
ITAPA 2013

***Big Data vo
verejnom sektore***

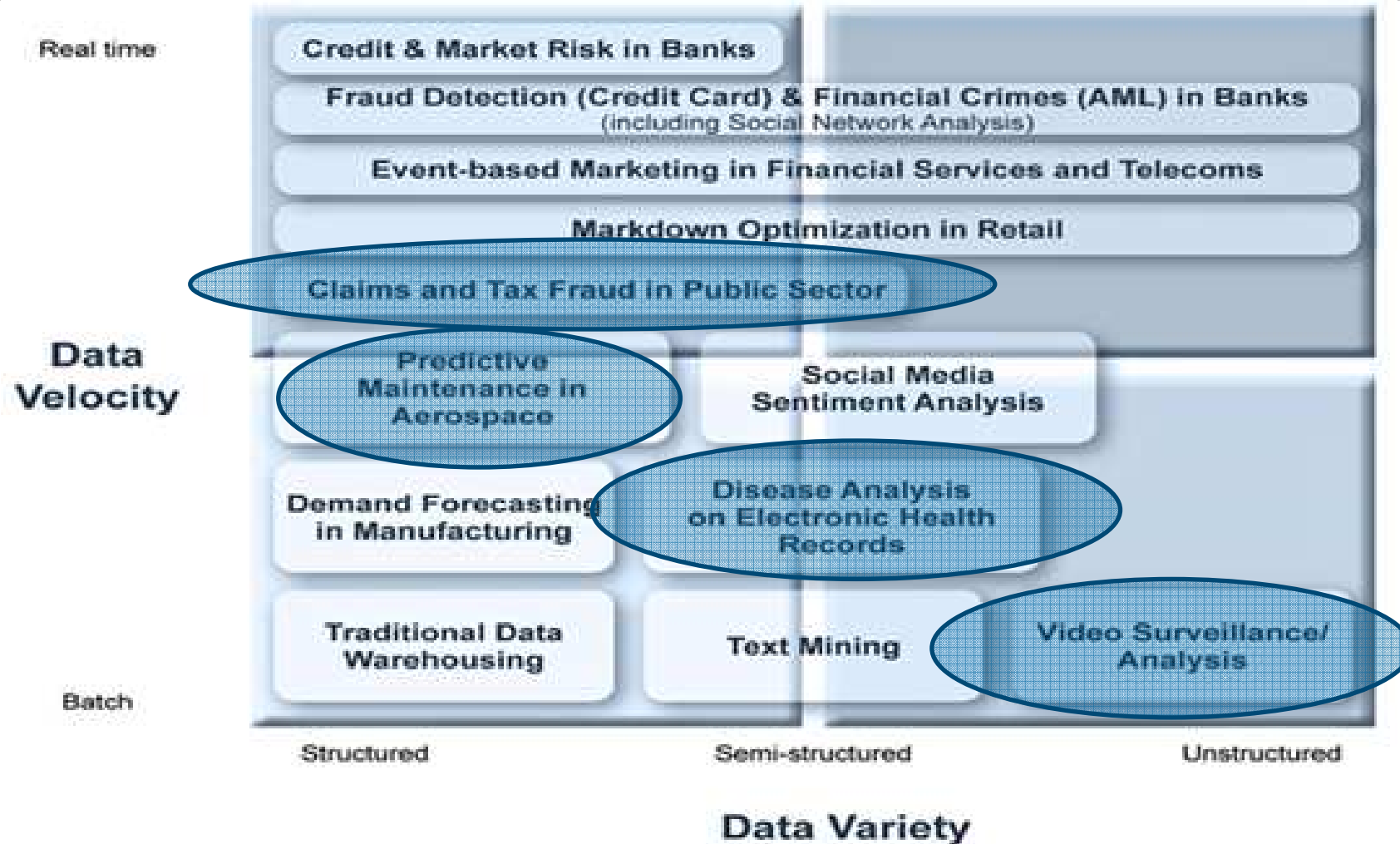
Milan Paštrnák

12-11-2013

Obsah prezentácie

- ▶ Big Data – problematika vo všeobecnosti
- ▶ Big Data a verejný sektor
- ▶ Atos a Big Data
- ▶ Project BIG
- ▶ Inovácie a výskum

