

SC8DI07171 D04.02.01.B: Common approach for the management of persistent URIs by EU institutions

Towards a common policy for the governance and management of persistent URIs by EU institutions

Study Report



Document Metadata

Property	Value
Date	2014-11-20
Status	for acceptance
Version	0.53
Authors	Makx Dekkers – AMI Consult Stijn Goedertier – PwC EU Services Nikolaos Loutas – PwC EU Services
Reviewed by	Pieter Breyne – PwC EU Services
Approved by	Willem Van Gemert – Publications Office of the EU Vassilios Peristeras – European Commission, DG Informatics Athanasios Karalopoulos - European Commission, DG Informatics

This study was prepared for the ISA Programme by:

PwC EU Services

Disclaimer:

The views expressed in this report are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.

The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof.

Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.

All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.

Document History

Version	Date	Description	Action
0.01	2013-12-11	Template & TOC creation	Creation
0.02 – 0.07	2013-12-18 – 2014-01-30	Update	Update
0.08	2014-01-30	Internal review	Review
0.09	2014-02-03	Update	Update
0.10	2014-02-06	Delivered for review	Delivered
0.11-0.13	2014-02-12	Update, incl. interview notes from CCL	Update
0.14	2014-02-14	Feedback from 2 nd URI Task Force Meeting	Internal
0.15-0.17	2014-02-26	Inclusion of interview input, requirements listed, design principles updates, scope section (1.2) added	Internal
0.18	2014-02-27	Internal review (partial)	Review
0.19	2014-02-27	Added section 6 on persistent URI service, plus few more notes in section 4, some general tidying up	internal
0.20 – 0.21	2014-03-03	Internal review	Review
0.22	2014-03-04	Delivered for review	Delivered
0.23	2014-03-10	Added todos following the URI Task Force meeting	Update
0.23	2014-03-10	Review by DG SANCO	Review
0.24	2014-03-15	Review by Yves Laffont	Review
0.25	2014-03-21	Addressed feedback by SANCO and DIGIT	Review
0.26	2014-03-22	The analysis of the current situation was moved to Annex I	Update
0.27 – 0.28	2014-03-24 – 2014-03-29	Various updates throughout the document	Update
0.29	2014-03-29	Added specification for the governance of URI sets	Update
0.30	2014-03-30	Added specification of the Persistent URI redirection service. Restructured the URI design rules.	Update
0.31	2014-03-31	Fixed some examples on the URI design rules	Update

Study on a common policy for the governance and management of persistent URIs by EU institutions

0.31	2014-03-31	Delivered for review to the URI Task Force	Delivered
0.32	2014-04-04	Updated based on input received during the URI Task Force meeting of 4 April 2014.	Update
0.33	2014-04-23	Review comments by DG SANCO, DG COMM and Publications Office	Review
0.34	2014-04-29	Revised version based on received input.	Update
0.35	2014-04-29	Delivered for acceptance	Delivered
0.35	2014-05-05	Review comments by EC Central Library	Review
0.36	2014-05-07	Inclusion of table in conclusions to capture agreements and open questions	Update
0.37	2014-05-12	Inclusion of agreements and open questions identified during the 6 th URI Task Force meeting of 8 May.	Update
0.38	2014-05-21	Minor updates in the next steps following the meeting of 21 May with DIGIT, DG COMM and ISA.	Acceptance
0.38	2014-06-06	Comments by Vassilios Peristeras	Review
0.39	2014-06-06	Comments addressed	Update
0.39	2014-06-06	Delivered for acceptance	Delivered
0.39	2014-06-10	Version formally accepted	Accepted
0.39	2014-06-23	Version carried further into SC8DI07171	Update
0.40	2014-06-23	Updated executive summary	Update
0.40	2014-06-25	Review by Willem Van Gemert	Review
0.41	2014-06-25	Updated time plan of next steps	Update
0.42 – 0.48	2014-07-04 through 2014-08-21	Analysis of alternatives	Update
0.49	2014-09-24	Reorganisation of document and inclusion of accepted proposals for the alternatives	Update
0.50	2014-09-26	Updated executive summary description of the persistent URI redirection service	Update
0.51	2014-10-08	Updated based on review comments by the Publications Office	Update
0.52 – 0.53	2014-11-18 – 2014-11-20	Resolved feedback from Task Force	Update

Table of Contents

EXECUTIVE SUMMARY	1
PROPOSED PERSISTENT URI POLICY	6
BACKGROUND AND OBJECTIVES	6
GOVERNANCE	6
URI PATTERN.....	7
URI PERSISTENCE.....	8
URI REDIRECTION SERVICE	8
1 INTRODUCTION	9
1.1 DEFINITION: UNIFORM RESOURCE IDENTIFIER (URI).....	9
1.2 CONTEXT: APPLICATION AND DATA INTEGRATION.....	11
1.3 THE NEED FOR A PERSISTENT URI POLICY	12
1.4 OBJECTIVES.....	13
1.5 APPROACH.....	13
1.6 GLOSSARY.....	13
2 BUSINESS CASE	16
2.1 STAKEHOLDER REQUESTS	16
2.2 EXISTING PROBLEMS AND NEED.....	17
2.3 POSSIBLE SOLUTION.....	18
2.4 EXPECTED BENEFITS	19
2.5 ESTIMATED COSTS	20
2.6 RISKS.....	20
3 GOVERNANCE AND MANAGEMENT OF PERSISTENT URI SETS	22
3.1 STAKEHOLDER REQUESTS	22
3.2 CANDIDATE HOSTS FOR URI GOVERNANCE	23
3.3 SCOPE CRITERIA	24
3.4 GOVERNANCE STRUCTURE.....	25
3.5 URI MANAGEMENT PROCESS.....	28
3.5.1 <i>Request a URI namespace</i>	<i>28</i>
3.5.2 <i>Request (co-)ownership of a URI namespace</i>	<i>29</i>
3.5.3 <i>Monitor URI namespaces.....</i>	<i>30</i>
3.5.4 <i>Monitor, evaluate, and improve the URI policy.....</i>	<i>30</i>
3.5.5 <i>Raise awareness on the URI policy</i>	<i>30</i>

**Study on a common policy for the governance and management of
persistent URIs by EU institutions**

4	DESIGN RULES FOR PERSISTENT URI SETS	31
4.1	STAKEHOLDER REQUESTS	31
4.2	FOLLOW THE PATTERN.....	33
4.2.1	<i>http://</i>	33
4.2.2	<i>{subdomain}</i>	33
4.2.3	<i>Europa.eu</i>	33
4.2.4	<i>{namespace string}</i>	33
4.2.5	<i>{local id}</i>	35
4.2.6	<i>{version}</i>	35
4.2.7	<i>{language}</i>	35
4.3	AVOID STATING OWNERSHIP OR BRANDING	35
4.4	REUSE EXISTING IDENTIFIERS.....	36
4.5	USE AUTO INCREMENT WITH CARE	36
4.6	BE CAREFUL WITH VERSION NUMBERS	37
4.7	AVOID USING QUERY STRINGS	37
4.8	BE CAREFUL WITH FILE EXTENSIONS.....	37
4.9	TREAT LANGUAGES AS VERSIONS	38
4.10	CHARACTER ENCODING AND MULTILINGUAL CONSIDERATIONS.....	39
4.11	IMPLEMENT 303 URIs FOR REAL-WORLD OBJECTS.....	40
4.12	USE A DEDICATED SERVICE.....	40
4.13	CONSIDER A FALL BACK SCENARIO	40
5	URI REDIRECTION SERVICE.....	41
5.1	USE CASE SCENARIO.....	41
5.2	STAKEHOLDER REQUESTS	42
5.3	EXISTING SOLUTIONS.....	43
5.4	HIGH-LEVEL SPECIFICATION	45
5.4.1	<i>Domain model</i>	45
5.4.2	<i>Use cases</i>	46
5.4.3	<i>Sequence diagrams</i>	47
6	CONCLUSION AND NEXT STEPS.....	49
6.1	AGREEMENTS	49
6.2	NEXT STEPS.....	50
7	ACKNOWLEDGEMENTS	51
8	BIBLIOGRAPHY.....	52
ANNEX I.	INTERVIEW QUESTIONS	54

ANNEX II.	ANALYSIS OF THE CURRENT SITUATION	56
II.1.	INTER-INSTITUTIONAL EDITORIAL COMMITTEE FOR INTERNET (CEIII)	56
II.2.	URI SETS MAINTAINED BY THE COUNCIL OF THE EUROPEAN UNION	57
II.3.	URI SETS MAINTAINED BY DG COMM	57
II.4.	URI SETS MAINTAINED BY DG CONNECT	58
II.5.	URI SETS MAINTAINED BY DG DIGIT	58
II.6.	URI SETS ON THE EU OPEN DATA PORTAL (EU ODP)	59
II.7.	URI SETS MAINTAINED BY THE EUROPEAN ENVIRONMENT AGENCY (EEA)	61
II.8.	URI SETS MAINTAINED BY DG EUROSTAT	61
II.9.	URI SETS MAINTAINED BY DG HEALTH AND CONSUMERS (DG SANCO)	62
II.10.	URI SETS MAINTAINED BY THE JOINT RESEARCH CENTRE (DG JRC)	63
II.11.	URI SETS MAINTAINED BY THE PUBLICATIONS OFFICE (PO)	63
ANNEX III.	ANALYSIS OF ALTERNATIVES	67
III.1.	SCOPE	67
III.2.	OPERATIONAL MODEL	68
III.3.	CENTRAL VERSUS LOCAL RESPONSIBILITIES	69
III.4.	RELATIONSHIP NAMESPACE – REGISTER	72
III.5.	GOVERNANCE LEVELS AND REPRESENTATION	73
III.6.	RESPONSIBILITY FOR PROPOSALS	74
III.7.	CO-OWNERSHIP OF NAMESPACES	75
III.8.	DESIGN RULES AND URI PATTERNS	76

List of Tables

TABLE 1 – GLOSSARY	13
TABLE 2 – STAKEHOLDER REQUESTS: BUSINESS CASE	16
TABLE 3 – STAKEHOLDER REQUESTS: GOVERNANCE OF URI SETS	22
TABLE 4 – EXISTING GOVERNANCE BODIES	23
TABLE 5 – STAKEHOLDER REQUESTS: DESIGN PRINCIPLES FOR URI SETS	31
TABLE 6 – STAKEHOLDER REQUESTS: PERSISTENT URI SERVICE	42
TABLE 7 – EXISTING SOLUTIONS FOR URI MANAGERMENTS	43
TABLE 8: AGREEMENTS AND OPEN QUESTIONS	49
TABLE 9 - ORGANISATIONS AND THEIR REPRESENTATIVES IN THE URI TASK FORCE	51
TABLE 10 – OVERVIEW OF STAKEHOLDER MEETINGS	55
TABLE 11 – URI SET OF THE ADMS CONTROLLED VOCABULARIES	59

Study on a common policy for the governance and management of persistent URIs by EU institutions

TABLE 12 – INSPIRE REGISTRY DATA SETS.....	63
TABLE 13 – URI SETS OF NALS.....	65
TABLE 14 – ANALYSIS OF ALTERNATIVES: WHAT TO DO WHEN THE LOCAL REGISTER IS ABSENT?.....	67
TABLE 15: ANALYSIS OF ALTERNATIVES: OPERATIONAL MODEL	68
TABLE 16 – ANALYSIS OF ALTERNATIVES: DIFFERENT GOVERNANCE STRUCTURES	69
TABLE 17 – ANALYSIS OF ALTERNATIVES: RELATIONSHIP NAMESPACE – REGISTER.....	72
TABLE 18 – ANALYSIS OF ALTERNATIVES: COMPOSITION OF THE GOVERNANCE BODIES	73
TABLE 19 – ANALYSIS OF ALTERNATIVES: RESPONSIBILITY FOR PROPOSALS	74
TABLE 20 – ANALYSIS OF ALTERNATIVES: CO-OWNERSHIP OF A URI NAMESPACE	75
TABLE 21 – ANALYSIS OF ALTERNATIVES: SUBDOMAIN	77
TABLE 22 – ANALYSIS OF ALTERNATIVES: URI PATTERN	78
TABLE 23 – ANALYSIS OF ALTERNATIVES FOR {/COLLECTION*}	81
TABLE 24 – ANALYSIS OF ALTERNATIVES FOR {TYPE}.....	82
TABLE 25 -- ANALYSIS OF ALTERNATIVES FOR {CONCEPT}	83
TABLE 26 – ANALYSIS OF ALTERNATIVES: URIS FOR DOIs	84

List of Figures

FIGURE 1: CENTRAL REGISTER OF URI NAMESPACEs – LOCAL REGISTERS OF RESOURCES	26
FIGURE 2: URI REDIRECTION SERVICE: CONCERNS OF THE THREE MAIN ACTORS.....	46
FIGURE 3: SEQUENCE DIAGRAM: EXTERNAL HTTP REDIRECTION (FORWARDING)	48
FIGURE 4: SEQUENCE DIAGRAM: INTERNAL HTTP REDIRECTION (PROXY FORWARDING)	48
FIGURE 5: SCREENSHOT: INFORMATION PROVIDERS GUIDE FOR EUROPA WEBSITES	57
FIGURE 6: EU OPEN DATA PORTAL (ODP): PROPOSED TEMPLATE FOR UNIFORM RESOURCE IDENTIFIERS (URIs) [BARGIOTTI, KÜSTER, & SCHMITZ, 2013]	60

EXECUTIVE SUMMARY

This report elaborates on a proposal by an informal, inter-institutional Task Force¹ on a common approach for the management of persistent, HTTP-based, Uniform Resource Identifiers (HTTP URIs) by EU institutions. The work is supported by the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission as part of its Action 1.1 on semantic interoperability.

In Chapter 1, an **HTTP URI** is defined as a compact sequence of characters that identifies a resource and that follows the HTTP URI scheme. HTTP URIs can be used both as an identifier to identify physical and abstract resources and as a link to get (information about) a resource. HTTP URIs are becoming increasingly important in the context of **application integration** and **data integration**. To unlock positive network effects via the use by third parties, HTTP URIs should abide among others the following principles:

1. **Persistence:** an HTTP URI should be a *permanent* identifier for a resource, it should be stable and not change or vanish over time;
2. **Common use:** different organisations or systems should commonly use *the same* HTTP URIs;
3. **Technology neutral:** it must be possible to obtain information about a resource in *various formats*; and
4. **Authoritativeness:** information about resources should come from *authentic sources*.

An example of positive network effects related to the use of HTTP URIs is the fact that the Archives of the Council of the European Union already rely on the persistence of CELLAR URIs and the NLEX identifier of the Official Journal. It therefore no longer needs to store a printed copy of the Official Journal, and only keeps an HTTP URI as a reference in its archives.

Chapter 2 elicits a **business case** for a common approach to the governance and management of persistent HTTP URIs by EU institutions. It enumerates existing problems, suggests a possible solution, and states the expected benefits:

- **Problem:** EU institutions, as well as third party public and private users, will only start relying on and using URIs minted by other institutions if there is a **credible persistent URI policy** behind it with sufficient service level guarantees, including guarantees for long-term persistence, resolvability, response times, information quality, etc. An analysis of the current situation (see Annex II) reveals that such a URI policy is currently lacking. The following problems can be detected:
 - **Changing namespaces and URIs:** URIs for important resources, such as official documents of the EU, are not always stable, due to organisational change or application renewal;

¹ The organisations and their representatives in the URI Task Force are listed in the acknowledgement on page 48.

- **Synonyms (duplicate identifiers):** there is a risk that two or more EU institutions mint overlapping URI sets (synonyms) to identify the same or similar resources;
 - **Homonyms:** the risk of homonyms – the use of the same HTTP URI to denote different concepts – can happen when a local identifier is being used to denote different concepts at different points in time;
 - **Several non-HTTP identifiers already exist:** For some resources, EU institutions already maintain non-HTTP URIs, such as Digital Object Identifiers, UUIDs, etc. These URIs could be mapped to HTTP URIs. However, what is currently lacking is a clear guidance on how this can be done;
 - **Fragmented namespaces (no design rules):** Currently EU institutions can mint their own URI sets, as there are no design rules for URI sets beyond the Website URL. This has led to a fragmentation of namespaces for URIs, many of which are not on the Europa.eu domain;
 - **Lack of service-level guarantees:** There is no commitment to service levels by the institution, still less a uniform service levels that they can aspire and adhere to;
 - **Lack of understanding of terms:** Often, the terminology leads to misunderstanding when people talk about different things using the same word or term with different meanings; and
 - **Ambiguous identifiers:** Because query strings in URLs often allow the search parameters to be in random order, such URLs can lead to ambiguous identifiers (synonyms).
- **Solution:** A possible solution lies in a **common persistent URI policy** aiming at improving persistence and service levels on HTTP URIs for high-value resources and encouraging the use of HTTP URIs for application and data integration. The common policy could be comprised of:
 1. **A common inter-institutional governance and management of URIs:** an inter-institutional URI management body with roles, responsibilities, and a decision mechanism;
 2. **Common design rules for persistent URI sets:** common rules and guidelines for the design of persistent URI sets by EU institutions; and
 3. **A persistent URI redirection service for the europa.eu domain:** a central Web service providing redirection and content negotiation mechanisms for persistent URI namespaces. This service would be responsible for the registration and management of persistent URI namespace strings and the forwarding of HTTP requests (**URI redirection**) towards the local register.

- **Benefits:** A persistent URI policy will lead to a situation where EU institutions start using each other's URIs and resources rather than duplicating similar information. This may bring the following benefits:
 - **Harmonisation and interoperability:** Due to the increased use of common identifiers for the same resources, there are less interoperability conflicts, hence fewer costs involved in the integration of data and applications;
 - **Disambiguation:** using URIs as common identifiers which resolves to a commonly agreed definition can avoid misunderstandings;
 - **Increased awareness:** EU institutions become better aware of the registers that are operated by other EU institutions;
 - **Speed and efficiency:** Setting up persistent URIs will take less time as EU institutions can make use of a common infrastructure (the Persistent URI redirection service) for registering a particular namespace for their register;
 - **Flexibility:** The Persistent URI redirection service makes it easier to put in practice a common URI policy for EU institutions while providing sufficient flexibility to local registers to manage their own resources.;
 - **Monitoring of service levels:** the service levels for dereferencing (resolving) URIs will be centrally monitored; and
 - **Clarification of roles:** EU institutions will be able to use this service that provides a clear set of rules for its use. This avoids overlaps in case of positive competence conflicts.

