahu.

Dostupná státní mapová díla zdrojem geodat

Státní mapová díla jsou definována v § 3 odst. 1 nařízení vlády č. 430/2006 Sb., o stanovení geodetických referenčních systémů a státních mapových děl závazných na území státu a zásadách jejich používání.

Státními mapovými díly jsou:

a) katastrální mapy,

b) „Státní mapa v měřítku 1 : 5 000“,

c) „Základní mapa České republiky v měřítcích 1 : 10 000, 1 : 25 000, 1 : 50 000, 1 : 100 000 a 1 : 200 000“,

d) „Mapa České republiky v měřítku 1 : 500 000“,

e) „Topografická mapa v měřítcích 1 : 25 000, 1 : 50 000, 1 : 100 000“,

f) „Vojenská mapa České republiky v měřítcích 1 : 250 000 a 1 : 500 000“.

Tematickými státními mapovými díly závaznými na území státu jsou tematická mapová díla vytvořená pro celé území státu na podkladě základních státních mapových děl. Státní mapová díla jsou vytvářena, obnovována a vydávána v grafické nebo digitální formě.

Digitální forma státních mapových děl

Přesnost a podrobnost státních mapových děl vytvářených v digitální formě je poplatná vlastnostem původního analogového díla a je závislá na použitých technologických postupech převodu těchto map do digitální formy.

1 Základní báze geografických dat

Základní mapa České republiky v měřítku 1 : 10 000 vyhotovovaná v letech 1972–1988 v geodetickém systému Jednotné trigonometrické sítě katastrální (S-JTSK) a Křovákově konformním kuželovém zobrazení v obecné poloze na Besselově elipsoidu a ve výškovém systému Balt – po vyrovnání vznikla transformací obsahu vojenských topografických map do nově určeného kladu mapových listů, redukcí obsahu a s použitím jiného klíče mapových značek. Bez zásadních úprav byl převeden výškopis v systému baltském – po vyrovnání znázorněný vrstevnicemi, výškovými kótami a relativními výškami terénních stupňů.

Údržba a obnova listů Základní mapy 1 : 10 000 byla prováděna v letech 1979–2000 výběrovým způsobem v závislosti na četnosti změn. Intravilány měst, oblasti velkých investičních celků a prostory povrchové důlní těžby byly aktualizovány v tříletých cyklech, v méně významných oblastech byl cyklus obnovy až 17 let. Změny byly doplňovány především analogovým fotogrammetrickým vyhodnocením leteckých měřických snímků a terénním topografickým šetřením nebo využitím geodetické dokumentace skutečného provedení sídlišť, dopravních a průmyslových staveb.

Absence digitálního ekvivalentu ZM ČR 1 : 10 000 pro potřeby státního informačního systému byla řešena usnesením vlády České republiky ze dne 8. září 1993 č. 492, ve kterém bylo předsedovi Českého úřadu zeměměřického a katastrálního (ČÚZK) uloženo vypracovat a předložit projekt Základní báze geografických dat (ZABAGED) k meziresortnímu projednání. Tato koncepce byla schválena a vyhlášena dne 1. listopadu 1994.

ZABAGED byla postupně naplněna v letech 1994–2003 po komplexní revizi tiskových podkladů Základní mapy ČR 1 : 10 000 (polohopisu, výškopisu, vodstva a areálů vybraných porostů a druhů využití půdy) včetně klasifikace objektů pro následnou digitalizaci a edici objektů podle katalogu objektů ZABAGED. Současně byla doplňována popisná data geografických prvků. Z digitalizovaných vrstevnic výškopisu ZM ČR 1 : 10 000 byl vytvořen digitální model reliéfu (DMR).

V období 2001–2005 byl prováděn 1. cyklus aktualizace ZABAGED na podkladě periodického leteckého snímkování celého státního území v tříletém intervalu pro prvotní účel vyhotovení digitálních ortofotomap pro potřeby integrovaného administrativního a kontrolního systému zemědělství (IACS).

Novela zákona č. 200/1994 Sb., o zeměměřictví, stanoví, že údaje ZABAGED jsou závazné pro tvorbu státních mapových děl v měřítku 1 : 10 000 a menším a pro tvorbu geografických informačních systémů (GIS) veřejné správy. Správce ZABAGED bude zmocněn vyžadovat změnové údaje od správních úřadů za účelem aktualizace vybraných popisných informací. Data ZABAGED se bezplatně poskytují orgánům veřejné správy.

2 Ortofotografické zobrazení území ČR

Projekt ortofotografického zobrazení území ČR byl zahájen v roce 1999 s cílem aktualizace ZABAGED. Ortofotografické zobrazení území je obecně velice perspektivní zdroj geodat především z důvodu možnosti operativního získání aktuálních informací z daného prostoru v poměrně výhodných cenových relacích. Jedná se o zpracování výsledků leteckého měřického snímkování, které se jednotně provádí v tříletém intervalu na území celého státu pro potřeby budovaného Integrovaného administrativního a kontrolního systému zemědělství (IACS).

Snímkování se provádí na základě veřejné soutěže podle jednotných zadávacích parametrů (od r. 2003 nálet ve směru os kladu mapových listů SM 5,60% podélný překryt, 25% příčný překryt, širokoúhlá kamera, konstanta komory $f = 150$ mm, formát snímku 230×230 mm, měřítko snímků 1 : 23 000, původně snímky černobílé – od roku 2003 barevné). Rozlišitelnost digitalizovaného leteckého měřického snímku je 0,5 m ve skutečnosti. Pro určení prvků vnější orientace se používá analytická nebo digitální aerotriangulace s přednáletovou signalizací bodů trigonometrické sítě.

Tvorba digitální ortofotomapy je nyní prováděna centrálně na pracovišti Zeměměřického úřadu v Pardubicích a zčásti též ve Vojenském geografickém a hydrometeorologickém úřadě v Dobrušce. Letecké měřické snímky se skenují na skeneru firmy Zeiss s rozlišením 20 mikrometrů v rovině snímku. Pro ortogonalizaci měřických snímků je z vrstevnicového modelu ZABAGED vytvořen 3D TIN model a programem Mosaic LH-Systems provedeno mozaikování ortofot do kladu listů SMO 5.

Data ortofotografického zobrazení území ČR byla do konce r. 2002 poskytována v digitální formě jako černobílý polotónový rastrový obraz v kladu listů ZM 10 000 (od r. 2003 jako barevný v kladu listů SMO 5). Ortogonalizovaný letecký snímek je aktuální ke dni leteckého snímkování. Data jsou v rastrovém formátu TIF s připojeným souborem souřadnic rohů mapových listů v S-JTSK. Data jsou poskytována s rozlišením cca 0,5 m a s hustotou 500 dpi, což umožňuje určit polohu jednoznačně identifikovatelného bodu s přesností okolo 1 m.

Značným kvalitativním skokem v tvorbě ortofotomap (vyšší prostorové rozlišení, dokonalé barevné vyrovnání, odstranění skenování fotografických snímků) bude přímé digitální letecké měřické snímkování multispektrálního charakteru, tj. současné pořízení černobílého rastrového obrazu území s extrémně vysokým rozlišením (daným velikostí pixelu 7,2 nebo 9 mikrometrů v rovině snímku proti současným 14 nebo 20 mikrometrům při skenování snímků na filmu), zároveň se snímáním složek barevného obrazu R, G, B, NIR (infračervené) s rozlišením 3x menším, které se použijí k „obarvení“ černobílého obrazu do formy přirozeně barevného nebo barevně infračerveného snímku s použitím principu „pan-

sharpening“. První zkoušky se v ČR realizují v roce 2006 a je reálný předpoklad hromadného nasazení nové technologie v roce 2008 nebo 2009.

3 Státní mapa 1 : 5 000

Od roku 1946 byly vyhotovovány Státní mapy ČSR 1 : 5 000 – hospodářské (SM 5 hospodářské) stolovou tachymetrií nebo užitím fotogrammetrických metod s připojením na polohové bodové pole v S-JTSK. Výškopisná složka mapy vyjádřená vrstevnicemi vznikla přímým měřením s připojením na jednotnou nivelační síť.

Protože kapacitní důvody neumožnily tvorbu SM 5 hospodářské tak, aby plnila veškeré požadavky veřejného zájmu, rozhodlo tehdejší Ministerstvo techniky po dohodě se Státním úřadem plánovacím o vyhotovení Státní mapy 1 : 5 000 – odvozené. Tato mapa byla pořízena na celém území českých zemí mimo prostory již existující SM-5 hospodářské. Jednalo se o téměř 16 tis. mapových listů.

Polohopis Státní mapy 1 : 5 000 – odvozené (SMO 5) byl vyhotoven fotomechanickou transformací polohopisu katastrální mapy do sekcí mapových listů Státní mapy 1 : 5 000. Generalizovaný polohopis katastrální mapy měl zajistit plynulý přechod nejen mezi mapovými listy, ale i na hranicích katastrálních území. Případný nesoulad byl eliminován posunem kopírovaných částí mapových listů katastrálních map. Výškopis byl přebírán z výškopisných příložných map nebo topografických map 1 : 25 000 (později též 1 : 10 000) obdobnou fotomechanickou transformací. Tím vlastně poprvé vzniklo na území Československa mapové dílo velkého měřítka v souvislém zobrazení.

