

# Privacy Enabled Big Data Analytics in the Cloud

Shelly Garion  
IBM Research - Haifa

Humboldt-Kolloquium  
„Brücken in die Zukunft:  
Deutsch-Israelische Wissenschaftsbeziehungen“  
21.–23. September 2016  
Tel Aviv, Israel

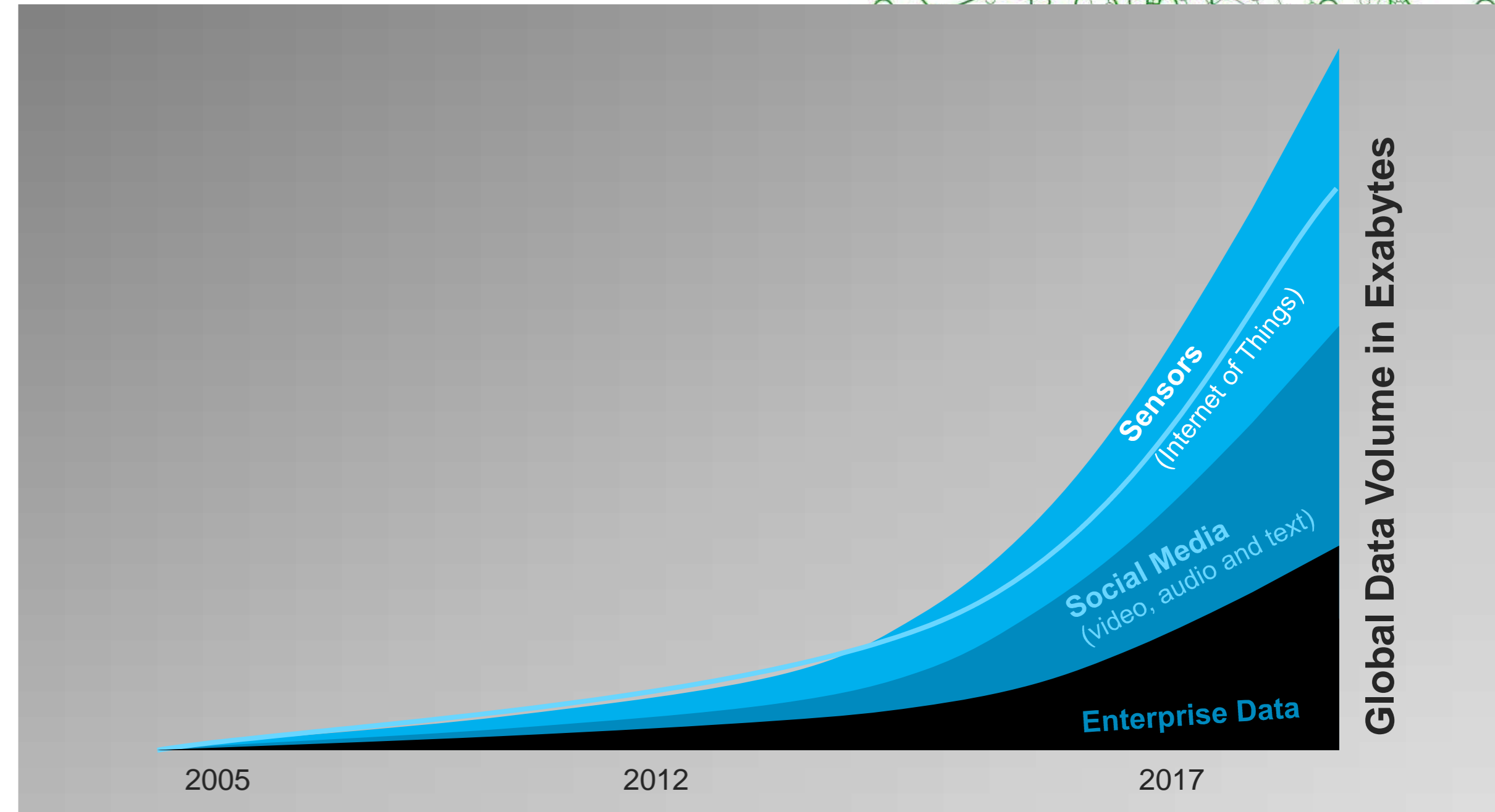
Humboldt Colloquium  
“Bridges to the Future:  
German-Israeli Scientific Relations”  
21–23 September 2016  
Tel Aviv, Israel



## Big Data Challenges

With the volumes of data in our world growing, we face new challenges:

- Where to store it?
- How to analyze it?
- What about privacy?