Chapter 3 proposes a solution for the **governance and management of persistent URI sets**. It defines scope criteria to identify resources which foremost merit persistent URIs. These scope criteria include:

1. **Authoritative source:** URI sets for resources for which an EU institution is the *authoritative source*.
2. **Commitment of persistence:** URI sets for resources for which an EU institution has an *obligation* or strong *commitment* to maintain information about them over a longer period of time.
3. **Inter-organisational character:** URI sets for resources that are relevant in the information exchange between organisations.
4. **Machine-readable information:** URI sets for information resources and non-information resources (e.g. places, staff, buildings) for which information *about them* can be meaningfully processed by machines.
5. **Existing register:** URI sets for which an EU institution maintains an official register and enforces local control over the identifying mechanism

Examples of resources that meet the aforementioned criteria are data models (e.g. the INSPIRE data specifications), reference data (e.g. the EuroVoc thesaurus), registers (e.g. the staff register), documents (e.g. Official Journal), and high-value datasets (e.g. the data of the Financial Transparency System).

Subsequently, Chapter 3 proposes a governance mechanism for persistent URI sets. It assumes the existence of one single register of URI namespaces under central governance and many registers of resources under local governance. As illustrated in the figure below, only the left-hand elements in the URI path are centrally governed. Lower levels (right-hand side elements) of the URI can be decided by the administration responsible for a certain subject matter or area. This has the advantage of a central decision-making entity that fosters consistency, persistence and manageability of *URI namespaces*, but allows for flexibility for registries of *resources* such that they can manage these resources without bottlenecks – i.e. it does not require information on resources to be stored in a central data store nor does it require local identifiers (local ids) to be attributed centrally. This additionally allows the minting of URIs at those places where the actual expertise exists. The chapter also sketches the high-level URI management processes, of which the **registration of a new URI namespace** on the Europa.eu domain is a key control point to ensure that the scope criteria, URI design rules and service level guarantees are met.



Chapter 4 outlines a number of **design rules for persistent URI sets**. Only new URI sets that are placed under common governance – and thus meet the scope criteria mentioned above – should follow the design rules. These design rules can be summarised as follows:

- Follow the pattern;
- Avoid stating ownership or branding;
- Reuse existing identifiers;
- Use auto-increment with care;
- Be careful with version numbers;
- Avoid using query strings;
- Be careful with file extensions;
- Treat languages as versions;
- Character encoding and multilingual considerations;
- Implement 303 URIs for real-world objects;
- Use a dedicated service; and

- Consider a fall-back scenario.

Chapter 5 contains the functional specifications of a **URI redirection service** responsible for the registration and management of persistent URI namespaces and the forwarding of HTTP requests (**URI redirection**) towards the local register. This is achieved by having a single domain root like uri.europa.eu and subsequent path elements of the URI namespace. The mechanism of URI redirection caters for both persistence and flexibility: should the underlying register change location, persistence of the URIs can be guaranteed by updating redirection rules configured for a particular URI namespace.

Chapter 6 concludes the study summarising on which topics in this report a consensus was reached among the members of the informal URI Task Force through a number of (virtual) meetings from January through November 2014.

To finalise this consensus, the following next steps are proposed:

- **Conduct pilots:** DG Employment is working together with the Publications Office to put in place the first persistent URI namespace for the European classification of Skills, Competences, Qualifications and Occupations (ESCO) on the URI namespace 'http://data.europa.eu/esco/'. Furthermore, DG SANCO, the Publications Office, and DIGIT are planning to conduct other pilots for health and consumer data, the Core Vocabularies, the European Legislation Identifier (ELI) and Digital Object Identifiers (DOIs);
- **Inform key stakeholders:** The study and proposed persistent URI policy will be presented to key stakeholders. This includes a presentation of Proposal for URI policy to Management Committee of Publications Office. This also includes further promotion and awareness raising activities in relevant fora;
- **Inter-service consultation:** an inter-service consultation will be launched at European Commission; and
- **Continuation of the study:** In preparation to operationalizing the policy a set of guidelines and best practices for persistence will be drafted. Furthermore, the study will be continued to draft short technical specifications for implementers, and identify additional patterns for structuring the local part persistent URIs based on best practices.

PROPOSED PERSISTENT URI POLICY

Background and objectives

The proposed **Persistent URI Policy** defines a common governance approach, a common URI design pattern, and a central URI redirection service for resources maintained by the European institutions.

The proposal is the outcome of an activity undertaken by an informal, inter-institutional Task Force on a common approach for the management of persistent, HTTP-based, Uniform Resource Identifiers (HTTP URIs) by EU institutions, supported by the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission as part of its Action 1.1 on semantic interoperability.

The proposed policy comprises three key aspects:

1. **Co-ordination** takes place on a **central level**, while **responsibility for identification and persistence of resources** lies on the **local level, i.e. at the institution that maintains the resources**. Assignment of URIs is done on the level of collections of resources that are relevant for external and internal re-use; the central co-ordination does not cover the individual resources in a collection.
2. **A common pattern for URIs** is established for the resource collections that fall under the policy.
3. **A central URI redirection service** responsible for the registration and management of persistent URI namespace strings and the forwarding of HTTP requests.

Governance

The main principle underlying the governance is that co-ordination takes place on the central level, while the identification of resources that fall under the common policy and the persistence of those resources is a local responsibility.

The proposed governance structure consists of three levels:

- The **URI Steering Committee**, consisting of representatives from all EU institutions and providing oversight to the work of the URI Committee. The URI Steering Committee decides on higher-level issues like updating of the policy or handling of special cases.
- The **URI Committee**, consisting of staff of one or more EU institutions that are dedicated to take decisions on URIs under this policy. The URI Committee reports to the URI Steering Committee on a regular basis and when special cases are encountered that cannot be handled on the basis of the existing policy.
- The **Technical Team**, consisting of staff of one or more EU institutions that are dedicated to perform the day-to-day operational tasks, implementing the decisions of the URI Committee.

This structure ensures that the implementation of the policy is light-weight while providing the opportunity to all EU institutions to jointly exercise oversight.

Any EU institution can request a persistent URI namespace string for a collection of resources and thereby place this collection of resources under the common URI policy. A set of **scope criteria** are used in the process of determining whether a collection falls under the policy:

- The institution that is responsible for the collection must be the **authoritative source** for the collection, i.e., the institution is the primary maintainer of the collection;
- The institution must have a **commitment of persistence** for the collection, i.e. the institution commits to ensuring that the assigned URIs will persistently resolve to the resources to which they are assigned;
- The collection must have an **inter-organisational character** in the sense that the resources are relevant for usage that spans more than one organisation;
- The resources must be associated with **machine-readable information** so that resolution of URIs leads to a response with data; and
- The institution must already have an **existing register** in which the resources are formally maintained, so that the maintenance of the collection is ensured.

URI pattern

The general pattern for URIs under the proposed policy is as follows:

<http://data.europa.eu/{namespace string}/{local id}>

The main aspects of this pattern are:

- All URIs under the policy are defined in the subdomain data.europa.eu;
- The namespace string is assigned by the central maintenance agency upon request and after evaluation of the scope criteria;
- The namespace string is an opaque string, i.e. a string of characters that has no intended meaning;
- In exceptional cases, when a collection already has an externally well-known acronym or when existing identifiers for the resources are based on a formal identifier standard, a meaningful string may be requested and assigned;
- The local institution is free to design and assign the local part (local id); best practice guidelines for the local part will be provided centrally.

URI persistence

URIs that have been assigned should be persistently maintained. This means in particular that a namespace string, either opaque or meaningful, has to remain valid for as long as there are references to the resources identified by the URIs in the namespace. In other words, once a namespace string has been assigned, it cannot be changed later.

The general rule is: **A URI that identifies a particular resource must permanently resolve to that same resource as long as there are references to that resource.**

URI redirection service

On the subdomain data.europa.eu, a central URI redirection service is installed that allows for URI redirection on the level of namespace strings. For each namespace string, redirection is configured using one or more redirection rules, for example:

```
http://data.europa.eu/1234/* redirects to  
http://publications.europa.eu/resource/authority/country/*
```

These redirections also enable suitable HTTP responses to be generated, e.g. 303 redirects for real-world entities.

The creation and maintenance of these redirection rules is initially done by the Technical Team, but can be distributed to local administrators in the future through a user-friendly maintenance interface.

1 INTRODUCTION

This report summarises the findings of a study on a common approach for the management of persistent Uniform Resource Identifiers (URIs) by EU institutions. It contains a business case, specifications for the governance and management of persistent URI sets, design rules for persistent URI sets, and functional specifications for a URI redirection service. The study was commissioned by the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission as part of its Action 1.1 on semantic interoperability. The study was conducted in collaboration with an informal, inter-institutional Task Force (the organisations and their representatives in the URI Task Force are listed in the acknowledgement on page 51), and builds on a previous proposal on persistent URIs prepared in 2012 – 2013 by the same Task Force [URI Task Force, 2013].

1.1 Definition: Uniform Resource Identifier (URI)

The Internet Engineering Task Force (IETF) provides the following definitions:

A **Uniform Resource Identifier** (URI) is a compact sequence of characters that identifies an abstract or physical resource (IETF [RFC3986](#)).

URIs are specified as being written only in ASCII characters, i.e. the Roman alphabet with no accents plus a few punctuation symbols. For reasons given later (section 4.10), the advice is that only ASCII characters, i.e. only URIs, should be used as persistent identifiers. It should be noted however that International Resource Identifiers, IRIs, can include a much broader range of characters including accented Roman characters, Greek and Cyrillic characters. The IETF defines them as follows:

An **International Resource Identifier** (IRI) is a sequence of characters from the Universal Character Set (Unicode/ISO 10646) characters that identifies an abstract or physical resource (IETF [RFC3987](#)).

URIs and IRIs can identify anything, including:

- **Web documents:** The most familiar related term is probably URL, Uniform Resource Location, which identifies a specific information resource on the World Wide Web, such as a Web page, a PDF document, an image etc. URLs all begin with `http://`.
- **Real-world things:** The more general term, Uniform Resource Identifier (URI), can identify *anything* whether it's something on the Web or not. As formulated in the W3C Note²: “*URIs identify not just Web documents, but also real-world objects like people and cars, and even abstract ideas and non-existing things like a mythical unicorn. We call these real-world objects or things.*”

² Cool URIs for the Semantic Web, <http://www.w3.org/TR/cooluris>

URIs and IRIs *can* take many forms. All classic numbering schemes, including ISBN numbers, DOIs or ISO numbers can be expressed using a URI scheme. URIs thus are broad and all encompassing. When used in the context of Linked Open Government Data URIs and IRIs almost always refer to identifiers that, like URLs, begin with `http://`.

An **HTTP URI is a URI that follows the HTTP scheme (RFC2616)**.

The most important feature of an **HTTP URI** is that it can be looked up (an action also known as *dereferencing* the URI): you can click it or put it into a Web browser and see what information it returns. If the URI identifies something like a Web page, then this is what will appear in the browser. But, what if the URI identifies a court, a lake or an abstract concept like 'the president?' In that situation, when a URI is looked up, the user will be redirected automatically to a document that *describes* the thing identified. For example:

<http://business.data.gov.uk/id/company/04285910>

identifies a print finishing company called Apple Binding. If a user dereferences that URI, s/he will be redirected (automatically) to a slightly different URI

<http://data.companieshouse.gov.uk/doc/company/04285910>

Notice that the `/id/` element in the first identifier has become `/doc/` in the second one which identifies a document describing the company. The domain name has changed too. Of the two, **the first URI is the important one because it identifies the thing itself** (in this case a company) which can always be looked up even if, in future, for whatever reason, it is configured to redirect to a different document describing it at a different location on the Web.

This highlights the **first important principle** of managing URIs: **persistence**. Developers of software that use those identifiers need to be confident that the identifier will persist, that is, it will remain unchanged for as long as the software is likely to remain in use. If so, then the software can use those URIs online without having to download and keep a local copy that might quickly go out of date.

The **second important principle** is that multiple datasets/data sources/applications **refer to the same resources using the same identifiers**. When this happens, each new dataset/ data source/application that uses the same URIs adds to the common pool of data and each individual dataset becomes more valuable as a result. This is a manifestation of the *network effect* most famously demonstrated by the development of the telephone. A single telephone is not much use, but as each new phone is connected to the network the value of the existing phones increases as they become more useful.

From a technical point of view we say that URIs have *global scope* – the same URI always identifies the same thing wherever you see it. As more and more people use and rely on the same set of URIs, the more important and valuable it becomes.

Finally, applications can usually access the data from a URI in any number of formats and languages. If you look up

<http://business.data.gov.uk/id/company/04285910>

using a normal Web browser you will get back an HTML Web page. However, that is not the only option. Browsers are just one example of what are known more generally as User Agents – that is, tools used by individuals to access information on the Web. There are many tools and applications besides the well-known Web browsers that do this and they consume data in different formats. Whenever a user agent makes a request for a Web resource it includes 'accept headers' that define what types of data it can process as well as the user's language preferences. Servers can be configured to return data in different formats and different languages depending on the values of the accept headers. In this way, a user agent that can only process, say, XML, can look up *exactly the same identifier* as above and receive XML data, not an HTML Web page. The data from UK Companies House is available in RDF/XML, JSON, XML, HTML, CSV and YAML, any of which will be returned by a user agent that asks for it from a URI like the example used. Similar processes can be used to make different translations of the same document available from the same URI.

This process, known as content negotiation, is what provides the **third important principle** of URIs: they should be **technology neutral**. A common way to override the content negotiation and access the data in a specific format using a Web browser, just add the appropriate file extension to the URI.

A **fourth principle**, relevant in the context of inter-institutional information exchange, relates to the **authoritativeness** of the data that is obtained by resolving URIs. EU institutions will only rely on data that comes from an authentic source.

1.2 Context: application and data integration

HTTP-based URIs are becoming increasingly important in the context of **application integration** and **data integration**.

1. **Application integration:** HTTP URIs are an important aspect of Representational State Transfer (ReST) design principles. Representational State Transfer (ReST) is a simple and scalable paradigm for Web service-based software architectures used for integrating large-scale Web applications of global Service Providers. It includes the following principles [Rodriguez, 2008] [Fielding, 2000]:
 - The use of HTTP URIs to identify resources on the Web;
 - The use of meaningful, *hackable*³, URIs with a directory-like structure;
 - Use of HTTP methods (post, get, put, delete) to manipulate resources (create, read, update, delete);
 - Avoiding the use of *verbs* in URI structures;
 - Stateless Web service;

³ A URI is hackable when after removing the leaf part of the URI (the part after the last '/'), the remaining URI is still resolvable.

- The use of JSON or XML as a paradigm for data exchange.
2. **Data integration:** Similarly, HTTP URIs are also increasingly used for data integration according to the design principle of “Linked Data”. Linked Data is a way of identifying, linking and accessing information on the Web according to the four **design principles** put forward by Tim Berners-Lee [TBL, 2006]:
- Use URIs as names for things;
 - Use HTTP URIs so that people can look up those names;
 - When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL); and
 - Include links to other URIs, so that they can discover more things.

Please note that the Linked Data design principles are consistent with the aforementioned RESTful design principles. The distinction between application integration and data integration is also not a crisp distinction to make.

Both **RESTful Web Services** and **Linked Data** enable the provision of “**application services**” and “**data services**” and conceive of the Web, including private intranets, as an ecosystem where providers and consumers can interconnect and integrate services and disparate datasets. It converts the Web from a “Web of documents” into a “Web of interconnected services and data”. Applied to e-Government, RESTful Web Services and Linked Data have the potential to lead to smarter and more efficient government services and applications, and to foster creativity and innovation in the digital economy.

1.3 The need for a persistent URI policy

Both when used for application integration (for example, for RESTful web services) or for data integration (for example using Linked Data principles) it is important that URIs don’t change.

The **Business Models on Linked Open Government Data** [BM4LOGD] study describes why persistent URIs are so important. In the Linked Data ecosystem, public administrations provide data as an online **service** to data users (i.e. citizens, businesses and other public administrations). Instead of downloading and processing a whole dataset, Linked Data allows a data user to retrieve specific information about the entity of interest, by resolving its URI. The data is provided in different machine-readable formats, ready to be linked and used with other data. To support this value proposition, providers must have a **URI policy** that lays down the expected service levels of the Web service; long-term persistence being one of the most important service guarantees. European Institutions should make their URI policy explicit, so that users can rely on services with confidence and other data providers can link to these URIs or reuse them to denote identical concepts.

The use of URIs as common identifiers to identify identical concepts in disparate datasets is a prerequisite to unlock the **positive network effects** of LOGD.

1.4 Objectives

The main objectives of this study are to identify the requirements, costs, benefits, risks and high-level features for **a common approach for the management and governance of persistent URIs by EU institutions** in order to meet the needs of data providers and consumers.

1. **Business case:** What are the benefits, costs, and risks for a common approach for the governance and management of persistent URIs by EU institutions?
2. **Persistent URI governance and management:** Which roles, responsibilities, policies, operating principles and practices, and decision-making processes should EU Institutions create to ensure that persistent URI sets are managed in accordance with the overall objectives of the EU?
3. **Design principles:** What are the *minimal* design principles that should be imposed to persistent URI sets on the europa.eu domain?
4. **Persistent URI management software:** What are the requirements for a URI redirection service for the europa.eu domain?

1.5 Approach

In this study, we performed the following steps:

1. **Analysis of the existing situation.** First, we made an overview of the current situation for URI sets that are used for identifying things, this is included in Annex II of the study.
2. **Formulation of the business case.** In Section 3, we identify the problem, the proposed solution, and the expected benefits and project cost.
3. **Identification of the requirements for URI governance and management.** These requirements include the naming conventions for URIs.
4. **Identification of the requirements for a URI redirection service for the europa.eu domain.** In order to offer a URI redirection service for the europa.eu domain, we need to outline the functional requirements.

