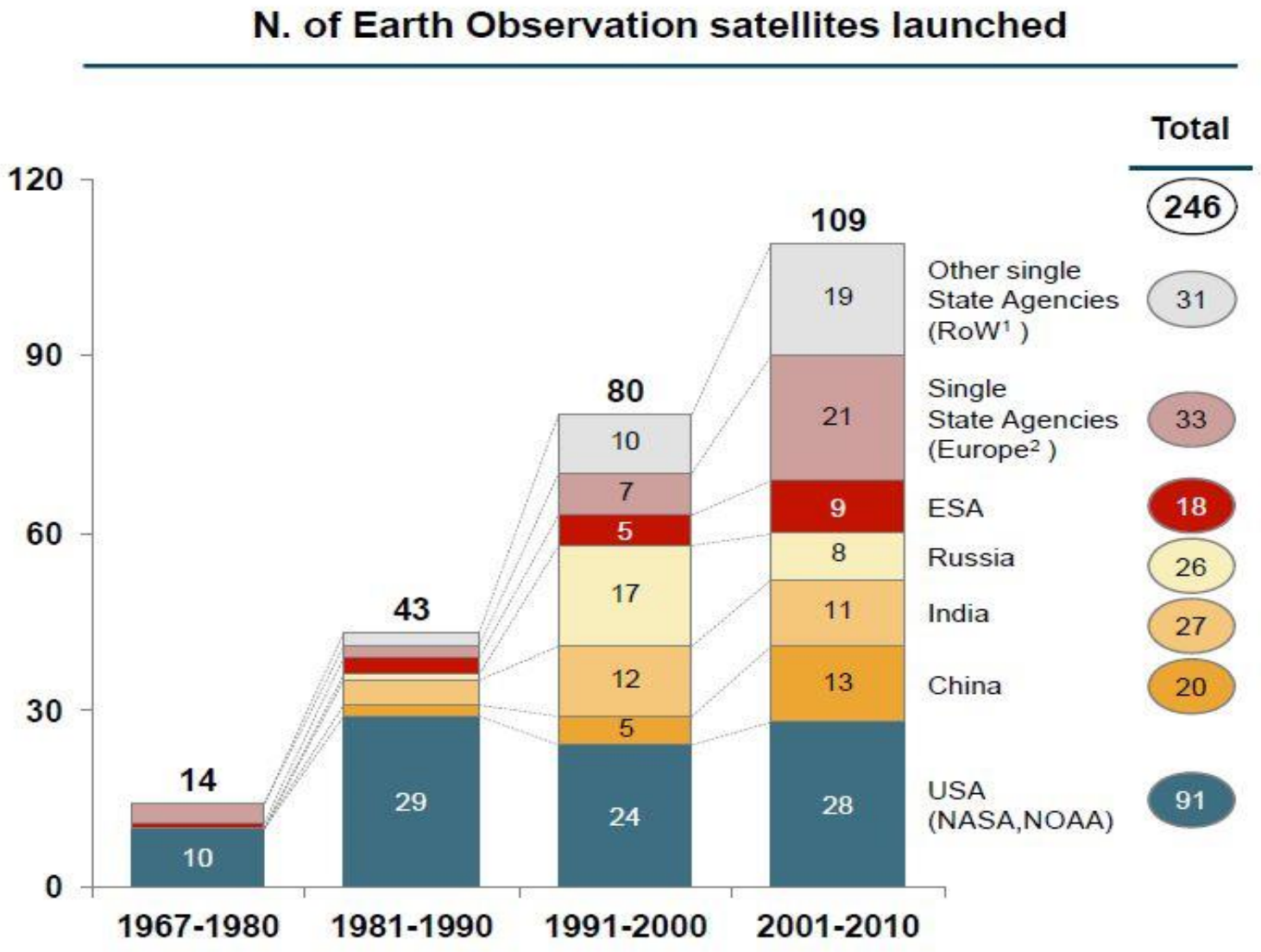
Building Virtual Earth Observatories Using Scientific Database and Semantic Web Technologies