Přesto, že tvorba SMO 5 měla být pouze dočasným řešením, je tato mapa užívána do dnešních dnů. Je totiž v současnosti stále nejpodrobnějším státním mapovým dílem velkého měřítka, které svými 16 193 mapovými listy pokrývá celé státní území v souvislém pravoúhlém kladu mapových listů v S-JTSK. O uživatelské oblibě svědčí i ten fakt, že od roku 1950 byly všechny mapové listy realizovány až v deseti vydáních a počet prodaných výtisků se stále pohybuje okolo 70 tis. ročně.

V padesátileté historii doznaly obsah i forma SMO 5 výrazných změn (např. kvalita a způsob reprodukce). Přesto však aktuálnost obsahu, přesnost polohopisu a výškopisu neodpovídá současným požadavkům kladeným na soudobé základní mapy. Mezi tyto základní nedostatky patří:

- polohopis přebíraný z platných katastrálních map vykazuje určitý nesoulad s aktuálním stavem v terénu, který je způsoben procesem vedení katastrálních map (majetkoprávně nedořešené změny, dosud nezaměřené nebo nezakreslené stavby, komunikace, úpravy vodních toků a nádrží, elektrické vedení, objekty uvnitř průmyslových závodů, státních drah, letišť apod.),

- přebíraný polohopis není homogenní z hlediska přesnosti a závisí na typu katastrálních map,

- výškopis je v současné době nejčastěji přebírán ze Základní mapy ČR 1 : 10 000, interval základních vrstevnic není jednotný a především v rovinatých územích nedostatečně vyjadřuje výškové poměry,
- u popisu není zohledněno standardizované názvosloví ZM ČR a často chybí významné místní popisy usnadňující orientaci v prostoru (popis významných budov, názvy ulic, veřejných prostranství apod.),
- obsah a grafická podoba je často poplatná výchozím podkladům a technologickému vybavení zpracovatelských organizací.

Odstranění některých výše uvedených nedostatků napomohla aplikace digitálních technologií a vznik **Státní mapy** 1 : 5 000 (SM 5) včetně stanovení zásad jejího vedení a obnovy.

Obsah SM 5 tvoří samostatné dílčí souborově orientované vrstvy ve vektorovém nebo rastrovém tvaru – katastrální, výškopisná a topografická. Tyto vrstvy jsou doplněny pro každý mapový list rámem mapového listu se souřadnicovou sítí včetně mimorámových údajů.

Vrstva katastrální obsahuje

- body základního polohového bodového pole (ZPBP) včetně zhušťovacích bodů (ZhB),
- body výškového bodového pole,
- hranice územních celků,
- další prvky polohopisu jako např. osy kolejí železničních tratí, lanové dráhy, hrany, koruny a střední dělící pás komunikací, mosty, propustky a tunely, břehové čáry vodních toků a vodní nádrže včetně vodohospodářských staveb, veřejné studny, nadzemní vedení VN a VVN, stožáry vysílacích a retranslačních stanic, zvonice, pomníky, památníky, boží muka, schodiště významných budov a další objekty, které jsou součástí v katastru evidovaných staveb,
- popis (označení bodů bodových polí, místní a pomístní názvosloví, jména ulic a veřejných prostranství ve vybraných obcích).

K doplnění katastrální vrstvy o body bodových polí se využívá databáze geodetických bodů. Místní a pomístní názvosloví se přebírá z databáze standardizovaných geografických jmen (GEONAMES). Pro druhová označení se používá atributů ZABAGED nebo Základní mapy ČR 1 : 10 000 (ZM ČR 10). Jména ulic a veřejných prostranství ve vybraných obcích jsou doplňována z plánů obcí a dalších informačních zdrojů obecních úřadů.

Vrstva výškopisu je tvořena

- vrstevnicemi ve vektorovém tvaru s intervalem 1 nebo 2 nebo 5 m,
- výškovými kótami a kótovanými body,
- smluvenými značkami terénních stupňů a skal.

Veškerá tato data jsou získána ze ZABAGED včetně vrstevnicového 3D modelu nebo digitální ZM ČR 10. Nadmořské výšky bodů ZPBP a ZhB se doplní podle dokumentace, zejména databáze geodetických bodů, nadmořské výšky vrstevnic jsou přebírány z 3D modelu reliéfu.

Podkladem pro **vrstvu topografickou** je digitální ortofoto s hustotou rastru minimálně 1 200 dpi, vytvářené pro aktualizaci ZABAGED nebo pro účely Ministerstva zemědělství (IACS).

Tvorbu a obnovu SM 5 zajišťují katastrální úřady, kde se též archivují tiskové podklady posledního vydání pro každý mapový list. Obsah SM 5 je poskytován po jednotlivých vrstvách, v libovolných kombinacích ve formě souborů digitálních dat nebo ve tvaru tiskových výstupů.

V prostorech, kde nejsou k dispozici podklady pro zpracování katastrální složky SM 5 ve vektorovém tvaru, se v současné době vyhotovuje **Státní mapa** 1 : 5 000 − **rastrová** (SM 5 R) s tím, že bude postupně nahrazována SM 5. **Katastrální složku SM 5 R** tvoří soubor rastrových dat, pořízený z tiskového podkladu polohopisu, s popisem posledního vydání SMO 5, lokalizovaný v souřadnicovém systému JTSK, oříznutý na rám mapového listu. **Výškopisnou složku SM 5 R** tvoří soubor rastrových dat z tiskového podkladu výškopisu posledního vydání SMO 5 připojený k souřadnicovému systému JTSK a upravený na rám mapového listu.

K tvorbě **topografické složky** se využije digitální ortofoto pořízené z výsledků posledního leteckého měřického snímkování. Soubory ortofot jsou vytvořeny pro jednotlivé mapové listy SM 5 nebo Základní mapy ČR 1 : 10 000 (od a do roku 2003). U mapových listů, do kterých zasahuje území sousedního státu, se ortofoto zobrazí jen na území České republiky. Ortofotografické zobrazení se ukončí na hraniční linii kresby katastrální složky a plocha vně oříznutí, ležící v prostoru mapového listu SM 5 R, se vyplní šedou barvou.

4 Katastrální mapa

Katastrální mapa je závazným státním mapovým dílem velkého měřítka, obsahuje body bodového pole, polohopis a popis. Předměty obsahu katastrální mapy se vyznačují standardizovanými mapovými značkami.

Obsahem bodového pole jsou všechny trvale stabilizované i trvale signalizované body polohového a výškového bodového pole včetně bodů přidružených k trigonometrickým a zhušťovacím bodům.

Předmětem polohopisu jsou hranice územně správních jednotek, hranice katastrálních území, hranice chráněných území a ochranných pásem, hranice evidovaných nemovitostí s odlišením hranic převzatých z map dřívějších pozemkových evidencí a zobrazení dalších prvků polohopisu. Hranice se v katastrální mapě zobrazují přímými spojnicemi lomových bodů, v odůvodněných případech lze použít kruhových oblouků. Kritériem generalizace obsahu katastrální mapy jsou předměty s délkou přímé spojnice lomových bodů na vlastnické hranici do 0,1 m, u ostatních prvků do 0,2 m. Pro zobrazení těchto tvarů platí kritérium 0,2 mm v měřítku mapy.

Popis katastrální mapy uvnitř mapového rámu tvoří čísla bodů polohového bodového pole, čísla hraničních znaků státní hranice, místní a pomístní názvosloví, označení parcel parcelními čísly a mapovými značkami. Vně mapového rámu je umístěn název Katastrální mapa, označení mapového listu včetně lokalizace ve správním členění státu, označení sousedních mapových listů, údaje o souřadnicovém systému a měřítku, údaje o vzniku katastrální mapy, tirážní údaje a okrajové náčrtky.

Projekt tvorby digitálního souboru geodetických informací

Vyhláška č. 190/1996 Sb., stanovuje tyto možné formy katastrální mapy:

- digitální katastrální mapa (DKM) s geometrickým a polohovým určením v národním referenčním geodetickém systému Jednotné trigonometrické sítě katastrální (S-JTSK) s přesností podrobných bodů polohopisu $m_{xy} \leq$ 0,14 m (kód kvality 3), nebo $m_{xy} \leq$ 0,26 m (kód kvality 4). DKM může obsahovat digitalizované podrobné body z katastrálních map grafických, charakterizované podle původu a měřítka mapy kódem kvality 6 ($m_{xy} \leq$ 0,21 m) nebo 7 ($m_{xy} \leq$ 0,42 m). Může obsahovat také digitalizované podrobné body z katastrálních map grafických, dosud vedených v sáhovém měřítku 1 : 2 880, charakterizované kódem kvality 8, pokud je nebylo možno s ohledem na provedený způsob obnovy katastrálního operátu určit přesnějším způsobem,

- katastrální mapa grafická s přesností dle předpisů platných v době jejího vzniku a se souřadnicovými soustavami podle těchto předpisů,

- katastrální mapa obnovená digitalizací grafických map uvedených pod písmenem b), s přesností podrobných bodů polohopisu $m_{xy} >$ 0,14 m (kód kvality 5, 6, 8 podle původu vzniku analogové mapy).