1.6 Glossary

The table below defines important terms and acronyms that are used throughout this report.

Table 1 – Glossary

Term	Description
CEiii	Inter-institutional Editorial Committee for Internet
DOI	Digital Object Identifier
EU Open Data Portal	Portal providing access to data produced by the institutions and

Study on a common policy for the governance and management of persistent URIs by EU institutions

	bodies of the European Union.
HTTP	Hyper Text Transfer Protocol
Information resource	According to the W3C document "Architecture of the World Wide Web, Volume One" ⁴ : A resource which has the property that all of its essential characteristics can be conveyed in a message. This usually applies to all resources that have a digital representation (such a Web page, digital image, text file, spreadsheet etc.)
Interoperability	According to the ISA Decision ⁵ , interoperability means the ability of disparate and diverse organisations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organisations, through the business processes they support, by means of the <i>exchange of data</i> between their respective ICT systems.
IRI	International Resource Identifier. A compact sequence of characters from the Universal Character Set (Unicode/ISO 10646) that identifies an abstract or physical resource (IETF RFC3987).
JSON	Java Script Object Notation
MDR	Metadata Registry of the Publications Office
mint URIs	The act of assigning a uniform resource identifier to a resource.
NAL	Named Authority Lists
Non-information resource	Term that has no formal definition, but is generally used to refer to resources that are not information resources (see above), such as 'real-world' entities like persons, organisations, places, languages, concepts etc. These resources themselves cannot be transmitted over a network connection, only information about them can be sent over the wire.
Open data	A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike (Open Knowledge Foundation Definition).
Registry (synonym: register)	A source of information under the control of an appointed entity.
Base Registry	A trusted, authentic source of information under the control of an appointed public administration or organisation appointed by

⁴ W3C. Architecture of the World Wide Web, Volume One. W3C Recommendation 15 December 2004. <http://www.w3.org/TR/webarch/#def-information-resource>

⁵ Article 2 (a) of the Decision No 922/2009/EC of the European Parliament and of the Council of 16 September 2009 on interoperability solutions for European public administrations (ISA) OJ L 260/20 (2009) <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:260:0020:0027:EN:PDF>

Study on a common policy for the governance and management of persistent URIs by EU institutions

(synonym: base register)	government. Base registries provide authentic sources of basic information on items such as persons, companies, vehicles, licences, buildings, locations and roads. (source: European Interoperability Framework ⁶).
RDF	Resource Description Framework, the technical standard from W3C that underpins Linked Data.
URI	Uniform Resource Identifier A Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or physical resource (IETF RFC3986).
URI management	We define URI management as the good practice of adopting policies, processes, and systems to plan, perform, evaluate, and improve the use and re-use of URI sets.
URI Governance	Governance comprises well-defined roles and responsibilities, cohesive policies and principles, and decision-making processes that define, govern and regulate the lifecycle of URI sets.
URI set	A collection of reference data published using URIs, about a set of related items, governed from a single source. An example of such a set could look like: http://europa.eu/id/policy/agriculture/corn/mon810 http://europa.eu/id/policy/agriculture/corn/mon810/prohibitions http://europa.eu/id/policy/agriculture/corn/mon810/permissions/France/FieldID#32927 All URIs concerning corn Monsanto 810 would be called a URI set
URL	A Uniform Resource Locator (URL) is the address of a specific Web site or file on the Internet.
URN	Uniform Resource Name

⁶ European Interoperability Framework, http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

2 BUSINESS CASE

This chapter elicits the problem, possible solution, costs, benefits, and risks for a common approach for the management and governance of persistent URIs by EU institutions.

2.1 Stakeholder requests

The stakeholder requests set out in the following table are derived from the interviews conducted as part of this study. For more details on current URI sets managed by these stakeholders see Annex II.

Table 2 – Stakeholder requests: business case

ID	Stakeholder request
B1	Problem – synonyms (duplicate identifiers): Some resources do not have a single owner, for example, data about forests are treated both by DG SANCO and DG AGRI, which effectively means that the two owners may mint different URIs for the same thing.
B2	<p>Problem – several:</p> <ul style="list-style-type: none"> • URIs change; • Several identifiers (URIs, DOI, etc.) already exist for certain resources. • Risk of homonyms: The same URI is assigned for different resources; • Human-users have little understanding on what a resource is about from just reading the URI; • Machines cannot infer relationships amongst resources based on the URI.
B3	Solution - URI Policy: the solution should consist of a common URI policy and enforcement mechanism that can ensure long-term persistence of HTTP URIs.
B4	<p>Solution – Persistence: Once a URI is created, it does not change. URIs cannot disappear or expire, even if the resource that they are denoting ceases to exist. In the latter case, the redirection can change and point to a new explanation about the resource not existing anymore.</p> <p>The impact of this requirement must be investigated further.</p>
B5	Solution – Governance: Clarify the roles and responsibilities in URI governance and maintenance of DGs OP, DIGIT, CNECT and COMM [URI Task Force, 2013].
B6	Solution – URI design rules: In 2012 – 2013, the URI Task Force [URI

	Task Force, 2013] recommendation was: <i>adopt a stable (five to ten years) common root and elements of URIs (e.g. open-data.europa.eu or data.europa.eu); and define a common Commission (or EU institutional) approach for minting URIs (URI root, branding, simplicity).</i>
B7	Solution - URIs for services: Cover also things such as Web service endpoints.
B8	Costs: the proposed solution entails a specific set of costs that should be included in the business case.
B9	Benefits: the use of common identifiers for resources can entail significant cost savings.
B10	Benefits: one additional benefit is that EU institutions become better aware of the registers that are operated by other EU institutions.
B11	Benefits: The benefits of interoperability and easy data integration should be explicitly mentioned.

2.2 Existing problems and need

HTTP URIs are becoming increasingly important in the information exchange between EU institutions. However, EU institutions, as well as third party public and private users, will only start relying and using URIs minted by other institutions if there is a **credible persistent URI policy** behind it with sufficient service level guarantees, including guarantees for long-term persistence, resolvability, response times, information quality, etc. An analysis of the current situation (see Annex II) reveals that such a URI policy is currently lacking. The following problems can be detected:

1. **Changing namespaces and URIs:** URIs for important resources, such as official documents of the EU, are not always stable. This happens due to organisational changes or application renewal. In these cases, the old URIs are no longer resolvable. As long as EU institutions do not have an explicit policy that prevents broken links, third-parties (including other institutions) will not rely on this.
2. **Synonyms (duplicate identifiers):** Currently there is no coordination among EU institutions upon the creation of URI sets. For some resources there may be more than one data owner among the EU institutions. Consequently, there is a risk that two or more EU institutions mint overlapping URI sets (synonyms) to identify the same or similar resources.
3. **Homonyms:** The risk of homonyms – the use of the same HTTP URI to denote different concepts – is less probable than the risk of synonyms, but not inconceivable. It can happen when a local identifier is being used to denote different concepts at different points in time; for example a code in a local register is given to a new, possibly unrelated resource.

4. **Several non-HTTP identifiers already exist:** For some resources, EU institutions already maintain non-HTTP URIs, such as Digital Object Identifiers, UUIDs, etc. These URIs could be mapped to HTTP URIs. However, what is currently lacking is a clear guidance on how this can be done.
5. **Fragmented namespaces (no design rules):** Currently EU institutions can mint their own URI sets within the bounds on website URLs set by Information Providers Guide for EUROPA websites⁷ and the Inter-institutional Editorial Committee for Internet (CEiii). There are **no design rules for URI sets** beyond the Website URL. The lack of coordination for URI sets does not give confidence to third-parties to use these URI sets. A common approach would provide guidelines that afford sufficient flexibility to allow URIs to be minted as required but with the guidance necessary to encourage URIs to be minted that can persist for the long term.
6. **Lack of service-level guarantees:** Currently EU institutions use their own sub-domains, virtual folders, etc. to make LOGD available. There is no commitment to service levels by the institution, still less a uniform service levels that they can aspire and adhere to. The lack of service-level guarantees has been identified as one of the major barrier to the reuse of LOGD [BM4LOGD].
7. **Lack of understanding of terms:** EC administration, but also other EU institutions have to talk to each other about sophisticated solutions to complex issues. Often, the terminology leads to misunderstanding when people talk about different things using the same word or term with different meanings.
8. **Ambiguous identifiers:** URLs used on the human readable Web do not always and by nature have all characteristics to serve as identifiers. They may be ambiguous. Query strings often do not take the order of the search terms into account. Depending on the implementation, most of the dynamic content can exchange the variables transported in the URL.
9. **Conway's law:** Conway's law states that organizations which design systems are constrained to produce designs which are copies of the communication structures of these organizations. As there is constant reorganisation, the persistence of URIs are endangered when using URIs that reflect organisational structure.

2.3 Possible solution

A possible solution lies in a **common persistent URI policy** aiming at improving persistence and service levels on HTTP URIs for high-value resources and encouraging the use of HTTP URIs for application and data integration. The common policy could be comprised of:

⁷ http://ec.europa.eu/ipg/index_en.htm

1. **Common inter-institutional governance and management:** an inter-institutional URI management body with roles, responsibilities, and a decision mechanism.
2. **Common design rules for persistent URI sets:** common rules and guidelines for the design of URI sets by EU institutions.
3. **URI redirection service for the europa.eu domain:** a central Web service providing redirection and content negotiation mechanisms for persistent URI namespaces. This service would be responsible for the registration and management of persistent URI namespaces and the forwarding of HTTP requests (**URI redirection**) towards the local register.

2.4 Expected benefits

The proposed solution could bring the following benefits:

- **Harmonisation and interoperability:** EU institutions (and other third parties) mint URIs according to harmonised URI design rules. Due to the increased service levels they can start reusing each other's resources rather than duplicating similar information. This can lead to considerable cost savings, as it will become much easier to integrate applications and data from different EU institutions. Due to the increased use of common identifiers for the same resources, there are less interoperability conflicts, hence fewer costs involved in the integration of data and applications.
- **Disambiguation:** If a term has a fixed URI and definition misunderstandings can be avoided by using the terms as defined. This will make discussion and information exchange more informed and will help understanding and manageability in complex situations.
- **Increased awareness:** EU institutions become better aware of the registers that are operated by other EU institutions. Friction between knowledge about an entity in one institution and the knowledge existing in another about the same entity can be reduced.
- **Speed and efficiency:** Setting up persistent URIs will take less time as EU institutions can make use of a common infrastructure (the Persistent URI service) for registering a particular namespace for their register;
- **Flexibility:** The Persistent URI service makes it *easier* to put in practice a common URI policy for EU institutions while providing sufficient *flexibility* to local registers to manage their own resources.
- **Monitoring of service levels:** the service levels for dereferencing (resolving) URIs will be centrally monitored; and
- **Clarification of roles:** EU institutions will be able to use this service that provides a clear set of rules for its use. This avoids overlaps in case of positive competence conflicts.

An example of the benefit of harmonisation and interoperability is the following. The Archives of the Council of the European Union already rely on the persistence of CELLAR URIs and the NLEX identifier of the Official Journal. Due to the fact that the archive can rely on the persistence of HTTP URIs, it no longer needs to store a

printed copy of the Official Journal, and only keeps an HTTP URI as a reference in its archives. Similar examples could be found from the Commission Historical Archives Service and document management systems.

2.5 Estimated costs

The following costs for developing a persistent URI policy for EU institutions have been identified:

- **Pre-study (this report):** effort from representatives in the URI Task Force + 30 man days support by external consultants;
- **Design & operation:** COMM made a rough estimate of cost factors to be considered if a common tool and governance for URIs is established:
 - **Establishment of URI design rules:** 1 FTE for the service responsible + other resources for horizontal services until completion (1 year);
 - **For the setup, development, maintenance, and promotion:** 1 FTE for the first year, or until launching + 1 FTE per year for daily management; it is assumed by this counts that the system will not be developed from scratch as e.g. the URL shortener of europa.eu already provides a good basis for a URI system. Development from scratch can be avoided.
 - **Dedicated server + software;**
 - **Governance and management:** ¼ - ½ FTE per year for governance and management by DGs and EU institutions.

The proposed solution assumes a central governance and management of the persistent URI namespaces and a local governance of the resources maintained in the local registers. In the assumption that a central, persistent URI service is deployed allowing URI forwarding, local registers do not incur significant *additional* costs.

2.6 Risks

The following risks exist:

- **Risk 1: Overly restrictive design principles for URI sets:** The limitations set by the design principles may hinder the creation (or maintenance) of new URI sets by EU institutions. In this case, institutions will not use persistent URIs or will try to circumvent the system.
Mitigation: the design principles should be evaluated on a systematic and periodic basis.
- **Risk 2: Complex governance mechanism slows down decision making:**
Mitigation: The design and provision of URIs should be as distributed as possible within a framework that offers the minimum amount of central control necessary to avoid clashes and lost URIs due to lack of permanence.

- **Risk 3: Persistent URI service** does not meet functional or non-functional requirements.

Mitigation: explicit requirements gathering process, prioritisation, and careful tool selection and/or evaluation.

- **Risk 4: Key stakeholders** do not acknowledge the need for a common approach.

Mitigation: ensure stakeholder buy-in through benefits realisation.

3 GOVERNANCE AND MANAGEMENT OF PERSISTENT URI SETS

This chapter describes the proposed governance mechanism for URI sets maintained by EU institutions.

3.1 Stakeholder requests

The stakeholder requests set out in the following table are derived from the interviews conducted as part of this study.

Table 3 – Stakeholder requests: governance of URI sets

ID	Stakeholder request
G1	Scope – criteria and list of resource types: the solution should consist of a set of criteria to determine whether a URI set is eligible for a common approach. In 2013, the URI Task Force also concluded that a list should be drafted of the types of resources that should identified by URIs, e.g. digital resources (datasets, documents, vocabularies, code lists), real-world entities (organisations, places, events), information artefacts (co-ordinate reference systems on maps, map layers, resource metadata) [Bhraonain & Georgiannakis, 2013].
G2	Scope – both public and private resources: The scope should cover both resources about which information is publicly accessible as well as resources about which information about them is only privately accessible. EU institutions need one approach for both. What is not-public today may be public tomorrow and vice versa.
G3	URI policy: the solution should consist of a common URI policy and an enforcement mechanism that can ensure long-term persistence of HTTP URIs.
G4	Efficiency via clear rules: most decisions on URI sets should be taken by applying clear and objective design rules and scope criteria. The use of a central system for the management of URI Sets could also facilitate enforcement and/or monitoring of the rules.
G5	Efficiency of the URI management process: Governance should guarantee a lean and agile URI management process.
G6	Alignment/harmonisation of URIs: the governance body should also encourage and require the creation of links to other existing URIs that are published by authoritative bodies <i>within</i> and <i>outside the EU institutions</i> . For example, creating alignment for the Named Authority Lists (NALs) ⁸ of the Publications Office of the EU with the authority files ⁹ of the US Library of Congress.
G7	Organisation structure – CEiii and IMMC: Existing inter-institutional governance bodies, primarily the CEiii but also the IMMC for structural metadata, could facilitate the decision making process.

⁸ <http://publications.europa.eu/mdr/authority/index.html>

⁹ <http://id.loc.gov/>

	Each recommendation should be accompanied by a thorough assessment of the recommendation's impact in terms of human, financial and technical resources. Regarding human resources the impact assessment shall contain a precise description of the skills set of the required profile(s). This goes beyond the scope of the present study.
G8	Organisation structure – central governance of the namespace, local governance of the identifier: the central common governance should only be about the left-hand side of the URI structure. EU institutions should have autonomy to manage URIs within the namespaces attributed to them.
G9	Responsibilities – process for continuous improvement: The URI management process should be monitored. If there is a need to have a namespace minted quickly the governance should not take too long. It is necessary to explore how to guarantee the quality and timeliness.
G10	Monitoring mechanism: A mechanism should be put in place to track both positive and negative consequences. If there is a rule that has a negative impact it should be removed or revised, based on feedback.

3.2 Candidate hosts for URI governance

The table below describes number of relevant inter-institutional governance bodies.

Table 4 – Existing governance bodies

Name	Description
EU Open Data Portal steering committee	<p>The EU Open Data Portal steering committee was founded in 2012 following the European Commission's reuse Decision¹⁰. Its role is to oversee the implementation of the EU Open Data Portal (http://open-data.europa.eu/). The EU Open Data Portal is the single point of access to the Commission's structured data in view of facilitating linking and reuse for commercial and non-commercial purposes. The data portal provides access as well to structured data of other EU institutions and bodies.</p> <p>The EU Open Data Portal steering committee is chaired by the Publications Office of the European Union. It consists of the Commission Secretariat-General, the Directorate-General for Communication, the Directorate-General for Communications Networks, Content and Technology, the Directorate-General for Informatics and several Directorates-General (JRC, ESTAT, ...) representing the data providers as well as other EU institutions including the European Parliament and the Council. Other institutions may be invited to join the committee at a later stage.</p> <p>The committee meets at least two times per year.</p>

¹⁰ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:330:0039:0042:EN:PDF>

Inter-institutional Editorial Committee for Internet (CEiii)	According to the IPG website ¹¹ the CEiii decides on the URL structures for websites on the europa.eu domain. The CEiii committee is managed and chaired by the European Commission, DG COMM, Unit A.5, responsible for EUROPA. The CEIII was created by the EU institutions in 2001 following a recommendation from the Inter-institutional Internet Task Force. It aims at setting up a permanent structure to coordinate the development of the institutions' websites. Its role is to facilitate the exchange of information and to create synergies between institutions and the other EU bodies on web matters. Each institution retains complete autonomy, however, over its own websites.
Inter-institutional Metadata Management Committee (IMMC)	Created in 2010, the Inter-Institutional Metadata Management Committee (IMMC) has the objective to promote the standardisation of structural metadata across EU Institutions, bodies, and services, to promote exchange, access and re-use of public sector information, and to provide metadata governance at the European Institutions level. The Publications Office assures the presidency and the secretariat of the IMMC. The IMMC was initially conceived to promote the exchange of information related to the EU decision-making process. The IMMC consists of representatives of General Secretariats and their equivalents in the EU Institutions. The IMMC uses the Metadata Registry (MDR) to register and maintains definition data (metadata elements, named authority lists, schemas, etc.) used by the different European Institutions involved in the legal decision making process. The following structural metadata is currently under IMMC governance: <ul style="list-style-type: none">• Named Authority Lists (Common Authority Tables/Value lists); and• IMMC Core Metadata element set.