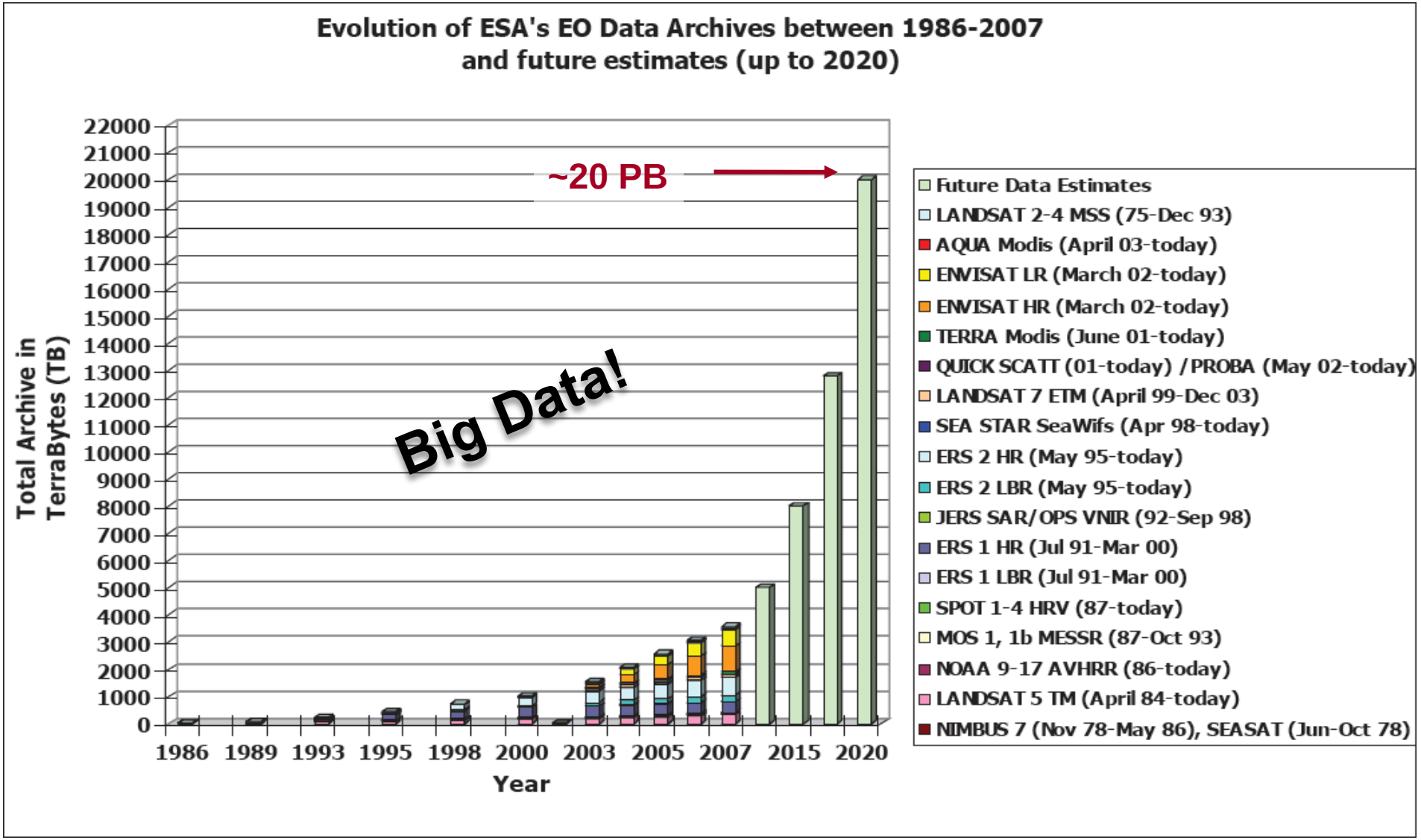
Kostis Kyzirakos
kostis@cwi.nl

Database Architectures group

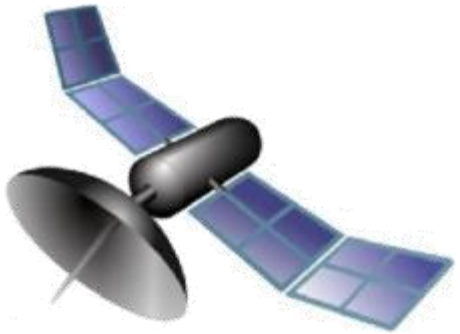
Motivation


N. of Earth Observation satellites launched
Total
120
90
60
30
0
14
43
80
109
246
10
29
24
28
10
7
5
17
12
5
19
21
9
8
11
13
Other single State Agencies (RoW¹)
31
Single State Agencies (Europe²)
33
ESA
18
Russia
26
India
27
China
20
USA (NASA,NOAA)
91
1967-1980
1981-1990
1991-2000
2001-2010

Motivation (cont'd)

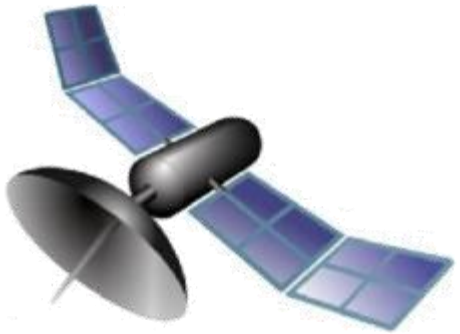


About Copernicus and Data: Geese and golden eggs



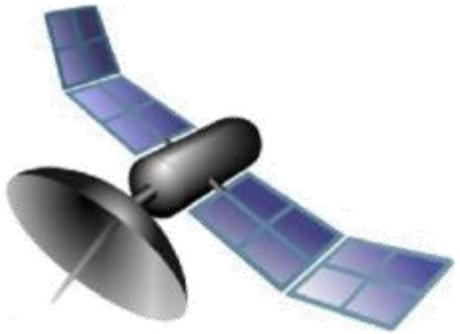
About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

About Copernicus and Data: Geese and golden eggs



About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

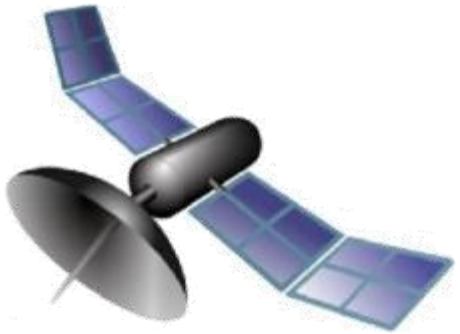
About Copernicus and Data: Geese and golden eggs



Country	Public sector body	PSI domain	Budget (M EUR)	PSI sales revenues (M EUR)	Cost-recovery ratio
Italy	Infocamere	Business register	**93.6**	**31**	31.31%
Netherlands	KvK	Business register	**243**	**6**	19.50%
United Kingdom	Companies House	Business register	**74.8**	**15.5**	20.73%
Austria	BEV	Geographic information	**85.0**	**22.5**	26.5%
Germany	BKG	Geographic information	**33.8**	**0.08**	0.24%
Germany	SenStadt	Geographic information	**9.1**	**0.945**	10.38%
Denmark	DECA	Geographic information	**31.6**	**0.26**	0.82%
Spain	IGN-CENIG	Geographic information	**52.0**	**2.1**	4.12%
Spain	Spanish Cadastre	Geographic information	**108.0**	**0**	0.00%
France	French cadastre	Geographic information	**162.5**	**0.9**	0.55%
Italy	Italian cadastre	Geographic information	**666.0**	**3.3**	0.50%
Netherlands	Dutch cadastre	Geographic information	**261.0**	**17.15**	6.57%
United Kingdom	Ordnance Survey	Geographic information	**127.0**	**21**	16.54%
Germany	DWD	Meteorological information	**214.9**	**2**	0.93%
Netherlands	KNMI	Meteorological information	**56.0**	**0.25**	0.45%
Norway	Met.no	Meteorological information	**58.0**	**0**	0.00%
Slovenia	ARSO	Meteorological information	**6.0**	**0.36**	6.00%
Spain	CENDOJ	Legal information	**9.0**	**1.5**	16.67%
France	DILA	Legal information	**135.0**	**0.9**	0.67%
France	SIRCOM	Fuel prices information	**1.1**	**0.179**	15.91%
Germany	DeStatis	Statistical information	**177.7**	**0.2**	0.11%

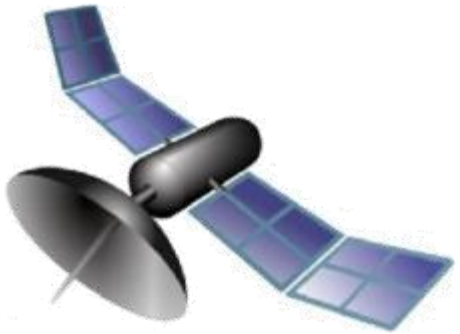
About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

About Copernicus and Data: Geese and golden eggs



About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

About Copernicus and Data: Geese and golden eggs

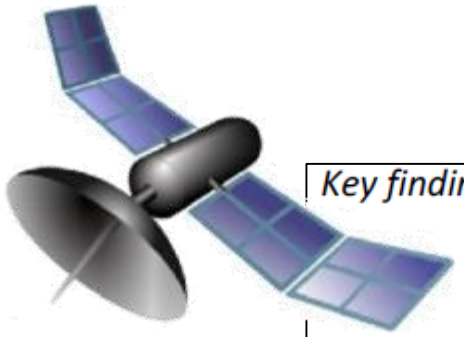


Dutch meteorological data case

	Dutch Meteorological Data
Introduction	This case was spotlighted in the *'Pricing of Public Sector Information Study'* (POPSIS study) prepared by **Deloitte and others** for the European Commission. It assessed models of supply and charging for PSI and implications of price changes.
Context	– In 1999 the Royal Netherlands Meteorological Institute (*het Koninklijk Meteorlogisch Instituut* (KNMI)) decided to cease all its commercial activities and encourage PSI re-use. Accordingly: o it approaches PSI provision for re-use as a public task; o it switched from a full cost-recovery pricing model to recovery of re-use facilitation costs only; consequently, the re-use system is self-financing; o the price of the full KNMI dataset has decreased by 80%; o its commercial arm was reorganised to facilitate privatisation; this was financed by public funding; o it does not enrich PSI beyond what is necessary to allow re-use by public and private re-users alike or produce 'value-added' products; and o data provision for academic purposes is on a marginal cost basis, provided that the results of the research are made publically available.
Key findings	– Easier access to PSI has led to the creation of a competitive and innovative private weather market. – The 1999 policy changes likely aided: o the 400% increase in turnover for private sector re-users; o the 250% increase in high-end users; o a rise in the employment activity of re-users of 300%; o an increase of over €35m on corporate tax returns; o the rise of new business models, offering free services to the public paid through advertising and innovative applications; and o improvements in data quality, professionalism in service delivery, and internal process efficiency gains of €3.5m through intensified use and feedback from re-users.
PSI domain	Meteorological data
Years of measurements	1999–2010
Year of publication	2011
Source	http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/11_2012/summary.pdf

