Universität Bielefeld

Universitätsbibliothek

COAR Interoperability Roadmap

Friedrich Summann

Universitätsbibliothek Bielefeld

Kolloquium Wissensinfrastruktur

30.1.2015

COAR

= Confederation of Open Access Repositories

[https://www.coar-repositories.org/]

- Enhances the visibility of research outputs
- Paves the road to interoperability
- Fosters knowledge exchange on repository issues
- Strengthens international open access implementation

Greater visibility and application of research through global networks of Open Access repositories

COAR – G × | UB Bielefe × | UB Bielefe × | UB Bielefe × | BASE: Adr × | Hochschu × | Nicht verfi ×

https://www.coar-repositories.org

Apps | Erste Schritte | Roadmap com... | http://www.revi... | http://ir.yic.ac.c... | Web Slice-Katal... | Aus Firefox imp...

COAR
Confederation of Open Access Repositories

Greater visibility and application of research through global networks of Open Access repositories

Login | English | Search

Activities | Community | News & Media | About

Home

Major international associations join together to support immediate open access to research articles: Endorse the Statement about embargo periods!

Subscribe to our feed

Visit us on Facebook

View our photos on Flickr

Videos by COAR on Vimeo

News & Announcements

JAN 20 COAR-SPARC Conference 2015 – Connecting research results, bridging communities, opening scholarship

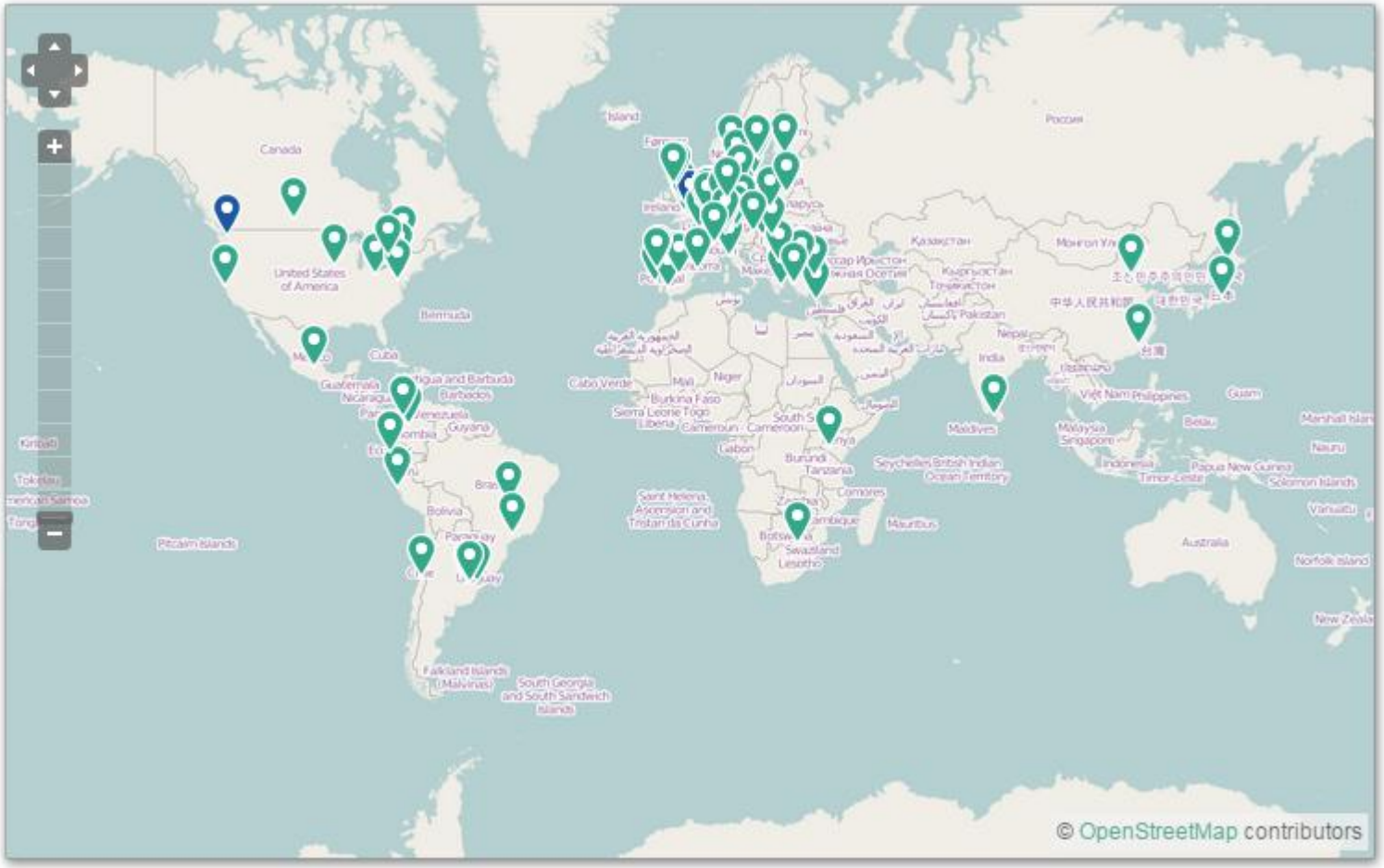
We are very pleased to announce our upcoming conference, jointly organized by COAR and SPARC, that will take place in the beautiful city of Porto,

Advocacy & Leadership

OCT 21 Major international associations reaffirm their support for immediate open access to research articles

On the occasion of Open Access Week, COAR and other international associations are reaffirming their support for immediate open access to research results.