3.3 Scope criteria

Policy choices will influence the effectiveness and persistence of URIs. This document proposes the following criteria to determine where **persistence** is likely to be of particular importance and where a **common approach** is most desirable.

1. **Authoritative source:** URI sets for resources for which an EU institution is the *authoritative source*.
2. **Commitment of persistence:** URI sets for resources for which an EU institution has an *obligation* or strong *commitment* to maintain information about them over a longer period of time. For such resources, EU institutions must be ready to underwrite a URI policy with specific service level guarantees, including guarantees on persistence.

¹¹ Information Providers Guide, http://ec.europa.eu/ipg/basics/urls/index_en.htm

3. **Inter-organisational character:** URI sets for resources that are relevant in the information exchange between organisations, for example between EU institutions or between an EU institution and a public administration in the Member States.
4. **Machine-readable information:** URI sets for information resources and non-information resources (e.g. places, staff, buildings) for which information *about them* can be meaningfully processed by machines. The latter means that URIs for non-information resources must resolve into a machine-readable document (an information resource) with information about the resource.
5. **Local register:** URI sets for which an EU institution maintains an official (possibly paper-based) register and enforces local control over the identifying mechanism. In such case, resources already have an identifier, called a **local id**. Resources for which no register exist are likely not to meet many of the aforementioned criteria.

Examples of resources that would meet the above-mentioned scope criteria could include:

- **Common data models and reference data:** URI sets for resources such as classes, properties, and concepts included in common data models and reference data. Such resources can greatly improve the interoperability of information exchanges. One example of this are the concepts in the Named Authority Lists (NALs) for corporate bodies, inter-institutional procedures, languages, countries, currencies maintained by the Inter-Institutional Metadata Maintenance Committee (IMMC): each of the concepts in the NALs has its own URI that can be reused in many different applications. Another example of this are the concepts in the EuroVoc thesaurus.
- **Documents related to the legislative process of the EU;**
- **High-value datasets of the EU:** the study 'high-value datasets by EU Institutions' [High-Value datasets, 2014] provides a definition of high-value datasets and identifies a set of 215 high-value datasets that bring transparency on the functioning of the European Union, have substantial social or economic value, or that that could lead to considerable cost reductions when shared with EU institutions.

3.4 Governance structure

Institutions interviewed for this study were interested in the idea of a centralised governance of URI namespaces with local governance of the registers:

- **Central register of URI namespaces:** the central register manages the URI namespaces. It could also be responsible for forwarding (redirecting) HTTP requests to local registers.
- **Local registers of resources:** the local registers manage the resources and attribute identifiers (local id) to these. Local registers may be established to host domain specific URIs (e.g. geospatial, statistics, legislation). Local URIs respect the centrally assigned URI structure. The local register also processes specific HTTP requests for this register.

This governance structure is also reflected in the URI pattern, as depicted in Figure 1. The left-hand side of the URI, the URI namespace, is registered in the central register, and the right-hand side (tail) is decided by the local register. The approach is similar to the UK guidance¹² but the URIs for the EU institutions use a simplified pattern. The different parts are discussed in Section 4.2. Decisions about the URI namespace are taken at the central level. Decisions about the tail, such as the {local id} are taken at the local level, under central guidance.



Figure 1: Central register of URI namespaces – local registers of resources

Two operational models have been considered (see also Figure 3 and Figure 4 in section 5.4.3):

1. **Forwarding (redirection):** in this model, a GET for a centrally administered URI results in a 30x redirect with a URI/URL on a local system. A 303 redirect is issued for a URI that identifies a non-digital resource (like a person, organisation, language, concept) while for digital resources (e.g. a contract notice, a legal text) a 307 redirect is issued. All subsequent communication takes place directly between HTTP client and local register without passing through the central service.
2. **Proxy forwarding:** in this model, a GET for a centrally administered URI for a non-digital resource results in a 303 redirect with a URI/URL on the central service. The next request from the HTTP client goes again to the central service. The central service then does a proxy forwarding to the local register which results in a response sent from the local register to the central service that passes the data through to the HTTP client. For a GET on a URI that identifies a digital resource, the central service does a proxy forward without the preliminary redirect.

The model that is used is the forwarding (redirection) model, as this:

- puts less stress on the central service because one request results in one response to the client (with a 3xx redirect) and nothing more;
- allows monitoring of number of accesses to a particular resource collection and allows checking that the tail conforms to established policy;

¹² <http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf>

- conserves visibility of the local register, indicating the responsible organisation for the information and providing context, which may be important in the case of human users.

A drawback is that it is not possible to monitor end-to-end performance because the communication between client and local register is direct and is not visible to the central service. However, it is still possible for the central service to implement a playback scenario (e.g. overnight) to test whether the requests lead to successful access of the resources.

In principle, a given URI namespace is associated with a *single* local register. However, as bodies which get a namespace allocated are free to create sub-namespaces, guidelines with “best practices” for the design of local paths should be developed.

If two or more EU institutions want to make use of the *same* URI namespace they must *organise* themselves as a single register to avoid the creation of homonyms and synonyms (**co-ownership of the URI namespace**). Similarly, if a register splits up its collection of resources, for instance due to a partial transfer of responsibilities from one EU institution to another, some resources would be migrated to a new register. To guarantee persistence, the existing and new register should find a solution to continue using the same namespace (**co-ownership of the URI namespace**) and persist previously existing URIs.

One organisation should not own the exclusive rights for a common namespace, such as “press releases”. This could be avoided by only assigning opaque namespaces (see section 4.2). Consequently, the request for a mnemonic namespace will require a more thorough analysis than the request for an opaque namespace.

The central governance of URI namespaces should have the following roles and responsibilities:

- **URI Steering Committee:**
 - **Responsibilities:** the URI Steering Committee takes decisions on the URI policy, the URI management process, the scope criteria, and design rules for URI sets. It also deals with requests for URI namespaces that require an exception to the criteria and rules.
 - **Composition:** the steering committee consists of representatives from all EU institutions.
- **URI Committee:**
 - **Responsibilities:** the URI Committee acts as a Secretariat and takes decisions on requests for URI namespaces by applying the scope criteria and design rules. It ensures the permanency of the central URI management.
 - **Composition:** the committee could consist of staff of one or more EU institutions that are dedicated to take decisions on URI namespaces.
- **URI Technical Team:**

- **Responsibilities:** the URI Technical Team provides technical support in the management and the monitoring of the URI namespaces.
- **Composition:** the team could consist of staff of one or more EU institutions that are dedicated to perform this work.

3.5 URI management process

The governance of URI sets should ensure that the following decisions are taken in the URI management process:

- **Register a URI namespace;**
- **Monitor the URI design rules and service levels;**
- **Monitor, evaluate, and improve the URI policy;**
- **Raise awareness on the URI policy.**

3.5.1 Request a URI namespace

The registration of a new URI namespace string for URI sets on the Europa.eu domain is a key control point to ensure that the scope criteria, URI design rules and service level guarantees are met. The process consists of the following steps:

1. Representatives of one or more **EU institutions** send a request for the registration of a new namespace string to the URI Committee. The request provides detailed information on the resources and the local register for which a URI namespace string is requested. The request also indicates how the scope criteria and URI design rules are met. The request also identifies other registers of EU institutions that contain *similar* resources or that could be linked with the local register.

Example: *The Publications Office wants to assign persistent URIs for contract notices (CNs) managed in its register Tenders Electronically Daily (TED). The Publications Office files a request with the URI Committee for a namespace string to be assigned to this resource collection. In special cases, the requesting institution may ask for a particular mnemonic to be assigned as namespace string.*

2. The **URI Committee** assigns the request to the URI Technical Team.
3. The **URI Technical Team** contacts the requesting **EU institution** to prepare a proposal, analysing how the resource collection meets the scope criteria. The proposal is elaborated in consultation with the requesting institution. In the general case, the Technical Team proposes an opaque namespace string; in the special case, the Technical Team will verify that the requested mnemonic does not clash with existing namespaces and that the mnemonic is already known by external parties. The proposal is submitted to the **URI Committee**.

Example: *The URI Technical Team verifies the scope criteria and URI design rules. It confirms among others that TED is the authoritative source for information on contract notices, that TED has a legal commitment to persist information on tender notices, that information on contract notices is used in an inter-institutional*

context, etc. The URI Technical Team then proposes to assign the opaque namespace string 'abc' to the resource collection.

4. The **URI Committee** considers the proposal; it involves the **URI Steering Committee** where an exception to the scope criteria or the URI pattern is requested. On the basis of these investigations, it is decided whether the URI namespace string can be attributed to resources in the local register. In this step, the URI Committee should also actively promote the use of existing URIs that are *already* published by authoritative bodies within and outside the EU institutions.

Example: *The URI Committee verifies that the Technical Team has followed the appropriate procedure. If this is the case, the URI Committee takes a positive decision and logs this decision together with the investigation.*

5. The **URI Technical Team** gives representatives of the EU institutions access to the Persistent URI (PURI) application and the right to configure the requested namespace.

Example: *The URI Technical Team configures the Persistent URI service to give the representatives of TED access rights to modify the namespace `http://data.europa.eu/abc/*`. Users are identified and authenticated using their ECAS account.*

6. The representatives of the **EU institution** configure URI redirection rules on the namespace. The use of redirection rules allows using the local register as the single and authoritative source of information – no information must be duplicated. At the same time, redirection rules give flexibility to change the underlying local register without changing the persistent URIs.

Example: *TED configures a redirection rule on the namespace string:*

```
http://data.europa.eu/abc/{$s1} redirect to  
http://ted.europa.eu/udl?uri=TED:NOTICE:{$s1}.
```

Applications (machines) can now retrieve machine-readable data from TED using a persistent identifier. Should 5 years later TED be migrated to another platform for tender registration, TED updates the redirection rule

```
http://data.europa.eu/abc/{$s1} redirect to  
http://cellar.europa.eu/TED:NOTICE:{$s1}.
```

Applications continue to use the same persistent URI to retrieve machine-readable information from TED.

3.5.2 Request (co-)ownership of a URI namespace

Where there is a case that two or more institutions want to use the same namespace string, the institutions need to arrange this on the organisational level. If such a case is detected, the institutions that want to share the namespace, they need to take bilateral action to agree which of the institutions takes operational responsibility for the namespace.

In practice, the one namespace – one register principle will always apply, which means that, if two organisations want to share a namespace, they should merge their registers, at either the organisational or the technology level.

3.5.3 Monitor URI namespaces

The URI Committee should also monitor the URI namespace. In particular, it should perform the following checks:

- **Monitor HTTP requests to the URI namespace:** the URI Technical Team should monitor the HTTP requests to URI namespaces to verify whether the URI design rules are followed and whether service levels are met.

Example: *The URI Technical Team discovers that TED is using file extensions in URIs. It can discover this through the logging and analysis of incoming HTTP requests on the namespace*

*<http://data.europa.eu/contractnotice/>**

Similarly, the URI Technical Team may discover via analysis of its log files that URIs that are frequently requested on the namespace of TED are no longer dereferenceable; an HTTP 404 error code is returned. The URI Technical Team reports these violations of the design rules and service level guarantees to the TED team.

- **Collect key performance indicators per URI namespace:** The URI Technical Team should collect key performance indicators based on the **Number of HTTP requests** per namespace with detailed statistics on the requestor, accept parameter, language parameter, and the final HTTP response code.
- **Monitor changes to the configuration of the URI namespace:** the URI Technical Team should keep a log of changes that are performed to URI namespaces, such as changes of redirection rules or access rights.

3.5.4 Monitor, evaluate, and improve the URI policy

The URI Steering Committee should monitor, evaluate and improve its URI policy, the scope criteria, and URI design rules.

3.5.5 Raise awareness on the URI policy

The URI Committee should actively promote the URI policy among EU institutions and third-parties. This can be done by maintaining a Website with information on the following topics:

- URI policy, scope criteria, URI design rules, URI management process, etc.;
- Existing URI namespaces under common governance; and
- Statistics on the Service Levels.

4 DESIGN RULES FOR PERSISTENT URI SETS

This chapter captures the rules that URI sets placed under common governance must follow.

4.1 Stakeholder requests

The stakeholder requests set out in the following table are derived from the interviews conducted as part of this study. The requirements in Table 5 are referred to in the remainder of this document by their identifier.

Table 5 – Stakeholder requests: design principles for URI sets

ID	Request
R0	Cost-benefit: Each design principle should have a positive expected cost-benefit relationship.
R1	The ability to link to data managed by others.
	Pattern – single subdomain: The URI Task Force proposes to have only a single subdomain (e.g. puri.europa.eu) – or no subdomain at all (e.g. europa.eu).
R2	Use existing identifiers: The design rules should allow minting URIs composed of an HTTP namespace and another (already existing) local identifier, such as DOIs, VIAF, ISNI etc.
R7	Guidance on the use common concepts/dictionaries as elements within URIs.
R8	Multilingual considerations – generic resource: In the first place, it is important to identify resources across different languages versions, i.e. concepts must be multilingual.
	Multilingual considerations – specific language version: language negotiation does not always work very well. Policy is to specify the language at the end of the URI, so as not to make the user think that the whole site is translated.
R10	HTTP only: Make the recommendation to use HTTP URIs explicit.
R11	URIs for language versions and formats of information resources: for information resources, there should be both different URIs for specific languages versions and formats as well as generic URIs for context-neutral resources.
R12	Versioning - Memento: Support for Memento ¹³ is desirable to capture the evolution of a resource over time.
R13	Legacy URIs: Legacy URIs should continue being supported and be accommodated in the principles.
R15	Europa.eu only: URIs for items for which an EU Institution is the official and

¹³ <http://mementoweb.org/>

Study on a common policy for the governance and management of persistent URIs by EU institutions

	authoritative source should be published on the europa.eu domain only.
R16	<p>No subdomain: Preference should be given to adding path elements rather than subdomains (i.e. europa.eu/abc rather than abc.europa.eu).</p> <p>This is also related to a recommendation formulated by the URI Task Force in 2013 [Bhraonain & Georgiannakis, 2013]: <i>agree on a restricted list of sub-spaces for a common URI root taking into account the technical implications (e.g. bottle-necks if de-referencing) of different solutions for partitioning sub-spaces or sub-domains (e.g. health.data.europa.eu or data.europa.eu/health).</i></p>
R17	<p>Versioning – latest version: URIs should be defined for the latest version and for specific versions of resources.</p>
R18	<p>Thematic approach - no statement of ownership: stating ownership makes URLs unstable on Europa.eu. A thematic or by categories approach for the structure of persistent URI sets would make more sense. However, it is very difficult to achieve this. Different institutions will have to share thematic URIs on the same basic namespace (e.g. /transport). Clear rules and governance should regulate this.</p> <p>This also related to a recommendation formulated by the URI Task Force in 2013 [Bhraonain & Georgiannakis, 2013]: <i>decide if the organisational structure responsible for a specific policy rather than a broad policy area (e.g. 'climate change' rather than 'environment') would govern and define the rules for the assignment of URIs in that sub-space and what would happen if the particular responsibility moves to another governing body.</i></p>
R19	<p>Left-hand side, central governance. Right-hand side, local governance. The central common governance should only be about the left-hand side of the URI structure. EU institutions should have autonomy to manager URIs within the namespaces attributed to them.</p>
R20	<p>Ensure retroactive 'fitting' with DGs that have already minted URIs [Ní Bhraonáin & Georgiannakis, 2013].</p>
R21	<p>Canonicalisation: There is a need to canonicalise the path elements outside the domain name part to disambiguate the identifiers.</p>
R22	<p>Granularity: Datasets can be published as a dataset presenting one hundred indicators or one hundred datasets each one with data points for only one indicator. The system should provide sufficient granularity to link data items to each other.</p>
R23	<p>Temporal aspects: The design rules should give some guidance on how to deal with temporal aspects, i.e. resources for which the meaning changes over time. For example, if the legal entity European Economic Community became the European Union, is it a different resource and does it therefore need a different identifier?</p> <p>Obviously, whether or not a resource should keep the same identifier, or becomes a new identifier is a matter to be dealt with by the local register.</p>

4.2 Follow the pattern

For *new* URI sets placed under common governance, the following generic, standard format should be used to build up a URI:

```
http://{subdomain}.europa.eu/{namespace string}/  
{local-id}/{version}/{language}
```

The segment `{version}/{language}` is proposed as an optional sample pattern, not a strict requirement.

Legacy URI sets placed under common governance should continue to persist and are therefore not required to follow this pattern.

4.2.1 *http://*

HTTP URIs should be used to provide persistent identifiers as described in the scope criteria. This is the primary way in which the required feature for linking from one application or dataset to another can be met.

Where necessary, HTTPS may also be used. Two URIs that differ only in that one is an HTTP URI and the other an HTTPS should only ever refer to the same resource.

An HTTP link header can indicate that both are equivalent and such an HTTP link header should be set where appropriate. If there is a preference for one version, e.g. a preference that HTTPS be used, then the link header on the HTTP version should be set to point to the HTTPS URI using the `@rel` value of 'self.' Where there is no preference for either version then a relationship type of 'duplicate' may be appropriate. The full list of registered `@rel` values is maintained by IANA¹⁴.

4.2.2 *{subdomain}*

New URI sets placed under common governance are placed under the subdomain **data.europa.eu**.