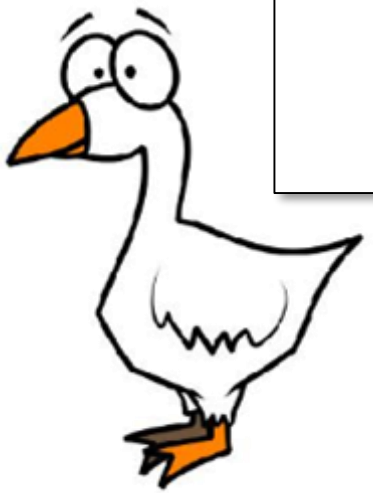
About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

About Copernicus and Data: Geese and golden eggs



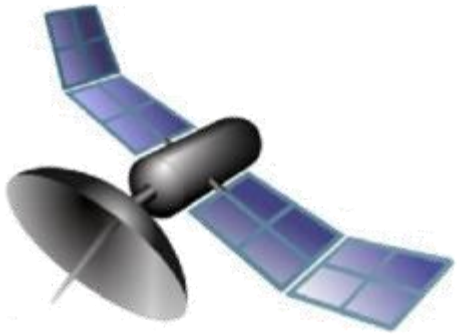
Dutch meteorological data case

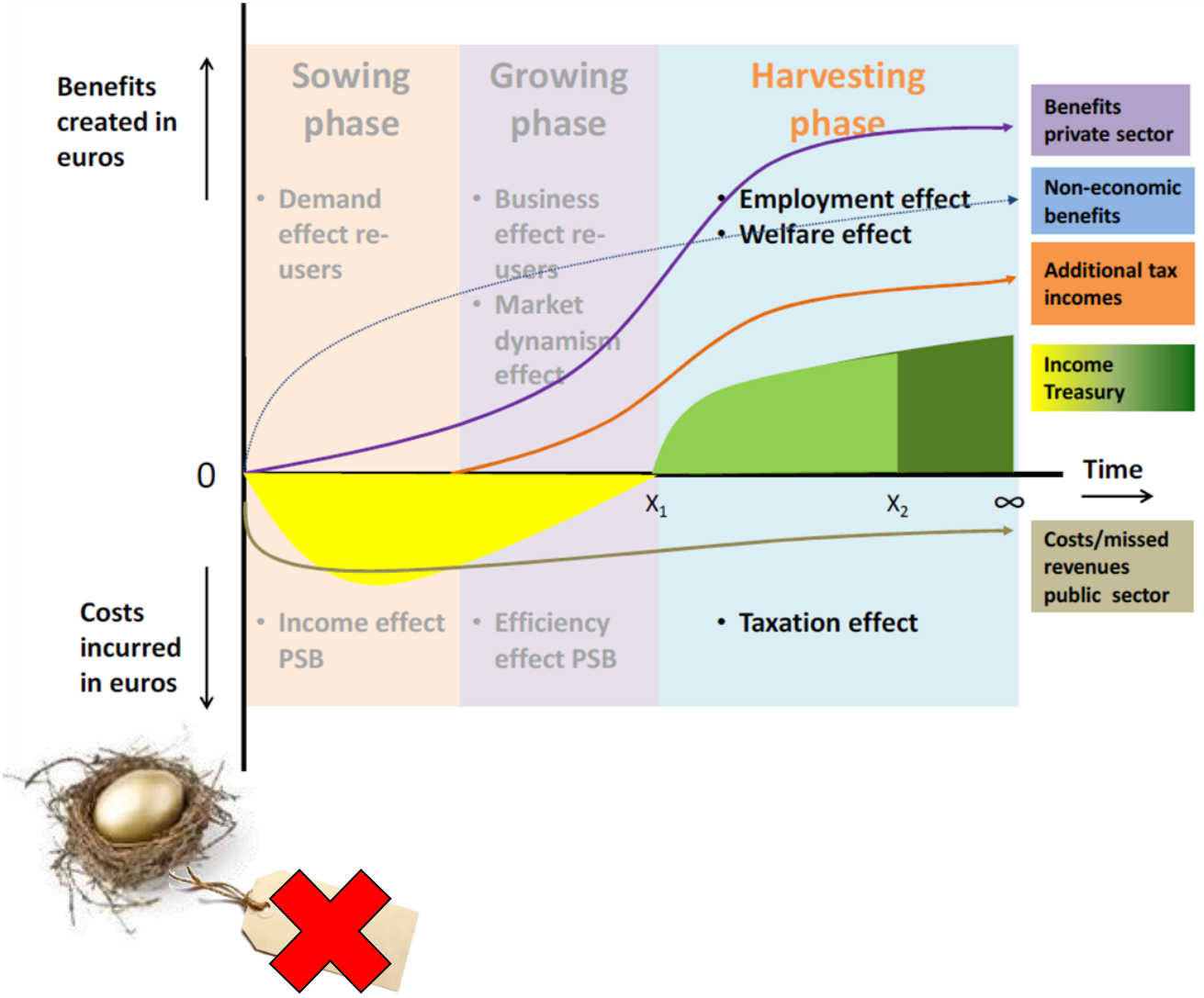
	Dutch Meteorological Data
Introduction	This case was spotlighted in the *'Pricing of Public Sector Information Study'* (POPSIS study) prepared by **Deloitte and others** for the European Commission.
Key findings	– Easier access to PSI has led to the creation of a competitive and innovative private weather market. – The 1999 policy changes likely aided: o the 400% increase in turnover for private sector re-users; o the 250% increase in high-end users; o a rise in the employment activity of re-users of 300%; o an increase of over €35m on corporate tax returns; o the rise of new business models, offering free services to the public paid through advertising and innovative applications; and o improvements in data quality, professionalism in service delivery, and internal process efficiency gains of €3.5m through intensified use and feedback from re-users.
PSI domain	Meteorological data
Years of measurements	1999–2010
Year of publication	2011
Source	http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/11_2012/summary.pdf