Úplná náhrada grafických katastrálních map mapami v digitální formě je jednou ze základních podmínek komplexní realizace a provozu informačního systému katastru nemovitostí (ISKN), který patří mezi nejobjemnější a nejsložitější

informační systémy státní správy. Tento systém by měl mimo jiné zajistit efektivní a bezpečný proces evidence pozemků a jejich převodů. Vysoké nároky je třeba klást na technickou spolehlivost systému a kvalitu dat.

Vlastní proces digitalizace souboru geodetických informací (SGI) není nijak lehký, uvědomíme-li si, jak technologicky a obsahově rozdílné jsou stávající analogové katastrální mapy. Digitální katastrální mapy by měly nést atributy jednotného státního mapového díla velkého měřítka. Proto je žádoucí technologicky sjednotit všechny typy DKM, i když mají v jednotlivých katastrálních územích rozdílnou technickou kvalitu (přesnost polohy bodu) v závislosti na kvalitě podkladové analogové katastrální mapy a měřických manuálů, nebo způsobu vedení katastrálního operátu (KO) v daném k.ú.

Soubor geodetických informací (SGI) by měl být zásadním způsobem transformován tak, aby se stal bází geodat s maximální podrobností. Deklarovaná relativní přesnost stávajících katastrálních map by měla být zachována a povýšena podle potřeb a požadavků dalších uživatelů na mapové dílo zobrazující předměty mapování polohově a geometricky spolehlivě. Takto by bylo možno kombinovat právní aspekty katastrální mapy s prostorovými elementy datových bází informačních systémů. Tím by byl také položen základ víceúčelového katastru, ke kterému směřuje řada evropských států. Data vedená v ISKN by měla být jedním z důležitých zdrojů informací o změnách při aktualizaci základní báze geografických dat (ZABAGED). Digitální SGI by se měl stát hlavním zdrojem dat pro tvorbu katastrální vrstvy digitální Státní mapy $1 : 5\,000$ (SM 5).

Metodika tvorby DKM v lokalitách sáhových map popsaná v [5] byla již předkládána se záměrem koncepčního řešení údržby DKM. Je zřejmé, že složitost procesu údržby DKM bude závislá na úrovni (typu) přepracované digitální katastrální mapy. Cílovým řešením je dosažení jednotné úrovně DKM vedené v prostředí ISKN. Časový horizont tohoto stavu však není reálný v několika příštích letech. Proto je navrhováno postupovat etapově tak, aby byla prioritní záležitostí především jednotnost metodiky údržby a přebírání výsledků zeměměřických činností pro KN s garantovanými parametry přesnosti i garancí majetkoprávních vztahů. I při etapovém řešení digitalizace je možné kontinuálně naplňovat základní registry ISVS požadovanými geoprvky (v úvodní etapě např. katastrální území[8] vymezené prvkem obsahu KM – vyrovnané katastrální hranice), které jsou vymezovány a garantovány katastrem nemovitostí.

V procesu vedení takto vytvořené DKM je kladen důraz na kvalitu šetření a zaměřování vyšetřených skutečností společně s odpovědným doplněním veškerých přepracovatelných výsledků, čímž se kontinuálně zvyšuje technická a právní spolehlivost katastrálního operátu (KO). Aby byly výsledky digitalizace využitelné v dohledně krátké době pro veškeré zeměměřické činnosti, které ze stá-

[8] **Katastrální území** – technická jednotka, kterou tvoří místopisně uzavřený a v katastru společně evidovaný soubor nemovitostí

vajícího KO vycházejí nebo na tento operát navazují, případně jej aktualizují, je navrženo etapové řešení digitalizace s prioritním důrazem na vytvoření souvislého zobrazení map PK na celém území státu, zaměřování změn v S-JTSK a vedení DKM ve formě hybridní digitální katastrální mapy, která umožní kontinuální zpřesňování obnoveného operátu KN. Zásadní je návrh vytvoření a plnění databáze pevných bodů (DB PB) pro širokou škálu činností. Podrobné informace byly publikovány např. v [5].

Tvorba katalogu geoprvků DKM

Jedním ze základních datových fondů pro naplnění a aktualizaci datových bází infrastruktury prostorových dat na národní úrovni je digitální katastrální mapa (DKM), vytvářená v procesu digitalizace souboru geodetických informací (D-SGI). Obsah katastrální mapy jako součásti SGI je stanoven zákonem č. 344/1992 Sb., o katastru nemovitostí ČR, a ve vyhlášce č. 190/1996 Sb., kde je uvedena podrobná specifikace jednotlivých předmětů obsahu katastrální mapy. Struktura DKM je popsána v předpisu Struktura a výměnný formát digitální katastrální mapy a souboru popisných informací katastru nemovitostí České republiky a dat BPEJ verze 1.3 (ze dne 24. listopadu 1999 č. j. 5270/1999-22). Tento předpis společně s vyhláškou č. 190/1996 Sb., tvoří podklad pro návrh katalogu geoprvků DKM. Návrh katalogu geoprvků vycházejícího z úrovně katastru nemovitostí byl publikován např. v [9].

Návrh směrnice Evropského parlamentu

Problematika základních fondů geodat je dnes intenzivně sledována i v rámci připravovaného evropského projektu INSPIRE (Infrastucture for Spatial Infrastructure in Europe). V těchto podkladech se definují tzv. referenční data (reference data), ekvivalentně užívána též základní data (core data), vymezená jako „data potřebná k identifikaci polohy (lokalizace) fyzického jevu (přírodního nebo umělého) sloužící pro zobrazování jiných informací v geoprostorovém kontextu. Referenční data jsou aplikačně nezávislá a poskytují objektivní obraz reálného světa.“ Na jiném místě jsou charakterizována jako „data poskytující jednoznačnou lokalizaci pro uživatelské informace, usnadňující slučování dat z více zdrojů a poskytující souvislosti umožňující lepší porozumění prezentovaným informacím“. V návrzích se předpokládá, že každý stát EU bude muset postupně zajistit dostupnost referenčních dat, a to:

- souřadnicové referenční systémy,
- geografické mřížové systémy,
- geografické názvy,

- administrativní jednotky,
- dopravní sítě,
- hydrografii,
- chráněná území,
- výškopis
- identifikátory nemovitostí (adresní body),
- katastrální parcely,
- pokryv území,
- ortofotografické zobrazení,
- budovy.

Projekt INSPIRE sleduje vytvoření funkčního informačního systému umožňujícího výběr uživatelsky potřebných datových bází na teritoriu Evropy (meta-informační resp. katalogové služby) a jejich publikaci uživateli nebo jejich přímé využití a modifikaci – informační a datové služby. Jako součást INSPIRE se explicitně sleduje i řešení problematiky referenčních dat (jejich vymezení), podmínek přístupu k datům a podmínek spojených s jejich využitím a standardizací.

Návrh směrnice[8] vychází z doporučení pracovních skupin INSPIRE. Návrh stanovuje licencování a peněžní politiku pro geodata a mechanismus jejich sdílení. Dále garantuje veřejnosti bezplatné nalezení datových sad a s nimi spojených služeb na základě metadat. Zdarma je zajištěno také prohlížení těchto dat s následujícími funkcemi: zobrazení (display), navigace (navigate), přiblížení (zoom in), oddálení (zoom out), posun zobrazení (pan) a překrytí (overlay) datových sad a zobrazení legendy a veškerého obsahu metadat. Stažení a kopie datových sad nebo jejich částí bude zpoplatněno.

Návrh dále obsahuje seznam a stručný popis témat prostorových dat, která by měla být součástí infrastruktury prostorových dat. V původním návrhu INSPIRE bylo 60 témat. Tato témata byla vybrána na základě konferencí, mítinků a konzultací nejen pracovních skupin, ale i dotčených organizací a členských zemí. Vzhledem k vysokým nákladům implementace šedesáti témat prostorových dat byla vytvořena pracovní skupina zaměřená na přezkoumání rozsahu těchto dat a jejich výsledek je popsán ve Scoping Paper [3]. Zredukování témat proběhlo ve velmi krátké době (2–3 měsíce) a výběr byl uskutečněn pouze pracovní skupinou, takže členské země a dotčené organizace neměly možnost ovlivnit obsah infrastruktury. V návrhu směrnice Evropského parlamentu byl výběr témat opět změněn. Je jich 31 a jsou rozdělena do 3 skupin (viz příloha I, příloha II, příloha III). Do katalogu geoprvků DKM nebyla zahrnuta témata ze 3. skupiny obsahující převážně tematická data vztahující se zejména k životnímu prostředí.

Porovnání katalogu geoprvků DKM a katalogu objektů ZABAGED s požadavky INSPIRE

V červenci 2004 vešel v platnost výše zmíněný návrh směrnice Evropského parlamentu [8], který obsahuje témata prostorových dat, jež by měla být součástí národní úrovně INSPIRE. Bylo provedeno porovnání požadavků návrhu směrnice s dříve navrženým katalogem geoprvků DKM a s katalogem objektů ZABAGED. Výsledek porovnání je uveden v tabulce 1.