COAR members and partners



COAR Members

China

- National Science Library, Chinese Academy of Sciences (CAS)
- Peking University Library
- Xiamen University Libraries

Colombia

- Red Nacional Académica de Tecnología Avanzada (RENATA)
- Universidad del Rosario
- Universidad EAFIT

Cyprus

- Cyprus University of Technology

Denmark

- Denmarks's Electronic Research Library (DEFF) on behalf of the Danish Agency for Libraries and Media

Ecuador

COAR members

⊞ Universidade do Minho	Braga, Portugal
⊞ Universidade do Porto	Porto, Portugal
⊞ Universitat de Barcelona - Centre de Recursos per a l'Aprenentatge i la Investigació (CRAI)	Barcelona, Spain
⊞ Universität Regensburg	Regensburg, Germany
⊞ Université de Lorraine Library	Villers-les-Nancy, France
⊞ Université du Luxembourg	Luxembourg, Luxembourg
⊞ Universiteit Gent	Gent, Belgium
⊞ University of Alberta Library	Edmonton, Alberta, Canada
⊞ University of Antwerp	Antwerpen, Belgium
⊞ University of Bielefeld	Bielefeld, Germany
⊞ University of California - Berkeley Law Library	Berkeley, United States
⊞ University of Debrecen	Debrecen, Hungary

COAR Partners

Partners

⊞ Bepress	Berkeley, CA, United States
⊞ euroCRIS	The Hague, Netherlands
⊞ LIBER	The Hague, Netherlands
⊞ Microsoft Research	Redmond, United States
⊞ Open Repository - BioMed Central	London, United Kingdom
⊞ SPARC Europe	The Hague, Netherlands
⊞ SPARC North America	Washington D.C., United States
⊞ The European Organization for Nuclear Research (CERN)	Geneva, Switzerland

Arbeitsgruppen mit Bielefelder Beteiligung

- COAR Working Group 2 Interoperability

- COAR Interest Group “Controlled Vocabularies for Repository Assets”

The Interest Group will be the new maintainer of the (in DRIVER and OpenAire defined) vocabulary and widen its applicability for global use. It will be hosted at COAR and made publicly available as a resource for the global repository community.

Moreover this Interest Group will provide an overview of related controlled vocabularies and open access indicators used in international and regional guidelines that are targeted at repositories. Using this review will allow to update and align controlled vocabulary for repositories at COAR.

COAR Interoperability Project

- Phase 1: The Case for Interoperability for Open Access Repositories."
- Phase 2: Open Discussion
- Phase 3: The Current State of Open Access Repository Interoperability (2012)
- Phase 4: **COAR Roadmap for Future Directions for Repositories Interoperability**

Concerned COAR Discussion Groups

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

The preparation activities

- Experts Feedback and Comments
- Internal Discussions
- Identifying the Interoperability Issues
- Extracting a Questionnaire
- Processing the Results
- The result: 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 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

January 2015

http://coar-repositories.org

Table of Contents

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 Ángel Sicilia**, University of Alcalá, Spain
- **Paul Vierkant**, Humboldt University of Berlin, Germany
- **Paul Walk**, University of Bath, UK

Executive Summary

In the past few years, Open Access repositories and their associated services have become an important component of the global e-research infrastructure. Increasingly, repositories are also being integrated with other systems, such as research administrative systems and with research data repositories, with the aim of providing a more integrated and seamless suite of services to various communities. Repositories can also be connected into networks (e.g. at the national or regional level) to support unified access to an open, aggregated collection of scholarship and related materials that machines can mine enabling researchers to work with content in new ways and allowing funders and institutions to track research outputs.

Scholarly communication is undergoing fundamental changes, in particular with new requirements for open access to research outputs, new forms of peer-review, and alternative methods for measuring impact. In parallel, technical developments, especially in communication and interface technologies facilitate bi-directional data exchange across related applications and systems. The aim of this roadmap is to identify important trends and their associated action points in order for the repository community to determine priorities for further investments in interoperability.

The roadmap process began with the compilation of a comprehensive list of interoperability issues derived from a broad discussion in the information, publishing and repository community. An Expert Advisory Panel was then asked to rate each issue according to its level of complexity and temporal relevance (or timing). This report presents the results of this process, ranking the issues according to these dimensions. The table below presents the key aspects in a two-dimensional structure.

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

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

The **goal of this roadmap** is to define the interoperability cornerstones for repositories according to their relevance and level of complexity, with particular attention paid to the following challenges:

- Technical
  1. implications for APIs, metadata formats, added-value services, vocabularies
  2. linking to other entities (publications, research data, project information, impact / statistics)
  3. sharing of digital assets
  4. coverage and support of digital assets beyond text
- Organizational
  1. roles and responsibilities of operation, support, development
- Legal
  1. issues on data exchange and re-use

- The **objectives** are to provide a detailed account of repository interoperability issues in order to:
-
- Describe all areas of interoperability for repositories in the future, including:
  - researchers needs and workflows (creating, reading, re-using, discovery and filtering as well as extracting of knowledge; front and backend),
  - funders and institutional interests,
  - integration with other infrastructures (e.g. disciplinary research infrastructures, authority services; other (incl. commercial) stakeholders (e.g. Google Scholar, Mendeley, ResearchGate, F1000));
- Identify levels of complexity, timing and importance for each key area of interoperability;
- Develop a list of priority issues for interoperability efforts for the repository community.

