# The COAR Interoperability Roadmap

Friedrich Summann
Universitätsbibliothek Bielefeld

Open Access Tage
Zürich 8 – 9 Sept. 2015

Betroffene COAR Discussion Groups

- Experts  Advisory Panel
- Roadmap Editorial Group
- COAR WG2 Interoperability

# Die Vorbereitungsschritte:

- Mail an Experten
- Feedback und Kommentare von Experten
- Interne Diskussionen Editorial Group
- Ableitung der Interoperability Themen
- Ausarbeitung Fragebogen
- Automatische Bearbeitung der Ergebnisse
- Strategische Einführung
- Das Ergebnis: **The Roadmap document**

## Questionnaire

- Issues
- Immediate relevance  (1-3 years)
- Future relevance (3-5 years)
- Complexity of implementing (low, medium, high
- Comments

| Issues | Immediate relevance (1-3 years) | Future relevance (3-5 years) | Complexity of implementing (low, medium, high | Comments |
|---|---|---|---|---|
| **Impact and Visibility** | | | | |
| **Strategic Benefit**<br><br>Supporting the Visibility of Repositories and their contents including their Relevance and Usage and Impact Metrics | x | | medium | |
| **Concrete Issues** | | | | |
| Supporting Search Engine Optimization (SEO) | x | | medium | Has to be adopted to the flexible strategies |
| *SEO methods are focused on optimizing the ranking of web sites and their contents in search engines.* | | | | |


*Promoting greater visibility and application of research through global networks of Open Access repositories*

# COAR Roadmap
# Future Directions for Repository Interoperability

Working Group 2: Repository Interoperability

February 2015

# Gliederung

**Lead Editors:**

- **Friedrich Summann,** Bielefeld University, Germany
- **Kathleen Shearer,** Confederation of Open Access Repositories (COAR), Canada

**Editors:**

- **Timo Borst,** Leibniz Information Center for Economics, Germany
- **Pablo de Castro,** EDINA National Data Centre Edinburgh, UK
- **Wolfram Horstmann,** University of Göttingen, Germany
- **Alicia López Medina,** National Distance Education University Madrid, Spain
- **Katharina Müller,** University of Göttingen, Germany
- **Maxie Putlitz**, University of Göttingen, Germany
- **Eloy Rodrigues**, University of Minho, Portugal
- **Jochen Schirrwagen,** Bielefeld University, Germany

**Experts and Reviewers:**

- **Isidro Aguillo**, CINDOC-CSIC, Spain
- **Ana Alice Baptista**, University of Minho, Portugal
- **Tom Beirender,** World Bank Group, USA
- **Daniel Beucke**, University of Göttingen, Germany
- **Sheridan Brown**, V4OA Project Consultant, UK
- **Donatella Castelli**, Italian National Research Council, Italy
- **Gernot Deinzer,** University of Regensburg, Germany
- **Patrick Hochstenbach,** Ghent University, Belgium
- **Maarten Hoogerwerf**, Data Archiving and Networked Services (DANS), The Netherlands
- **Keith G. Jeffery** Consultant, UK
- **Johannes Keizer**, Food and Agriculture Organization of the United Nations, Italy
- **Thomas Krichel**, Long Island University, USA
- **Clifford Lynch**, Coalition for Networked Information (CNI), USA
- **Devika Madalli**, Indian Statistical Institute, India
- **Salvatore Mele**, CERN, Switzerland
- **Susan Reilly**, LIBER (Association of European Research Libraries), The Netherlands
- **Frank Scholze**, Karlsruhe Institute of Technology (KIT), Germany
- **Miguel Angel Sicilia**, University of Alcalá, Spain
- **Paul Vierkant**, Humboldt University of Berlin, Germany
- **Paul Walk**, University of Bath, UK

# Ziele

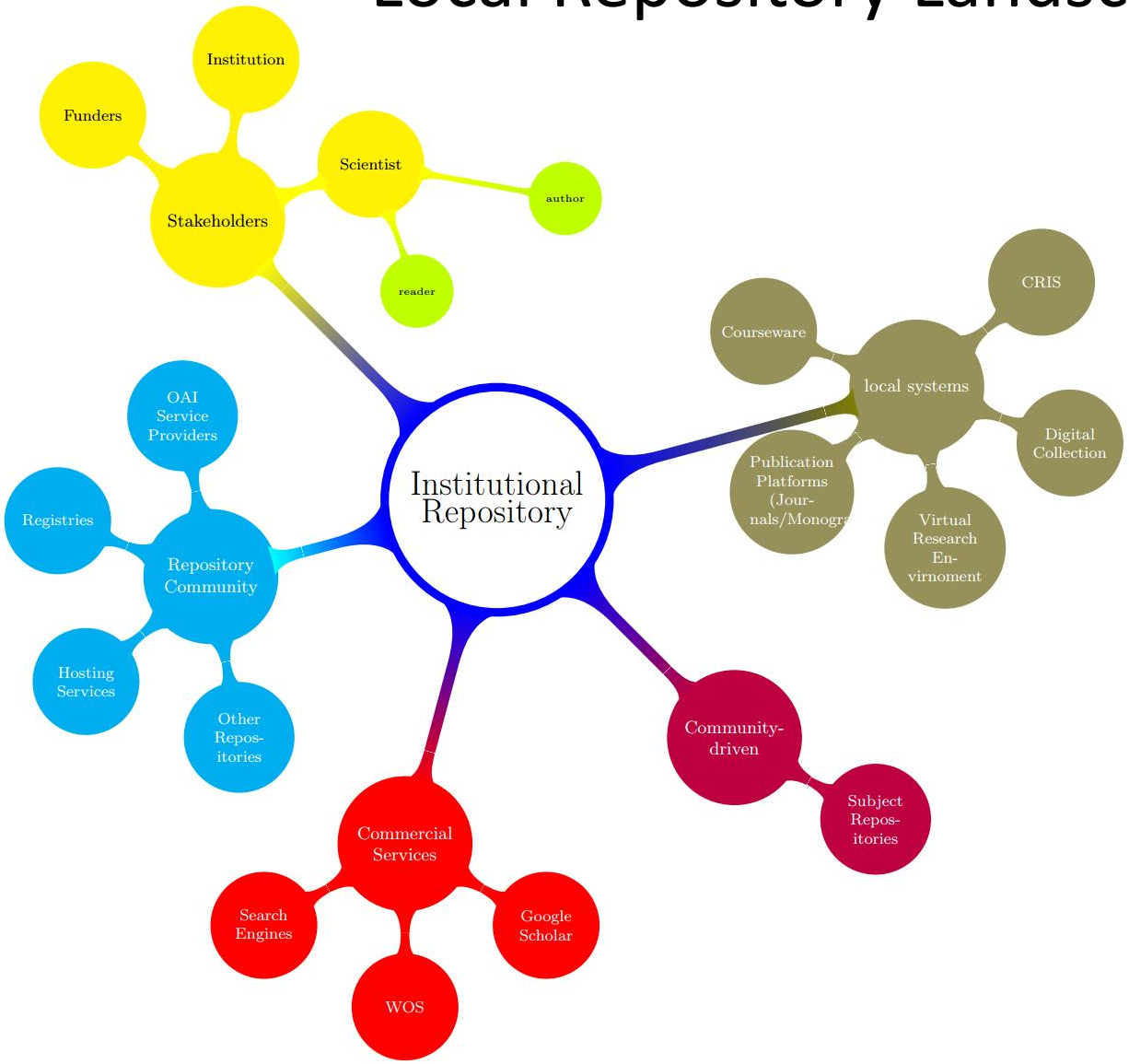
- Identifizieren der Nutzer-Wünsche und der Arbeitsabläufe
- Identifizieren der Interessen der weiteren Beteiligten (Institutionen, Förderorganizationen, Verlage, Informationsanbieter …)
- Ermittlung der erforderlichen Schnittstellen-Technologie
- Definition einer Prioritätenliste der Aktionsfelder

| Stakeholder | Users requirements |
|---|---|
| Researcher as an author | • Easy metadata feeds (including re-using existing data)<br>• Upload documents easily<br>• Easy and comfortable creation of complex data relations<br>• Automatic addition of linked data<br>• High visibility of his digital objects/documents/scientific profile and relations<br>• Easy embedding of publications in different working environments (personal publication lists, virtual research environments, etc.)<br>• Comfortable creating of complex documents (enhanced publications)<br>• Transparent usage statistics (download and citation frequencies)<br>• Easy storage and publishing solutions for articles, journals, monographs, working papers |
| Researcher as reader/end user | • Open Access to publications<br>• Visible references of their publications in secondary environments<br>• Comfortable search tools<br>• Visualized complex information on publication relationships (to other (similar or recommended) publications, to related research data)<br>• Transparent bibliometric information<br>• Stable document links<br>• Stable and safe document storage (Long-term preservation) |
| Institution | • Exposure of their affiliated publication output (institutional bibliography)<br>• Document and report research output information for assessment and compliance monitoring |
| Funder | • Assess impact of funded research outcome<br>• Provide open access to research outputs<br>• Track and monitor research outputs |
| External stakeholder *(publisher, information company, service provider)* | • Comprehensive, high quality, and standardized metadata information on publications and research data in order to reuse them |