Tabulka 1 Porovnání DKM, ZABAGED a INSPIRE

Témata prostorových dat	Výskyt v katalogu		Poznámka
	geoprvků DKM	objektů ZABAGED	
Souřadnicové referenční systémy	ano	ano	Systémy pro jednoznačné vyjádření prostorových referencí jakožto množiny souřadnic (x, y, z) nebo šířka, délka a výška založené na geodetickém a výškovém systému.
Geografické mřížové systémy	ne	ne	Pravidelné mříže s několika rozlišeními se společným počátečním bodem a standardizovaným umístěním a rozměrem buněk mříže.
Geografické názvy	ano	ano	Názvy ploch, regionů, lokalit, velkoměst, městských čtvrtí, měst a sídel nebo geografické či topografické jevy veřejného nebo historického zájmu.
Administrativní jednotky	ano	ano	Státní území rozdělené na administrativní jednotky pro místní, regionální a státní správu. Administrativní jednotky jsou odděleny administrativními hranicemi. Také zahrnují hranice státního území a pobřežní linii.
Dopravní sítě	ano	ano	Silniční, železniční, vzdušné a vodní dopravní sítě a s nimi spojená infrastruktura. Zahrnuje spoje mezi různými sítěmi. Také obsahuje trans-evropskou dopravní síť definovanou v usnesení 1692/96/EC25 ve znění pozdějších předpisů.

Témata prostorových dat	Výskyt v katalogu		Poznámka
	geoprvků DKM	objektů ZABAGED	
Hydrografie	ano	ano	Hydrografické prvky, přírodní i umělé, jsou řeky, jezera, přechodná vodstva, nádrže, zvodnělé vrstvy, kanály nebo další vodní tělesa. Tyto prvky odpovídají vodní síti a jsou provázány s ostatními sítěmi. Hydrografie zahrnuje povodí řek definované ve Směrnici 2000/60/EC.
Chráněná území	ano	ano	Plocha vyhrazená nebo regulovaná a vedená k zachování určitých chráněných objektů.
Výškopis	ne	ano	Digitální výškový model pro povrchy země, ledu a oceánu. Zahrnuje zemský výškopis, měření hloubek a břehovou čáru.
Identifikátory nemovitostí	ano	ne	Identifikátor nemovitostí, obvykle podle názvu ulice, čísla budovy a kódu poštovního obvodu.
Katastrální parcely	ano	ne	Plocha definovaná katastrálními hranicemi s určitým právním statutem vlastnictví.
Pokryv území	ano	ano	Fyzický a biologický pokryv zemského povrchu zhrnující umělé povrchy, zemědělské plochy, lesy, přírodní plochy, mokřiny, vodní tělesa.
Ortofotografické zobrazení	ne	ano	Prostorově určená obrazová data zemského povrchu, ze senzorů dopravovaných družicemi či letadly.
Budovy	ano	ano	Geografické umístění budov.

Závěr

Geografické informační systémy, které ještě na počátku devadesátých let minulého století byly nástrojem několika tisíc specialistů se v současné době stávají běžnou součástí v procesu řízení společnosti, výroby, vzdělávání, ale zasahují stále častěji do každodenního života občanů. Má-li společnost na tyto trendy relevantně reagovat, musí zajistit prioritně využití těchto technologií v řízení veřejné správy. Dostupnost informací a služeb veřejné správy v elektronické podobě je nejen jednou z priorit Státní informační politiky ČR, ale zejména významnou

službou pro obyvatele České republiky, kterým by měla usnadnit kontakt s veřejnou správou.

Řada projektů ISVS a naplňování základních bází geodat je koncepčně vázáno na ISKN resp. proces digitalizace SGI. Proces digitalizace katastrálních map je problematika velice komplexní, která zajisté ovlivní funkčnost celého KN na následující dlouhé období, a tím i majetkoprávní garance nemovitého majetku většiny občanů ČR – majitelů nemovitostí.

S výše uvedenými referenčními geoprostorovými daty je v současné době spojena řada problémů, jako např. nejasné nebo dosud nespecifikované vzájemné vazby, pomalá realizace projektů nebo neexistence požadovaných dat na celém území ČR. Dále se jedná především o úroveň dosažitelné podrobnosti a přesnosti referenčních geodat, homogenizaci stávajících geodat a duplicitní financování sběru stejných kategorií geodat. Geodata mají integrační charakter a tvoří společný základní obsah většiny tematických či aplikačních datových bází využívaných v prostorově orientovaných rozhodovacích procesech veřejné správy i mimo ni. Uspokojivě není řešen ani mechanismus kontinuální aktualizace a důsledné respektování norem ISO řady 19100 – Geografická informace.

Je nezbytně nutné prosadit fundamentální zásadu, že základní geoprostorová data představují v působnosti veřejné správy strategický národní zdroj mimořádné hodnoty, která je dána relativně vysokými náklady na jejich pořízení, aktualizaci a správu, hrazenou převážně ze státního rozpočtu. Proto je nezbytné nejen koordinovat ochranu těchto dat (osobních dat, dat významných pro obranu státu apod.), ale také se zasadit o jedinečnou datovou reprezentaci každého potřebného prvku reálného světa.

Správa a poskytování geodat ve vazbě na jejich kvalitativní různorodost, objem a především při požadavku na kontinuální aktuálnost není v současné době myslitelná bez technologií webových mapových služeb (Web Map Service – WMS) pro sdílení geodat v distribuovaném prostředí Internetu. Uživatelé takto mohou sdílet datové sady geodat bez nutnosti mít příslušná data na svém počítači nebo serveru. Služby WMS budou postupně doplňovány a rozšířeny službami Web Feature Service (WFS). Pouze takto lze do budoucna uvažovat o vybudování geoinformační infrastruktury v národním, ale i mezinárodním měřítku.

Veškerý vývoj geografických informačních systémů je podmíněn lidským faktorem obsluhy a především uživatelů těchto systémů. Proto je zcela zásadní položit důraz na vzdělávání a výzkum v těchto oblastech.

Literatura

[1] *Národní geoinformační infrastruktura České republiky. Program rozvoje v letech 2001–2005.* Nemoforum, 2001.

[2] *Terminologický výkladový slovník pojmů z oblasti geoinformací.* Věstník ÚVIS, 2001, roč. II, částka 3.

[3] *Task Force Scoping – C. STEENMANS: INSPIRE Scoping Paper.* Final version 24. 3. 2004.

[4] MARTINEK, M., HOJDAR, J., ČADA, V. Základní datové báze geodat. 10. konference GIS... Ostrava. VŠB-TU Ostrava a CAGI. 27.–29. 1. 2003. In sborník konference na http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2003/Sborník/ [online]. 14. 9. 2006.

[5] ČADA, V. *Návod pro obnovu katastrálního operátu přepracováním ze systému stabilního katastru.* Praha : ČÚZK, 2001.

[6] ČADA, V. Koncepce vedení a údržby digitálního souboru geodetických informací. *40. Geodetické informační dny. Sborník příspěvků.* Brno : 2004, str. 113–125. ISBN 80-864-31-5.

[7] *Zákon č. 319/2004 Sb. kterým se mění zákon č. 200/1994 Sb., o zeměměřictví a o změně a doplnění některých zákonů souvisejících s jeho zavedením, ve znění pozdějších předpisů.*

[8] *Proposal for a Directive of the European Parliament and of the Council establishing an Infrastructure for Spatial Information in the Community (INSPIRE).* COM(2004) 516 final, kód Rady 11781/04, 2004/0175 (COD).

[9] ČADA, V., MILDORF, T. Vymezení základních geoprostorových dat na úrovni pozemkového datového modelu. GIS Ostrava 2005. In sborník konference na http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2005/Sborník/ [online]. 14. 9. 2006.

[10] ČADA, V., MÍKA, S., ŠÍMA, J. Desetileté výročí studijního programu „Geomatika“ na FAV. In *Univerzitní listy.* leden 2006, ZČU v Plzni, 2006. http://noviny.zcu.cz/archiv/UN1_06.pdf [online]. 14. 9. 2006.

Současnost a budoucnost „Mapového portálu města Plzně“ http://gis.plzen-city.cz

Stanislav Štangl

E-mail: stangl@sitmp.cz

Pokud hledáte aktualizované mapy Plzně, podívejte se na oficiální stránky Magistrátu města Plzně (MMP) na **Mapový portál města Plzně**. Jedná se o průvodce mapami, aplikacemi a službami v GIS, který slouží ke snadné orientaci všem uživatelům. Je k dispozici na adrese http://gis.plzen-city.cz nebo je přístupný z hlavních stránek Informačního serveru města Plzně (http://info.plzen-city.cz) pod odkazem „Mapový portál“.