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) • Upload documents easily • Easy and comfortable creation of complex data relations • Automatic addition of linked data • High visibility of his digital objects/documents/scientific profile and relations • Easy embedding of publications in different working environments (personal publication lists, virtual research environments, etc.) • Comfortable creating of complex documents (enhanced publications) • Transparent usage statistics (download and citation frequencies) • Easy storage and publishing solutions for articles, journals, monographs, working papers
Researcher as reader/end user	• Open Access to publications • Visible references of their publications in secondary environments • Comfortable search tools • Visualized complex information on publication relationships (to other (similar or recommended) publications, to related research data) • Transparent bibliometric information • Stable document links • Stable and safe document storage (Long-term preservation)
Institution	• Exposure of their affiliated publication output (institutional bibliography) • Exposure of related institutional research information (projects, prizes • Document and report research output information for assessment and compliance monitoring
Funder	• Assess impact of funded research outcome • Provide open access to research outputs • 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

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

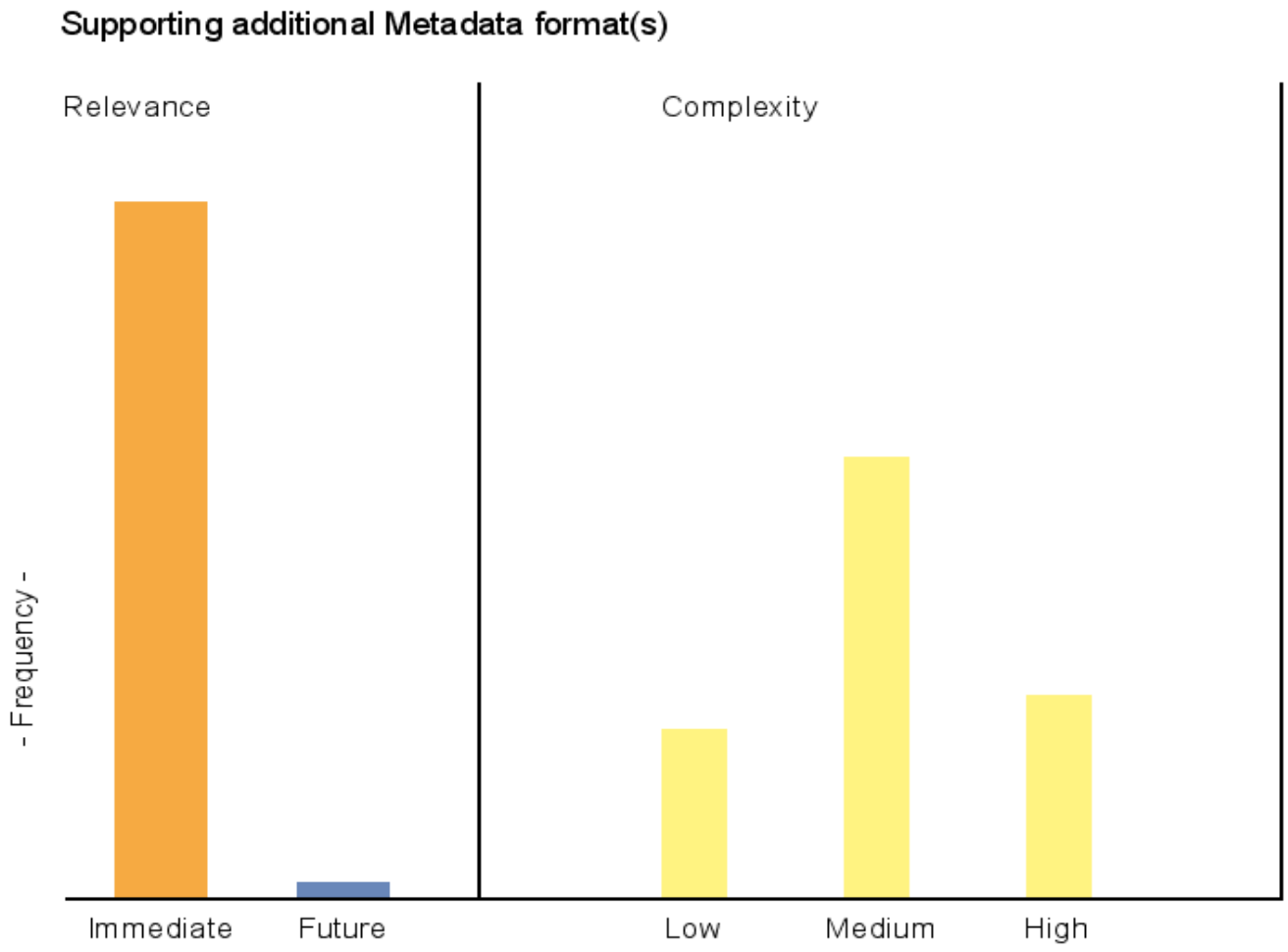
- Validating Repository Metadata
- Processing Related Full-text
- De Duplication

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.