## Where to store Big Data? Cloud Object Store

**Object**

Data  
(unstructured / semi-structured)

Metadata



- An object encapsulates data and metadata
- High capacity, low cost
- Written once and not modified
- Accessed through RESTful HTTP
- Ideal for mobile, natural for cloud
- Unlike block storage, directly accessed by developers

## How to analyze Big Data? Apache Spark



Fast and general open-source engine for large-scale data processing

- Run programs up to 100x faster than Hadoop MapReduce in memory, or 10x faster on disk
- Write applications quickly in Scala, Python, Java, or R
- Combines SQL, streaming, and complex analytics (machine learning, graph processing)
- Runs on Apache Mesos, Hadoop YARN cluster manager, standalone, or in the cloud. It can access diverse data sources (HDFS, S3, Openstack Swift Object Store)
- Built by a wide set of developers from over 200 companies. Since 2009, more than 1000 developers have contributed to it

<https://spark.apache.org/>

## What about Privacy?



Consent Manager Service

From European Union Data Protection Directive

- **Personal data** – “any information relating to an identified or identifiable **natural person**”
- Data subject (person) has the right to be informed when personal data is being processed
- **Consent** – Data may be processed only when the data subject has given consent
- **Proportionality** – The data processed and the time for which it is stored should be no more than required for the stated **purpose**

**Data Security**  
“Protecting data from destructive forces and from the unwanted actions of unauthorized users.”  
Wikipedia

**Data Privacy**  
Ensuring that personal data is used and stored only as is needed to provide the approved user services.

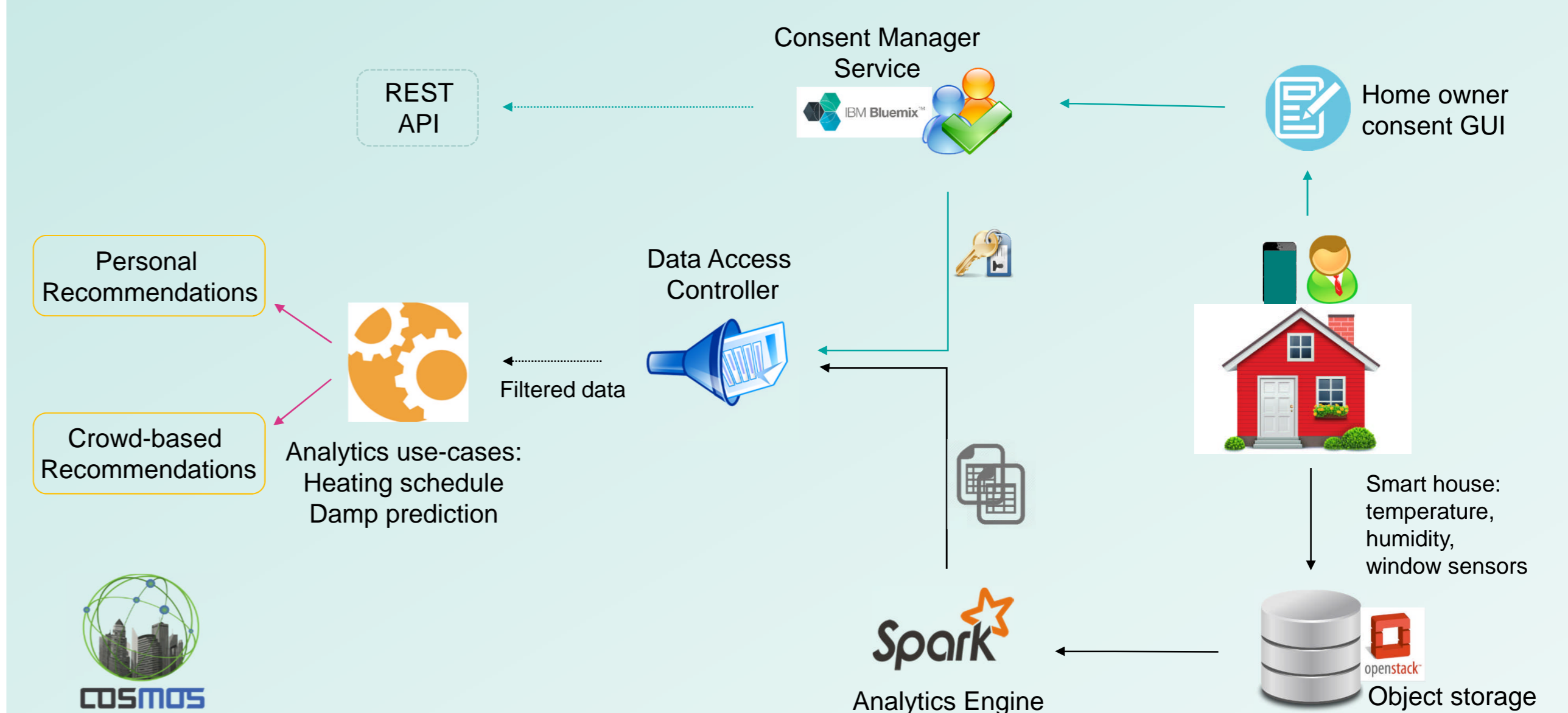
## COSMOS



Vision: Enable ‘things’ to interact with each other based on shared experience, trust, reputation, etc.

- Funding: EU FP7 for 3 years
- Started: Sept 2013
- Coordinator: ATOS
- Technical partners: IBM, NTUA, Univ Surrey, Siemens, ATOS
- Use case partners: Hildebrand/Camden, EMT Madrid Bus Transport / Madrid Council, III Taiwan – Smart Cities use cases

## Privacy Enabled Analytics Architecture for COSMOS



For more information, contact us: Shelly Garion shelly@il.ibm.com