About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

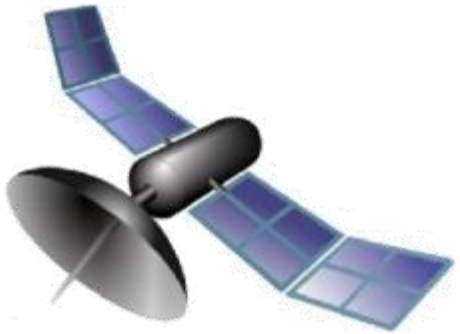
About Copernicus and Data: Geese and golden eggs





About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

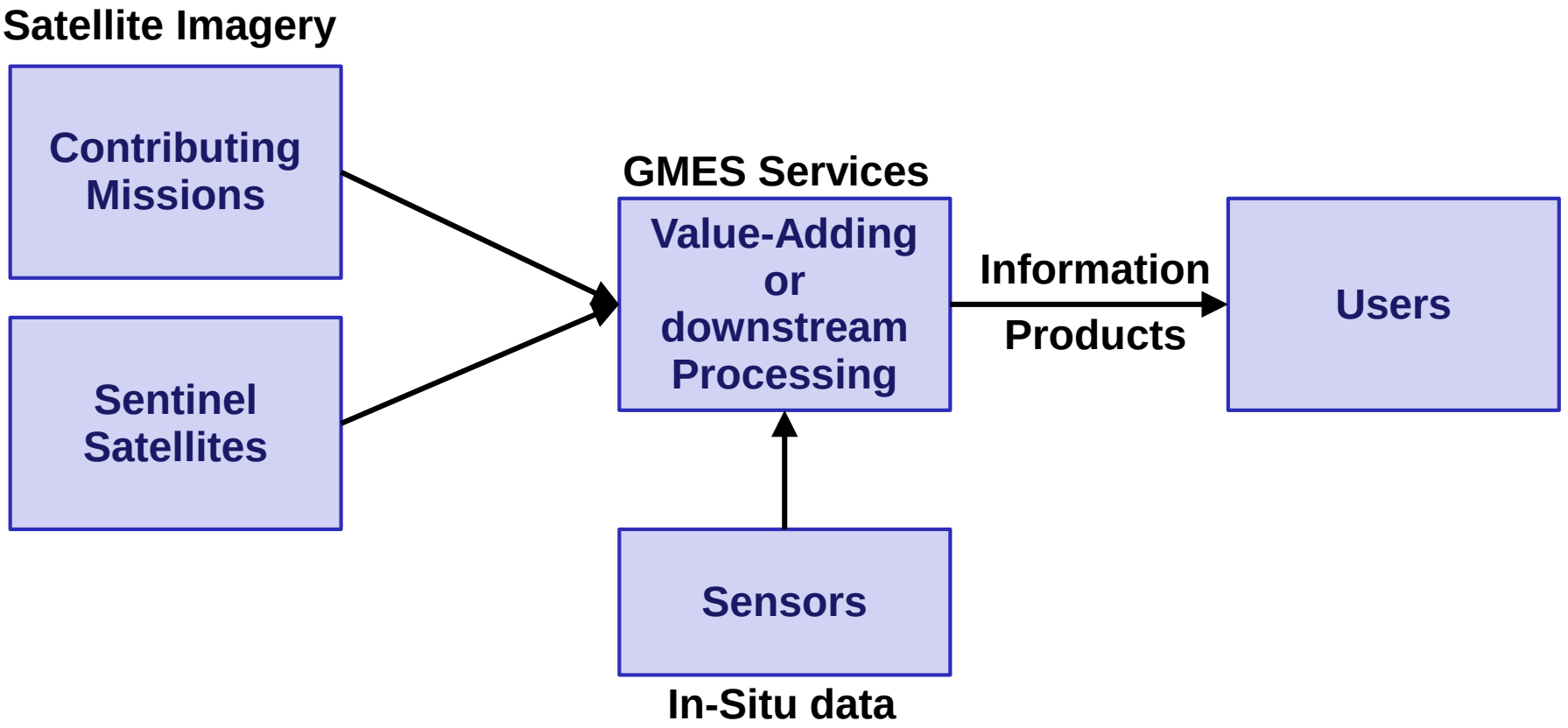
About Copernicus and Data: Geese and golden eggs



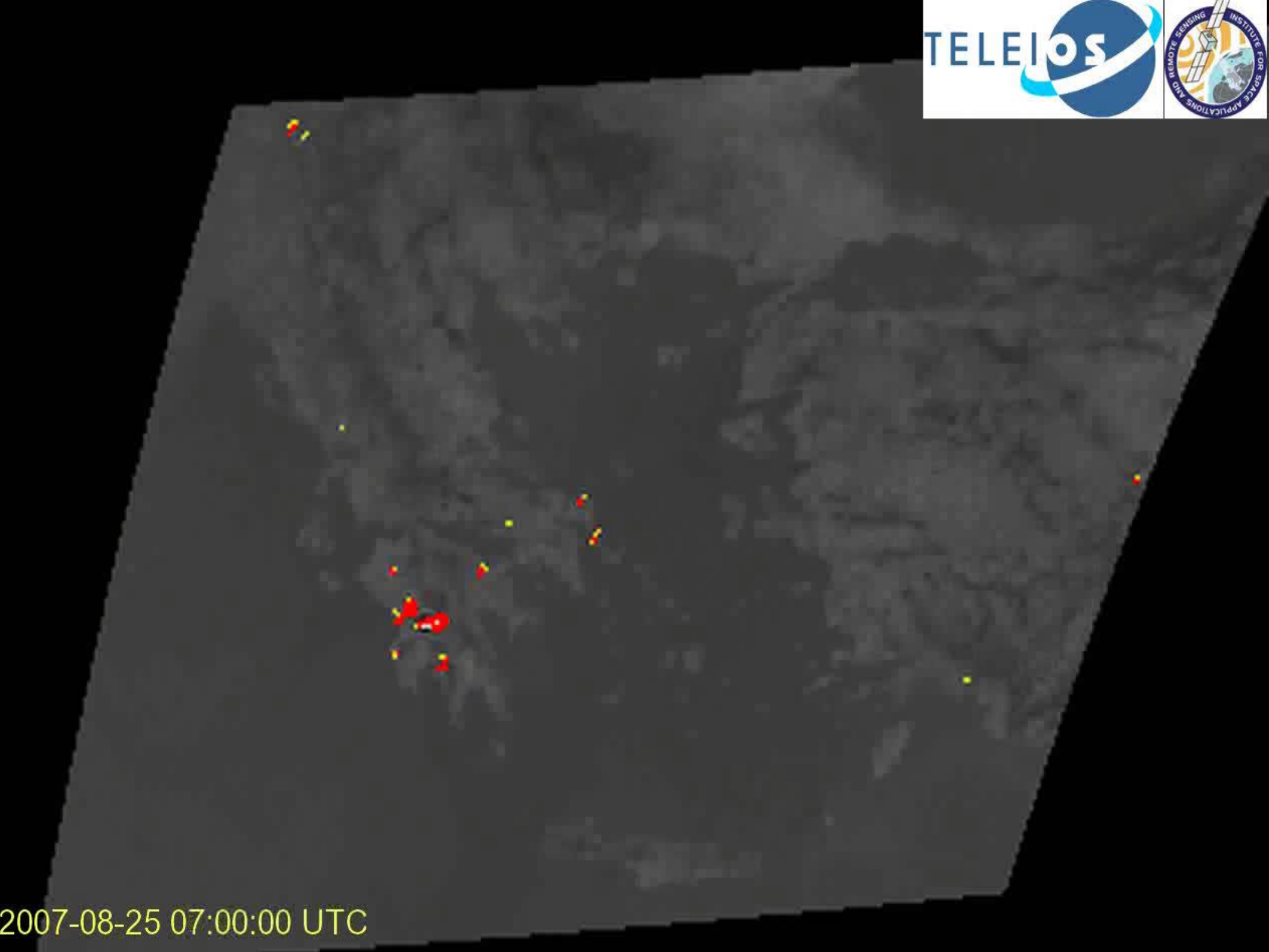
Free and Open Data Policy for the Sentinel Satellite Data