Comments

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

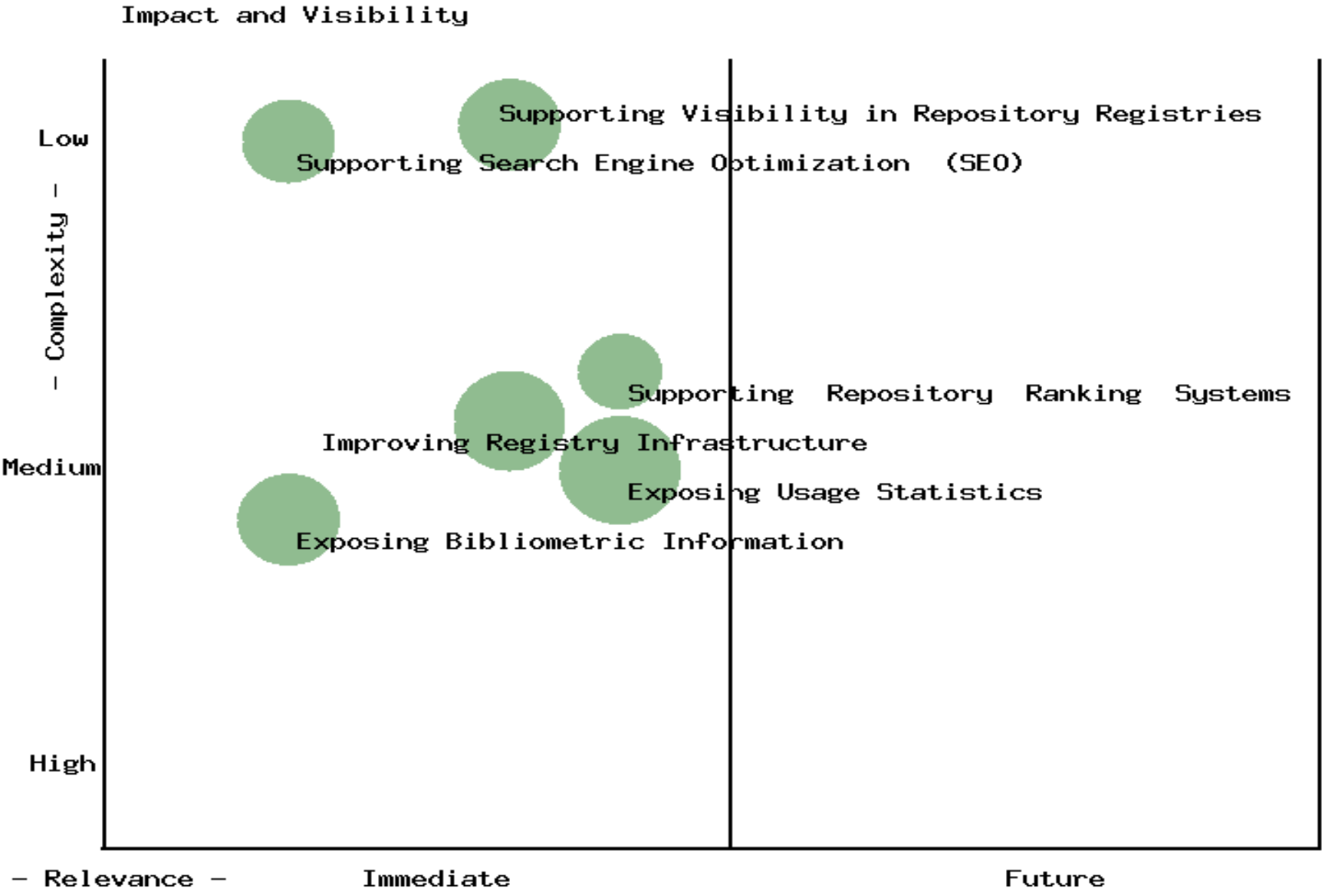
"If DC as a generic format is not good enough, then it needs to be improved or replaced. We don't want additional formats for the same purpose."

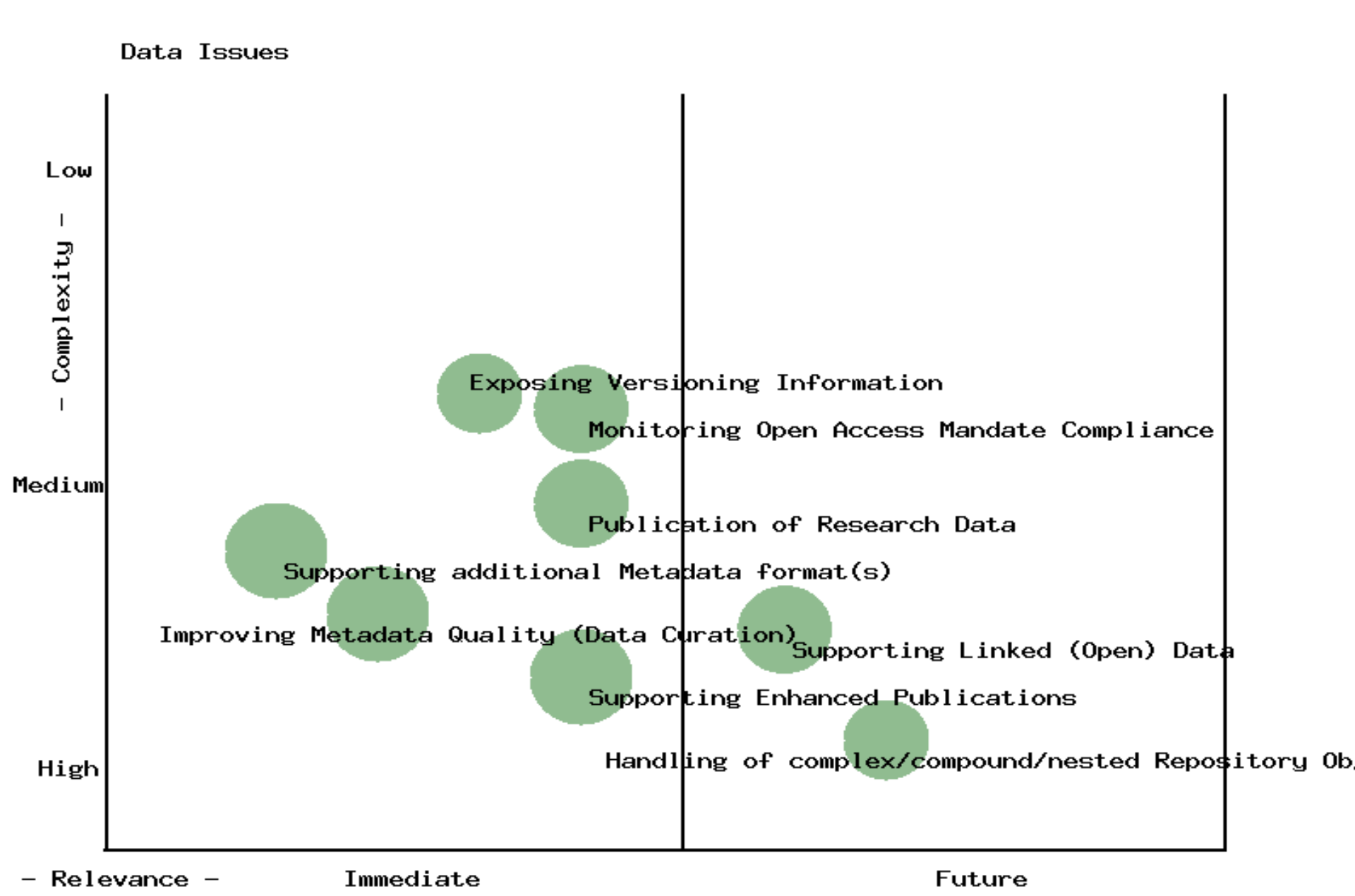
"Depends on community and complexity of additional format"