# Institutionelle Repositorien – verbundene Systeme

- Aggregator Services
- Bibliographic Management Tools
- Current Research Information Systems (CRISs)
- Digital Collections
- Discipline-based Repositories
- E-Learning Systems
- Hosting Services
- Internet Search Engines
- Local Library Systems (catalogues)
- Publication Management Systems
- Publishing Systems (journals, monographs)
- Research Data Repositories
- Virtual Research Environments (VREs)
- Other Global Services and Players

# Local Repository Landscape



35 Issues grouped in

- Key Aspect: Impact and Visibility
- Key Aspect: Data Issues
- Key Aspect: Validation and Aggregation
- Key Aspect: Usability
- Key Aspect: Sustainability
- Key Aspect: Technical  Issues

## Impact and Visibility

- Supporting Search Engine Optimization (SEO)
- Supporting Repository Ranking Systems
- Exposing Usage Statistics
- Exposing Bibliometric Information
- Supporting Visibility in Repository Registries
- Improving Registry Infrastructure

## Usability

- Supporting Authorization and Authentication
- Supporting Embedding Services
- Exposing Publication Lists
- Exposing Citation Formats
- Supporting Data Export Functions
- Integrating Availability Services
- Supporting Author Identification Systems
- Supporting Institutional Services
- Extending End-User Usability
- Extending Usage of Visualization Tools

## Sustainability

- Improving Platform Stability
- Supporting Long-term Preservation and Archiving
- Exposing Persistent Identifiers
- Integrating different Persistent Identifiers

## Data Issues

- Supporting additional Metadata Format(s)
- Improving Metadata Quality (Data Curation)
- Supporting Enhanced Publications
- Supporting Linked (Open) Data
- Publication of Research Data
- Handling of Complex/Compound/Nested Repository Objects
- Monitoring Open Access Mandate Compliance
- Exposing Versioning Information

## Validation and Aggregation

- Validating Repository Metadata
- Processing Related Full-text
- De Duplication

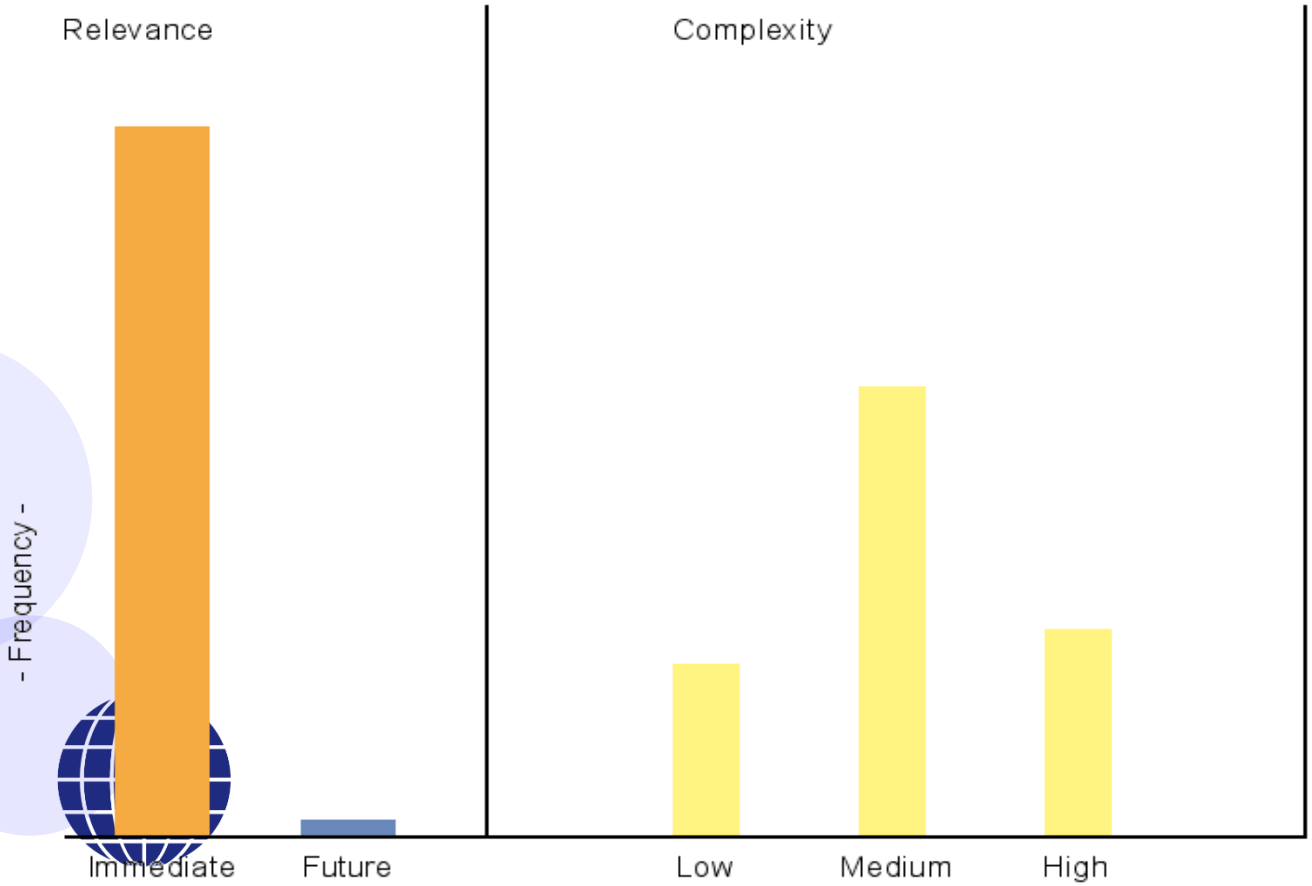
## Technical Issues

- Defining Architectural Recommendations for Repositories and their Interoperability
- Extending/Replacing Metadata Exposition Protocols
- Supporting OAI Service Provider Usage
- Supporting Deposit Protocols

## Issue: Supporting additional Metadata format(s)

Currently repositories deliver metadata mostly via OAI-PMH in Dublin Core format as mandatory and some of them support a broad variety of extended formats. Since DC is interrelated with a limited number of tags and a certain vagueness of interpretation there is a strong need to agree for alternative, more convenient metadata formats offering finer granularity. Potential formats to be considered (and depending on the purpose) are MODS, METS, MARC, CERIF and others.

Supporting additional Metadata format(s)



## Comments

*"Broader discussion among repo stakeholders, guidelines and training needed"*

*"If DC as a generic format is not good enough, then it needs to be improved or replaced. We don't want additional formats for the same purpose."*

COAR Office at Goettingen State and University Library
Platz der Göttinger Sieben 1, D-37073 Göttingen, Germany, Tel. +49 551 39 22215, Fax +49 551 39 5222
office@coar-repositories.org