About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

GMES Value Chain

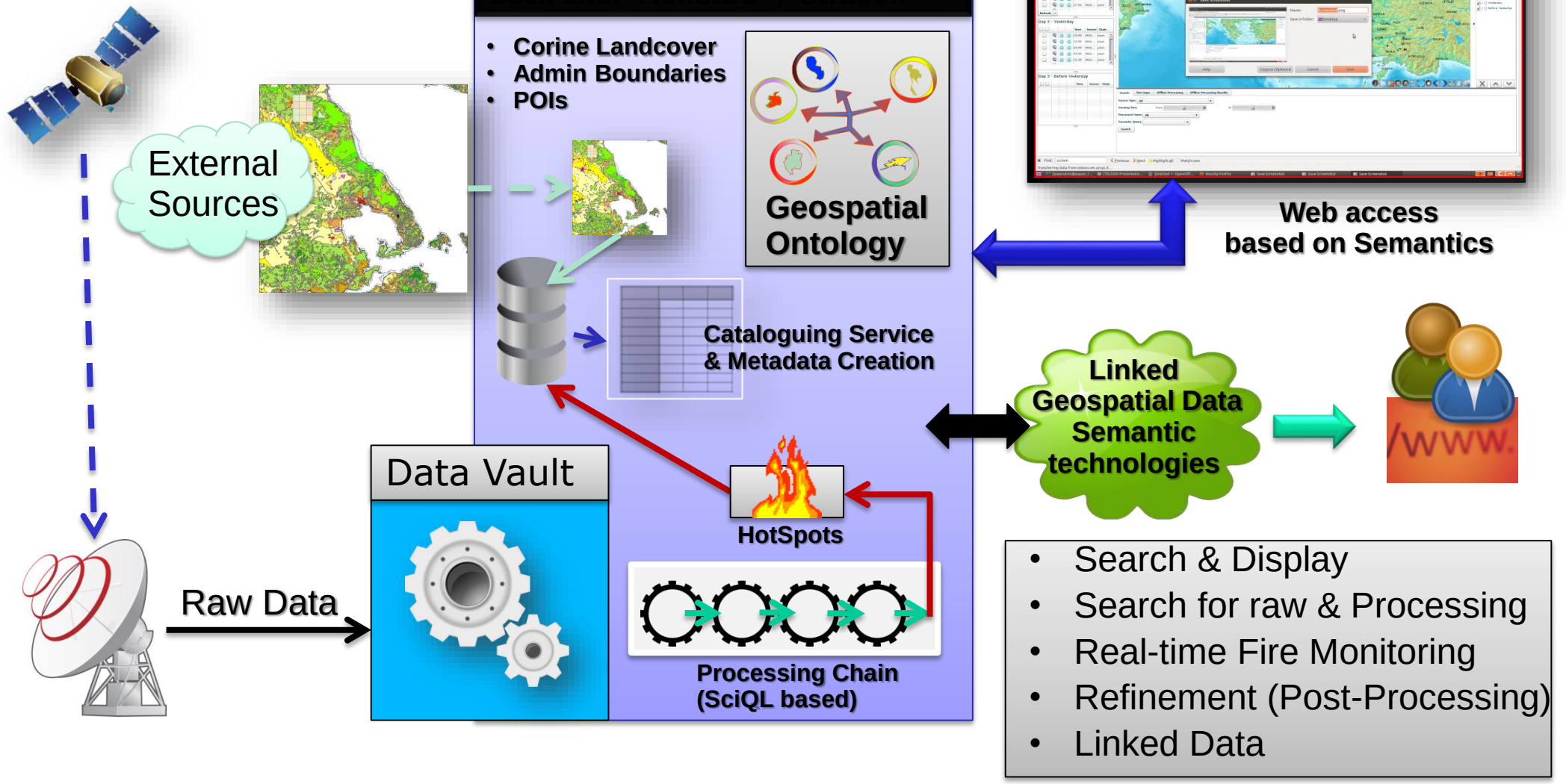


About gmes and data geese and golden eggs
http://esamultimedia.esa.int/docs/EarthObservation/Open_Data_Study_Final_Report.pdf

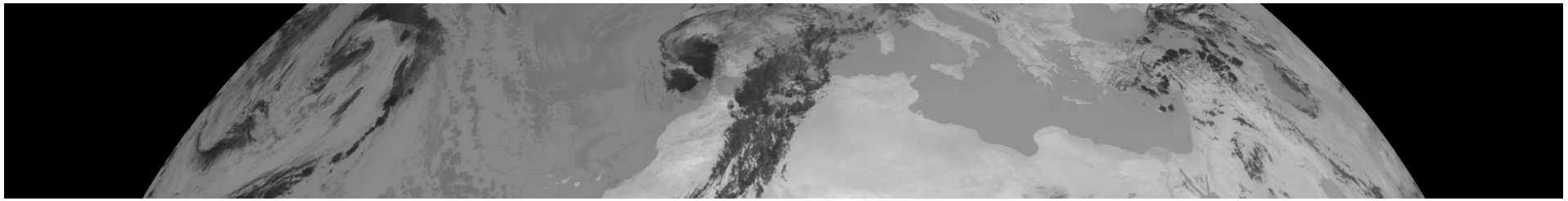

TELEIOS
INSTITUTE FOR SPACE APPLICATIONS AND REMOTE SENSING
2007-08-25 07:00:00 UTC

Fire monitoring application

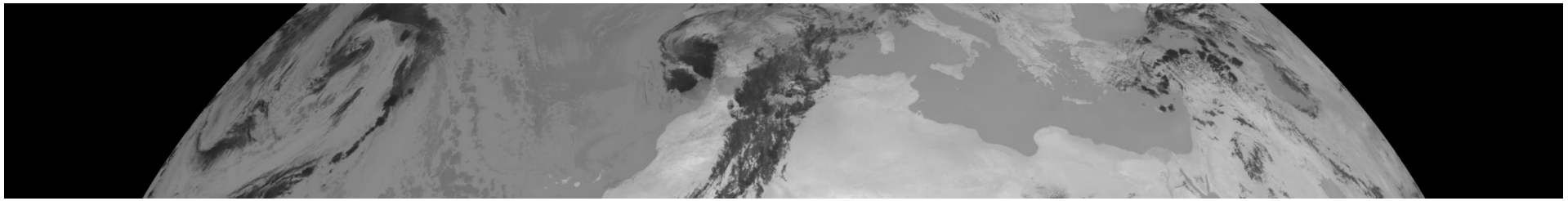
Advancements - Integration of the TELEIOS technologies