Big Data – otázka viacerých „V“



Verejný sektor a téma Big Data

▶ **Verejný sektor pre svoju činnosť**

- využíva stále sa zväčšujúci objem dát,
- má v rukách legislatívu, ktorá súčasne musí riešiť možnosť prístupu k týmto dátam
- musí garantovať ochranu súkromia poskytovateľov dát

▶ **Základný hodnotový reťazec v oblasti Big Data:**

- Získavanie dát
- Analýza dát
- Správa dát
- Využitie dát

▶ **Technické aspekty:**

- Požadované funkcionality
- Dostupné technológie
- Otvorené otázky pre ďalšie R&D

Prečo je pre verejný sektor táto téma dôležitá:

- ▶ prediktívne analýzy pre plánovacie účely vo verejných službách
 - ▶ OGB (OPEN GOVERNMENT DATA) – viac k tejto téme aj počas Stredoeurópskeho okrúhleho stola
 - ▶ personalizácia verejných služieb
 - ▶ odhaľovanie potenciálu správania, porušujúceho právo (zločin)
 - ▶ oblasť informatickej bezpečnosti
 - ▶ manažment predikcie nebezpečenstiev a prírodných katastrof, riadenia zvládnutia týchto udalostí
-
- ▶ *Výstupy: konsolidované informácie z každej oblasti slúžia potom na vytvorenie konsolidovanej technologickej perspektívy (vývoj nových technológií) a stanovania optimálnej technickej road-mapy.*

Big Data – verejný sektor je producentom i spracovateľom stále väčšieho objemu dát , zároveň je tiež ich garantom

Téma BIG DATA je tiež súčasťou výskumu a vývoja, financovaného z prostriedkov EÚ (napr. 7.rámcový program). Jedným z reprezentatívnych projektov, zastrešujúcich túto tému, je projekt BIG – Big Data Public Private Forum.

Ciele projektu sú:

- Vytvorenie prostredia poskytujúcu korektné informácie a technologický pohľad na vývoj v technológiách Big Data
- Zameranie sa na kľúčových „decision takers“ v EU; ako vo vedeckých komunitách tak aj v IT technologických firmách
- Propagácia prvých reálnych použití „big data“ technológii v praxi
- Identifikácia existujúcich bariér ako sú rôzne regulácie a normy spomaľujúce nasadenia big data technológii a jednotný postup pri ich riešení v EU

Big Data- techniky a technológie

Vývoj vhodných technológií pre oblasť BIG DATA vo verejnom sektore sa pozostáva z postupov podľa toho, či sa venujeme algoritmom práce s veľkými dátami alebo samotným technológiám (nástrojom), s ktorými budeme pracovať.

- ▶ V štandardných aplikáciách je potreba vývoja analýz dát a vývoja technológií pre získavanie dát a použitie vhodných nástrojov na zabezpečenie čo najpresnejšej interpretácie prostredníctvom analyzovaných dát.
- ▶ V „security“ aplikáciách sa požiadavky viac sústreďujú na technológie, umožňujúce využitie veľkých súborov dát. Požadované technológie pre spravovanie a uchovávanie dát sú približne rovnaké v oboch aplikáciách.

Roadmap v téme BIG DATA vo verejnej správe

- ▶ **Projekt BIG (a ďalšie...) majú ambíciu na základe reálneho vývoja v oblasti nárastu dát, vývoja nových technológií, poznania legislatívy, najmä v oblasti ochrany súkromia a osobných údajov pripraviť všeobecne akceptovateľnú cestovnú mapu („roadmap“) výskumu. Takýto projekt má zmysel len vtedy, ak jeho výstupom bude schopnosť spoločnosti využiť poznatky z tejto oblasti informatiky pre riadenie a plánovanie všetkých procesov, za ktoré je verejná správa zodpovedná.**
- ▶ **Využitím generických aplikácií potom bude možné dosiahnuť,** aby technológie a všeobecné aplikácie veľkých dát vo verejnom sektore, mohli byť použité v rôznych scenároch v súkromnej sfére (napr. v množstve „smart“ tém, ako smart city, smart grid...), kde verejná správa poskytne dáta súkromnej.