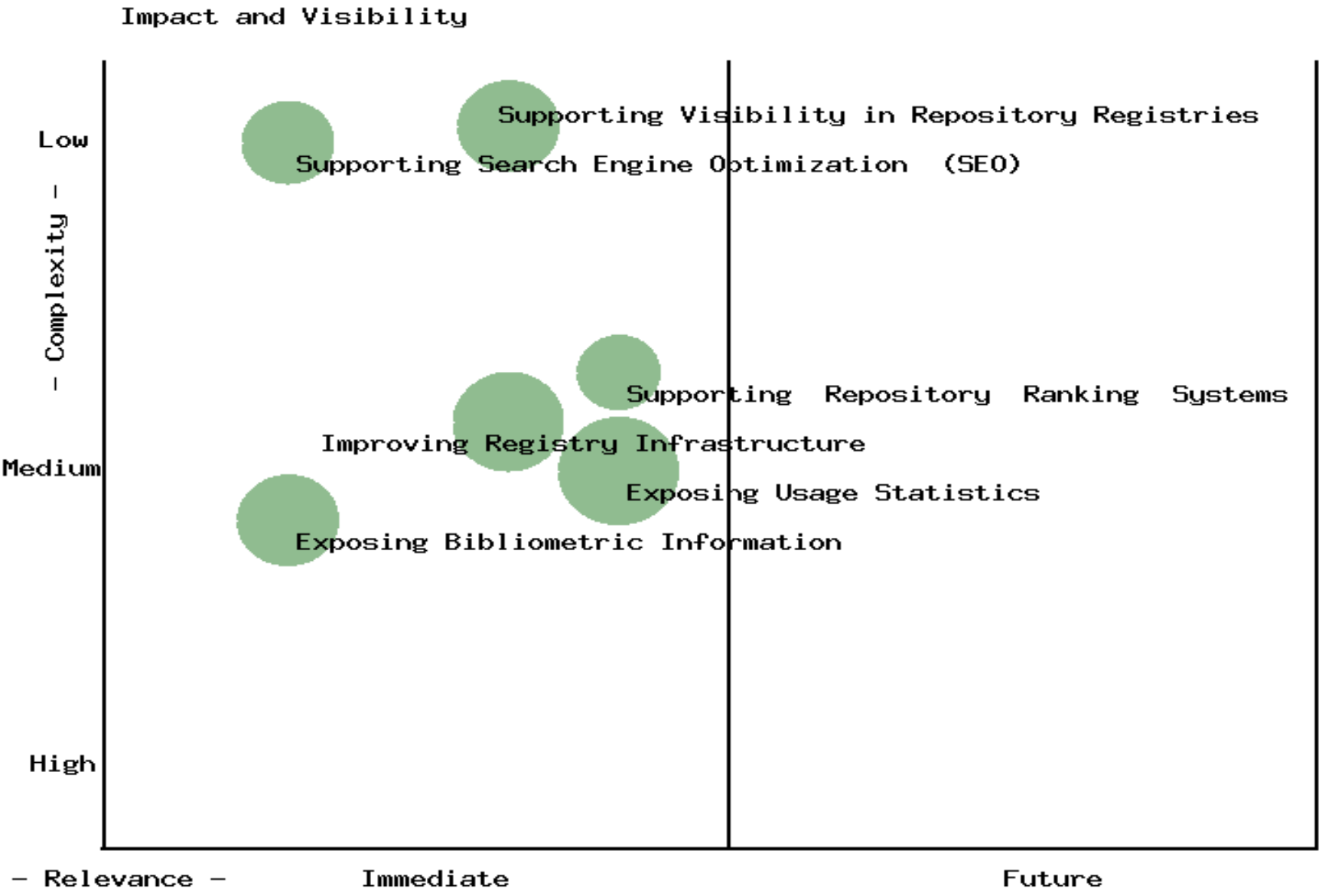


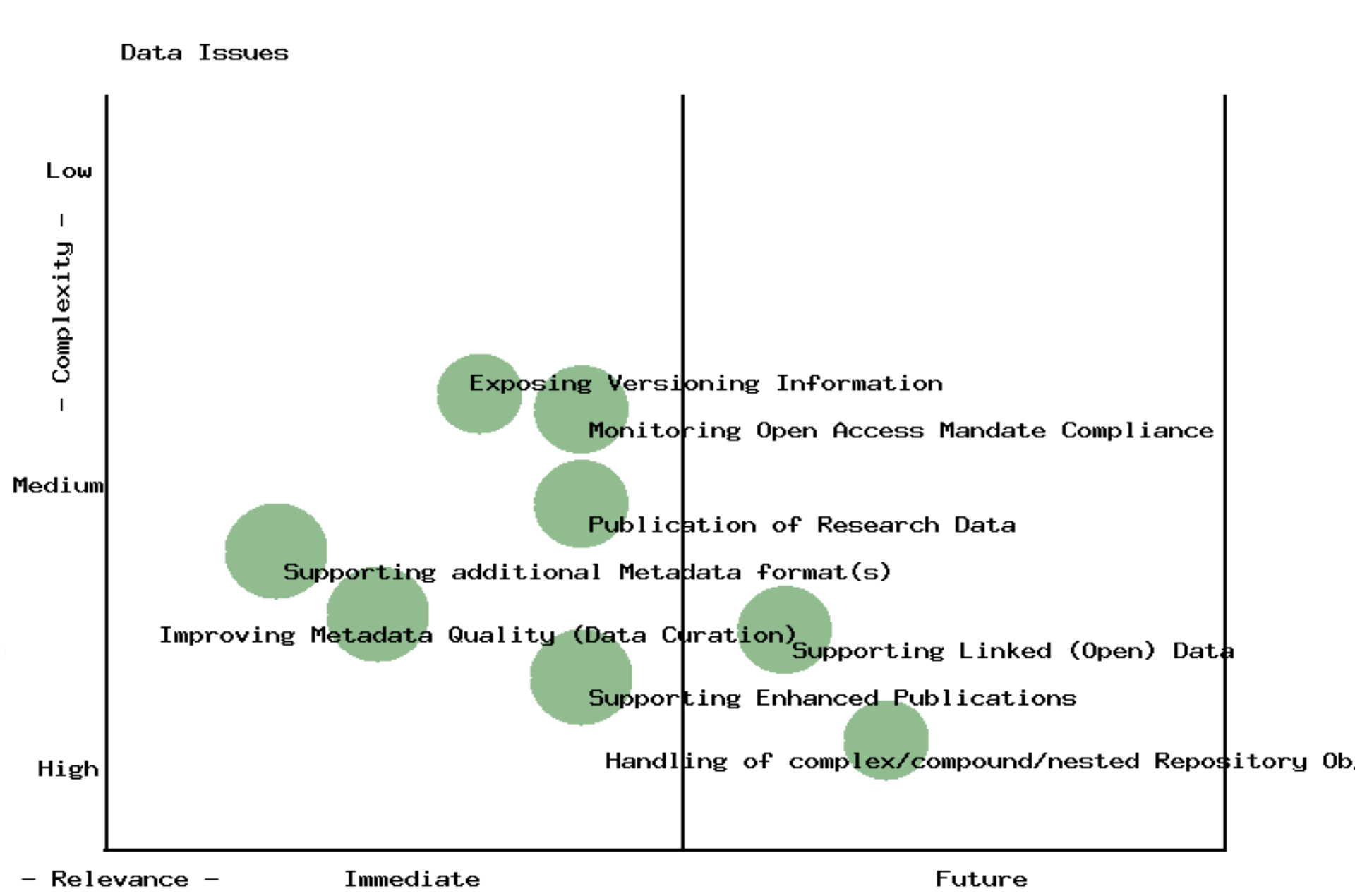
*"Depends on community and complexity of additional format"*

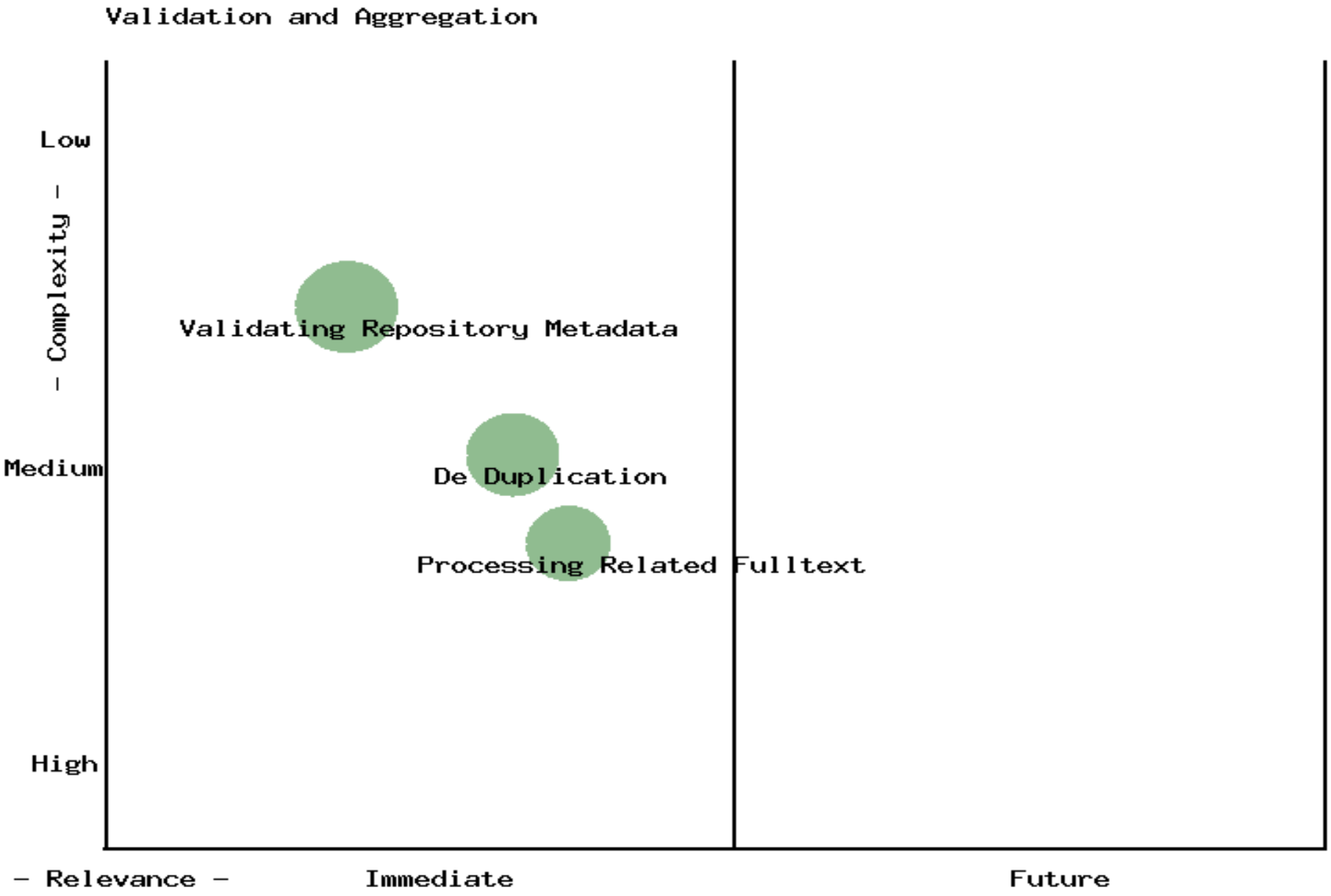
*"Adding more standards when they bring richness and detail is a key step to move forward in the current situation. The complexity of course depends on number and complexity of the new adoptions. DC is no longer useful for advancing in the field."*