Na jediné adrese zde naleznete mimo jiné odkazy na aplikaci GSHTML s mapami zaměřenými na konkrétní téma (turistika, doprava, služby, životní prostředí, územní členění apod.) nebo odkazy na aplikaci GSWEB s kompletními veřejnými daty a propracovanými uživatelskými funkcemi pro prohlížení map. Pravděpodobně vás zaujmou i speciální aplikace využívající GIS data (3D model, Povodňový model, dynamicky generované www stránky). V nabídce GIS portálu je obsažen popis Geografického informačního systému města Plzně, včetně uvedení výhod systému, technologické struktury, seznamu dostupných dat, ukázek možností systému, způsobu vedení pasportů a jednotlivých kategorií dat. Portál obsahuje některé odkazy na užitečné adresy s podobnou tematikou na Katastr nemovitostí ČÚZK, na mapy Krajského úřadu Plzeňského kraje, Západočeskou univerzitu, spolupracující firmy apod.

GIS portál a aplikace v něm prezentované jsou využívány pracovníky města, městských obvodů, příspěvkových organizací zřízených městem, finančního úřadu, katastrálního úřadu, krajského úřadu, správci inženýrských sítí, projektanty, geodety, realitními kancelářemi, soukromými firmami, studenty vysokých škol a samozřejmě veřejností. Počty přistupujících uživatelů poukazují na oblibu a hojné využívání aplikací a od nasazení systému v roce 2001 se počet přístupů neustále zvyšuje.

Správa, údržba a vývoj řešení je realizován vlastními silami pracovníků Správy GIS na Správě informačních technologií města Plzně (SITMP). O kvalitní práci vypovídají získaná ocenění: „Geoaplikace 2004“, kde GIS portál města Plzně získal zvláštní cenu odborné poroty v kategorii A – GIS pro lepší služby veřejné správy a „Geoaplikace 2001“ – 1. místo získal Geografický informační systém města Plzně s aplikací GSWEB v kategorii C – města a obce.

Obr. 1 Ukázka úvodní stránky mapového portálu města Plzně

Cíle GIS města Plzně

Provozovaný GIS je součástí komplexního Informačního systému města Plzně a plní celou řadu požadavků. Základním požadavkem na funkčnost GIS bylo vytvoření centrálního skladu geografických dat (implementován v r. 2000), který chápeme jako stabilní databázi, v níž jsou uloženy všechny grafické objekty společně s popisnými informacemi. K duplikaci dat nedochází a každý prvek respektuje „stavovou logiku“ životního cyklu geografických objektů včetně možnosti historizace. Sklad podporuje požadovanou bezešvost map.

Důležitým předpokladem je správa, úprava datového modelu a rozvíjení systému vlastními silami. Prostředky databáze lze přidělovat veškeré přístupy a využívat všech jejích výhod (role, trigery, pohledy atd.). Nedílnou součástí systému je zavedení číselníků, subtabulek a předdefinovaných hodnot.

Hlavním cílem je poskytování geografických dat v digitální podobě uživatelům prostřednictvím internetových aplikací GSWEB, GSHTML jako jednotného grafického prostředí města. Architektura GIS přináší všestranné využití

GIS v rámci integrovaného informačního systému města, což dokazují jeho propojení s jinými aplikacemi (ekonomický systém SAP R/3, Komplexní datová báze, Správa sídelní zeleně, BMS – mostní hospodářství aj.).

Používaná GIS technologie současně využívá mapovou službu WMS (Web Map Service), která umožňuje prezentovat prostorová data z jiných externích WMS serverů na Internetu dle standardu Open Geospatial Consortia – Web Map Service (ISO 19128). Uživatel tak získává možnost využívat vybrané mapové poklady pro území celé ČR (popř. EU) a WMS efektivně řeší integraci dat z různých zdrojů.

Obr. 2 Ukázka propojení intranetového klienta GSWEB s Nahlížením do katastru nemovitostí (ČÚZK)

Dostupnost geoinformací

Na stránkách o GIS města Plzně se pravidelně zobrazují aktuální informace, jež se týkají zejména rozvoje systému a zpřístupnění nových mapových vrstev. Konkrétní informace o aktuálnosti, původu grafických dat a jejich přístup-

nosti pro veřejnost jsou k dispozici na hlavní stránce aplikace GSWEB, popř. na http://gis.plzen-city.cz/ogis/data.asp. V databázi jsou vedeny desítky grafických vrstev (map), které jsou tematicky členěny do kategorií, např. KATASTRÁLNÍ MAPA, TECHNICKÁ MAPA, ÚZEMNÍ PLÁNOVÁNÍ, RASTROVÉ PODKLADY. Na aktualizaci dat se podílí pracovníci z různých organizačních složek města a změny jsou okamžitě promítnuty v aplikaci GSWEB.

Bohužel ne všechna data jsou na internetu k dispozici z důvodu bezpečnosti nebo licenčních či smluvních omezení. Veřejnosti zůstávají skryté například inženýrské sítě velkých správců a parcelní čísla katastrální mapy. Proto město vytvořilo „Pravidla pro poskytování výstupů z datových souborů GIS města Plzně“, která definují jednotný postup a podmínky pro poskytování výstupů. Všechny dokumenty jsou k dispozici ve formátu PDF nebo DOC na internetové adrese: http://gis.plzen-city.cz/ogis/ v oddíle Služby Správy GIS¿Pravidla pro poskytování výstupů. Bezplatné poskytování výřezů digitálních dat využívají hojně (více než 100 předání) geodetické a projekční kanceláře pro zpracování zakázek zadávaných městem. Stejně tak se poskytují bezplatně vzorky dat studentům pro výukové účely.

Poskytování dalších skužeb

Město provozuje vlastní rozsáhlou informační optickou síť MisNet a PilsNet, která zajišťuje datové propojení jednotlivých budov MMP, úřadů městských obvodů 1–10, příspěvkových organizací města Plzně a organizací zřízených městem např. Plzeňské městské dopravní podniky, Plzeňský holding, Městská policie, některé základní školy atd. Tato síť propojuje také objekty spravované Krajským úřadem Plzeňského kraje, Západočeskou univerzitu, Lékařskou fakultu UK a dále objekty Ministerstva práce a sociálních věcí, Fakultní nemocnici Plzeň a další instituce. Datové propojení usnadňuje komunikaci (elektronická pošta, internet, intranet), umožňuje využívat síťové aplikace a sdílet data (např. databázové ekonomické a grafické úlohy). V současné době je v síti zapojeno 116 objektů a délka optické sítě přesahuje 65 km optických kabelů.

V GISu je veden kompletní pasport této optické sítě a žadateli–stavebníkovi se v zájmovém území (staveništi, trase) poskytuje vyjádření o existenci podzemních sdělovacích vedení městské informační sítě (MISNet). Pro tvorbu a aktualizaci Digitální technické mapy města Plzně potvrzuje Správce technické mapy (Správa GIS SITMP) převzetí geodetické dokumentace (digitálních dat – Zaměření skutečného provedení stavby). Město pořizuje a aktualizuje hlavně digitální povrchovou situaci a na vyžádání ji poskytuje stavebníkovi nebo geodetovi.

Město má k dispozici barevné ortofotomapy z leteckého snímkování mezi lety 1996–2005 a vždy po několika letech se snaží pořídit aktuální snímkování. které dokumentují obrovský stavební a průmyslový rozvoj města a okolí. Mezi hlavní

stavby patří dálniční obchvat s tunelem Valík, komplex průmyslové zóny Borská pole nebo rozsáhlé rekonstrukce komunikací a stavby nových mostů. Všechny snímky ortofotomap jsou k dispozici uživatelům internetu, ale pokud dáváte přednost knižnímu vydání, můžete si zakoupit publikace Atlas ortofotomap Plzně z roku 1998 nebo Atlas ortofotomap Plzeň a okolí z roku 2004.

Závěrem lze shrnout, že mapový portál města Plzně představuje jedinečné řešení v oblasti GIS pro města, které slouží k obsahově řízenému přístupu různých skupin uživatelů ke GIS. Uživatel přistupuje k aplikacím GIS dle svého profilu (úředník, státní orgán, občan, investor, turista) na jediné adrese a je přehlednou formou směrován k požadovaným informacím. Cílem projektu je poskytovat informace co nejširšímu počtu uživatelů podle tematiky, komfortu obsluhy a procesní podpory. Bohatý fond geografických dat je výsledkem fungující spolupráce mezi organizacemi města a je patrně nejrozsáhlejší v oblasti měst v ČR. Internetové řešení a rychlá optická síť nabízí přístup neomezenému počtu uživatelů.

Mapový server Správy NP Podyjí v kontextu přeshraniční spolupráce

Martin Kouřil

E-mail: kouril@nppodyji.cz

Podyjí – země nepopsaná

Krajina mezi Znojmem a Vranovem odedávna přitahovala pozornost. Romantiky počátku 19. století počínaje, přes početné členstvo turistických spolků posledních let habsburské říše až po prvorepublikové trampy a skauty předznamenávající moderní pojetí turistiky. Tehdy bylo střední Podyjí vnímáno především esteticky a pocitově a teprve jakoby z pozadí zaznívaly ojedinělé hlasy přírodovědců. upozorňující na mimořádný význam této oblasti v evropském kontextu.