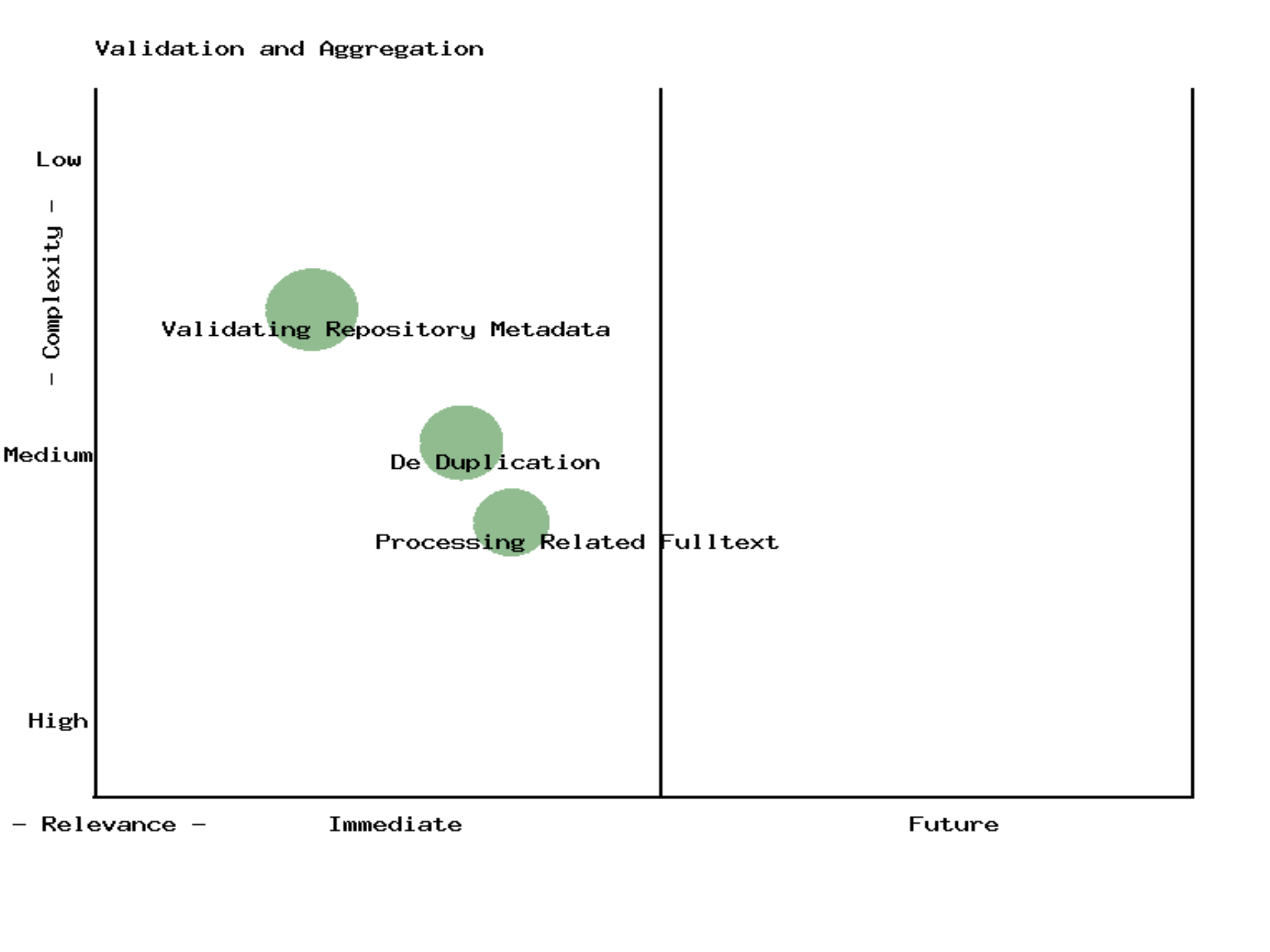
"Adding more standards when they bring richness and detail is a key step to move forward in the current situation. The complexity of course depends on number and complexity of the new adoptions. DC is no longer useful for advancing in the field."