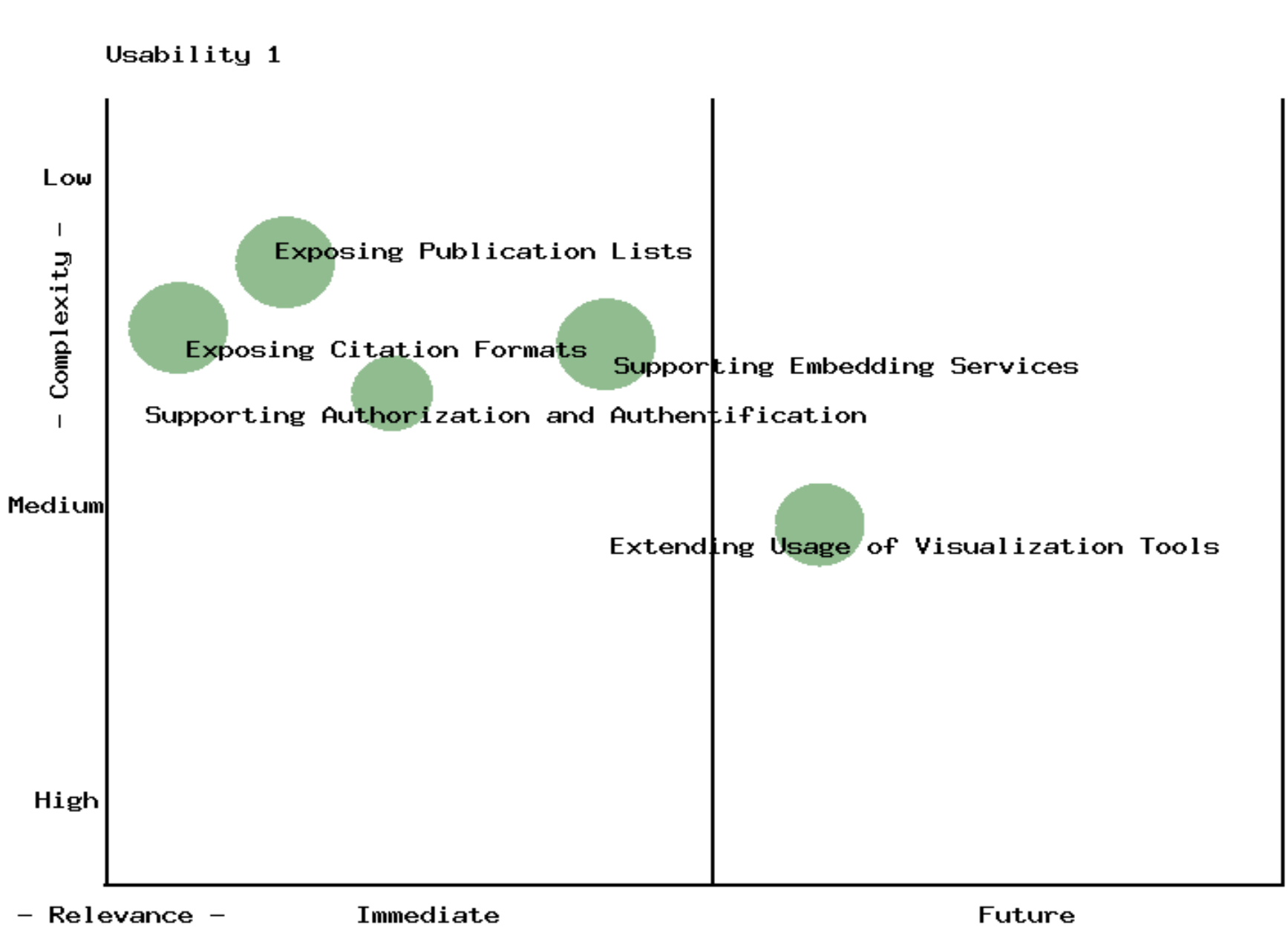
| | Short term | Medium term | Long term |
|---|---|---|---|
| **Low Complexity** | • Exposing Citation Formats<br>• Supporting Data Export Functions<br>• Supporting Author Identification Systems<br>• Supporting Search Engine Optimization (SEO)<br>• Exposing Publication Lists<br>• Integrating Different Persistent Identifiers | • Exposing Persistent Identifiers<br>• Supporting Authorization and Authentication<br>• Improving Platform Stability<br>• Supporting Institutional Services<br>• Extending End-User Usability<br>• Validating Repository Metadata<br>• Supporting Visibility in Repository Registries<br>• Supporting OAI Service Provider Usage<br>• Integrating Availability Services<br>• Supporting Embedding Services<br>• Supporting Repository Ranking Systems | |

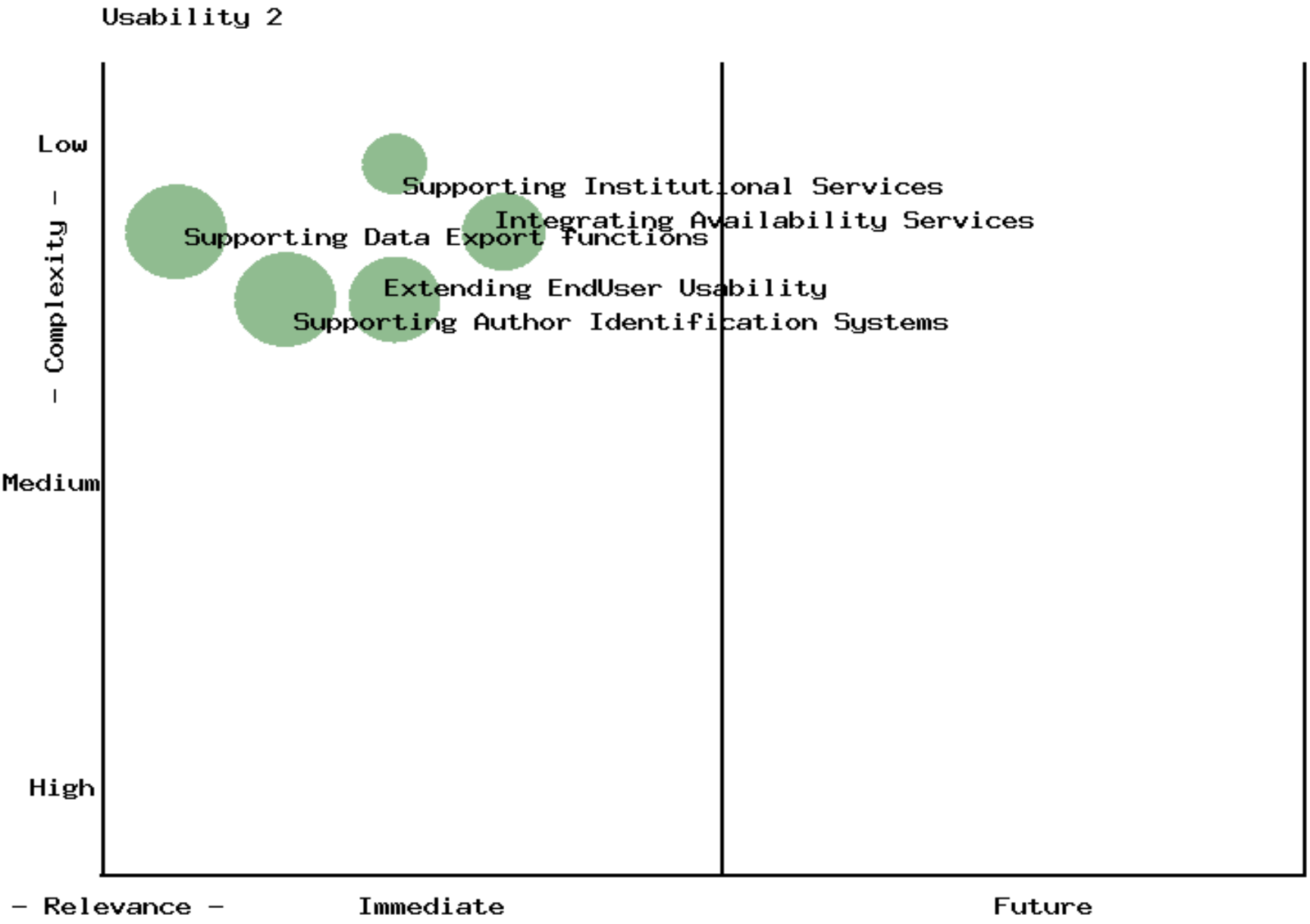
| | | | |
|---|---|---|---|
| **Moderate Complexity** | • Exposing Bibliometric Information | • Exposing Versioning Information<br>• De Duplication<br>• Improving Registry Infrastructure<br>• Monitoring Open Access Mandate Compliance<br>• | |
| **High Complexity** | • Exposing Usage Statistics<br>• Supporting Additional Metadata Format(s) | • Publication of Research Data<br>• Improving Metadata Quality (Data Curation)<br>• Processing Related Fulltext<br>• Supporting Deposit Protocols<br>• Defining Architectural Recommendations for Repositories and their Interoperability<br>• Supporting Enhanced Publications | • Extending Usage of Visualization Tools<br>• Supporting Linked (Open) Data<br>• Extending/Replacing Metadata Exposition Protocols<br>• Handling of Complex/ Compound/Nested Repository Objects<br>• Supporting Long-term Preservation and Archiving |