Uzavření většiny tohoto území do nepřístupného hraničního pásma v 50. letech minulého století znamenalo nejen podstatné omezení vlivů člověka na přírodní procesy, ale vedlo rovněž k přerušení kontinuity poznávání tohoto území odbornou i laickou veřejností. Pádem *železné opony* v listopadu 1989 se nejen v Podyjí otevřely vpravdě nové obzory, a každý si mohl na vlastní oči zažít ten kolumbovský pocit objevení nové země. Nepočetné vědecké práce z období před rokem 1950, poukazující na mimořádnou biodiverzitu v Podyjí, dávaly tušit enormní zájem veřejnosti i vědeckých a vzdělávacích institucí o tuto oblast. Kompaktnost území kolem kaňonu Dyje a přiměřená rozloha byly předpokladem pro uchopení Podyjí jako ideálního modelu pro výzkumné úkoly ze všech oblastí přírodních věd. Bylo nesporné, že v krátké době dojde k toku velkého množství dat *ven* – k člověku, a posléze i *dovnitř* – k přírodě.

Jak na data

Při vyhlášení Národního parku Podyjí v r. 1991 a ustavení Správy NP Podyjí, coby instituce pečující o toto území, byla zohledněna budoucí potřeba pohybu velkého množství dat a již v začátcích se přistoupilo k budování pracoviště *geografického informačního systému* (GIS). Je na místě připomenout, že se tak událo na samém počátku masivního využití výpočetní techniky, kdy pojmy z oblasti

IT teprve hledaly svůj konkrétní význam. Pro mnohé pojem *informační systém* představoval počítač s nějakým programem a to podstatné – tedy problematika dat, jejich zdrojů, struktury, principy aktualizace a publikace – zůstávalo často v pozadí.

Z dnešního pohledu je podstatné, že v čele tehdejšího Ministerstva životního prostředí (MŽP), coby zřizovatele Správ národních parků, stanuli osvícení lidé, kteří správně pochopili význam dat. Věděli, že se bude z valné části jednat o data geograficky orientovaná a jelikož jim pojem GIS nebyl neznámý, prozíravě se zasadili o vybavení všech pracovišť spadajících pod dikci MŽP jednotným softwarovým vybavením – produktem ArcInfo firmy ESRI. Tento krok, jak se později ukázalo, byl impulsem, ze kterého dodnes žije komunita uživatelů GIS v ČR, bez ohledu na to, jaký software používá. Lze konstatovat, že mezinárodním kontextu je dnes Česká republika na špici mezi evropskými zeměmi nejen v používání vyspělých technologií GIS ale i ve správném chápání tohoto pojmu.

Pracoviště Informatiky a GIS Správy NP Podyjí začalo prakticky fungovat počátkem r. 1992 a personálně bylo (a do dnešních dnů je) obsazeno 1 pracovníkem. Na svou dobu bylo vybaveno špičkově – PC 486, 8 MB RAM 340 MB HD, tabletem A2 a barevnou tiskárnou A3.

V prvních dvou letech probíhalo strategické plánování výstavby GIS a byla pořízena některá zkušební digitální data GIS. Hledaly se vhodné postupy pořizování dat, datové formáty, a měřítka zpracovávaných map.

Již v průběhu r. 1993 vzniklo několik projektů, jejichž výsledky se používají dodnes. Je to především digitální mapa hospodářského rozdělení lesa (pilotní projekt pro ČR), a síťové botanické mapování. K rozvoji GIS organizace přispěl i původní český software Topol, jehož snadné ovládání vedlo k rozhodnutí pořizovat data především vlastními silami a dodnes se používá v aplikacích lesního hospodářství.

Nakukování k sousedům

Geografická poloha NP Podyjí na hranici s Rakouskem vedla brzy k silné potřebě *vidět na druhou stranu údolí* – tedy k potřebě mít k dispozici data i z rakouské strany. Situace byla komplikovaná tím, že na rakouské straně v té době neexistoval partner, se kterým by bylo možno navázat spolupráci. Základní mapa ČR 1 : 10 000, která se obecně používá jako referenční podklad pro podrobné mapování témat z oblasti životního prostředí, je v tomto případě slepá, a pro tehdejší neaktuálnost se ukázala jako nevhodná i pro mapování jevů na české straně. Vznikla proto svépomocí vlastní sada bezešvých topografických vrstev vycházejících z kombinací tématických map různých měřítek počínaje katastrální mapou a konče vojenskými mapami 1 : 25 000, které jako jediné vhodné zachycují území Rakouska.

Do tohoto období patří i první orotfotomapa z r. 1999 pořízena Správou NP Podyjí, která zahrnovala i území připravovaného NP Thayatal na rakouské straně Podyjí a byla zamýšlena jako první vklad do společného bilaterálního GIS.

Konečně spolupráce

Vyhlášení NP Thayatal na rakouské straně Podyjí v r. 2000 bylo z české strany dlouho očekávaným aktem. První konkrétní smlouvou, navazující na obecnou proklamaci o spolupráci obou Správ národních parků, byla *Smlouva o vzájemném sdílení a poskytování dat*, podepsaná počátkem r. 2001. Byl tak vytvořen právní rámec pro budování bilaterálního GIS Podyjí-Thayatal. Nastala doba diskuzí především technického charakteru, řešící rozdílnosti datových formátů a vzájemnou transformaci národních souřadnicových systémů.

Již v těchto fázích se ukázalo, že je třeba věnovat značné úsilí ke sjednocení pohledů nejen v oblasti ochrany přírody, ale i v oblasti sdílení dat. Rozdílné pohledy přitom nebyly způsobeny neochotou se dohodnout, ale především odlišným formálním postavením obou Správ NP v rámci právního prostředí jednotlivých zemí. Zatímco Správa NP Podyjí je v mnohém klasický úřad vykonávající zároveň management (přímé hospodaření) v lesích NP Podyjí a vybraných nelesních plochách i koordinující výzkum celé oblasti, Správa NP Thayatal je zaměřena více na služby v oblasti turistiky a enviromentálního vzdělávání. Z těchto pozic plyne mimo jiné i odlišná potřeba dat. Zatímco česká strana si již patnáctým rokem buduje svůj GIS na vlastním pracovišti, Rakušané používají služeb a zdrojů dolnorakouské vlády – pracoviště NÖGIS v St. Pölten a pro zpracování vlastních dat mají externího spolupracovníka. Potřeba speciálních tématických dat je na rakouské straně viditelně menší a tak větší aktivita směrem ke společnému GIS přichází přirozeně z české strany.

Slepá ulička – ReGeo

V letech 2002–2004 se obě strany zúčastnili projektu financovaného z evropských fondů s názvem ReGeo (http://www.regeo.org), jehož účelem bylo ověřit možnosti vytvoření turistického informačního systému chráněných oblastí na bázi nejmodernějších informačních technologií, mimo jiné i s použitím mapového serveru. V projektu byla jako modelová vybrána 4 evropská chráněná území a NP Podyjí a NP Thayatal zde figurovala na počátku projektu jako dvě oddělené testovací oblasti. Bohužel se lídra projektu – Univerzitu ve Freiburgu – nepodařilo přesvědčit o výhodnosti pojmout území NP Podyjí-Thayatal jako jeden celek. A tak se původní cíl, se kterým česká strana do projektu vstupovala – vybudování bi-

laterálního GIS, nepodařilo uskutečnit. Navíc se do tohoto projektu nepodařilo prosadit řádově levnější řešení, které navrhovala česká strana a dal se přednost vývoji sice autonomního, ale těžkopádného a již ve své době technologicky zastaralého mapového serveru. Jediné, co z projektu zůstalo, bylo poučení, že příště takhle už ne. A pak vlastní sebevědomí – praktické potvrzení toho, že v trendech vývoje GIS jsme na technologické špičce a za řešení, která máme k dispozici, se vůbec nemusíme stydět. Z toho vyplynula vůle pokračovat jinou, a to vlastní cestou.

Světlo na konci tunelu – MapServer

Společně s potřebou sdílení dat s rakouskou stranou se v posledních letech objevil další úkol – zjednodušit poskytování a publikování vlastních dat GIS široké veřejnosti. Zároveň vzrostla i potřeba rychlého přístupu k datům GIS uvnitř Správy NP Podyjí. Poté, co ve vnitřním chodu organizace přestalo stačit nasazení několika lokálních desktop ArcView a ArcGIS, bylo rozhodnuto, že je na čase nasadit mapový server. A to pro přístup k datům uvnitř i vně organizace.

Volba technologie byla celkem jednoznačná. Přesto, že zde funguje letitá spokojenost s produkty rodiny ArcInfo, bylo zamítnuto cenově náročné řešení na bázi ArcIMS a pozornost se obrátila na OpenSource produkt MapServer vyvinutý na Minessotské universitě. Výborné reference na samotný produkt i na firmu HSRS, která je renomovaným implementátorem tohoto produktu v českém prostředí, byly zárukou, že se jedná o správnou volbu. Možnost financovat implementaci mapového serveru pomocí projektu *Interreg II.A* bylo impulsem pro započetí prací.