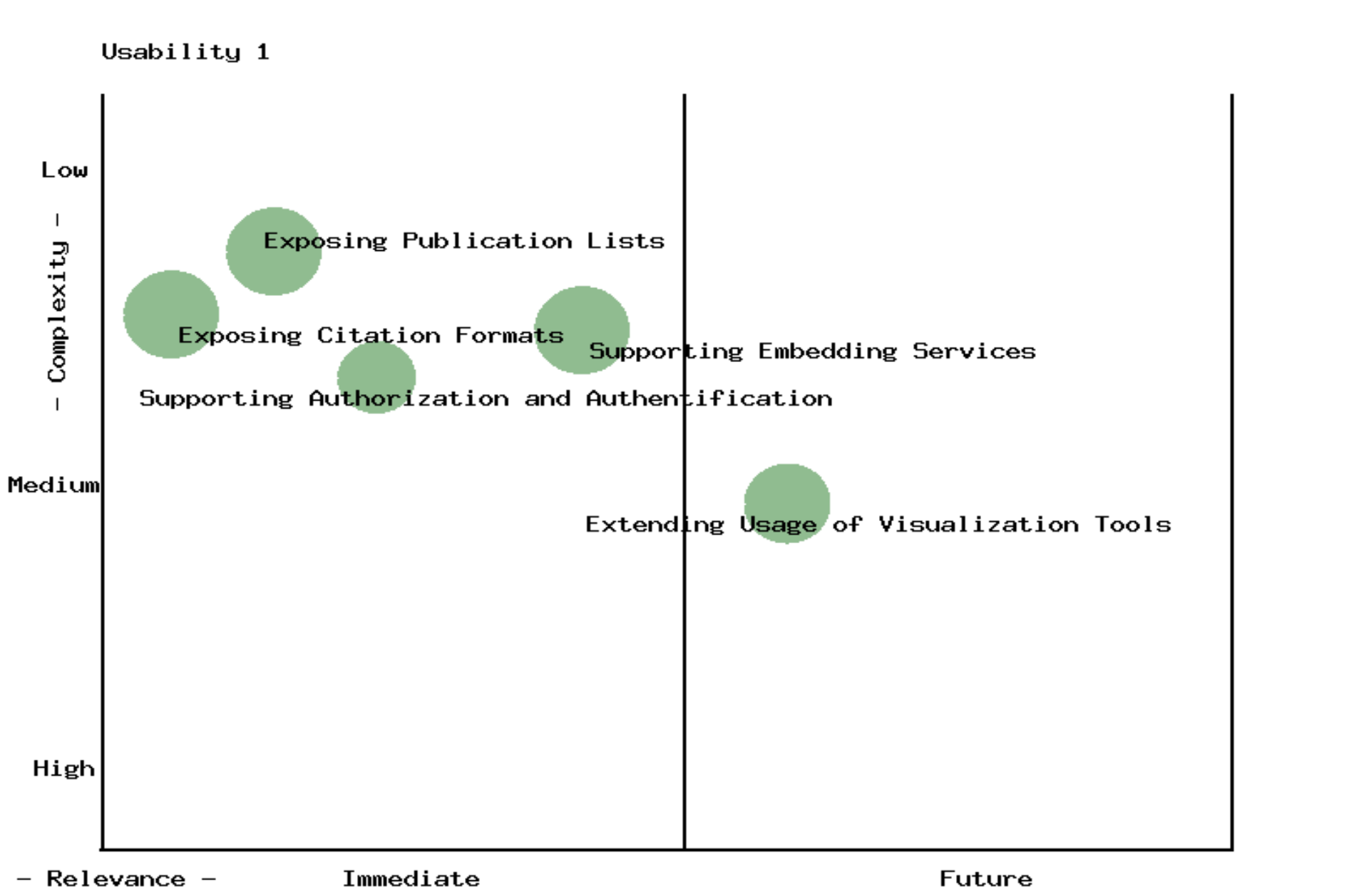
	Short term	Medium term	Long term
Low Complexity	• Exposing Citation Formats • Supporting Data Export Functions • Supporting Author Identification Systems • Supporting Search Engine Optimization (SEO) • Exposing Publication Lists • Integrating Different Persistent Identifiers	• Exposing Persistent Identifiers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