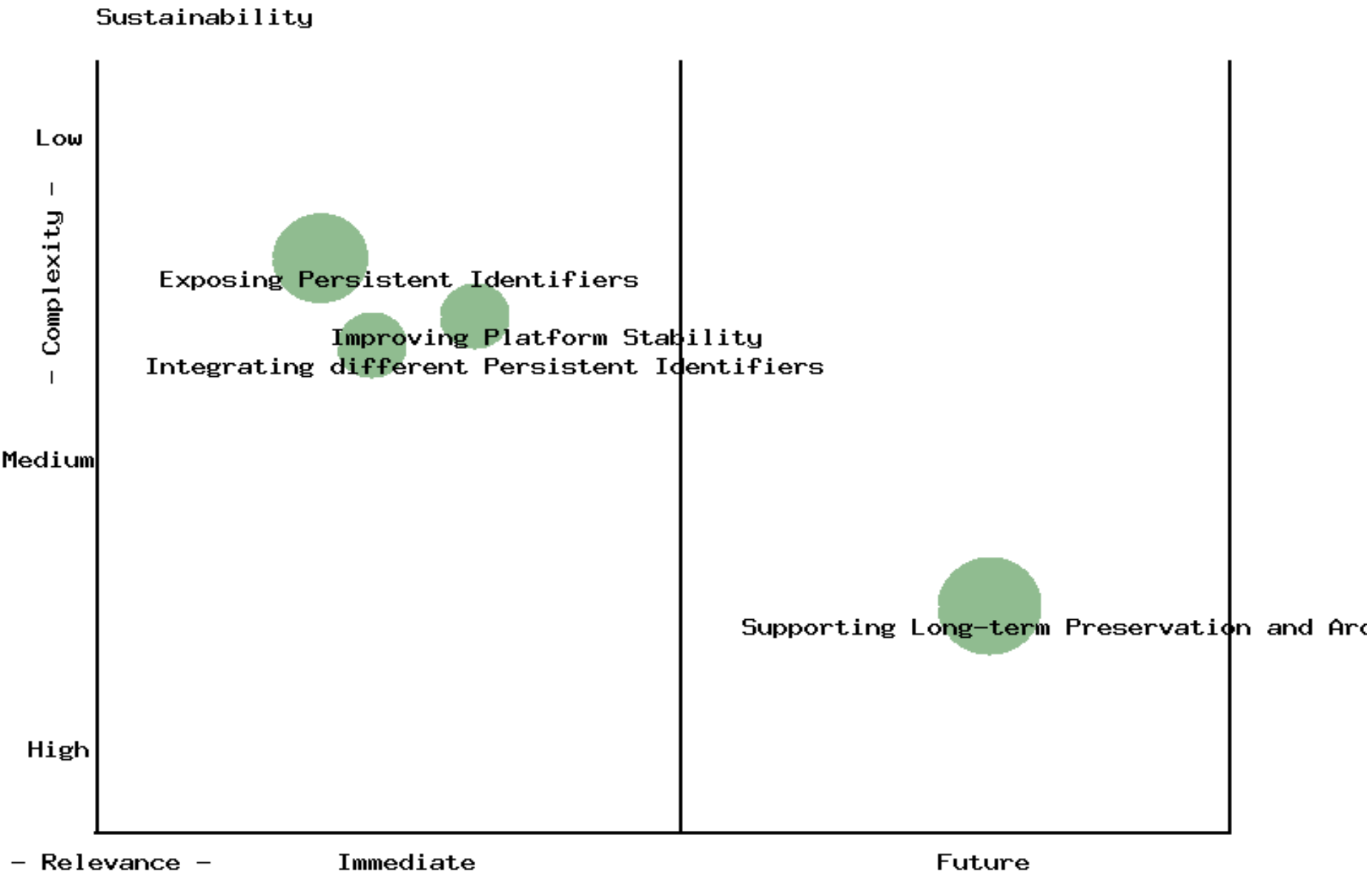

Impact and Visibility
Low
- Complexity -
Medium
High
Supporting Visibility in Repository Registries
Supporting Search Engine Optimization (SEO)
Supporting Repository Ranking Systems
Improving Registry Infrastructure
Exposing Usage Statistics
Exposing Bibliometric Information
- Relevance -
Immediate
Future

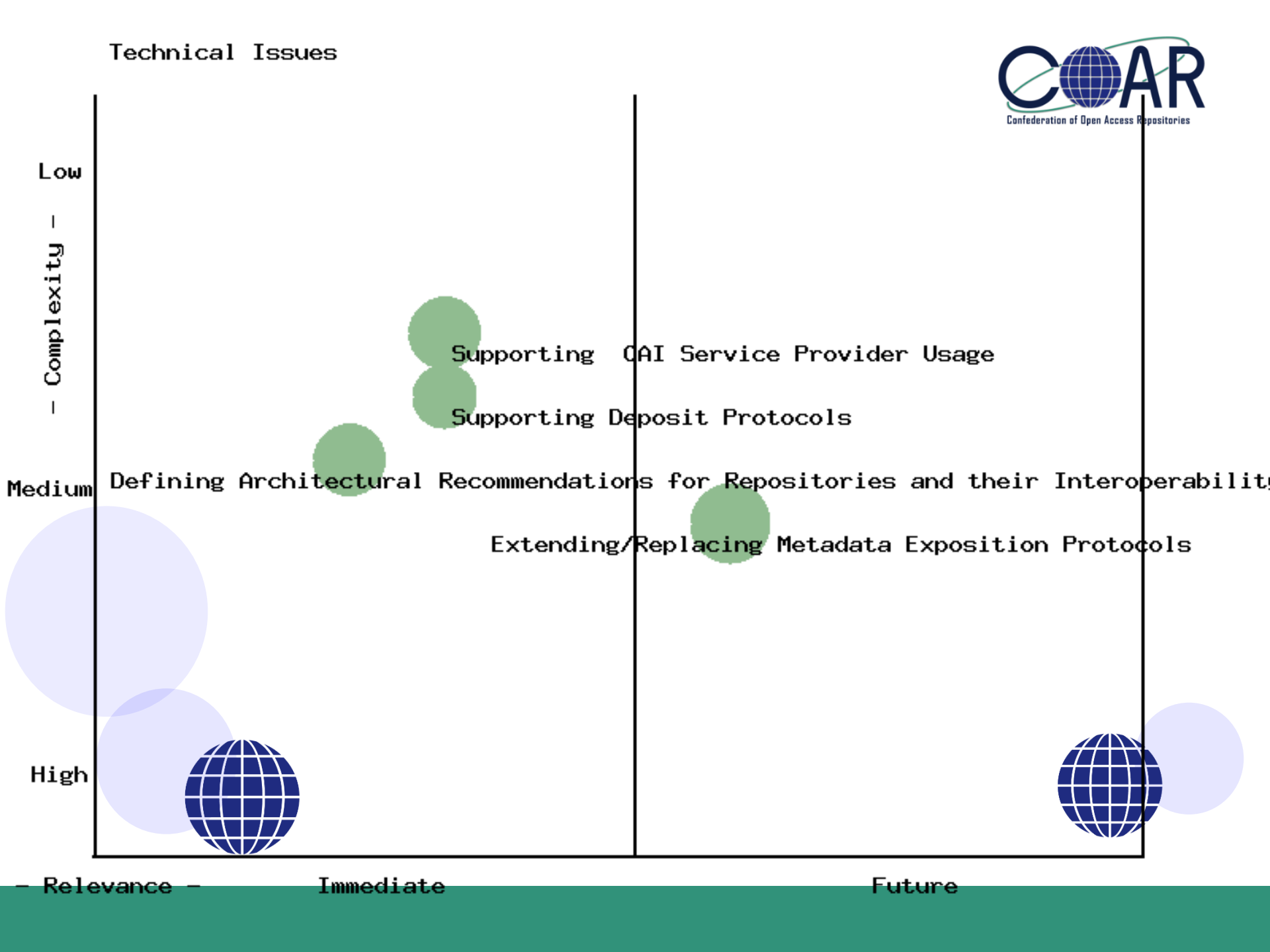

Data Issues
Low
- Complexity -
Medium
High
Exposing Versioning Information
Monitoring Open Access Mandate Compliance
Publication of Research Data
Supporting additional Metadata format(s)
Improving Metadata Quality (Data Curation)
Supporting Linked (Open) Data
Supporting Enhanced Publications
Handling of complex/compound/nested Repository Ob
- Relevance -
Immediate
Future


Validation and Aggregation
Low
- Complexity -
Medium
High
Validating Repository Metadata
De Duplication
Processing Related Fulltext
- Relevance -
Immediate
Future


Usability 1
Low
- Complexity -
Medium
High
Exposing Publication Lists
Exposing Citation Formats
Supporting Embedding Services
Supporting Authorization and Authentification
Extending Usage of Visualization Tools
- Relevance -
Immediate
Future


Usability 2
Low
- Complexity -
Medium
High
Supporting Institutional Services
Integrating Availability Services
Supporting Data Export functions
Extending EndUser Usability
Supporting Author Identification Systems
- Relevance -
Immediate
Future


Sustainability
Low
- Complexity -
Medium
High
Exposing Persistent Identifiers
Improving Platform Stability
Integrating different Persistent Identifiers
Supporting Long-term Preservation and Arc
- Relevance -
Immediate
Future


Technical Issues
Low
- Complexity -
Medium
High
Supporting OAI Service Provider Usage
Supporting Deposit Protocols
Defining Architectural Recommendations for Repositories and their Interoperabilit
Extending/Replacing Metadata Exposition Protocols
- Relevance -
Immediate
Future

## 5 Conclusion

Confederation of Open Access Repositories

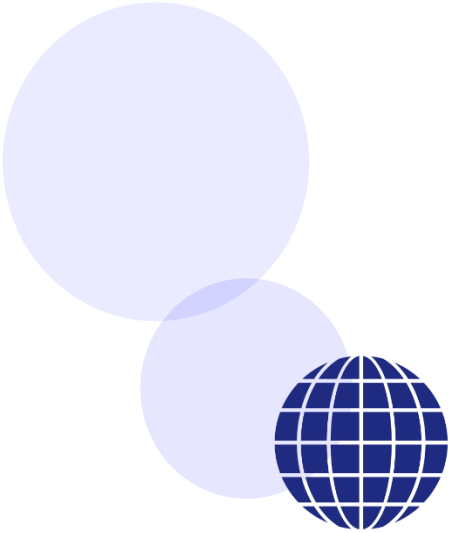
COAR will consider various paths for improving interoperability in the priority areas:

- What work is involved in ensuring interoperability in priority areas?
- Which stakeholders must be included in implementation and how can we best engage them in these activities? Particularly important will be the participation of the repository platform developers, as this is an essential strategy for widespread adoption.