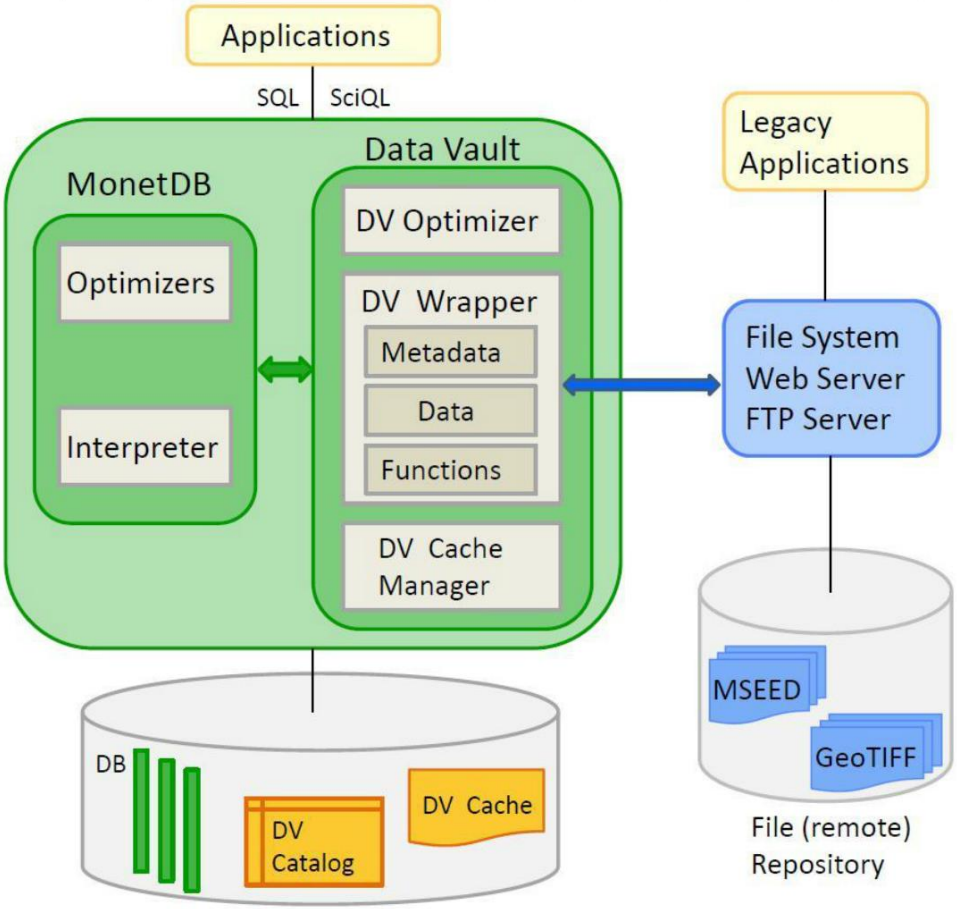


Data Acquisition



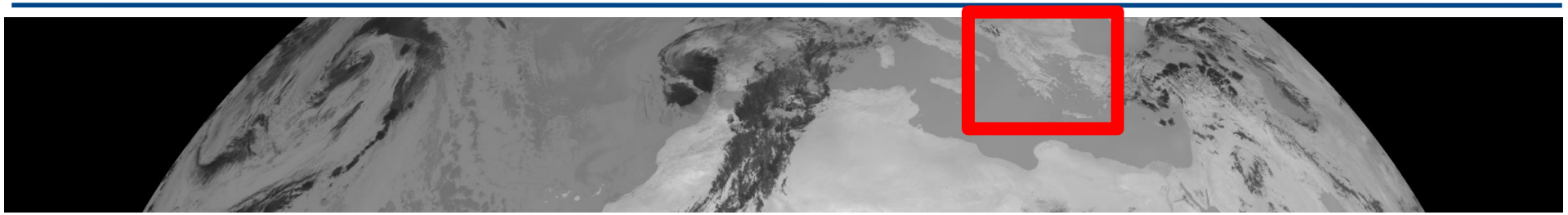
Data Ingestion: Data Vaults



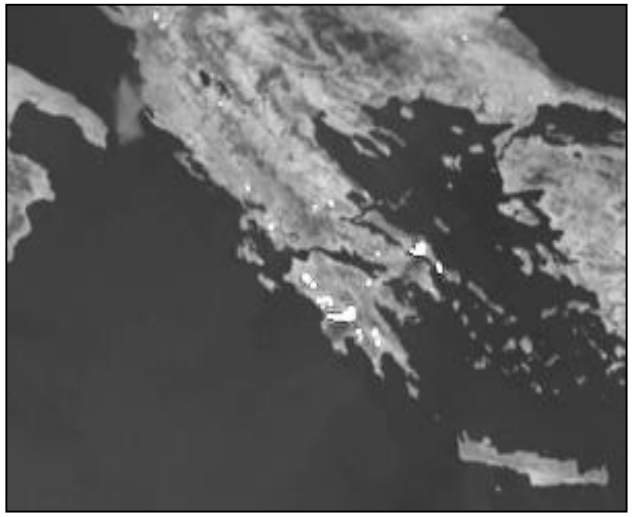


[SSDBM `12, `13]

