



# *Blue Paper on Data Management*

*Norbert Meyer, PSNC*

***e-IRG workshop, Amsterdam, December 4, 2012***

**e-IRG**  
e-Infrastructure  
Reflection Group

**Advancing the European e-Infrastructures**

**Integrating the Scientific and Public Data Level**

**Intensifying the Cooperation with Stakeholders**

# Co-operation

Recall: e-IRG workshop in 2011 invited ESFRI cluster projects

BioMedBridges, DASISH, ENVRI, CRISP **(pilot projects)**

also DC-NET, PaNdata, ITER

# Importance of requirements

Large datasets

Restricted access

Metadata structure

Federated AAI

Accounting

Data provenance

Integration

Interoperation

High trust and security

Reliability

Access

Advanced search functionality

Data preservation

Interpretation of data

Unified access

High quality

Simplified access

Searching information

Important stakeholders

# Topics covered by the Blue Paper on DM

## 1. Grand Challenges and ESFRI Requirements

## 2. Data e-Infrastructure

## 3. Reliability and Replications

Norbert Meyer (PSNC)

Johannes Reetz (MPG)

John Kennedy (MPG)

Maciej Brzeźniak (PSNC)

## 4. Metadata

## 5. Unified Access and Interoperability

## 6. Security

Gera Pronk (Surfnet), Daan Broeder

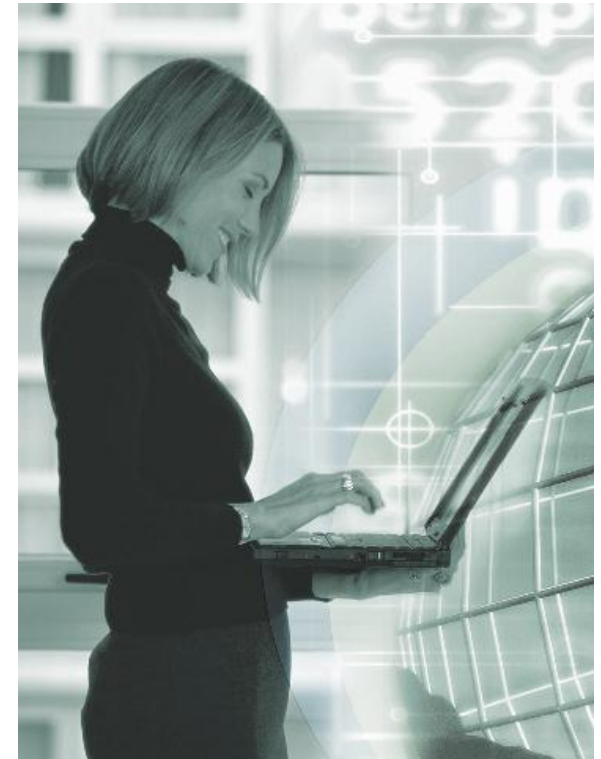
Angelos Bilas (Forth)

Steven Newhouse (EGI),

Sergio Andreozzi (EGI)