Roadmap podľa projektu BIG

- 1. Analýza postupov pre vyhľadávanie dát a ich spracovanie v reálnom čase:** vychádza sa z postupov získavania dát z činností v súkromnom sektore (energetika, telekomunikácie, finančníctvo, údaje z trhového prostredia....) a tieto použiť na spracovanie veľkých objemov dát a ich štrukturalizáciu aj pre využitie verejného sektora. Zdroje: Open data, historické údaje, médiá a sociálne siete, náhodné a neštruktúrované dáta.
- 2. Analýza dát:** prístup k heterogénnym zdrojom dát, štandardizácia jazykov , disponibilnosť súborov dát verejných informácií s významnými informáciami, ktoré umožnia uskutočniť tieto analýzy.

Open Government Data: základom je slobodný prístup k informáciám z verejných organizácií vo forme, ktorá napomôže neustálej akumulácii a jednoduchému prístupu k verejne prístupným dátam (nielen poskytovanie „surových dát“).

Road map podľa projektu BIG- sociálne siete

- ▶ Analýza nehomogénnych dát z médií a sociálnych sietí: predmetom výskumu je adaptabilita sociálnych sietí, zlepšenie škálovateľnosti a kvalifikácia oblasti použitia
- ▶ Hlavné výzvy: adaptovať modely meniacich sa dát zo sociálnych sietí a iných neštruktúrovaných obsahov a správne ich aplikovať na konkrétne oblasti
- ▶ Aplikácie: Plánovanie verejných služieb a ponúknuť tvorcom politik zaviesť nové služby

Big Data pre potreby verejnej správy

Veľké úsilie v oblasti Big Data sa venuje predovšetkým identifikácii, aké sú súčasné najväčšie výzvy pre túto významnú oblasť informatiky. Prínosom bude, ak samotná verejná správa bude schopná nadefinovať svoje požiadavky pre oblasť využitia súborov veľkých dát.

Výzvy a perspektívy pre tému „veľkých dát“:

- Ako sa môžeme vyrovnáť s pribúdajúcim objemom veľkých dát ?
- Je možné spoločne inteligentne spojiť mnohostranné zdroje dát ?
- Môžu naše IT systémy tejto potrebe vyhovieť a tieto dáta teraz pravdivo realizovať?

+ Čo prináša téma Big Data nové a kde leží jej pridaná hodnota ?
Príprava nových výkonných technológií, ktoré sme predtým nemali.

Road map podľa projektu BIG

- ▶ Spracovanie senzorových dát a obrazov v reálnom čase
(surveillance and security applications like physical detection in large open spaces as infrastructures...).

Táto technológia nie je výhradne určená pre verejný sektor ale tento sektor je zodpovedný za zabezpečenie základných služieb prostredníctvom kritickej infraštruktúry, rovnako aj bezpečnosti pre občanov.

Road map: Prekonať kultúru „dátového sila“

- ▶ Prediktívne analýzy pre plánovacie účely na základe spracovania súborov veľkých dát
- ▶ Na ich základe je možné definovať pevne stanovené požiadavky na zdroje a robiť rozhodnutia, založené na podkladoch zo skutočných údajov (a verifikovateľnej interpretácie) ohľadne verejných výdavkov, predpovedí zdrojov, napr. investícií potrebných do vzdelania, sociálne opatrenia (dôchodky), škôl, verejnej dopravy, zabezpečenia potravín...
- ▶ Potrebný prístup k národným a medzinárodným štatistickým dátam (napr. Svetová banka, OSN, OECD, SŠÚ, EUROSTAT..) z nasledovných oblastí, medzi inými : demografia, census, mobilita, vzdelávanie, zdravotníctvo, bývanie, podnikanie & trhy (objemy, ceny), zamestnanosť....

Dôležitý je prístup k dátovým súborom naprieč rôznymi organizáciami verejnej správy, aby bolo možné uskutočniť vhodnú agregáciu dát pre daný subjekt, hoc tieto zdroje, získané z verejnej správy môžu byť vedené pod rôznou identifikáciou

Inovacie a smerovanie- R&D

- ▶ **KVALITA VSTUPNÝCH DÁT** – kvalita zdrojov – získavanie veľkých objemov dát s prihliadnutím na kvalitu dát
- ▶ **ARCHITEKTÚRA VEĽKÝCH DÁT** – základné problémy výskumu, súvisiace s prispôsobivosťou, výkonnosťou a odozvou na analýzy veľkých objemov dát, ako napr. architektúra BIG DATA alebo vývoj nových a optimalizovaných dátových štruktúr a algoritmov
- ▶ **VIZUALIZÁCIA DÁT** - proces získavania a použitia poznatkov o veľkých objemov dát sa rýchlo vyvíja, paralelne sú potrebné nové vizualizácie a vzory dopytovania
- ▶ **ANALÝZA VIDEO a AUDIO DÁT** – problematika analýzy video a audio dát; napr. na hraničných kontrolách.