Na jaře letošního roku vznikly první obrysy mapového serveru. Obsahem jsou 4 základní aplikace: *Turistika, Příroda a Výzkum, Péče o krajinu* a *Katastrální mapa*. Aplikace *Turistika* je postavena jako aplikace na technologii DHTML, umožňující snadnější integraci přímo do webu organizace. Zároveň jako jediná zahrnuje i území na západ až po Bítov a Uherčice, jako přirozenou společnou turistickou destinaci s NP Podyjí. Ostatní tři aplikace jsou postaveny na technologii Java appletů, která umožňuje především větší možnosti editace. Toto řešení bylo zvoleno právě s ohledem na využití i ve vnitřním provozu.

Tématická bohatost obsahu je dána poměrně malým rozsahem území a velkým množstvím dat, které má Správa NP Podyjí k dispozici. Za pozornost stojí např. *detailní mapování drobných sakrálních staveb*, už jen proto, že se v prvním pohledu tématicky vymyká z problematiky ochrany přírody, v širším chápání krajiny, jako prostoru po staletí modelované člověkem, sem však toto téma určitě patří.

Čtyři aplikace mapového serveru jsou v rámci projektu *Interaktivní turistická mapa NP Podyjí* doplněny o efektní aplikaci *GeoShow 3D*, která umožňuje

interaktivní pohyb nad 3D modelem krajiny s připojenými informačními prvky různých typů, včetně např. panoramatických snímků. Tato 3D aplikace je obsahově totožná s aplikací *Turistika* na mapovém serveru. Technologicky se jedná o jediný datový soubor o velikosti několika GB, který je možné distribuovat na DVD, nebo interaktivně spouštět v prohlížeči. K tomu je však nutný hosting na speciálním serveru a stažení klientského programu coby prohlížečky. Pro připojení je pak nutná linka nejméně 128 kb/s.

Při přípravě dat pro mapový server byla velká pozornost věnována právě rozhodnutí, která data jsou volně poskytnutelná, která jsou publikovatelná bez omezení a která jsou pouze pro interní potřebu. V principu je na data, jejichž správcem je Správa NP Podyjí a která nejsou omezena jinými licenčními právy, pohlíženo jako na veřejný statek a omezení jejich použití na interní potřebu připadá pouze na taková data, jejichž zveřejnění by ohrozilo samotnou existenci popisovaného jevu. Typickým příkladem je bodové mapování výskytu vzácných rostlin a živočichů. Rozdělení přístupu na veřejnou a neveřejnou část mapového serveru je realizováno na úrovni jednotlivých mapových vrstev definicí uživatelských práv přidělením uživatelského jména a hesla.

MapServer umožňuje sdílet data prostřednictvím WMS služeb jiných mapových serverů. Jako externí data jsou načítány vybrané služby z mapového serveru CENIA (především podkladová data) a oboustranná výměna dat je připravena k realizaci s mapovým serverem Městského úřadu Znojmo. V případě zájmu o volná, či publikovatelná data, je Správa NP Podyjí připravena poskytnout tato data i dalším zájemcům přímo, nebo jako WMS službu. Pro účely orientace v datech Správy NP Podyjí jsou data postupně popisována ve veřejně přístupném metadatovém katalogu MICKA.

Mapový server je spravován pracovištěm GIS Správy NP Podyjí a to jako logické vyústění běžné práce uvnitř organizace a v tomto ohledu se nepředpokládá nárůst pracnosti při operaci s daty. Po ověření funkčnosti mapového serveru se počítá se zprovozněním jeho německé a anglické mutace. Přístup do neveřejných částí mapového serveru bude k dispozici i pracovníkům Správy NP Thayatal. Souvislost zobrazení území je zajištěno pomocí DMU25 – podkladu, který dostatečně pokrývá území obou národních parků a pomocí společné ortofotomapy. Vzhledem ke skutečnosti, že data GIS z velké části pocházejí ze zdrojů na české straně, jsou v současnosti v případě potřeby rakouská data offline transformována do systému JTSK. Praxe ukáže, zda v budoucnu bude nutno využít online transformaci souřadnicových systémů, kterou MapServer umožňuje. Problém spočívá v přesnosti této transformace – diference pohybující se v rozsahu přibližně 15 m jsou momentálně zbytečným kompromisem. Problematiku transformací by bylo patrně možno řešit provozováním společného mapového serveru „na půlce cesty“ – v souřadnicovém systému WGS-84, což je však vzhledem k současnému poměru datových zdrojů nepraktické.

Spuštění společného mapového serveru je možno pokládat za uzavření jedné z kapitol téměř patnáctiletého budování GIS krajiny středního Podyjí. Končí jedna etapa, poznamenaná překotným vývojem technologií i občasným dobrodružným tápáním v neprobádaných vodách IT. Nastávající etapa bude pravděpodobně vyžadovat spíše racionálnější, než pocitové uvažování, neboť přichází doba jasných kompetencí k datům, přísných licenčních práv a náročných uživatelů.

Nezbývá nic jiného, než vyslovit přání, aby nový společný mapový server v této době obstál a přispěl nejen k ke zvýšení informovanosti zájemců o tento krásný kout obou našich zemí, ale stal se zároveň i komunikačním nástrojem ke hledání optimálního způsobu péče o bilaterální národní park Podyjí-Thayatal.

Open Source GIS – uDig, GeoTools

Jan Ježek

E-mail: h.jezek@centrum.cz

Abstrakt

Open Source GIS pokrývá většinu oblastí pro správu geografických dat. Vedle již zaběhlých produktů vyvíjených v jazyce C (Grass, UMN Mapserver) vznikají nové možnosti postavené na objektově orientovaných jazycích. Příspěvek se zabývá základním popisem a strukturou dostupných knihoven a produktů v jazyce JAVA a také používanou standardizací v této oblasti. Zvláštní pozornost je věnována produktu uDig a knihovně GeoTools, vyvíjených firmou Refractions Research. Vedle samotného popisu jsou představeny také možnosti spolupráce a rozšíření těchto projektů. Článek se zaměřuje především na technologické aspekty a je určen především těm, kteří hledají způsob, jakým využít Open Source produkt k jeho rozšíření a uplanění jak samostatně, tak i v jiných aplikacích.

1 Specifikace Open GIS konsorcia (OGC)

Jako ve většina oblastí IT i GIS procházejí procesem standardizace. Důvodem je rozvoj webových služeb a s tím spojený tlak na maximální interoperabilitu. Zásadní vliv na specifikace a standardizaci v GIS má několik konsorcií. Mezi nejdůležitější subjekty, které se přímo standardizací v Geoinformatice zabývají, patří ISO (International Organization for Standardization) a OGC (Open Geospatial Consortium). Tato konsorcia vyvíjí specifikace aplikačních rozhraní a protokolů, které umožňují interoperabilitu v rámci aplikací, prostorových dat a služeb tzv. „geoprocessingu".

OGC je mezinárodní průmyslové neziskové konsorcium více než 300 obchodních společností, univerzit a vládních organizací, které společně usilují o interoperabilitu v oblasti Geografických informačních systémů a tzv. „Location Base" službách. OGC bylo založeno v roce 1994.

2 Vývoj Open Source GIS

2.1 Projekty v jazyce C

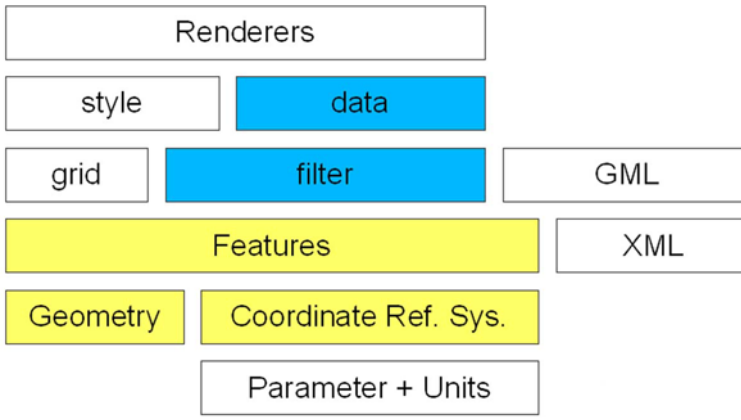
Obecně platí, že projekty v jazyce C jsou mnohem vyzrálejší, a to především díky delšímu časového období jejich vývoje. Základem těchto projektů jsou softwarové knihovny PROJ a GDAL určené pro práci se souřadnicovými systémy a datovými formáty. Knihovny lze stáhnout například jako produkt FWTools. (http://fwtools.maptools.org/).

Mezi nejstarší desktop aplikace patří GIS GRASS, který dodnes představuje jeden z nejúspěšnějších Open Source projektů vůbec.

2.2 Projekty v jazyce JAVA

Projekty v jazyce JAVA probíhají vývojem a momentálně proto nemůžou ještě příliš konkurovat těm v jazyce C. Přesto se zde vyvíjí komplexní řešení všech částí GIS produktů. Existuje několik nezávislých projektů, ale také komplexní řešení na bázi knihoven a jejich implementací do desktop i do server GIS aplikací. Hlavní představitelům se budeme věnovat v následujícím textu.