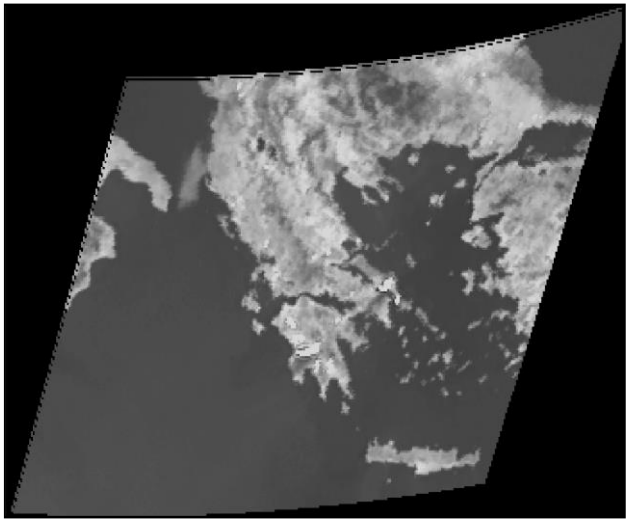
Data Processing



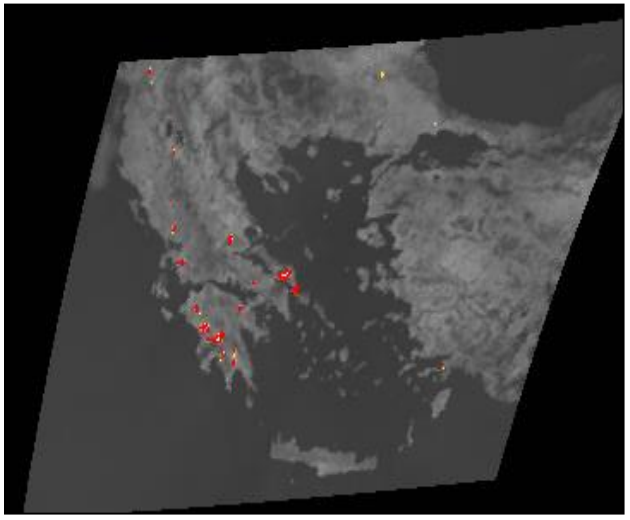
Cropping



Georeferencing



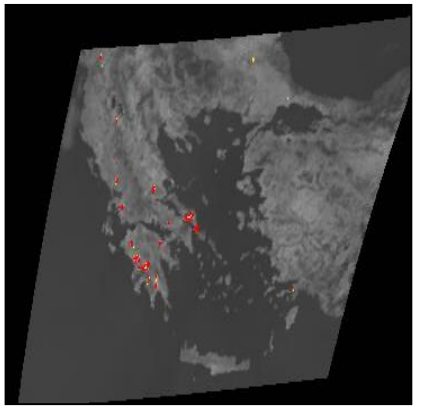
Classification



SciQL: A Query Language for Science Applications
Classification

[SIGMOD '13]

Classification

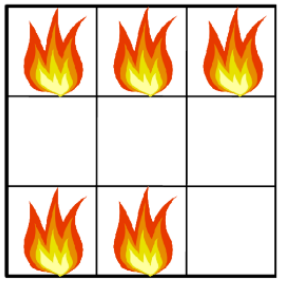


```
SET SCHEMA rs;

DECLARE size_x SMALLINT, size_y SMALLINT;
SET size_x = (SELECT MAX(x) + 1 FROM rs.image1);
SET size_y = (SELECT MAX(y) + 1 FROM rs.image1);

CREATE ARRAY fire (x SMALLINT DIMENSION[size_x], y SMALLINT DIMENSION[size_y], f INT);

INSERT INTO fire (
  SELECT b3.x, b3.y, 1
  FROM rs.image1 AS b3, rs.image2 AS b4, rs.image3 AS b7, rs.image4 AS msk
  WHERE b3.x =  b4.x AND b3.y =  b4.y -- join the images
    AND b3.x =  b7.x AND b3.y =  b7.y -- join the images
    AND b3.x = msk.x AND b3.y = msk.y -- join the images
    AND msk.intensity = 1 -- cloud- & water-mask
    AND b3.intensity > 0 AND b4.intensity > 0 AND b7.intensity > 0
    AND b4.intensity <= 60 -- indexNIR
    AND (b3.intensity + b4.intensity) / 2 <= 50 -- indexALBEDO
    AND (CAST(b4.intensity - b7.intensity AS REAL) / (b4.intensity + b7.intensity) +
1.0) * 127.5 <= 126.0 -- indexNBR
);
```



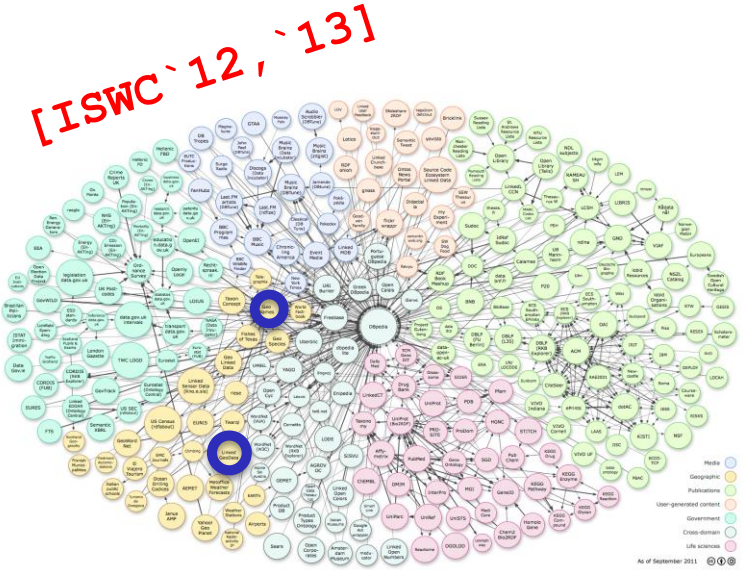
Improvements

- Improving the fire monitoring service using Semantic Web technologies
  - **Representing** fire related products using ontologies
  - **Enriching products** with linked geospatial data
  - **Improving accuracy** with respect to:
    - Underlying land cover/land use
    - Persistence in time

[ISWC 2012 Semantic Web Challenge 3'rd place winner]