For legacy URI sets placed under common governance, it is important that `{subdomain}` may be something that is sector-specific such as 'transport' or 'environment' but should not be the name of a particular DG or department since these are liable to change frequently over time. Placing existing URI sets under common governance may still be useful, so as to increase coordination and avoid problems such as for example those of overlapping URI sets (synonyms) belonging to different sectors.

4.2.3 *Europa.eu*

For new URI sets placed under common governance, the domain must be Europa.eu.

4.2.4 *{namespace string}*

For new URI sets placed under common governance, a namespace string will be assigned. Two cases can be distinguished:

¹⁴ <https://www.iana.org/assignments/link-relations/link-relations.xhtml>

1. The general case where an **opaque** string is assigned to the namespace. In this case, a string is assigned to a resource collection for which a namespace is requested.

Examples:

`http://data.europa.eu/ab/{local-id}...`

`http://data.europa.eu/g4/{local-id}...`

`http://data.europa.eu/1a/{local-id}...`

Namespace strings should be encoded using an ASCII 0-9, a-z (lower case only) character set. Initially, two characters could be used (allowing 36^2 equal to 1.296 combinations). This could be extended at a later point in time to three characters (allowing 36^3 , equal to 46.656, combinations) or four characters (1.679.616 combinations).

2. The special case where the namespace string assigned is a **mnemonic** that is related to the nature of the resource collection.

This will only apply to cases where a mnemonic is already externally known for a particular collection, or where standard identifiers have already been assigned to the resources.

Examples:

`http://data.europa.eu/eli/{local-id}...` for resources identified by a European Legislation Identifier

`http://data.europa.eu/esco/{local-id}...` for resources under the European Skills, Competences, Qualifications and Occupations register

`http://data.europa.eu/doi/{local-id}...` for resources that already have Digital Object Identifiers

It is recommended that an HTTP request for a URI without local-id part, so with the bare namespace only, e.g. `http://data.europa.eu/eli/` or `http://data.europa.eu/abc/`, results in an HTTP response with a description of the resource collection. This HTTP response should take the form of a human-readable webpage describing the collection in the local registry, or machine-readable descriptive metadata about the collection in the local registry. Whether the HTTP response is in a human-readable or machine-readable format is ideally determined via the accept parameter in the HTTP request (this mechanism is called content negotiation).

URIs that have been assigned should be persistently maintained. This means in particular that a namespace string, either opaque or meaningful, has to remain valid for as long as there are references to the resources identified by the URIs in the namespace. In other words, once a namespace string has been assigned, it cannot be changed later.

4.2.5 *{local id}*

The final element of the URI, the `{local id}`, is a specific item. These might be things like:

- 3611
- FR
- DIGIT
- Aminopeptidase

i.e. 3611 is a concept in Eurovoc, FR is a specific country, DIGIT is an institution and Aminopeptidase is a specific enzyme.

The decision to use `{local id}` within a URI is taken at the *local* level – the governance of the registry that contains information about a resource – with some central guidance.

4.2.6 *{version}*

As a sample pattern, the optional segment `{version}` could be used for determining the version of a resource (see section 4.6 for more details).

The decision to use `{version}` within a URI is taken at the *local* level – the governance of the registry that contains information about a resource – with some central guidance.

4.2.7 *{language}*

As a sample pattern, the optional segment `{language}` could be used for determining the language of a resource (see section 4.9 for more details). To indicate the language, the codes of the languages Named Authority List¹⁵ should be used. These codes are aligned with ISO 639-3¹⁶ three-letter codes. ISO 639-3 is an extension of ISO 639-2/T¹⁷ which uses the same three-letter codes but for a smaller collection of languages.

The decision to use `{language}` within a URI is taken at the *local* level – the governance of the registry that contains information about a resource – with some central guidance.

4.3 Avoid stating ownership or branding

The pattern outlined in section 4.2 does not include any indication of ownership or branding. In the general case, a namespace string will be opaque, and the mnemonic namespace strings are intended to only be used when the string is already known externally or is the name of a standard identifier.

¹⁵ Publications Office of the European Union. Metadata Registry. Authorities. Languages. <http://publications.europa.eu/mdr/authority/language/>

¹⁶ SIL.org. ISO 639-3. <http://sil.org/iso639-3/>

¹⁷ Library of Congress. ISO 639-2. <http://www.loc.gov/standards/iso639-2/>

4.4 Reuse existing identifiers

Where resources are already uniquely identified, e.g. using DOIs, VIAFs or other sorts of identifiers, those identifiers should be incorporated into the URI. For example, there are established codes for EC buildings so the URIs for those buildings should be included in their URI. Possible examples, for the case that EC buildings have been assigned an opaque namespace string 'xyz', are:

- <http://data.europa.eu/xyz/CCAB>
- <http://data.europa.eu/xyz/BERL>
- <http://data.europa.eu/xyz/MO34>

It is worth noting that other identifiers, besides URIs, should also be developed with persistence in mind. For instance, if the city of Brussels decided to change the name of Rue Montoyer to something else, the URI <http://data.europa.eu/xyz/MO34> would still identify the same building but the building itself would have a different address.

However, this structure is easily extended to identify specific rooms so that the room C2 in the Albert Borschette conference centre might be identified as

- <http://data.europa.eu/xyz/CCAB/C2>

Caution: when reusing an identifier, it is essential to reuse it without changing the original semantics. For example, *Aminopeptidase* identifies a specific enzyme and should not be used as a shorthand for *Lactococcus lactis* which is the organism from which it is obtained. In fact, the organism itself will have a different identifier.

Digital Object Identifiers (DOIs) in particular are in widespread use across the European institutions. The Publications Office of the EU is a DOI Registration Agency. DOI-based identifiers should follow the same pattern as shown in section 4.2.4.

4.5 Use auto increment with care

Minting new URIs for large collections of resources will need to be automated and the process must be guaranteed to produce unique identifiers. One way to do this might be to simply increment a counter as each new URI is minted. Imagine that picking up on the previous example, enzymes were given integer identifiers rather than use their names. In that case, the following could be possible URIs for two different enzymes in a resource collection that was assigned the opaque namespace string 'a2':

- <http://data.europa.eu/a2/123456>
- <http://data.europa.eu/a2/123457>

Although this approach is perfectly feasible, we would recommend it only if one of the following is true:

- the process will never be repeated;
- the process can be repeated to create exactly the same URIs for the same input data with new URIs minted only for new items.

4.6 Be careful with version numbers

Although concept schemes, ontologies, taxonomies and vocabularies are likely to go through iterative cycles of change, version numbers and status information should not be included in the URIs. Rather, URIs should remain stable between versions and new ones minted for new terms. URIs may be deprecated and their use discouraged but they should nevertheless be maintained both in terms of the actual URI and the resource they identify. Once minted, a URI should never be deleted.

Where URIs identify resources that are versioned, new URIs should be minted for each new version of the resource as well as a URI for the latest version. For example, imagine monthly spending data. It might have a URI under an opaque namespace string 'e6', such as

- <http://data.europa.eu/e6/123456/latest>

That URI would identify the latest available monthly spending data and so although the URI is persistent, the resource identified would change each month. Each specific month's spending data would have its own URI such as

- <http://data.europa.eu/e6/123456/2013/11>
- <http://data.europa.eu/e6/123456/2013/12>
- <http://data.europa.eu/e6/123456/2014/01>
- etc.

The use of Memento¹⁸ for versioning over time is an option. Impact in terms of resources would need to be analysed.

4.7 Avoid using query strings

Query strings (e.g. '?param=value') are usually used in URIs as keys to look up terms in a database. This is brittle since it relies on a particular implementation, i.e. the technical infrastructure used. By conforming to the URI template of section 4.2 and not exposing any of the underlying technical infrastructure in the URI string itself, persistence is much easier to maintain. Redirections can be handled internally to the server without triggering an extra HTTP request (that results in the target URI being shown in a browser's address bar).

Furthermore, it is common to assume that the order of parameter/value pairs in a query string is unimportant since in implementations this is frequently true, if not always true. However, from a URI perspective order *does* matter and maintaining a system with mappings between every possible combination of parameter/value pairs would be extremely cumbersome.

4.8 Be careful with file extensions

For similar reasons to the previous point, file extensions that are technology-specific and do not add meaning should not be included in the URI. Particularly

¹⁸ <http://mementoweb.org/>

those file extensions that stem from the technology used on the Web Application Server such as .cfm, aspx or .py should be avoided.

That said, it is good practice to provide URIs that point directly to specific formats. Appending the URI with file extensions such as .csv or .xml would return data in the format usually associated with that file extension. So to return to the example in section 4.6, dataset distributions of the January 2014 spending data in CSV or RDF format could be identified by:

- <http://data.europa.eu/g4/123456/2014/01.csv>
- <http://data.europa.eu/g4/123456/2014/01.rdf>

Alternatively, the data itself might be available in multiple formats and the one returned when the URI was dereferenced could depend on content negotiation. In this case, dataset distributions could be retrieved using a generic URI, and the desired format would be included as a parameter in the accept-parameter of the HTTP header. In this example, this means that if the following URI were requested with accept parameter 'application/rdf+xml'

- <http://data.europa.eu/g4/123456/2014/01>

the URI would redirect to

- <http://data.europa.eu/g4/123456/2014/01.rdf>

N.B. For emphasis, URIs are dumb strings and no portion has any semantics beyond the functional {scheme}://{authority}/{path}. It is the HTTP Response Headers that include authoritative information about the content type. Therefore, if a URI like <http://example.com/foo.pdf> returns an XML file, that is technically not wrong IF the HTTP header declares the content type to be application/xml. It may be awkward, unusual, surprising and unhelpful... but technically it is not wrong.

4.9 Treat languages as versions

A lot of data does not have an associated human language and many data formats allow textual labels to be provided in multiple languages. Therefore, for many resources identified by persistent URIs, there will be no need to consider identifying its language in their URIs. However, many documents do have an associated language (the ubiquity of modern technology, especially the Web, has radically altered the definition of the word 'document' such that it now includes audio visual content an, Web pages that contain multiple elements and that are updated many times a day as well as the more traditional notion of a static document, whether it be a physical printed document or its electronic counterpart). Where resources are available in multiple languages the same approach as for different formats and versions should be used, that is, mint a URI for each version with the language at the rightmost position. For example, the annual spending report might have a URI such as

- <http://data.europa.eu/1s/123456/2013>

and this would be available in multiple languages. Servers can be configured to return the correct language version in just the same way that they can return the user's preferred document format but to identify specific language versions, specific URIs should be minted such as

- <http://europa.eu/1s/123456/2013/eng>
- <http://europa.eu/1s/123456/2013/fra>
- <http://europa.eu/1s/123456/2013/deu>

Appending URIs with `{/lang}` retains the structure to the left of the language in a machine-processable way. Dereferencing the non-specific URI might usefully provide a list of the different language versions available. This is not so easily arranged if the URIs end with, for example, `/2013_en`.

A potential problem arises if resources are available in multiple languages AND multiple formats. It is anticipated that in this eventuality, a practical method can be devised for including both variants as separate path segments in the URI according to the particular circumstance. In the examples immediately above, appending each of those URIs with `.pdf` or `.doc` would perhaps be appropriate.

4.10 Character encoding and multilingual considerations

As noted, URIs are dumb strings that carry no semantics beyond the function of identifying resources. Internationalised Resource Identifiers, IRIs, allow the use of any character in the international character set and so can be written in any script, not just the unaccented Latin characters in ASCII. However, it is useful to developers and those managing URI sets if they are human readable.

The examples in the preceding sections have all used English words as URI segments. This is not a hard and fast rule – path segments can be written in any language, however, the strong advice is to stick to the Latin alphabet without accents, i.e. ASCII-7 characters, as defined for URIs as distinct from IRIs. This is strongly encouraged for three reasons:

- tool support for IRIs is less well developed than for ASCII-only URIs;
- using characters that look superficially the same as ASCII characters can lead to unintended errors or deliberate attacks;
- where the need to use non-ASCII characters overrides the advice not to, such characters should be percent encoded rather than written as the character itself (e.g. `%e9` rather than `é`). This avoids issues like multiple UTF-8 encoding of accentuated characters (e.g. `'è'` can be encoded as `'è'` or `'e`'`) that may lead to different Punycode¹⁹-encoded URIs.

Whatever language is used for the path segments in the URI, labels for resources should be available in as many languages as possible. Note that a single resource can have any number of labels, differentiated by their language tag, and these labels can be selectively presented to end users through software. As a concrete example, the DCAT vocabulary uses English terms in its URIs but at the time of writing provides labels in French, Spanish, Arabic, Greek and Japanese with more on the way.

¹⁹ IETF. RFC3492. Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA). <https://www.ietf.org/rfc/rfc3492.txt>

4.11 Implement 303 URIs for real-world objects

When de-referenced, URIs that identify real-world objects that cannot be transmitted as a series of bytes (such as buildings, places and people) should redirect using HTTP response code 303 to a document that describes the object. An example of this is depicted in Figure 3 (Section 5.4.3).

4.12 Use a dedicated service

The data owner should run its own local register and Web service where it is publishing information about its resources. It should do this as part of its regular operations. Multiple, dedicated, local registry services that are capable of processing specific HTTP requests for the resources in this register, and that are easily transferable are the greatest guarantors of persistence.

4.13 Consider a fall back scenario

Local registry services will be set up and managed by specific EU institutions and will be subject to the policy decisions of that institution. In the event that a decision may be taken to close a service at a later date, it is worth considering establishing an agreement with another agency or third party that would come into force if the original service could no longer be supported. Any agreement with a third party fall back or other external provider should be bound to a guarantee of persistence of the URIs.

5 URI REDIRECTION SERVICE

This chapter captures the high-level requirements for a URI redirection service, used for the management of persistent URI namespaces and the redirection of HTTP requests to the (local) registries.

5.1 Use case scenario

As explained in Section 3.4, the common approach for persistent URIs for EU institutions assumes a **central register of URI namespaces** and many **local registers of resources**. The central register manages the URI namespaces and could be responsible for forwarding (redirecting) HTTP requests to correct local registers. The local registers manage the resources, attribute identifiers (local id) to these resources and deal with specific HTTP requests.

The Persistent URI redirection service could be used to support the URI management processes described in Section 3.5. The below use case scenario illustrates the interaction between the central register for URI namespaces (the redirection service) and local register of resources (Tenders Electronically Daily - TED).

Use case scenario

1. Tenders Electronically Daily (TED) wants to assign persistent URIs to contract notices (CN). TED uses the **URI redirection service** to file a request with the URI Committee to claim the URI namespace `http://europa.eu/id/contract-notice/*`.
2. The URI Committee verifies the request and uses the **URI redirection service** to approve the request and grant TED access to the URI space
3. TED uses the **URI redirection service** to configure a redirection rule on the namespace:

```
http://data.europa.eu/contract-notice/{s1} redirect to
http://ted.europa.eu/udl?uri=TED:NOTICE:{s1}.
```

Applications (machines) can now retrieve machine-readable data from TED using a *persistent* URI.

6. The URI Committee uses the **URI redirection service** to monitor the incoming HTTP requests on the namespace. The URI Technical Team discovers that TED is using file extensions in URIs. Similarly, the URI Technical Team may discover that URIs on the namespace of TED are no longer dereferenceable; an HTTP 404 error code is returned. The URI Technical Team reports these violations of the design rules and service level guarantees to the TED team. The TED team fixes these comments.

7. 5 years later, TED is migrated to another platform for tender registration, implementing new regulations by the EU. TED updates the redirection rule in the **URI redirection service**

```
http://data.europa.eu/contract-notice/{s1} redirect to
```

`http://cellar.europa.eu/TED:NOTICE:{$s1}`.

A Persistent URI redirection service contributes to the following benefits:

- **Flexibility:** The service makes it *easier* to put in practice a common URI policy for EU institutions while providing sufficient *flexibility* to local registers to manage their own resources.
- **Speed and efficiency:** Setting up persistent URIs will take less time as EU institutions can make use of a common infrastructure for registering a particular namespace for their register; and
- **Monitoring of service levels:** the service levels for dereferencing (resolving) URIs will be centrally monitored.

5.2 Stakeholder requests

Table 6 summarises the requirements for a persistent URI service, as elicited from the interviews conducted for this study.

Table 6 – Stakeholder requests: persistent URI service

ID	Request
S1	Authentication and access control: authentication via the European Commission Authentication Service (ECAS), which also allows users from EU Institutions such as the European Parliament or the Council of the EU to authenticate themselves with their corporate account. The application should also allow to grant access to users to a particular URI namespace, delegate access etc.
S2	Interface – API with authentication key: The Persistent URI service should allow the programmatic creation and maintenance of URI namespaces for URI sets. Similar to the existing URL shortener, the service should expose an API with an authentication key.
S3	Monitoring – key performance indicators: automatic control of name space guidelines, regular verification of links, possibility to put on quarantine 404 files, e-mail notification to owners.
S4	Monitoring – Web analytics: It should be possible to do Web Analytics to monitor requests for the persistent URIs.
S5	Namespace configuration – internal or external redirection: It should be possible to have the redirection visible to Web clients (external redirection: HTTP status code 30X) or invisible (internal redirection HTTP status code 20X). URIs for non-information resources would always require a 303 redirection upon resolution.
S6	Namespace configuration – metadata: the system must expose metadata about URIs and the resources they identify.

S7	Namespace configuration – multiple (sub)domains. The Persistent URI service should be capable of managing namespaces for more than one (sub) domain. The URI design rules currently foresee that new persistent URIs should be located on a single subdomain of Europa.eu. However, this would exclude legacy URIs to be managed through the Persistent URI service.
S8	Namespace configuration – original HTTP referrer: The referrer in the http header should be maintained. This is needed for Web Analytics. The central redirection is just another entry point into the website.
S9	Namespace configuration – redirection rules: Possibility to create URIs manually and on the fly - guidelines for connecting IS with the central system.
S10	Namespace configuration – URI pattern: namespace depth level to be defined.
S11	Non-functional – Open source: if possible the solution should reuse existing open-source components or where a solution is developed its source code should be made available for reuse under an open licence.
S12	Namespace configuration and redirection – HTTPs. The service should cater for both HTTP (internal/external redirection) and HTTPS URIs (internal redirection). URIs that differ only in that one begins http and the other with https should be considered as equivalent.
S13	Namespace configuration – Search Engine Optimization (SEO): For URI sets that identify web pages, the impact on page ranks should be restricted to a minimum. Normally the resources for which URIs are assigned are non-information resources (concepts, buildings, people, places) that are used in software, not ones that will be searched for by humans using a Web browser and search engine. Therefore it is suggested that whilst the service should not actively work against SEO, it should not be a major requirement.
S14	Mapping/Linking: In addition to mint new URIs and redirect URIs, we need a way to map identifiers to each other. This requirement may not be met by a central Persistent URI service, but could be implemented on the local registers that foresee alignment with other authoritative registers.