In terms of next steps, COAR will:

1. Disseminate the roadmap and its results to COAR members and the broader community of stakeholders, in particular:
    a. Regional/National Repository Networks
    b. Repository Platform Communities
    c. Repository Managers
    d. Other related stakeholders (e.g. research administrative communities, publishers)
2. Build support and awareness of the benefits and need for interoperability
3. Support dialogue and progress towards the adoption of common approaches across regions and stakeholder communities
4. Develop and undertake strategies for implementing standards in repositories

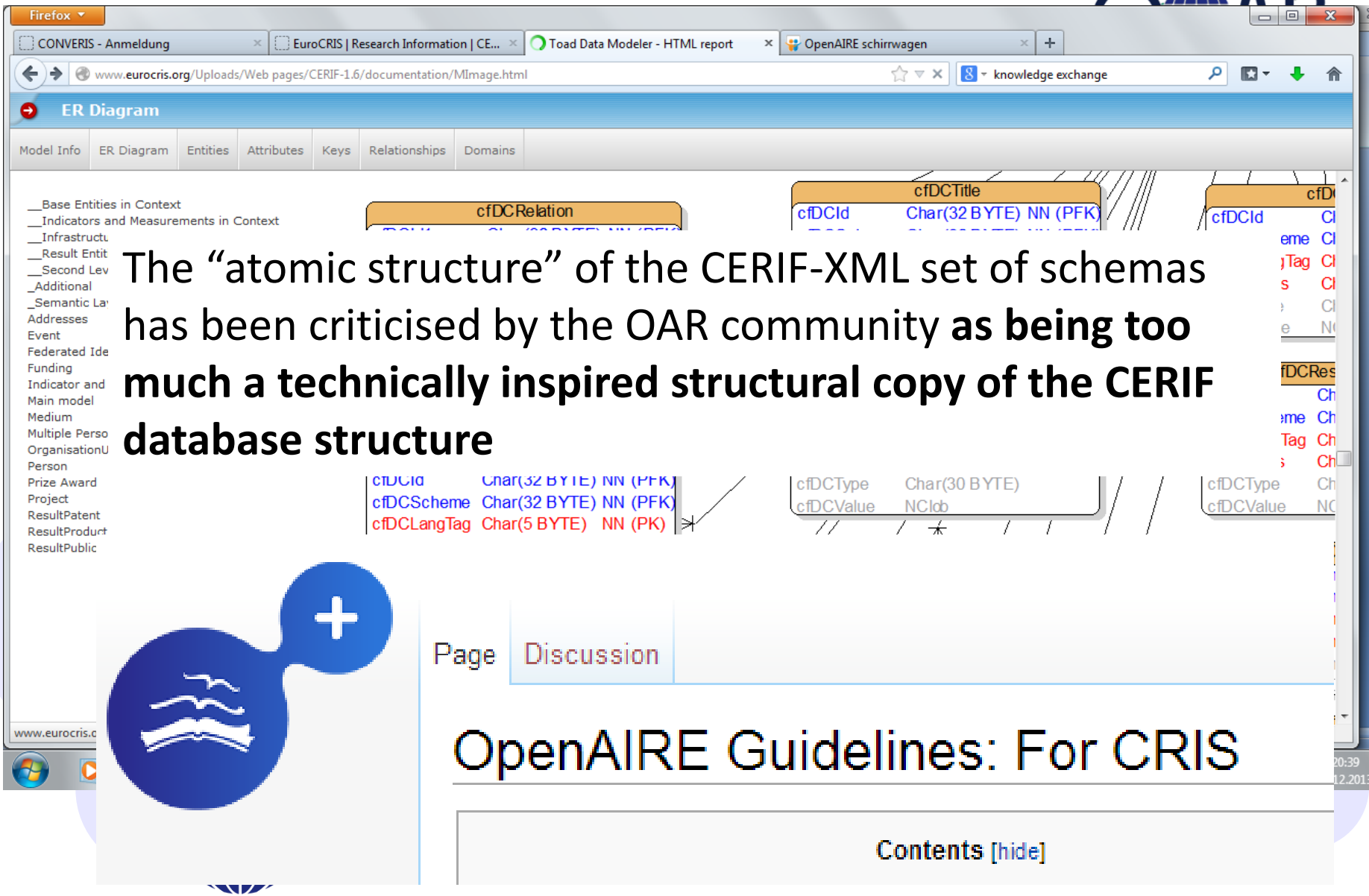
Clearly, as a global organization, COAR has an important role to play in connecting these various communities and coalescing around some best practices. In addition, COAR can coordinate the essential efforts for preparing underlying definitions, recommendations and guidelines to assist the development and implementation process.



# IR Interoperability

- The special case: CRIS-IR interoperability
- The common intersection: Publications
- The challenges:
  - The format question (CERIF – DC/MODS/MARC)
  - Related  Services

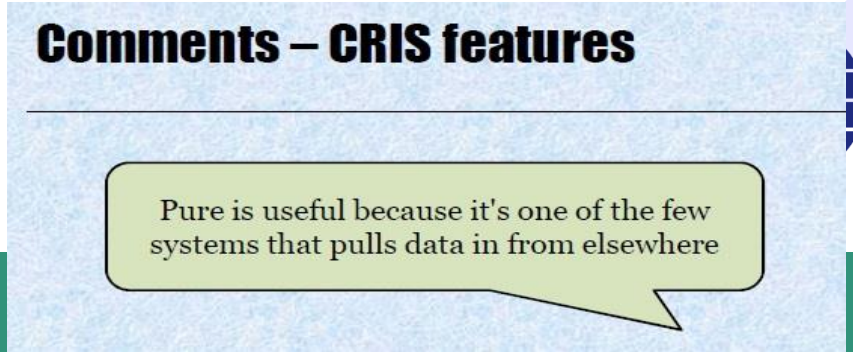
| CRIS | IR |
| --- | --- |
| Internal | Externally-orientated |
| Research Office | Library |
| Research Information Mgt. | Open Access |
| Metadata | Fulltext |
| Automatic Harvesting | Self/Archive/Mediate Deposit |
| CERIF | DC METS MODS |
| Proprietary Software | Open Source |




The “atomic structure” of the CERIF-XML set of schemas has been criticised by the OAR community **as being too much a technically inspired structural copy of the CERIF database structure**

# Why some institutions prefer to run CRISs

- Very comprehensive research information collection
- Oriented towards research reporting and evaluation (REF in the UK a big driver for CRISs)
- CRISs allow to deal with research funding at internal level -- very valued feature by Research Office managers
- CRISs will automatically harvest most institutional publications from external databases (it's about references not necessarily full-text)
- CRISs can now work as repositories: CRIS-as-IR use case
- CRISs will link to admin systems like HR or Finance
- CRISs (often) use CERIF as metadata standard -- very comprehensive and will ensure interoperabiity with funder systems



# Why other institutions will keep using their repositories

• Emphasis on showcasing and offering Open Access to the institutional research output

• CRISs are very expensive and difficult to implement

• Repositories can rather effectively (and easily) be managed from the Library

• Repositories' publication-centered metadata model may be enhanced to collect additional info: IR-as-CRIS use case

• There are now funder mandates requiring Open Access for the next national research asses

**EPrints: a hybrid CRIS/repository**

Share

Carr, Leslie (2010) EPrints: a hybrid CRIS/repository. At *Workshop on CRIS, CERIF and Institutional Repositories, Rome, IT, 10 - 11 May 2010.* 2pp.

http://eprints.soton.ac.uk/271048/

# Lösungen

- CRIS-Systeme mit Publikationsnachweis (**CRIS-as-IR**)
- IR mit Projektinformationen (**IR-as-CRIS**)
- Kopplungen CRIS – IR (**CRIS-IR-Interop**)

# CRIS Systeme im Repository Bereich

- PURE (Elsevier)
- Converis (Thomson-Reuters)
- Symplectic
- DSpace IRIS (Open Source)
- Proprietäre Lösungen/Erweiterungen der Repository-Software