3 GeoTools



Obr. 1 GeoTool

Geoltools představuje Open Source JAVA GIS toolkit (pod licencí LGPL) pro vývoj GIS produktů s velkým respektem k OGC specifikacím. Důraz je kladen také na modularitu celého systému tak, aby uživatel mohl využívat jen ty

části, které skutečně potřebuje. Pohodlný přístup k požadovaným součástem zaručuje nástroj pro správu softwaru (software project management) – MAVEN (http://maven.apache.org/). Struktura částí GeoTools je patrná na obr. 1. GeoTools se soustřeďují pouze na funkcionalitu a práci s geografickým daty – nepředstavují knihovnu grafických komponent pro tvorbu uživatelského rozhraní.

GeoTools mají maximálně otevřený proces vývoje, k dispozici je veřejný mail list, a to jak pro uživatele, tak pro vývojáře. Další možností komukace je IRC chat, kterého se ve smluveném čase účastní většina hlavnách vývojářů. GeoTools jsou závislé ještě na těchto knihovnách:

- GeoApi – knihovna JAVA rozhraní podle specifikací OGC, jejich implementací je zaručeno splnění specifikací (http://docs.codehaus.org/display/GEO).
- JTS (Java Topology Suit) – knihovna pro základní topologické funce (http://www.vividsolutions.com/jts/jtshome.htm). Knihovna obsahuje základní objekty jako bod, polygon atd., avšak nerespektuje specifikace OGC. V současnosti je snaha o minimalizaci závislostí na této knihovně.

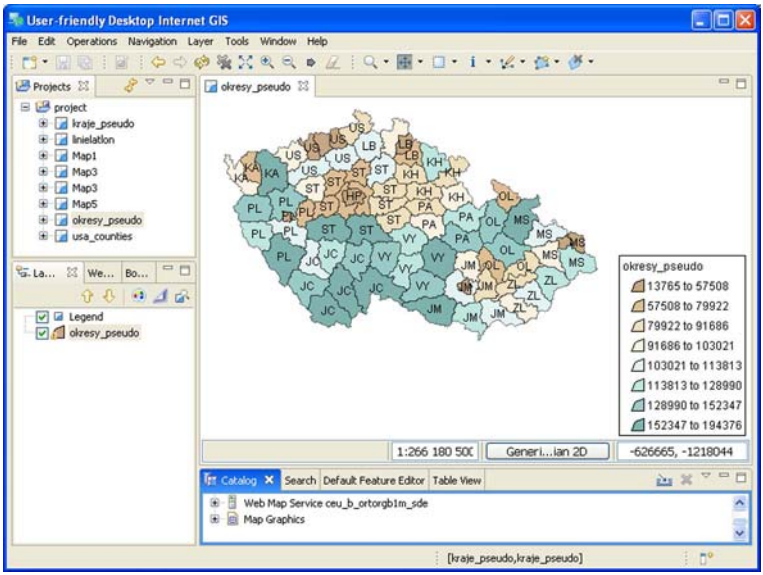
4 uDig

uDig (User-friendly Desktop Internet GIS) představuje desktop GIS produkt (pod licencí LGPL), s důrazem na internetové mapové služby a OGC specifikace. Spolu s aplikacemi Geserver, GeoTools i PostGIS stojí i za vývojem uDigu především firma Refractions Research. uDig je postaven na knihovně GeoTools a na platformě Eclipse Rich Client Platform. Tím je zaručena variabnilta celého systému, který se skládá z oddělených součastí (pluginů) a lze tedy jednoduše přidávat další plug-in moduly, případně také stávající moduly využívat v jiných aplikacích postavených na RCP. Platforma RCP zároveň přináší knihovny SWT a JFACE pro tvorbu uživatelského rozhraní. Za vývojem těchto grafických knihoven stojí firma IBM. Ukázka pracovního prostředí viz obr. 2. uDig i GeoTools využívají pro uložení zdrojového kódu společný SVN server http://svn.geotools.org/.

V Současnosti uDig poskytuje tuto funkcionalitu:

- WFS client read/write umožňuje jak prohlížení, tak editaci dat poskytovaných prostřednictvím služby WFS a WFS-T.
- WMS client umožňuje prohlížení dat zprostředkovyných pomocí WMS služby.
- Podporuje Styled Layer Descriptor (SLD), umožňuje barevnou tematizaci grafických podkladů (přidělení barvy prvku dle hodnoty jeho atributu).

- Podpora tiskového výstupu.
- Podpora standardních GIS formátů.
- Podpora práce se souřadnicovými systémy.
- Podpora připojení databází – PostGIS, OracleSpatial, ArcSDE a MySQL.
- uDig je nezávislý na platformě Windows, OS/X a Linux.



Obr. 2 uDig

5 Geoserver

Geoserver (pod licencí LGPL) je implementací Web Feature Server specification OpenGIS konsorcia založené na jazyce JAVA (J2EE). Aplikace je postavena na knihovně Geotools, což umožňuje oddělenou správu základní logiky. Z technického hlediska se jedná o webovou aplikaci založenou na JSP a servletech fungující pod některým z aplikačních serverů (např. Tomcat). Oproti nejrozšířenější obdobné aplikaci UMN Mapserver vyniká především jednodušší instalací i obsluhou. V současnosti umožňuje serverovat tyto datové formáty:

- Oracle Spatial
- ArcSDE

- PostGIS
- ESRI Shape Files

Tato data jsou zpřístupněna jako služby WFS (standard pro serverování vektorových dat), WMS (standard pro serverování rastrových dat) nebo WFS-T (stadart umožňující i editaci dat). Zajímavostí je plánovaná podpora služby WCS (Web Coverage Service), která umožňuje serverovat multidimenzionální rastrová data, např. rastrová data spolu s informací o nadmořské výšce pixelu (digitální model terénu). Dalším výhodou oproti kunkurenčním produktům je možnost serverovat data ve formátu KML, a tak je zobrazovat v aplikaci Google Earth.

6 Závěr

Popisované projety jsou v dnešní době ve fázi vývoje. Lze najít mnoho ktitiků těchto technologií, ale rozhodně se jedná o jeden z nejdynamičteji se rozvíjejcích počinů v oblasti Open Source GIS. Důkazem je změna reputace, kterou si tyto projekty vysloužily během posledních dvou let, kdy od neznámého experimentu tvůrci došli k respektovanému řešení ve většině komunit zaměřených na Open Souce GIS.

Literatura

[1] uDig home page, http://udig.refractions.net

[2] GeoTools home page, http://geotools.codehaus.org/

[3] Geoserver home page, http://docs.codehaus.org/

[4] Open Geospatial Consortium home page, http://www.opengeospatial.org/

GPS, Windows Mobile 5.0 a Virtual Earth

Štěpán Bechynský

E-mail: Stepan.Bechynsky@microsoft.com

Pro sběr GIS, resp. GPS dat lze v současné době velmi snadno využít kapesní počítače (PDA). Na našem trhu je několik modelů od různých výrobců obsahujících integrovaný GPS modul. Také je dostatek externích modulů připojitelných přes Bluetooth.

Poslední verze operačního systému Windows Mobile 5.0, určeného pro kapesní počítače, obsahuje API pro komunikaci s GPS modulem. Odpadá tak pracné dekódování dat, která poskytuje GPS modul. GPS moduly se převážně chovají jako zařízení připojené na sériovém portu. Problém pak nastává, pokud potřebujete, aby s jedním GPS modulem komunikovalo více aplikací. Zde jsme omezeni vlastnostmi sériového portu. První aplikace, která si COM port s GPS modulem otevře, tak ho zároveň zablokuje a ostatní aplikace se k němu již nedostanou. Tento problém také GPS API řeší. Nevýhodou je, že aplikace musí být psána tak, aby k GPS přistupovala pomocí GPS API a ne přes sériový port, jak bylo zvykem. GPS API není dostupné u starších verzí Windows Mobile.

Pro vývoj aplikací na platformě Windows Mobile je potřeba nainstalovat Windows Mobile SDK, které je zdarma k dispozici na http://www.microsoft.com. Součástí WM SDK je ukázková aplikace využívající GPS API.

K zobrazení nasbíraných dat lze využít jak komerční GIS aplikace tak mapy dostupné na Internetu. Jednou z nich je Virtual Earth. Popis VE SDK je na adrese http://dev.live.com/virtualearth/sdk/. Většina mapových serverů, které jsou na Internetu, funguje na podobném principu. Ke stránce, kde se má mapa zobrazit, je nejdříve je třeba připojit JavaScript soubor, který obsahuje potřebné objekty. Pak je potřeba vytvořit místo pro mapu. To bývá element HTML div s atributem id. Pak už se jen vytvoří objekt mapy a předá se mu id elementu, který slouží jako kontejner pro mapu.

Ukázka kódu

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<script src="http://dev.virtualearth.net/mapcontrol/v3/mapcontrol.js"
```

```
     type="text/javascript"></script>
<title>Demo 1</title>
</head>

<body>
<div id="mapa" style="width:1024px; height:728px;position:absolute">
    </div>
<script type="text/javascript">
//<![CDATA[
        var map = new VEMap(‘mapa‘);
        map.LoadMap(new VELatLong(50.04780278, 14.45501111), // Sred mapy
         15, // zoom 1 - 19
         VEMapStyle.Road, // Styl - cesty, letecky snimek, ptaci pohled,
             cesty + letecky snimek
         false);
// ]]>
</script>
</body>
</html>
```