Záver a smerovanie

V každom z komponentov veľkého dátového hodnotového reťazca existuje veľa rôznych funkčných aplikácií s rôznymi technológiami (existujúcimi alebo vo vývoji). Otvorenou otázkou pre výskum je, čo je potrebné nastaviť, aby budúce využitie týchto technológií (súčasných alebo vo vývoji), bolo v súlade s výstupmi, ktoré od nich očakáva verejný sektor.

Oblasť verejnej správy zas musí prijať také opatrenia, ktoré v tak širokom segmentovanom sektore (EU, národný, regionálny) s rôznymi cieľmi i kultúrou zabezpečia podmienky na obojstrannú spoluprácu s týmto kľúčovým výskumom v oblasti informatiky.

Konsolidovať tiež prítomnosť tých istých dát na rôznych miestach verejnej správy (napr. konsolidácia a up-date dát registrovaných na rôznych miestach VS).

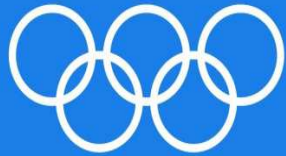
Big Data examples in the Olympics 2012

- Supporting 36 Olympic competitions and 21 Paralympic competitions
- 30% more competition data than ever before for media worldwide.
- Combining all results from all 36 Competition Venues in a single Data Feed.
 - 1.8M incoming messages, 53M distributed to customers worldwide.
- Providing results, schedule, weather, transport , biographies and other essential info from all 15,000 athletes to the Olympic Family via the Atos Info+ and myInfo+ systems.
- More than 20,000 pictures and biographies for Athletes and Officials.
- More than 200,000 pictures and records for all Accredited persons.



Atos

Worldwide IT Partner



2020

A vision of social networks **now** and **then**

London 2012 will be the favorite, most liked and trended sporting event in history. But this is just the start. By the time of the 2020 Games, social networks will be integral to every aspect of our lives and an enormous generator of sporting data

2012

4 billion

estimated global audience for London 2012

57%

of the world's population

2020

5 billion

expected to rise to at least 5bn by 2020, as two thirds of the world's population gain access to mobile broadband

66%

of the world's population

2012

1,300,000 likes

of London 2012 on Facebook

2020

5.2 million likes

estimated for the 2020 Games

2012

20,000 tweets/second

estimated peak during London 2012

2020

660,000 messages/second

via single social interface

2012

1.3 billion

total active users of Facebook, YouTube and Twitter

2020

4.3 billion

estimated total subscribers to a social interface by 2020

2012

700,000 followers

of London 2012 on Twitter

2020

2.8 million followers

estimate for the 2020 Games



2012



16 million

Most followed Olympic athlete is LeBron James (Basketball)

2020

64 MILLION

Who will be the most followed athlete?

2012

1 billion

people expected to share London 2012 content via social networks

2020

3000%

forecast increase in data traffic by 2020

2012



25%

proportion of viewers who subscribe to a mobile broadband network

2020



100% estimated by 2020

hub.olympic.org

2012

1,000

athletes joining the Olympic Athletes Hub at hub.olympic.org

2020

10,500

all athletes will be on the hub by 2020

2012

Official London 2012 pages exist on

Facebook Twitter YouTube & Flickr

2020

There will be

One Games channel

that you will access through your social network

Growth in data traffic from 2012 to 2020

2012

30x

2020

Kontakt:

milan.pastrnak@atos.net

roman.behul@atos.net

Atos, the Atos logo, Atos Consulting, Atos Worldline, Atos Sphere, Atos Cloud and Atos Worldgrid are registered trademarks of Atos SE.
May 2013

© 2013 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.

12-11-2013