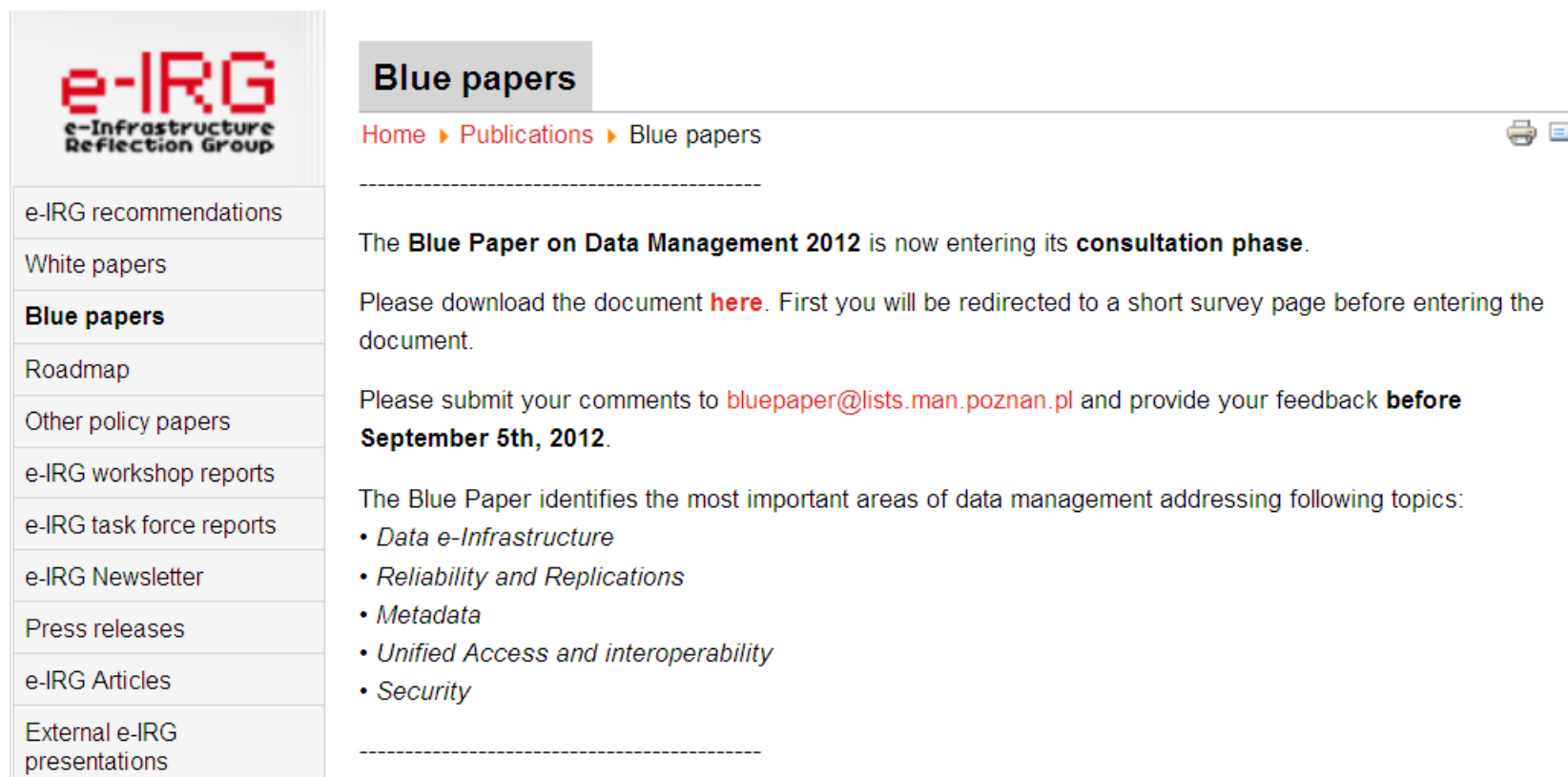
# Identified stakeholders

- End user
- Data owner
- Infrastructure provider
- Service provider
- Computer science researchers (on database management)
- Policy maker



# Public consultation process

<http://www.e-irg.eu/publications/blue-papers.html>



**e-IRG**  
e-Infrastructure  
Reflection Group

e-IRG recommendations  
White papers  
**Blue papers**  
Roadmap  
Other policy papers  
e-IRG workshop reports  
e-IRG task force reports  
e-IRG Newsletter  
Press releases  
e-IRG Articles  
External e-IRG presentations

## Blue papers

Home > Publications > Blue papers

The **Blue Paper on Data Management 2012** is now entering its **consultation phase**.

Please download the document [here](#). First you will be redirected to a short survey page before entering the document.

Please submit your comments to [bluepaper@lists.man.poznan.pl](mailto:bluepaper@lists.man.poznan.pl) and provide your feedback **before September 5th, 2012**.

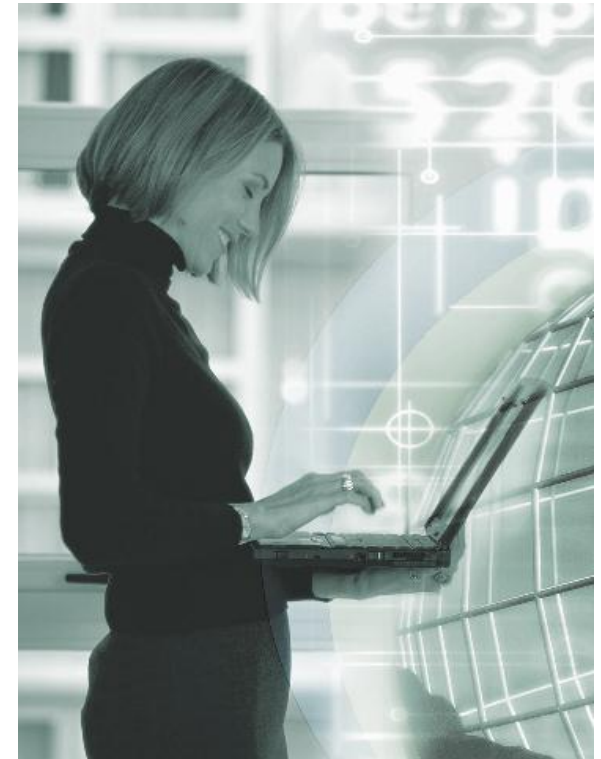
The Blue Paper identifies the most important areas of data management addressing following topics:

- *Data e-Infrastructure*
- *Reliability and Replications*
- *Metadata*
- *Unified Access and interoperability*
- *Security*

**July – September, 2012**

# Feedback

- **User Community Support Team, EGI.eu**
- **IGE project (LRZ, Germany)**
- Cloud Software Program  
[www.cloudsoftwareprogram.org](http://www.cloudsoftwareprogram.org)
- OpenAIRE
- BioMedBridges Consortium
- PaNData ODI, CRISP and EuroFEL
  
- ENVRI (ICOS, EPOS, EMSO, LifeWatch, Eiscat-3D, Euro-Argo, IAGOS, SIOS, etc. )





# RECOMMENDATIONS

# Data e-Infrastructure - recommendations

- Define **business cases and requirements** for ESFRI projects related to data access and data infrastructure
- **Define cross related requirements** generalized for all ESFRI communities
- Provide a current state of the data infrastructure within ESFRI
- Ask the service providers to provide a **sustainability policy** and to state what they can offer ESFRI projects
- Need a **redundant data infrastructure** where strong data centres form a backbone for data access and preservation
- Define a role of each ESFRI project. For the near future, we have to assume a high degree of **specialisation**, where the roles of service/application provider, infrastructure provider and data owner/producer will be separated. Only really big organisations will still be able to consolidate all of these roles..

## Communities and interoperability

- Establish **metadata service managers** and give them a greater role in supporting 'newcomers' with their specific requirements to join, and setting up cross disciplinary metadata search
- Give recommendations for distributing responsibilities wrt.
  - metadata quality
  - providing guidance for interoperability (across disciplines)
  - providing metadata services (across disciplines)

## Best practices

- Enable easy and **standardised metadata**
- Establish **federated data catalogues**
- Pay attention to aspects of granularity. The need for describing sets of resources (data-sets) at different levels of granularity calls for different metadata schemas.

- Investigate on all ESFRI project the security requirements
- Check if the community is ready to use a **federated authentication process** and the cost of the transition phase
  - **Federated authentication process:** The user should be able to handle the authentication process with a similar user experience as with the most common web applications and leverage existing institutional or national electronic identities (such as the **European Citizen Card**) to access institutional or community attribute servers to gain access to distributed data and services.
- **Implement data encryption:** data-centric, file-level encryption that is portable across all computing platforms and operating systems should be available to users as a way to increase data protection, confidentiality and integrity in transit and at rest.
- **Influence the EU Data Protection directive under revision** which should address the aspects which are important especially from the data owner and end user point of view.

- Blue Paper recommendations point to need for policy-development at EC and Member Countries levels
- We proposed new **ESFRI+e-IRG WG** to look into policy implications, such as
  - What would be required from the ESFRI Member States at the policy level to accommodate for the recommendations?
  - What changes should be implemented at the EC level?
  - What recommendations related to RIs in HORIZON 2020 can be given?
  - What aspects from the report should be taken into account in future assessments of RIs.

# Open questions from the public consultation process

1. **business model:** who should pay for what? you only recommend to look at business cases. in such a document you would need to discuss this point at least that investment is needed if data is to be kept at all.
2. **Standards:** who enforces them? what is the motivation to use them? who pays the people holding the standards and assuring the quality?
3. **common interfaces:** who defines these? the market does not do it, although some efforts are there to standardize on interfaces, data is always ignored because there is money to be made with nonstandard interfaces. who pays the people enforcing and managing the common interface?

# Open questions from the public consultation process (cont.)

4. **trust network:** who are the really trusted providers of authentication and authorization information? can it really be decentralized? or should there be some kind of passport office? this is very important for medical data, personal data, and all other data that needs to be protected.

**Good topics for further investigations (?)**

# Thank you!

Blue Paper report available on  
[www.e-irg.eu/publications](http://www.e-irg.eu/publications)



e-IRG “Blue Paper” on  
Data Management

---