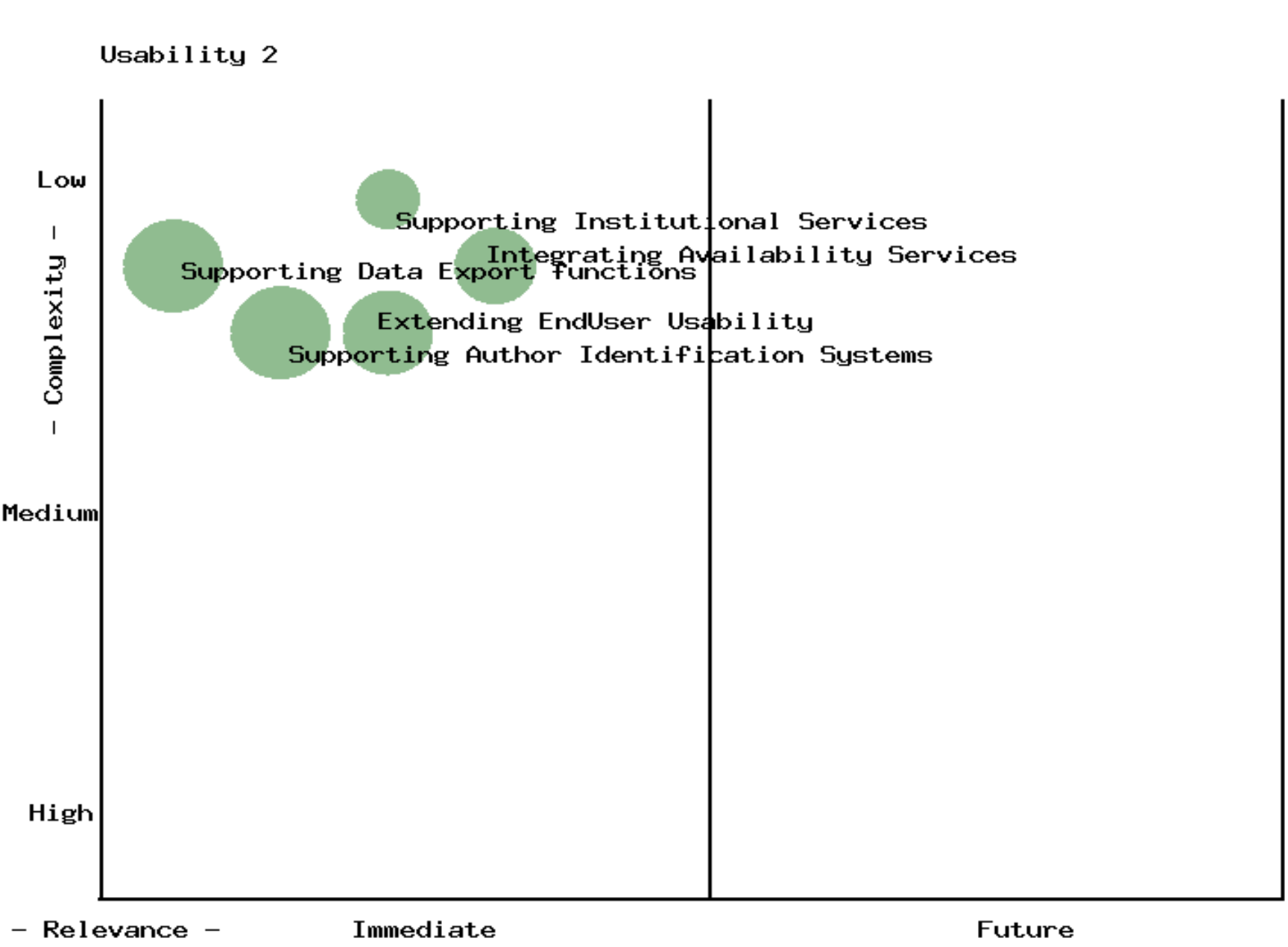
Moderate Complexity	• Exposing Bibliometric Information	• Exposing Versioning Information • De Duplication • Improving Registry Infrastructure • Monitoring Open Access Mandate Compliance •	
High Complexity	• Exposing Usage Statistics • Supporting Additional Metadata Format(s)	• Publication of Research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