# Vorhandene CRIS-Installationen in Europa (Stand: Mai 2013, euroCRIS)

- UK – 57
  - (Pure , Pure-DSpace, Pure-Eprints,
  - Converis-DSpace, Symplectic-DSpace, etc)
- Dänemark – 45
  - (Pure)
- Deutschland – 4
  - (Pure, Converis)
- Italien – 3
  - (Dspace IRIS)

# Three basic use cases: CRIS-as-IR

CRIS replaces repository and by becoming OAI-PMH-compliant it's able to perform its role (at a basic level now, but evolving)


Pure Queen's University Belfast

# Three basic use cases: IR-as-CRIS

IR data model and architecture is extended to cover additional aspects such as researchers, organisations and funding information (projects, grants, funder programmes), thus becoming able to play a basic CRIS role for research reporting purposes

**Enlighten: Research and APC funding workflows at the University of Glasgow**

Nixon, W.J., Ashworth, S., and McCutcheon, V. (2013) *Enlighten: Research and APC funding workflows at the University of Glasgow.* Insights: the UKSG journal, 26 (2). pp. 159-167. ISSN 2048-7754 (doi:10.1629/2048-7754.80)

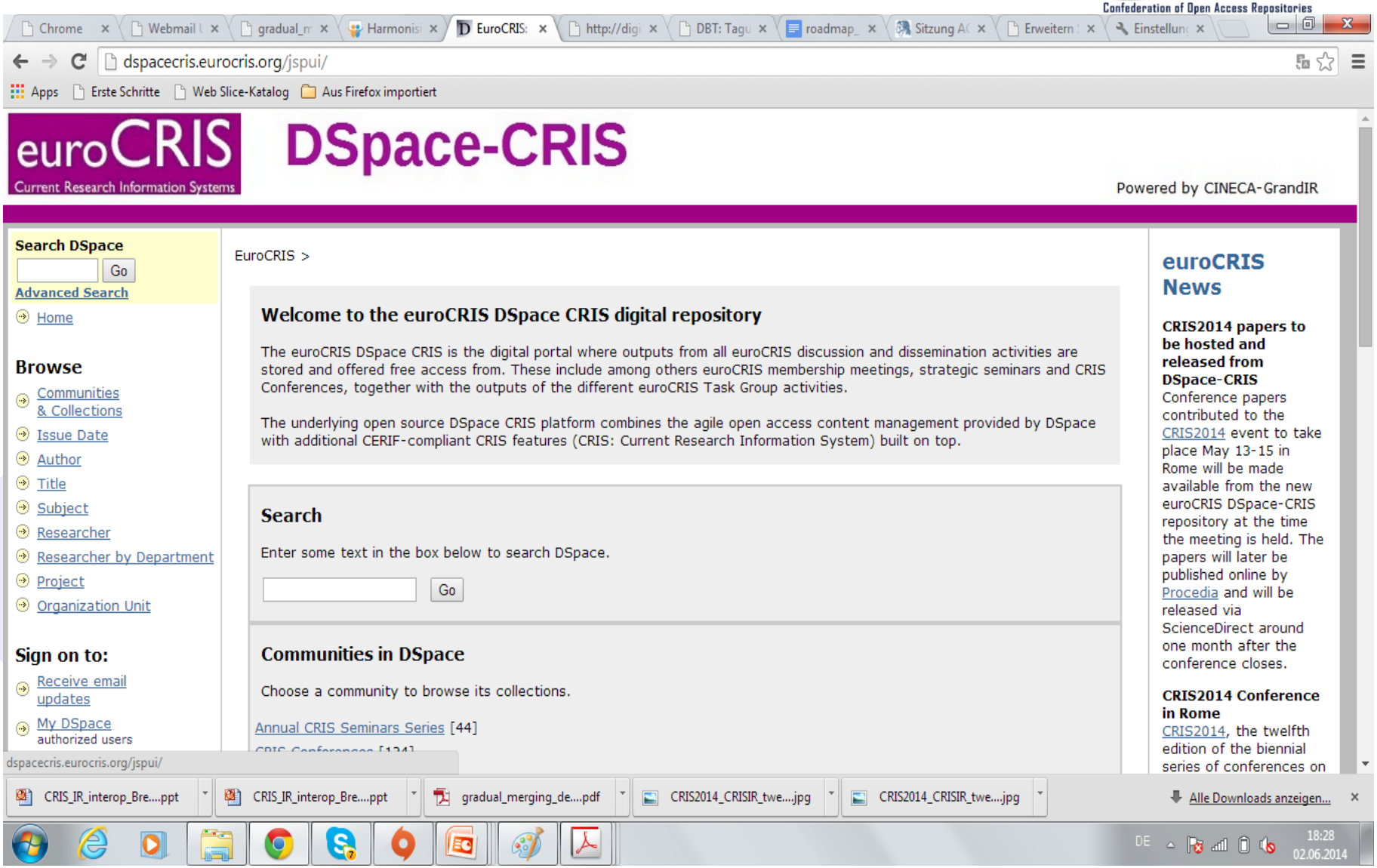
| | |
|---|---|
| **Status:** | Published |
| **Refereed:** | Yes |
| **Glasgow Author(s):** | Padgett, Prof Miles and Dienerowitz, Dr Maria |
| **Authors:** | Dienerowitz, M., Gibson, G., Dienerowitz, F., and Padgett, M. |
| **College/School:** | College of Science and Engineering > School of Physics and Astronomy |
| **Journal Name:** | Journal of Optics |
| **Journal Abbr.:** | J. Opt. |
| **ISSN:** | 2040-8978 |
| **ISSN (Online):** | 2040-8986 |
| **Copyright Holders:** | Copyright © 2012 European Optical Society |
| **First Published:** | First published in Journal of Optics 14(4):045003 |
| **Publisher Policy:** | Reproduced in accordance with the copyright policy of the publisher |

University Staff: Request a correction | Enlighten Editors: Update this record

**Funder and Project Information**

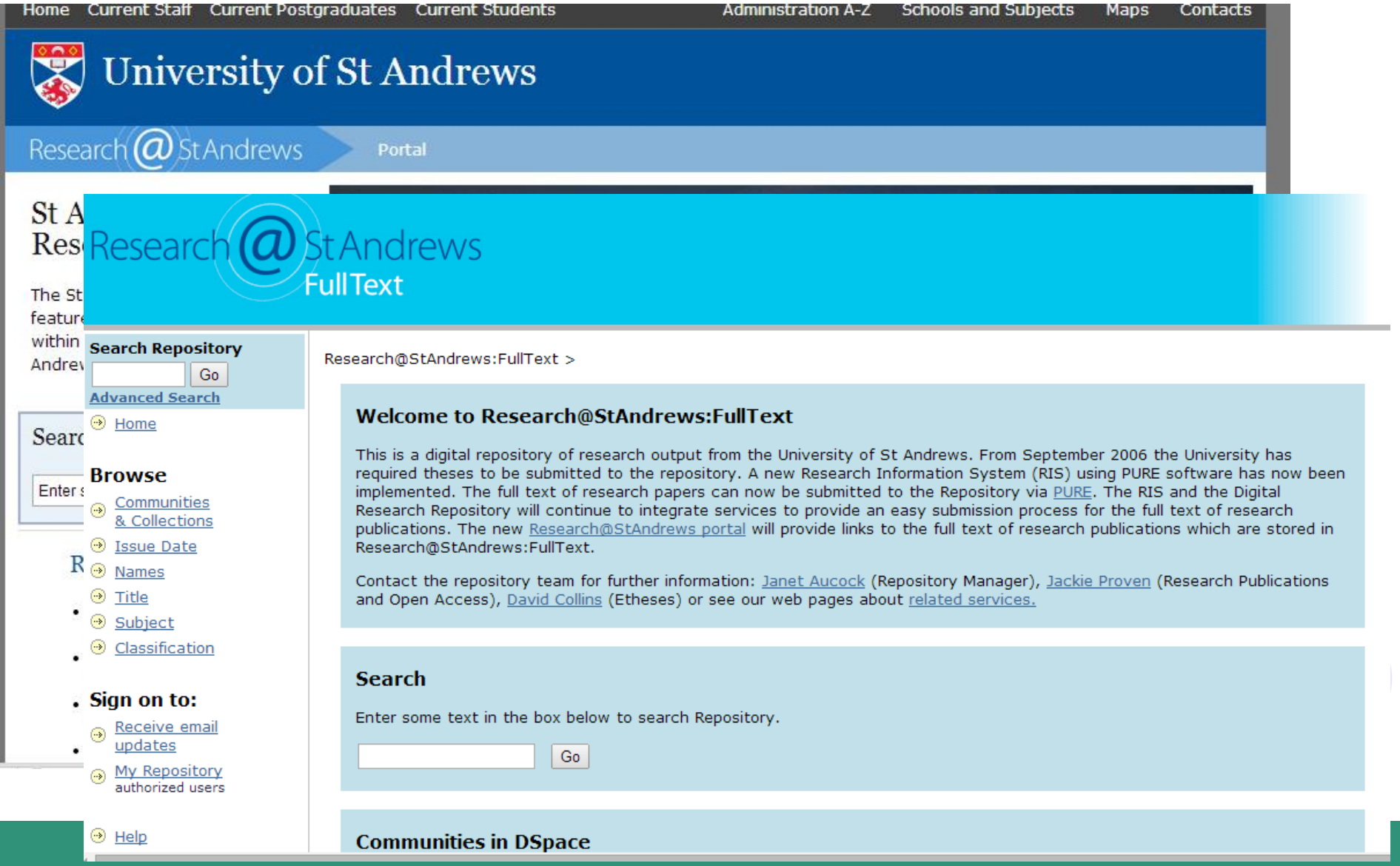
| Project Code | Award No | Project Name | Principal Investigator | Funder's Name | Funder Ref | Lead Dept |
|---|---|---|---|---|---|---|
| 48257 | 1 | Multi-object, high-throughput, spectro-microscopy: Life Sciences Interface | Miles Padgett | Engineering & Physical Sciences Research Council (EPSRC) | EP/H007636/1 | Physics and Astronomy |