5.3 Existing solutions

Table 7 lists a number of existing solutions for URI management that are available for reuse.

Table 7 – Existing solutions for URI managements

Name	Description
------	-------------

Apache HTTP Server	<p>The Apache HTTP Server is an open-source HTTP server. It is compliant with current HTTP standard and extensible through modules. Apache server is the most straight forward option for the Persistent URI service. It offers complete features for the configuration of redirection rules using regular expressions. These redirection rules, the security and authentication needed can be developed using virtual directories and .htaccess files.</p>
Purlz (purl.org)	<p>The best known example of a persistent URI service is purl.org that is operated by the OCLC. The features are:</p> <ul style="list-style-type: none">• Authorisation• Content negotiation• Redirection rules (partial redirect)• SLA monitor <p>Purl.org is based on the open source purlz software²⁰, which contains an elaborate authorisation model that allows access control by users and by groups, per URI and per directory. Although the software is configurable, there is no possibility for content negotiation, nor are there redirection 'rules.', Purlz does foresee in <i>partial redirects</i>²¹, which can be seen as a form of redirection rules.</p>
UKGovLD Registry	<p>Under the auspices of the UK Government Linked Data Working Group, Epimorphics has created open source software for registries²². This is in use by the World Meteorological Organisation²³ and the UK Environment Agency²⁴ and offers many important features including:</p> <ul style="list-style-type: none">• Authorisation• Content negotiation• Redirection (external redirection)• Proxy forwarding (internal redirection)• Request filtering/blocking• Response from the internal data store• Auditing• Version management• Scalability options• Free text search• Bulk RDF upload at installation time and/or subsequently• Extra tooling for CSV to RDF conversion

²⁰ <http://www.purlz.org/>

²¹ <http://purl.oclc.org/docs/faq.html#toc1.9>

²² <https://github.com/UKGovLD/ukl-registry-poc>

²³ <http://codes.wmo.int/>

²⁴ <http://environment.data.gov.uk/registry/>

PID	The Persistent Identifier Service ²⁵ uses an approach to intercept all incoming HTTP requests at the Apache HTTP web server level and then performs a set of user-defined actions, such as, HTTP header manipulation, redirects, proxying requests, delegating resolution to another service, etc. Unlike software like Purlz it does not provide a mechanism for access control that is configurable at the level of the URI namespace.
Commission URL Shortener	DG COMM runs a URL shortener ²⁶ service similar to those offered by bit.ly, is.gd and t.co. It represents an existing service within the EC infrastructure that could be extended to form a persistent URI forwarding application.
NetKernel Enterprise	NetKernel Enterprise ²⁷ is the commercial enhanced version of the open-source NetKernel engine which is used by the Purlz server. It provides an abstraction called "Resource Oriented Computing" for building scalable services. Functionalities include caching, load balancing, request visualization, and visual functional programming.

5.4 High-level specification

This section contains a number of high-level requirements that provide a minimal functional specification for the URI redirection service.

5.4.1 Domain model

The Persistent URI service should have a domain model that contains at least the following entities:

- **Resource:** an information resource (e.g. document, Web page, a dataset distribution) or a real world thing (e.g. a building, a person, a legislative procedure, a description of a dataset).
- **URI:** an identifier for a resource.
- **URI namespace:** A URI namespace is associated with a collection of resources in a local register.
- **Local register:** a trusted authentic source of resources under the control of an appointed EU institution.
- **Redirection rule:** a pattern for the redirection of URIs (e.g. using regular expressions) within a given namespace to another namespace. The simple redirection rule involves the redirection from one URI to another.
- **HTTP client:** any Web-based application (e.g. Browser) that can issue HTTP requests.

²⁵ <https://www.seegrid.csiro.au/wiki/Siss/PIDService>

²⁶ <https://webgate.ec.europa.eu/europa-hub/help/url-shortener>

²⁷ <http://www.1060research.com/products/#netkernel>

5.4.2 Use cases

The Persistent URI service has the following actors, as depicted in Figure 2:

- **Administrator:** the administrator administers the URI redirection service. The most important privilege is to approve the registration of a new URI namespace.
- **Namespace owner:** a namespace owner owns a local registry and one or more persistent URI namespaces for the resources in the registry.
- **User:** any non-authenticated user (human or machine) who submits HTTP requests to persistent URIs using a browser or other HTTP client.

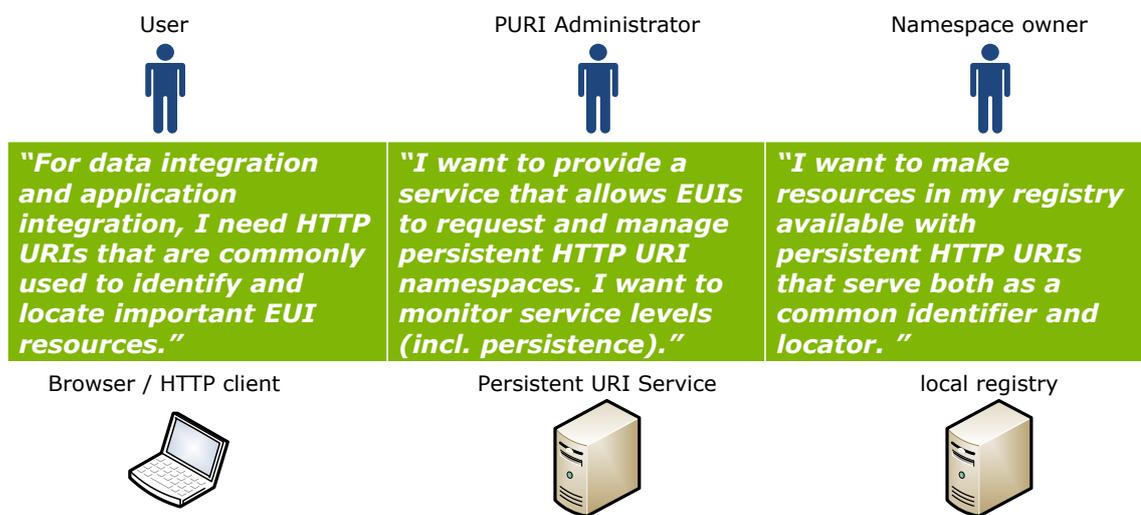


Figure 2: URI redirection service: concerns of the three main actors

The Persistent URI service should implement a number of high-level use cases:

- **Propose a URI namespace:** the goal of this use case is for a namespace owner to propose a URI namespace string, associated with a collection of resources in a local register. Upon completion, a URI namespace has a status 'proposed'. The register prevents the proposal of URI namespaces that do not follow the URI pattern stipulated in the URI design rules.
- **Approve a URI namespace:** the goal of this use case is for an administrator to approve a proposed URI namespace. Only administrators can update the status of a URI namespace, for example from 'proposed' to 'approved'. Upon approval, the namespace owner who proposed the URI namespace is granted "owner" access rights for that namespace. He also has the possibility to delegate access rights to other users.
- **Configure a redirection rule:** the goal of this use case is to configure redirection rules such that HTTP requests are correctly forwarded to the local register that is associated with a particular URI namespace. Any authenticated user who has access rights to modify a URI namespace can create, read, update, and delete redirection rules associated with that URI

namespace. A redirection rule has among others the following configuration parameters:

- **HTTP request parameters:** specific conditions for the parameters in the HTTP requests. For example, the condition that the HTTP accept parameter should be equal to `'Accept: application/rdf+xml'` for the rule to apply.
- **Input URI pattern:** a regular expression with parameters that should match with the URL string on the HTTP request for the rule to apply.
- **Output URI pattern:** a parameterised string that encodes the destination URI to which the request is redirected.
- **Internal or external redirection:** The redirection can be internal or external, as explained in Section 5.4.3.
- **Response status code:** the HTTP status code of the response (e.g. 303 for redirection of non-information resources);
- **HTTP response parameters:** any HTTP response header fields that should be set in the HTTP response. For example, the response header field should contain `Content-Type: application/rdf+xml`.
- **Submit HTTP requests:** the goal of this use case is for users to submit HTTP requests to the Persistent URI Service, which are redirected to the corresponding local registry, allowing users to retrieve information about the resources identified by Persistent URIs:
- **Log HTTP requests:** both an administrator and "owner" of a URI namespace can obtain a log of all HTTP requests per URI namespace.

5.4.3 Sequence diagrams

Figure 3 and Figure 4 illustrate how the URI redirection from the Persistent URI service takes place in the case of external and internal redirection:

- **External HTTP redirection** (forwarding): An HTTP client issues an HTTP request to retrieve information about a contract notice, identified by URI <http://data.europa.eu/contract-notice/229842-2014>. The Persistent URI service redirects the client to retrieve a document on the TED server <http://ted.europa.eu/udl?uri=TED:NOTICE:229842-2014>.
- **Internal HTTP redirection** (proxy forwarding):): An HTTP client issues an HTTP request to retrieve information about a contract notice, identified by URI <http://data.europa.eu/contract-notice/229842-2014>. In the background the Persistent URI service acts as a proxy and has internally forwarded the request to the TED server, this is also called proxy forwarding. Please note that in this case the HTTP client always receives an HTTP 303 redirection response code. This is because for non-information resources (such as contract notices) it is a good practice to provide HTTP 303 redirection, as discussed in Section 4.11.

Study on a common policy for the governance and management of persistent URIs by EU institutions

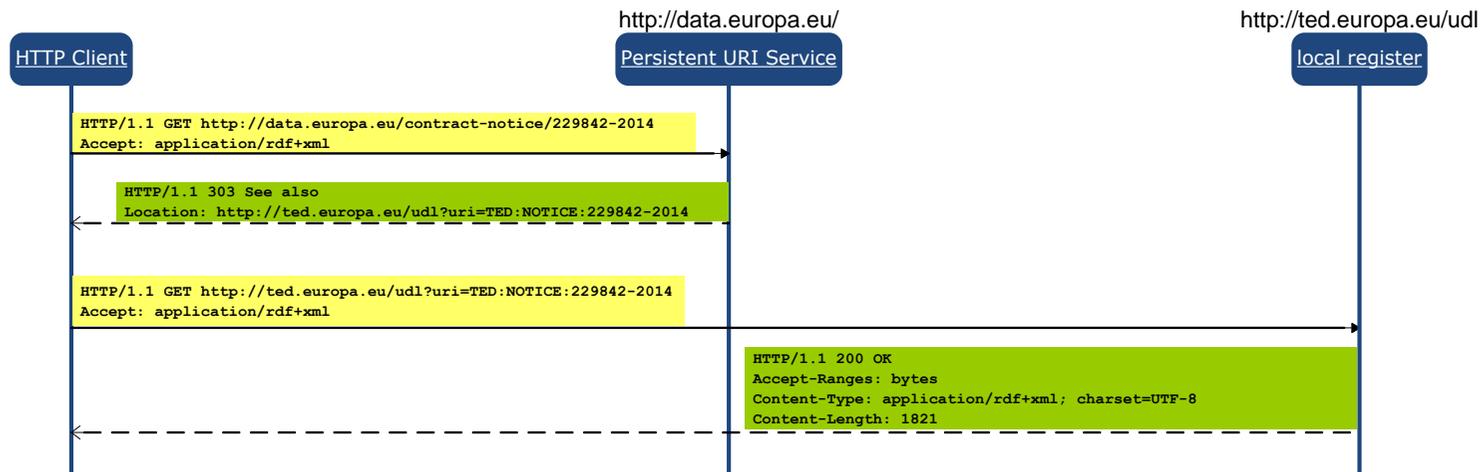


Figure 3: Sequence diagram: external HTTP redirection (forwarding)

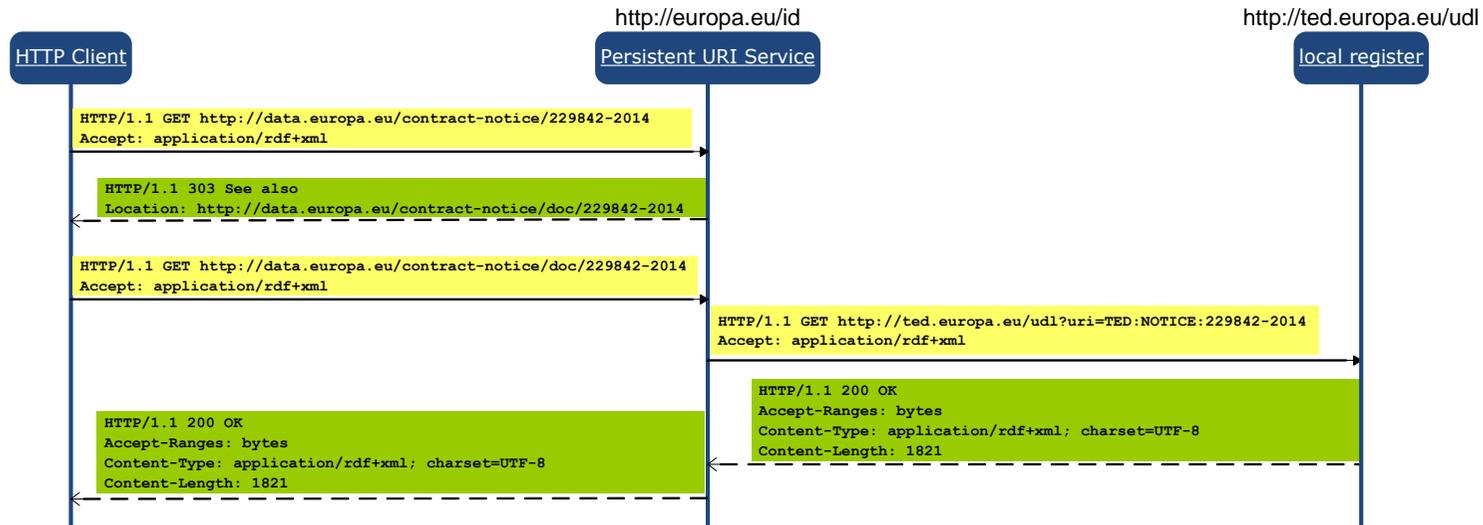


Figure 4: Sequence diagram: internal HTTP redirection (proxy forwarding)

6 CONCLUSION AND NEXT STEPS

This report summarises the business case and specifications for a common approach for the governance and management of persistent URIs by EU institutions. It results from the work of an informal, inter-institutional Task Force, and builds on a previous proposal on persistent URIs prepared in the period 2012 – 2013 by the same Task Force. The report consists of:

- a business case;
- specifications for the governance and management of persistent URI sets;
- design rules for persistent URI sets; and
- functional specifications for a persistent URI service.

6.1 Agreements

The table below summarises on which topics in this report a consensus was reached among the members of the informal URI Task Force through a number of (virtual) meetings from January through November 2014.

Table 8: Agreements and open questions

Topic	Status (agreement / open question)
Section 2 Business case	
• Problems: the identified problems in Section 2.2 justify the need for common policy on persistent URIs.	Agreement
• Solutions: The proposed solutions listed in Section 2.3 address the needs.	Agreement
Section 3 Governance and management of persistent URI sets	
• Scope criteria: The scope criteria listed in Section 3.3 help decide whether a URI set should be placed under common governance.	Agreement
• URI governance structure: The governance structure described in Section 3.4 (one central register of URI namespaces and many local registers of resources) is adequate for the governance of centrally co-ordinated namespaces	
• URI management process: The URI management process (for URI sets under central governance) described in Section 3.5 provides a good basis for the management of the URI namespaces.	Agreement
Section 4 Design rules for persistent URI sets	

<ul style="list-style-type: none"> • New persistent URI sets: The design rules for persistent URI sets should only be mandatory for <i>new</i> URI sets that are <i>placed under common governance</i>. Legacy URI sets placed under common governance do not need to comply with these rules. 	Agreement
<ul style="list-style-type: none"> • Design rules: The proposed design rules in Section 4 are accepted. The rules for the tail-section are suggested patterns, and have the modality of guideline only. 	Agreement
Section 5 URI redirection service	
<ul style="list-style-type: none"> • Functional specifications: The persistent URI redirection service described in Section 5 should help maintain a central register of URI namespaces and provide a redirection mechanism (URI forwarding or proxy forwarding) for HTTP requests. 	Agreement

6.2 Next steps

To finalise this consensus, the following next steps are proposed:

- **Conduct pilots:** DG Employment is working together with the Publications Office to put in place the first persistent URI namespace for the European classification of Skills, Competences, Qualifications and Occupations (ESCO) on the URI namespace 'http://data.europa.eu/esco/'. Furthermore, DG SANCO, the Publications Office, and DIGIT are planning to conduct other pilots for health and consumer data, the Core Vocabularies, the European Legislation Identifier (ELI) and Digital Object Identifiers (DOIs);
- **Inform key stakeholders:** The study and proposed persistent URI policy will be presented to key stakeholders. This includes a presentation of Proposal for URI policy to Management Committee of Publications Office. This also includes further promotion and awareness raising activities in relevant fora;
- **Inter-service consultation:** an inter-service consultation will be launched at European Commission; and
- **Continuation of the study:** In preparation to operationalizing the policy a set of guidelines and best practices for persistence will be drafted. Furthermore, the study will be continued to draft short technical specifications for implementers, and identify additional patterns for structuring the local part persistent URIs based on best practices.