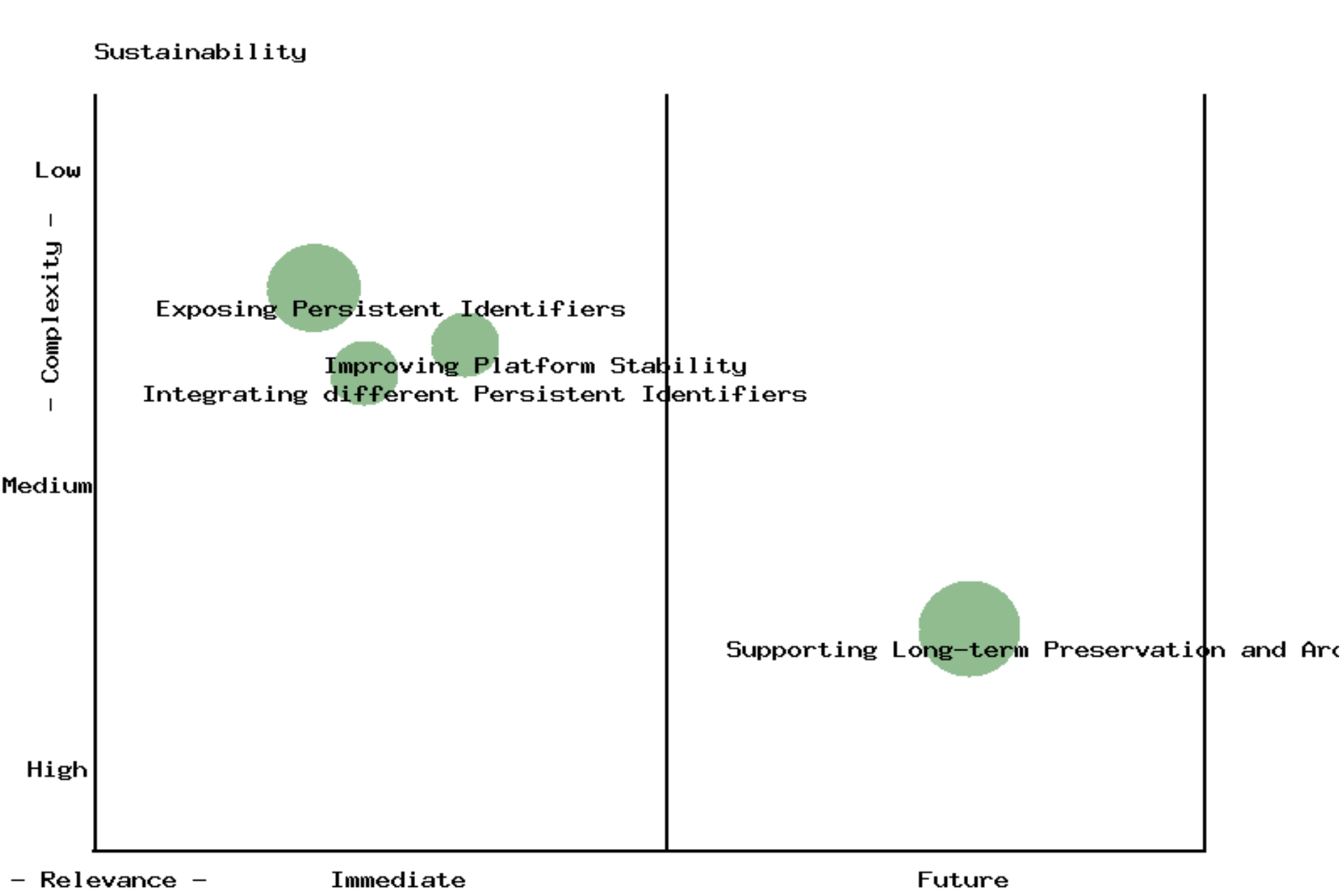

Impact and Visibility
Supporting Visibility in Repository Registries
Supporting Search Engine Optimization (SEO)
Supporting Repository Ranking Systems
Improving Registry Infrastructure
Exposing Usage Statistics
Exposing Bibliometric Information
Low
- Complexity -
Medium
High
- Relevance -
Immediate
Future

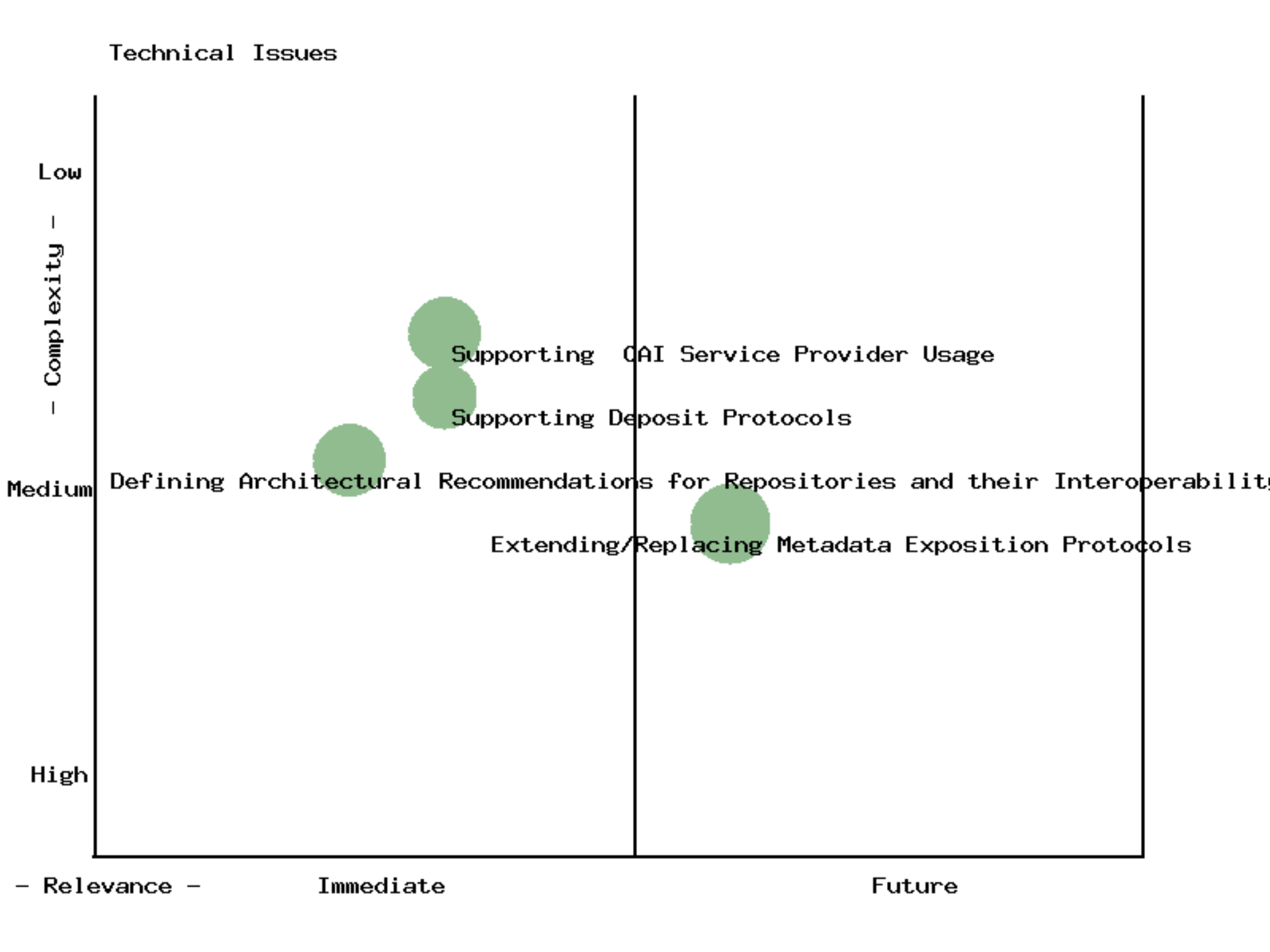

Data Issues
Low
- Complexity -
Medium
High
Exposing Versioning Information
Monitoring Open Access Mandate Compliance
Publication of Research Data
Supporting additional Metadata format(s)
Improving Metadata Quality (Data Curation)
Supporting Linked (Open) Data
Supporting Enhanced Publications
Handling 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 CAI 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.