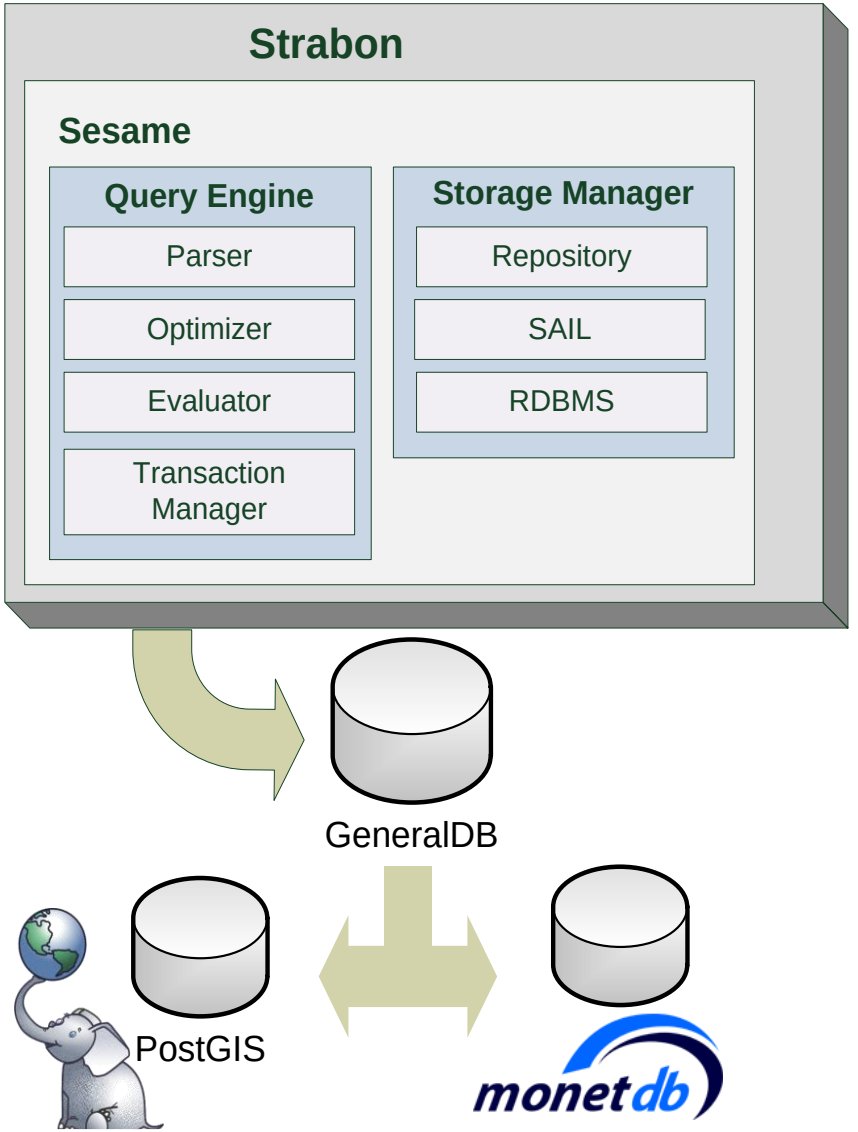
http://bit.ly/FiresInGreece

Strabon
A Scalable Geospatial RDF Store



stRDF graphs

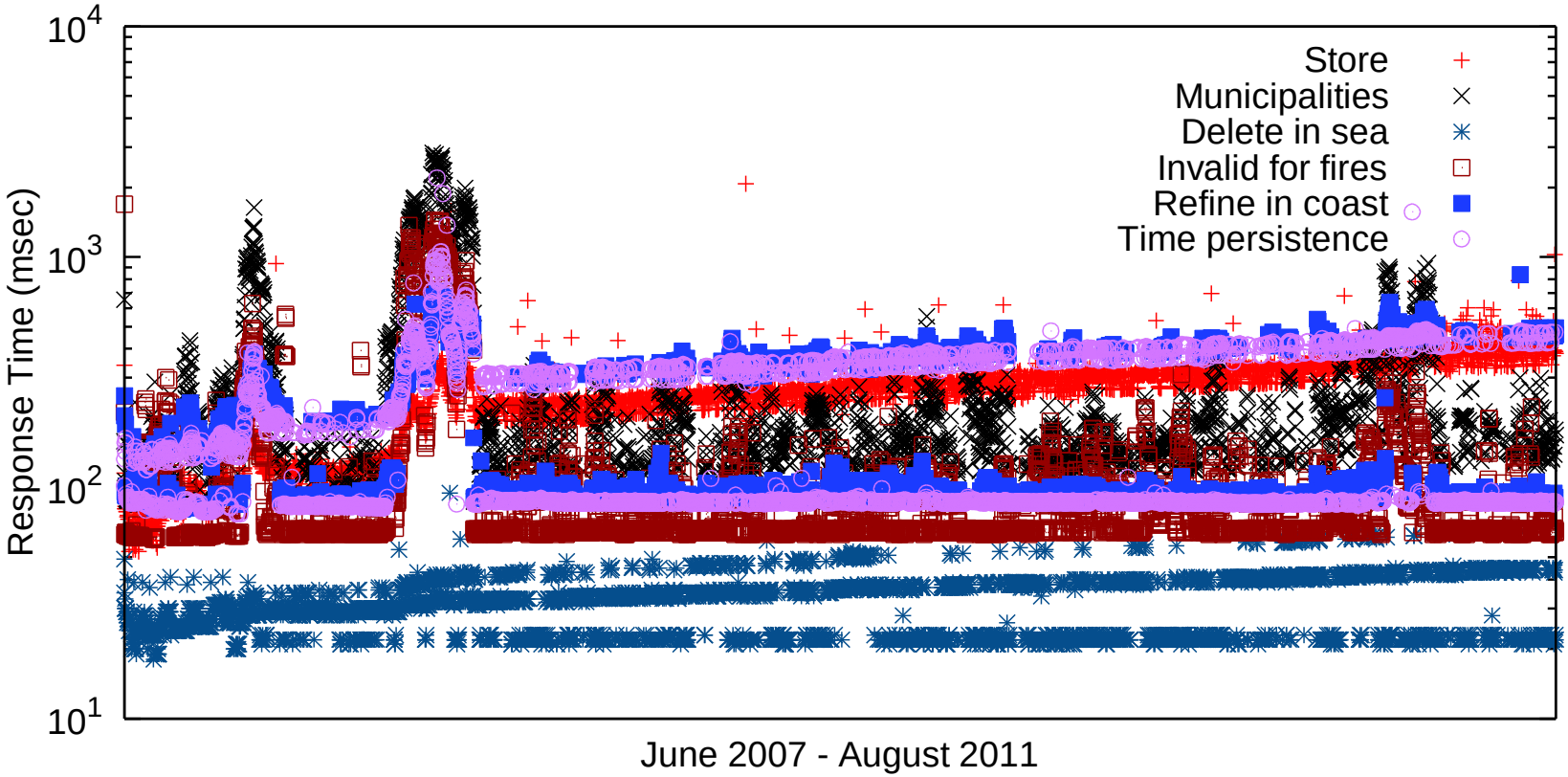
stSPARQL/ GeoSPARQL queries



Fire Monitoring Service

Preliminary evaluation – Processing times

- Product ingestion, processing and refinement is completed in less than 12 seconds
- More refinement operations to be added later given the five minutes time frame



Discussion

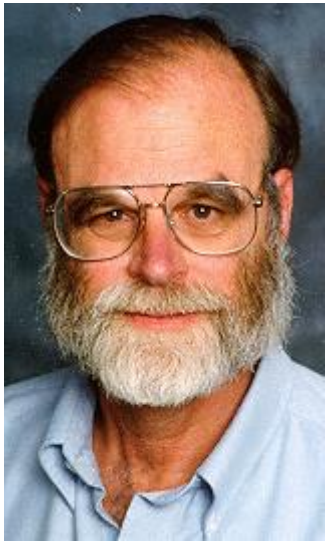
- Use **higher-level languages**, stop worrying about how to store and manage metadata, just **focus** on the actual **processing**
- Express common Earth Observation operations easily using the **SciQL** and **stSPARQL/GeoSPARQL queries** instead of using a lengthy **C** program
- **Rapid prototyping** and new refinement modules without the need to recompile everything

Following the advice of Jim Gray

"**How can you, a computer scientist, engage with a domain scientist (or group)?**

…

Well, first you have to learn a bit of their language. This generally involves mastering the introductory text for that domain – it is painful and you can skip this step if you are in a hurry, but you will end up doing this work in any case. Doing it early is the most efficient way. In parallel you have to form a working relationship with the domain experts (scientists.) You need to put in enough face time so that they are not surprised to see you. This goes hand-in hand with developing a common language. The converse of this, the domain scientists you are working with need to explore some of the things done in computer science and in other disciplines so that they have a sense of what is possible and what is almost possible."



Quote from article:
Where the Rubber Meets the Sky: Bridging the Gap between Databases and Science.
Jim Gray, Alexander S. Szalay. MSR-TR-2004-110, October 2004.
IEEE Data Engineering Bulletin, December 2004, Vol. 27.4, pp. 3-11.

Thank you for your attention!

Kostis Kyzirakos
kostis@cwi.nl