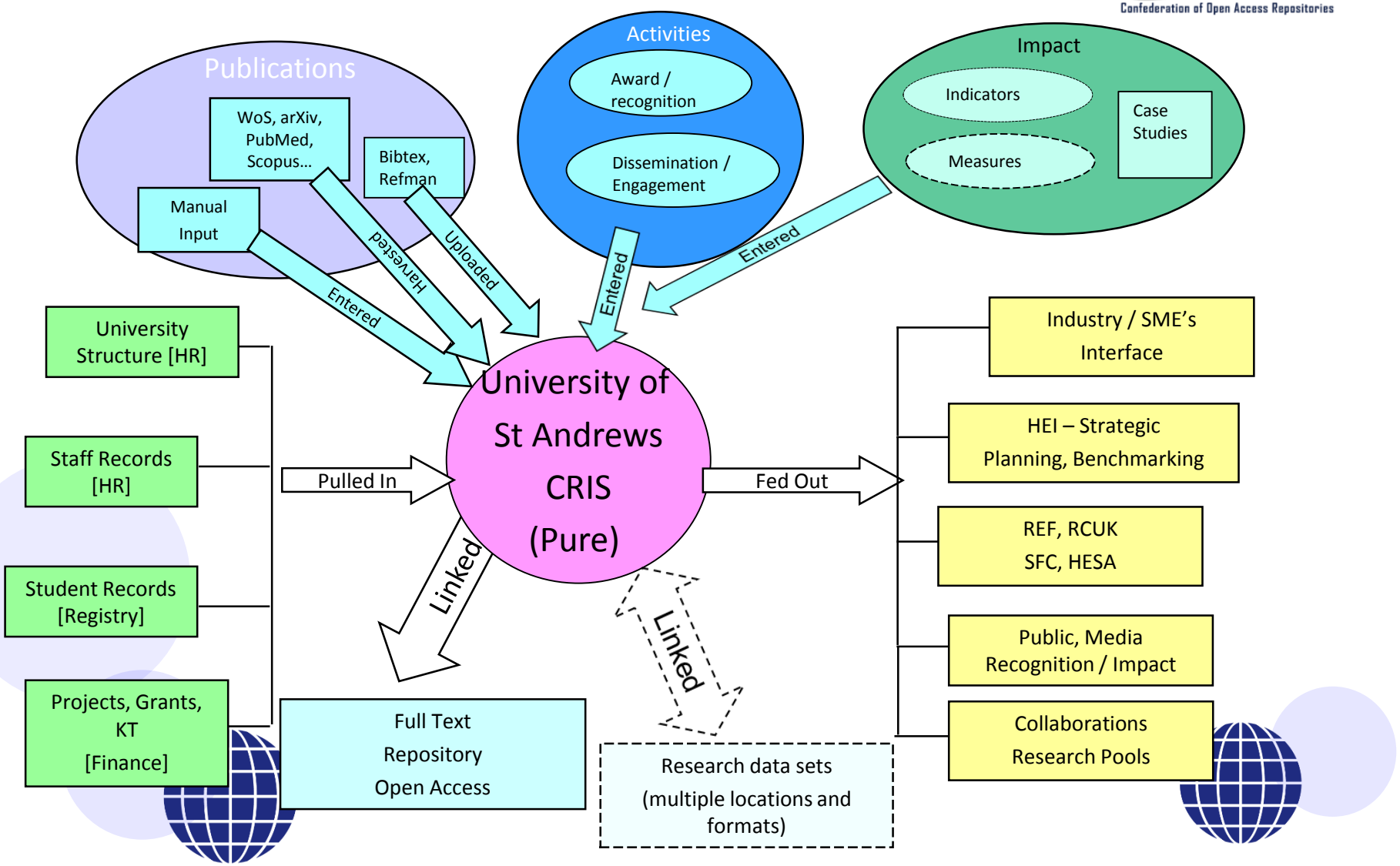
# Three basic use cases: IR-as-CRIS



# Three basic use cases: IR-as-CRIS

# Three basic use cases: CRIS-IR-Interop




Publications
WoS, arXiv, PubMed, Scopus…
Bibtex, Refman
Manual Input
Harvested
Uploaded
Entered
Activities
Award / recognition
Dissemination / Engagement
Entered
Impact
Indicators
Measures
Case Studies
Entered
University Structure [HR]
Staff Records [HR]
Student Records [Registry]
Projects, Grants, KT [Finance]
Pulled In
University of St Andrews CRIS (Pure)
Fed Out
Linked
Linked
Full Text Repository Open Access
Research data sets (multiple locations and formats)
Industry / SME's Interface
HEI – Strategic Planning, Benchmarking
REF, RCUK SFC, HESA
Public, Media Recognition / Impact
Collaborations Research Pools

# Die Schlüssel-Herausforderungen (aus IR-Sicht):

- Das Verhältnis CERIF – bibliographische Formate
- Integration von Services (Interoperabilität)

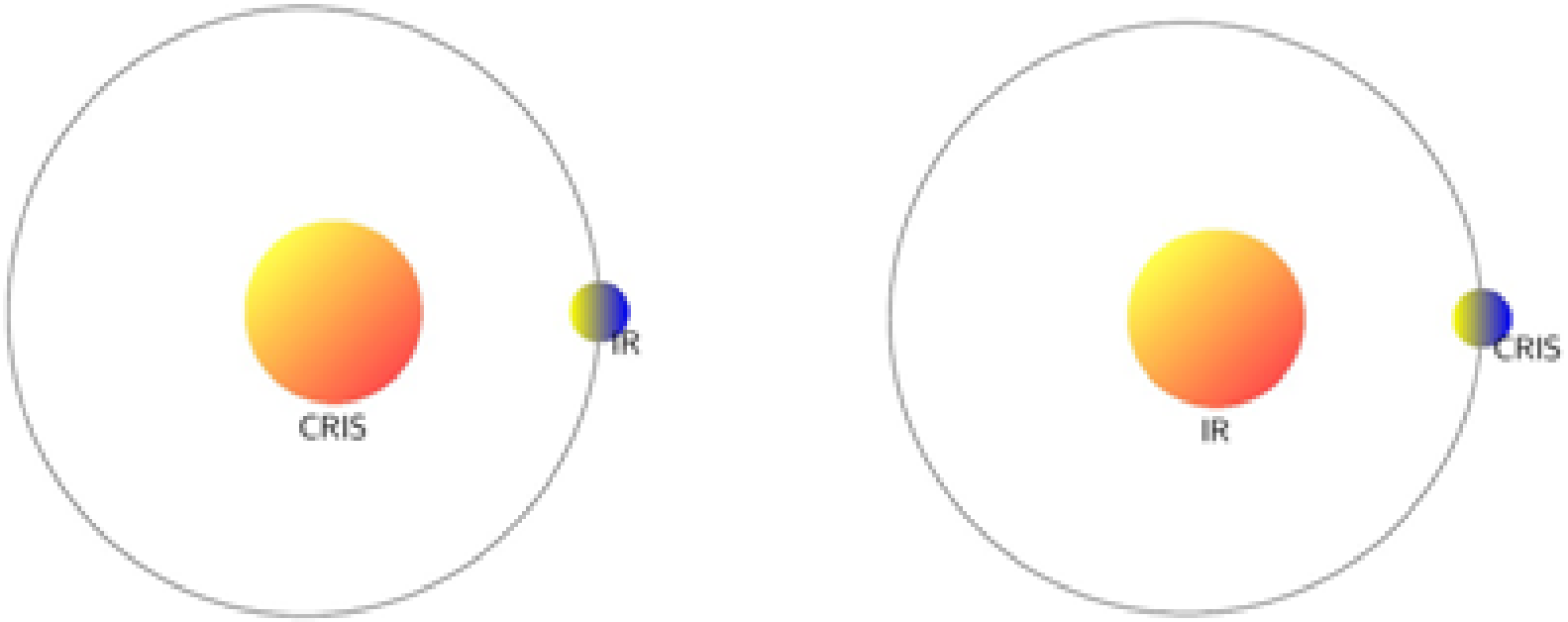
## Dringend empfohlen:

**Kommunikation zum Erfahrungsaustauch und zur Bündelung der Aktivitäten**

# Zusammenfassung

- Nicht unbedingt ein Gegensatz – Interoperabilität ist das Kernthema
- Kommunikation ist die Schlüsselstrategie

# “Not quite a dichotomy – interoperability the key feature”



“Not quite a dichotomy – interoperability the key feature”

# Danke!