7 ACKNOWLEDGEMENTS

Specific acknowledgement is due to:

Table 9 - Organisations and their representatives in the URI Task Force

Person	Organisation
Carlo Marzocchi	Council of the EU
Lucie Verachten	Council of the EU
Ángeles Noguerol Álvarez	European Commission, DG COMM
Laurent-Jacques Corveleyn	European Commission, DG COMM
Stephan Jaeger	European Commission, DG COMM
Stefano Abruzzini	European Commission, DG CONNECT
Stefano Bertolo	European Commission, DG CONNECT
Henri Putseys	European Commission, DG DIGIT
Rosa Ordinana Calabuig	European Commission, DG DIGIT
Athanasios Karalopoulos	European Commission, DG DIGIT, ISA Programme
Susanne Wigard	European Commission, DG DIGIT, ISA Programme
Vassilios Peristeras	European Commission, DG DIGIT, ISA Programme
Giorgos Georgiannakis	European Commission, DG Health and Consumers (SANCO)
Ruth Ní Bhraonáin	European Commission, DG Health and Consumers (SANCO)
Anders Friis-Christensen	European Commission, DG JRC
Andrea Perego	European Commission, DG JRC
Angelo Quaglia	European Commission, DG JRC
Jesus Hermida	European Commission, DG JRC
Yannis Kanellopoulos	European Commission, DG JRC
Gerard Hilbert	European Parliament, DG ITEC
Carol Bream	European Commission, EAC
Carol Riccalton	Publications Office
Jose Pastor Camarasa	Publications Office
Marc Wilhelm Küster	Publications Office
Mylona Polyxeni	Publications Office
Norbert Hohn	Publications Office
Peter Schmitz	Publications Office
Willem van Gemert	Publications Office

8 BIBLIOGRAPHY

- [BM4LOGD] Interoperability solutions for European public administrations (ISA). Archer, P., Dekkers, M., Goedertier, S., & Loutas, N. (2013). *Study on business models for Linked Open Government Data (BM4LOGD)*. European Commission, ISA Programme. <https://joinup.ec.europa.eu/node/72473>
- [DCAT-AP] Interoperability solutions for European public administrations (ISA). (2013). DCAT application profile for data portals in Europe. https://joinup.ec.europa.eu/asset/dcat_application_profile/description
- [EuroVoc] Publication Office. (2013, June 28). *EuroVoc, the EU's multilingual thesaurus*. Retrieved December 11, 2013, from European Union Open Data Portal: <http://open-data.europa.eu/en/data/dataset/I57spZ6CMZMFAIvArDiHaw>
- [EuroVoc] Publications Office. (2010). *EuroVoc, a path to the semantic web*. Publications Office.
- [Fielding, 2000]. Architectural Styles and the Design of Network-based Software Architectures.
- [High-Value datasets, 2014] Interoperability solutions for European public administrations (ISA). De Keyzer M., Bargiotti L., Goedertier S., Loutas N., & Van Geel F (2014). Report on high-value datasets from EU institutions. European Commission, ISA Programme.
- [IPG, 2013] *Information Providers Guide: THE EU Internet Handbook: Web addresses / URLs*. Retrieved December 11, 2013, from European Commission: http://ec.europa.eu/ipg/basics/urls/index_en.htm
- [MDR] Publications Office of the European Union. (2013, December 06). *MDR - Authorities*. Retrieved December 11, 2013, from MDR Metadata Registry: <http://publications.europa.eu/mdr/authority/index.html>
- [ODP Proposal, 2013] Bargiotti, L., Küster, M., & Schmitz, P. (2013, 03 07). EU Open Data Portal (ODP): Template for Uniform Resource Identifiers (URIs) - Version: 0.9.6 (draft).
- [Open Data] European Commission. (2011, December 12). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions. *Open data - An engine for innovation, growth and transparent governance*. Brussels, Belgium.
- [OpenDefinition] Open Knowledge Definition. (n.d.). *Open Definition*. Retrieved 11 20, 2013, from Open Definition: <http://opendefinition.org/>
- [PURI] *10 Rules for Persistent URIs*. European Commission, ISA Programme. Archer, P., Goedertier, S. & Loutas, N. (2012, December 17). <https://joinup.ec.europa.eu/node/53858>

- [RFC3986] *Uniform Resource Identifier (URI): Generic Syntax*. T. Berners-Lee, R. Fielding and L. Masinter (2005). <http://tools.ietf.org/html/rfc3986>
- [Rodriguez, 2008]. *RESTful Web services: The basics*. Retrieved from IBM developerWorks:
<https://www.ibm.com/developerworks/webservices/library/ws-restful/>
- [SANCO, 2013] European Commission. (2012). *DG SANCO Data Inventory*. Retrieved December 11, 2013, from European Commission:
http://ec.europa.eu/dgs/health_consumer/information_systems/
- [TBL, 2006] Linked Data Design Issues. 2006.
<http://www.w3.org/DesignIssues/LinkedData.html>
- [URI Task Force, 2013] Ní Bhraonáin, R., & Georgiannakis, G. (2013, 02 26). Draft proposal from the informal working group on a short-term approach to persistent uniform identifiers (URIs).
<https://webgate.ec.europa.eu/CITnet/confluence/display/PURI/Suggestions>

Annex I. INTERVIEW QUESTIONS

The following questions were asked during individual interviews with members of the URI Task Force, representing their institutions.

Business case

- Do you agree with the business case? Do you see anything missing?
- Which costs would a coordinated effort entail (inter-institutional governance, URI service)?
- What should be the scope of the common approach? For which URI sets do we need a common approach?

Design principles

- Which URI structure to impose? E.g. one sub-domain (purl.europa.org) or multiple? Should URI sets share a common level-3 domain, such as data.europa.eu? Or should we use different subdomains? Should every URI set created by EU institutions be on the Europa.eu domain?
- What should be strict rules and what should be guidelines?
- How to deal with legacy URI sets?
- What comments do you have on the proposed URI template?

Governance

- Should there be a permanent inter-institutional governance of URI sets?
- Is there an existing governance body that could assume this task?

URI management software

- A single service instance or multiple instances (linked to the level-3 domains)?
- Configuration on Web server or customisable solution?
- Open-source or proprietary?
- Authorisation and delegation mechanisms?
- Do you agree with the "URI Management" features? Which additional requirements do you have?
- One instance or multiple instances (e.g. per sub-domain)?
- What should be the relationship with the Inter-institutional Editorial Committee for Internet (CEiii)?
- When does a URI set require a common governance and should be subjected to inter-institutional governance?

To gather the requirements for EU institutions, the following stakeholder meetings were held.

Table 10 – Overview of stakeholder meetings

Date	Meeting
2014-31-01	1st virtual meeting of URI Task Force
2014-02-06	Individual Interview with Commission Central Library
2014-02-14	2nd virtual meeting of the URI Task Force
2014-02-14	Individual interview with DG JRC
2014-02-18	Individual Interview with the Publications Office
2014-02-19	Individual Interview with the Council of the EU
2014-02-21	Individual Interview with DG COMM
2014-02-26	Individual Interview with European Parliament
2014-02-25	Individual Interview with DIGIT.A.3
2014-03-07	3rd virtual meeting of the URI Task Force
2014-03-13	Individual Interview with DG SANCO
2014-03-14	4th virtual meeting of the URI Task Force
2014-03-18	Individual Interview with DG CONNECT
2014-04-04	5th virtual meeting of the URI Task Force
2014-05-08	6th virtual meeting of the URI Task Force
2014-07-10	1 st Virtual meeting of subgroup to discuss alternatives
2014-08-28	2 nd Virtual meeting of subgroup to discuss alternatives
2014-09-19	Face-to-face meeting of the URI Task Force
2014-11-13	7th virtual meeting of the URI Task Force

Annex II. ANALYSIS OF THE CURRENT SITUATION

This section describes the current usage of HTTP-based URI sets by EU Institutions.

II.1. Inter-institutional Editorial Committee for Internet (CEiii)

The Information Providers Guide for EUROPA websites²⁸ (European Commission, 2010) provides accurate information about the current rules, governance, and management of Web Addresses (URLs) for EU institutions and agencies. However, the Information Provider Guidelines do not specify particular requirements for the management of URIs (or IRIS).

The Information Providers Guide mentions the following rules:

1. Following a decision of the Secretariat-General, all official websites of the European Institutions and Agencies must use a URL address in the second level domain **europa.eu** with the following syntax: **(ec.)europa.eu/sitename**. The second level domain europa.eu is managed by the Commission.
2. The creation of third-level domains follows the rules issued by the **Inter-institutional Editorial Committee for Internet (CEiii)**. These include among others:
 - a. **Websites of EU Institutions:** The names of the institutions and agencies' sites are composed by adding a third level identification (for example for the Commission ec.europa.eu or the Parliament europarl.europa.eu).
 - b. **Special websites:** Third level names other than Institutions' names can be attributed for special sites, **subject to the approval of the CEiii**. The creation of such third level names should nevertheless be the exception. Preference must be given to the creation of sub-sites within the europa.eu or ec.europa.eu domain.
 - c. **Virtual folders and 302 redirection:** Usage of short names in virtual folders with 302 redirections is allowed. Example: <http://ec.europa.eu/roaming> redirecting to http://ec.europa.eu/information_society/activities/roaming/index_en.htm
 - d. **URL shortener service:** Long URLs are transformed into shorter versions with a random code and remain accessible through these URLs.

Currently EU institutions can mint their own URI sets within the boundaries set by Information Providers Guide for EUROPA websites on website URLs and the Inter-institutional Editorial Committee for Internet (CEiii). There is no requirement that URI sets should follow design principles beyond the Website URL. The lack of design

²⁸ Information Providers Guide http://ec.europa.eu/ipg/basics/urls/index_en.htm

principles and governance for URI sets does not give confidence to third parties to reuse these URI sets.

The screenshot shows the 'The EU Internet Handbook' website. The main content area is titled 'Web addresses / URLs'. A red box labeled 'IPG rule' contains the text: 'All official websites of the European Institutions and Agencies must use URL address in the second level domain **europa.eu** with the following syntax: **(ec.)europa.eu/sitename**.' Below this is a link 'View all IPG Rules'. To the right, a box titled 'On this page' lists several topics: Domain names, europa.eu is the common domain of the EU institutions, Naming rules in the europa.eu domain, Use of short names, Syntax of URLs and its promotion, URL structure, Multiple word names, and Request a web address/URL. Below the main content, a section titled 'Domain names' states: 'europa.eu is the common domain of the EU institutions'.

Figure 5: Screenshot: Information Providers Guide for EUROPA websites

II.2. URI sets maintained by the Council of the European Union

The Council of the European Union is expecting to create URI sets for entities and datasets for which it is the authoritative source, these include council meetings and important datasets (such as votes by Council members) around these meetings and (archival) documents.

In 2010, the Council started a digitisation project for the [Archives of the Council of the European Union](#). Archival documents are being digitised and will soon be available on line.

The historical paper archives are transferred to the European University Institute (EUI) in Florence after 30 years. It is the intention not to *duplicate* digitised documents, but rather to rely on the given HTTP URIs as uniform, unique, and permanent identifiers for documents.

II.3. URI Sets Maintained by DG COMM

DG COMM maintains a number of documents that follow a pattern, such as the RAPID repository of press releases. For example

http://europa.eu/rapid/press-release_IP-14-195_en.htm

This identifies a specific press release in a specific format in a specific language (*Debate on the future of Europe: Commissioner Barnier and French Minister Repentin will debate issues with citizens at Plaine Saint Denis*, in English, published as a Web page). The same press release is available in French and German and this is reflected in the last two characters of the URI before the final dot. (i.e. *_de.htm* and *_fr.htm*). However, these are not designed to be persistent and changed in the recent revamp of the RAPID system.

A similar pattern is used in the Eurobarometer archives. Again, A URI such as

http://ec.europa.eu/public_opinion/archives/ebs/ebs_130_en.pdf

identifies a specific language version of a specific format of a specific document (that may or may not be available in different languages, this one is not). There is no common identifier for DG COMM resources for an abstract resource that can then be accessed in specific languages and/or formats. Indeed, DG COMM resources are only ever made available in a single format and does not create URIs for a conceptual resource and its concretisation in specific languages and formats.

II.4. URI sets maintained by DG CONNECT

DG CONNECT maintains a semantic data repository, currently only dedicated to statistical data for the Digital Agenda Scoreboard, but that could be used for publishing other non-statistical datasets. Due to timing constraints and practical arrangements the resource URIs are not located on Europa.eu. The following URI is an example:

http://semantic.digital-agenda-data.eu/data/digital-agenda-scoreboard-key-indicators/p_iuse/ent_all_xfin/pc_emp/AT/2013

The URIs are dereferenceable.

DG CONNECT does not mint URI sets on europa.eu; this may reduce confidence of third party users wishing to reuse these URI sets.

II.5. URI sets maintained by DG DIGIT

DG DIGIT has created two URI sets for Asset Description Metadata Schema (ADMS) and Core Public Service Vocabulary (CPSV):

1. The original specification of the [ADMS](#)²⁹ includes six controlled vocabularies (see Table 11). The URI pattern for each term is the same:

`http://purl.org/adms/{ConceptScheme}/{Concept}`

where:

`{ConceptScheme}` is the controlled vocabulary and `{Concept}` is the specific term within the vocabulary.

As Table 11 shows, the URIs of the vocabularies themselves use a version number instead of the `{Concept}` element.

2. The CPSV was developed to enable the exchange of information about public sector services. It includes a basic model of the service's inputs and outputs together with links to the legislative and policy framework in which the service operates.

Like the ADMS controlled vocabularies, the CPSV uses the permanent URL service operated by OCLC (purl.org) as the guarantor of the persistence of its URIs and it follows a similar URI pattern of

²⁹ https://joinup.ec.europa.eu/asset/adms/asset_release/adms

`http://purl.org/vocab/{vocabName}#{vocabTerm}`

Like many RDF vocabularies, terms in the CPSV are defined using fragment identifiers rather than path elements (i.e. the separator is a # character, not a /). Fragment identifiers are not transmitted across the wire so that an HTTP request for any term in the vocabulary is equivalent to a request for `http://purl.org/vocab/cpsv` (it is the browser that then deals with the fragment after the # character).

Table 11 – URI set of the ADMS Controlled Vocabularies

ADMS Controlled Vocabulary	URI
Asset Type	<code>http://purl.org/adms/assettype/1.0</code>
Interoperability Level	<code>http://purl.org/adms/interoperabilitylevel/1.0</code>
License Type	<code>http://purl.org/adms/licencetype/1.0</code>
Publisher Type	<code>http://purl.org/adms/publishertype/1.0</code>
Representation Technique	<code>http://purl.org/adms/representationtechnique/1.0</code>
Status	<code>http://purl.org/adms/status/1.0</code>

The URI sets of the ADMS controlled vocabularies and the Core Public Service Vocabulary are not based on the europa.eu second-level domain, but on the purl.org domain. The lack of a formal governance model for purl.org does not give confidence to third-parties to re-use these URI sets. It is not clear whether the European Commission will maintain them in the future.

II.6. URI sets on the EU Open Data Portal (EU ODP)

The EU ODP offers a single point of access to data of EU institutions and other EU bodies. It also provides an identifying mechanism for the dataset once listed on the EU ODP. The proposed URI template for the EU ODP is as follows [Bargiotti, Küster, & Schmitz, 2013]:

`http://open-data.europa.eu/{odp-entity-type}/{odp-entity-name,representation-UUID*}`

Where:

`{odp-entity-type}`: "dataset"

`{odp-entity-name}`: stable, concise and unique name of the dataset given by the data provider

`{representation-UUID}`: optionally, an arbitrary stable Universally Unique Identifier (UUID) of an individual representation / resource of the data set, preceded by the string "resource/"

Example 1: URI of a dataset

`odp-entity-type` := "dataset"

`odp-entity-name` := "fp7-projects-2012"

<http://open-data.europa.eu/dataset/fp7-projects-2012>

Example 2: Reference to representation of a data set

odp-entity-type := "dataset"

odp-entity-name := "fp7-projects-2012"

representation-UUID := ("resource", "8015a116-09f0-4515-adb8-faf40139ac9c")

<http://open-data.europa.eu/dataset/fp7-projects-2012/resource/8015a116-09f0-4515-adb8-faf40139ac9c>

It may not be needed to use a UUID to identify a particular distribution of a dataset. The more usual approach is simply to append a file extension to the URI of the dataset so that if we take the dataset URI to be <http://open-data.europa.eu/dataset/fp-projects-2012> then the CSV distribution of that dataset could simply be <http://open-data.europa.eu/distribution/fp-projects-2012.csv>.

The URI set proposal for the EU ODP [Bargiotti, Küster, & Schmitz, 2013] also states that URIs defined following the URI template for datasets should point to the metadata description of a dataset exposed by the ODP. Whenever a client requests another *representation* of a resource, like for instance an MS-Excel or XML file, he is redirected to the specific distribution URL of the data provider.

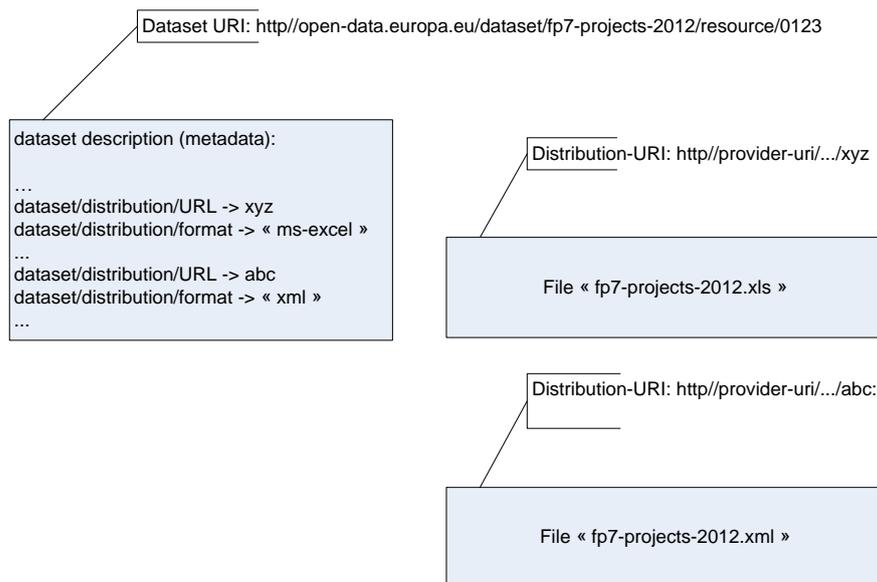


Figure 6: EU Open Data Portal (ODP): Proposed template for Uniform Resource Identifiers (URIs) [Bargiotti, Küster, & Schmitz, 2013]

According to DCAT, a dataset is an abstract concept and the URI for it may not be directly dereferenceable. As in the ODP model above, it has the separate notion of distributions and there may be multiple such distributions of the same dataset. Metadata about a dataset is provided in a catalogue record and this has its own URI.

II.7. URI sets maintained by the European Environment Agency (EEA)

Many of the important datasets held and managed by the EEA have been published as Linked Data for many years. For example, by browsing through the [GEMET thesaurus](#)³⁰ a user may navigate to the term 'adsorption.' The URL shown in the browser is

<http://www.eionet.europa.eu/gemet/concept?cp=120&langcode=en&ns=1>

At the same time, at the bottom of the page there is a persistent URI:

<http://www.eionet.europa.eu/gemet/concept/120>.

So that a human can see that 'adsorption' corresponds to the concept number 120 in the GEMET thesaurus. Dereferencing that persistent URI will lead (after redirection) to the relevant Web page. There is no content negotiation in place, but it is possible to download the underlying RDF data and there is a SPARQL endpoint giving access to all the available data.

The European Environment Information and Observation Network run by the EEA continues to develop its linked data platform and its URIs are clearly designed for persistence and, rightly, separated from the technology that returns data from them.

II.8. URI sets maintained by DG EUROSTAT

Eurostat does not mint *official* URI sets for its statistical data or structural metadata. However, through the LATC project, an important amount of [Eurostat data](#)³¹ was made available as 5 star Linked Open Data. In each case the URIs used follow the same pattern:

<http://eurostat.linked-statistics.org/{type}/{subtype}>

with 3 possible values for {type}:

- dic (dictionaries)
- dsd (data structure definitions)
- data (the data itself)

The data is statistical in nature so that, following the norms of that discipline, it is correct to talk about 'observations' – i.e. specific data points. These are all encoded thus:

<http://eurostat.linked-statistics.org/data/{dataset}#{dimension1},{dimensionN}>

Although the LATC project has come to an end, the Eurostat linked data continues to be updated weekly and the underlying system is likely to be updated throughout the year 2014,³² i.e. this remains an active undertaking.

³⁰ <http://www.eionet.europa.eu/gemet/>

³¹ <http://eurostat.linked-statistics.org/>

³² <https://twitter.com/cygri/status/421617713442611200>

Eurostat does not mint URI sets. However, the URI sets on the eurostat.linked-statistics.org domain, i.e. not a subdomain of europa.eu; this may reduce confidence of third party users wishing to reuse these URI sets.

II.9. URI sets maintained by DG Health and Consumers (DG SANCO)

DG SANCO is one of the pioneers in EC-wide initiatives for persistent URIs [Bhraonain & Georgiannakis, 2013]. As reported in BM4LOGD, DG SANCO is operating an experimental LOGD service, which serves the following purposes:

- **Flexible data integration:** serving both external reusers and also intra-organisational needs.
- **Increase in data quality:** Linked Data helps DG Health & Consumers identify quality problems in legacy data and act on it.
- **Development of new services:** 2 Linked Data apps have already been developed, the RDFa maker and the Forest Reproduction Material client application to support the automatic publication of the common registry of primary material of forest reproductive material from the local MS registries.

The URI pattern followed by DG SANCO (see below) is the same as the one from the original/last URI report received from the Publications Office and is based on best practice. DG SANCO does not identify datasets, but identifies specific data resources within the datasets.

<http://open-data.europa.eu/data/dataset/<catalogue-record-name>>

while the resources within the datasets follow the following pattern:

http://ec.europa.eu/semantic_webgate/dataset/{concept}pesticides/resource/{local id}

where {concept} denotes the different types of resources, e.g. additive, pesticide etc.

The URIs of DG SANCO are dereferenceable and navigable, if you click one URI you get information about another. There are multiple endpoints/technologies in place allowing the exploitation of the URIs and the respective data. Among others, DG SANCO has made a SPARQL endpoint available at http://ec.europa.eu/semantic_webgate/query.

Other possible distributions of the data include CSV, RDF, RDFa and JSON formats which are machine readable and can be oriented for client applications that can consume this data.

As reported in the BM4LOGD study, all the data that DG SANCO makes publicly available is accompanied by an EU disclaimer³³, which clarifies among others that the data is:

- of a general nature only and is not intended to address the specific circumstances of any particular individual or entity;

³³ EU disclaimer – legal notice, http://ec.europa.eu/geninfo/legal_notices_en.htm

- not necessarily comprehensive, complete, accurate or up to date;
- sometimes linked to external sites over which the Commission services have no control and for which the Commission assumes no responsibility;
- not professional or legal advice (if you need specific advice, you should always consult a suitably qualified professional).

The open data of DG Health & Consumers is hosted on Circa-BC³⁴ and is covered by the SLAs of that system. However, there is no SLA in place yet for the LOGD infrastructure.

II.10. URI sets maintained by the Joint Research Centre (DG JRC)

The [INSPIRE registry](#)³⁵ is managed by DG JRC. It serves as a central access point for identifiers, including their labels and descriptions. At the time of writing, the first 4 URI sets are in place and all the URIs are dereferenceable, returning data in one of 4 available formats via content negotiation (HTML, JSON, XML or ATOM). However, since the registry is still in its infancy stage, certain aspects still need to be taken care of. For example, multilingual labels have not been added yet.

At the time of writing there are four datasets in the registry (Table 12) all of which use the familiar URI pattern:

`http://inspire.ec.europa.eu/{type}/{subtype}`

where:

`{type}` indicates the dataset and `{subtype}` indicates the specific item in the registry.

Table 12 – INSPIRE Registry Data Sets

INSPIRE registry	URI
INSPIRE theme register	<code>http://inspire.ec.europa.eu/theme/</code>
INSPIRE application schema register	<code>http://inspire.ec.europa.eu/applicationschema/</code>
INSPIRE code list register	<code>http://inspire.ec.europa.eu/codelist/</code>
INSPIRE feature concept dictionary	<code>http://inspire.ec.europa.eu/featureconcept/</code>

II.11. URI sets maintained by the Publications Office (PO)

The Publications Office attributes several sets of URIs to different types of data, including:

1. **CELLAR** is the descriptive metadata and content repository of the PO. Publishing information and committing to its long term preservation and

³⁴ <https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp>

³⁵ INSPIRE registry: <http://inspire.ec.europa.eu/registry>

stability has always been an important aspect of the OP's work. In CELLAR, the URIs themselves are designed carefully for long term management and stability. Every URI begins with the same pattern:

`http://publications.europa.eu/{type}/{subtype}`

where there are among others the following possible values for `{type}`:

1. `resource`, for content and metadata resources;
2. `ontology`, for schemas;
3. `webapi`, for Web API services.

For example, editions of the Official Journal all have URIs beginning with:

`http://publications.europa.eu/resource/oj/`

where 'oj' acts as the subtype.

The OP makes extensive use of content negotiation (section 2.3) and language negotiation. Many items published by the OP are 'works' within the FRBR sense of the word. Each work has its own URI and CELLAR returns a specific manifestation of that work based on the HTTP Request headers.

CELLAR makes use of its HTTP server's native support for content negotiation for the first of these but not the second. That is, the server inspects the Accept header in the HTTP request and returns HTML (for humans), RDF or XML. (All HTTP requests include information about the device making the request. In the case of a Web browser, this will include the type of browser, operating system and language preferences). One of these three formats is always returned. HTTP is less deterministic for languages so that if the request header specifies a language in which the particular work is not available, the server response can vary between implementations. The canonical response is either 'No Acceptable Variant' or 'Multiple Choices' - neither of which may be helpful for some users and so CELLAR uses its own software to always return a representation of the work. Each manifestation of a work, that is, a particular version of the requested resource in a specific language and specific data format, has its own URI and this can of course be accessed directly.

2. **Named Authority Lists (NALS)** serve for the harmonisation of codes and labels used by the PO and other EU institutions and bodies to promote the exchange of information at the interinstitutional level (Publications Office of the European Union, 2013). The URI pattern for the NALS is:

<http://publications.europa.eu/resource/authority/{name}>

NALS are encoded in a variety of formats, including a SKOS concept schemes. NALS are made available by the PO itself through the Metadata Registry (MDR) and are also accessible via the EU ODP³⁶. The table below lists the URI entries for the identified NALS in the MDR.

³⁶ European Union Open Data Portal for NALS in the MDR:

Table 13 – URI sets of NALs

NAL	URI
Corporate bodies	http://publications.europa.eu/resource/authority/corporate-body
Countries	http://publications.europa.eu/resource/authority/country/
File types	http://publications.europa.eu/resource/authority/file-type/
Interinstitutional procedures	http://publications.europa.eu/resource/authority/procedure/
Languages	http://publications.europa.eu/resource/authority/language/
Multilingual	http://publications.europa.eu/resource/authority/multilingual/
Resource types	http://publications.europa.eu/resource/authority/resource-type
Roles	http://publications.europa.eu/resource/authority/role/
Treaties	http://publications.europa.eu/resource/authority/treaty/
Currencies	http://publications.europa.eu/resource/authority/currency/
Places	http://publications.europa.eu/resource/authority/place/

3. **Eurovoc** is a multilingual thesaurus describing several EU activities (Publications Office, 2013). The format for URI sets for EuroVoc is:

<http://eurovoc.europa.eu/{item}> where:

item is a unique number per concept or language (Publications Office, 2010).

Some examples:

The URI that identifies the concept for Climate Change is: <http://eurovoc.europa.eu/5482>.

The same concept in a specific language, e.g. Spanish, is <http://eurovoc.europa.eu/125206>.

4. The PO also disseminates legal information through the **EUR-Lex website**.³⁷ URIs identifying documents falling within this category are built using different rules. In certain cases the CELEX number is used, see:

HTML: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31966L0162:EN:HTML>

In other cases the natural number of the document is used instead:

<http://open-data.europa.eu/en/data/publisher/publ>

³⁷ The domain name of EUR-Lex is: <http://eur-lex.europa.eu>. The domain name of new EUR-Lex is <http://new.eur-lex.europa.eu>

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2013:0574:FIN:EN:PDF>

In addition, there are URIs for documents published in the Official Journal:

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=DD:I:1965-1966:31966L0162:EN:PDF>

5. European Legislation Identifier (ELI)

The European Legislation Identifier (ELI) aims to provide common guidelines for the identification of legal information through the use of http URIs both at European and national level³⁸. ELI is composed of three pillars:

- A HTTP URI template to identify in a unique and persistent way legal information;
- A set of metadata to describe legal information and
- A metadata schema (ontology).

An example of the URI pattern is as follows (each element being optional without a predefined order):

```
/eli/{jurisdiction}/{agent}/{sub-agent}/{year}/{month}/{day}/{type}/{natural identifier}/{level 1...}/{point in time}/ {version}/{language}
```

At the time of writing the PO is finalising the analysis for the implementation of ELI.

6. European Case Law Identifier (ECLI)

The European Case Law Identifier identifies identify in a unique way case law at both European and national level. The main difference compared to ELI is that it is not dereferenceable.

The PO has already implemented the 'ECLI' metadata field in order to enable users to look for case-law by using the ECLI identifier.

The URI sets attributed by the PO have different level-three domains and use different URI structures. Nonetheless, the OP is making its best effort to create a stable URI set, designed, managed and published with longevity in mind. The OP has a track record of maintaining URIs for more than 10 years and cannot today see any reason why the URIs embodied in CELLAR will not persist. The subdomain, publications.europa.eu, is as stable as any can be and was chosen deliberately for that reason. Even if the name of the institution were to change, the subdomain is sufficiently generic that it could easily survive. This would not be the case if, for example, the name CELLAR, i.e. the project that created it, had been used as the subdomain.

³⁸ Official Journal C 325, 26.10.2012 <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2012:325:FULL:EN:PDF>

Annex III. ANALYSIS OF ALTERNATIVES

This Annex contains a reflection of the work that was carried out by a sub-group of the Persistent URI Task Force during the months of July and August 2014, identifying alternatives for the proposed persistent URI policy according to various aspects. For each alternative, the pros and cons are listed.

III.1. Scope

In section 3.3, five scope criteria are listed. The fifth criterion is:

Local register: *URI sets for which an EU institution maintains an official (possibly paper-based) register and enforces local control over the identifying mechanism. In such case, resources already have an identifier, called a **local id**. Resources for which no register exist are likely not to meet many of the aforementioned criteria.*

Table 14 – Analysis of alternatives: what to do when the local register is absent?

<p>1. No URI namespace without local register</p>	<p>Resources for which no local register (even no “paper-based” register) exists are likely not to have an authoritative source and no machine-readable information may be available for them.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Scope restricted to high-value resources and resources that are already well maintained, reducing start-up cost, and most likely covering most relevant resources <p>Cons:</p> <ul style="list-style-type: none"> - May miss out on relevant resources (e.g. historical information) for which a local register could be created.
<p>2. URI namespace for any collection as requested by owner</p>	<p>Resource collections may exist that have not been formally described and registered. The central URI namespace could be a starting point for building a register, either formally in a local register or in an external catalogue.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Larger collection of resources addressable through the central URI application, capturing additional high-quality resources - Support for building local or external registers where they don't exist yet <p>Cons:</p> <ul style="list-style-type: none"> - Potentially difficult to determine quality and coherence of the collection if it hasn't been previously managed

Based on this analysis, the following proposal was accepted after discussion in the Persistent URI Task Force:

Keep the scope to only include collections for which a publically accessible local register exists. Consider opening up for additional collections later.

III.2. Operational model

In section 3.4, two possible operational models are described: a **Forwarding (redirection)** model in which a central application dispatches requests for URIs to a local handler and disengages itself from further communication, and a **Proxy forwarding** model in which a central application acts as an intermediary between the requester and the local register during all phases of communication.

Table 15: Analysis of alternatives: operational model

1. Forwarding (redirection)	<p>The central service issues a 30x redirect for each GET request on a centrally administered URI, then communication between client and local register takes place between those two without involvement of the central service.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Less stress on the central service because one request results in one response to the client (with a 30x redirect) and nothing more. - Allows monitoring of number of accesses to a particular resource collection and allows checking that the tail conforms to established policy. - Conserves visibility of the local register, indicating the responsible organisation for the information and providing context, this may be important in the case of human users. <p>Cons:</p> <ul style="list-style-type: none"> - No end-to-end service monitoring possible because the communication between client and local register is direct and is not visible to the central service.
2. Proxy forwarding	<p>The central service issues redirects that come back to the central service and all communication between client and local register passes through the central service.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Allows end-to-end monitoring because all communication between client and local register passes through the central service. <p>Cons:</p> <ul style="list-style-type: none"> - Creates overhead on the central service. - No visibility of the local register.

Based on this analysis, the following proposal was accepted after discussion in the Persistent URI Task Force:

Implement the forwarding (redirection) model as this puts less stress on the central service. Special considerations should be made regarding scalability of the solution.

III.3. Central versus local responsibilities

In section 3.4, the governance model is outlined with central co-ordination of URI namespace and local responsibility for the maintenance of a local register of resources.

Table 16 – Analysis of alternatives: different governance structures

1. central register of URI namespaces – local register of resources	<p>For URI sets placed under common governance (scope criteria), the URI namespace is decided centrally. Decisions about the tail, such as the {local id} are taken at the local level, under central guidance.</p> <p>Example:</p> <ul style="list-style-type: none">- <i>Type & Concept identifier & local identifier</i> http://data.europa.eu/id/vessel/*- <i>Concept identifier & local identifier</i> http://data.europa.eu/vehicle/* <p>Pros:</p> <ul style="list-style-type: none">- Harmonisation and interoperability: a central decision making body coordinates URI namespaces avoiding synonyms and fragmentation of namespaces; the central decision making body would not assign a parallel namespace for a register of vehicles if one already exists.- Division of roles and responsibilities: a central decision making body coordinates responsibilities in cases of positive competence conflicts; if multiple requests for namespace for similar sets of resources (e.g. vehicles) are received, the central decision body would create a single namespace and assign the responsibility for the namespace to one of the requesting organisations.- Determining whether a collection falls within the scope criteria is done once for the owner and collection; when validated, the owner of the collection can add URIs for the individual resources without central involvement.- The URI includes no indication of ownership – so changes of ownership will not result in changes of the URIs. It only contains an indicator about the type of resource.- Speed and efficiency for managing redirection rules via a <i>central</i> Persistent URI service; a namespace per collection keep the redirection tables small, reducing
--	--

	<p>maintenance overhead.</p> <ul style="list-style-type: none">- Workload is distributed: collection owners can maintain redirection rules themselves, under central guidance as necessary.- Namespace-based redirection allows for moving of collections; only a single rule has to be updated to point to another location.- Monitor service levels centrally via a <i>central</i> Persistent URI service. <p>Cons:</p> <ul style="list-style-type: none">- Loss of local control over the URI namespace creates interdependency: EU institutions need to request ownership over individual URI namespaces for resources that fall within the scope criteria of the inter-institutional URI policy.- The inclusion of a string that identifies the type of resource (linked to the local register) adds meaning in the URI (which is susceptible to change over time) and leads to longer URIs.
<p>2. Central register of short URIs</p>	<p>The central URI register assigns random numbers through an algorithm that ensures uniqueness within the context of the central register.</p> <p>Example:</p> <p>http://data.europa.eu/123456789</p> <p>Pros:</p> <ul style="list-style-type: none">- Harmonisation and interoperability: a central decision making body decides for which collections central URIs can be assigned; the central body will determine if the collection meets the scope criteria and does not overlap with other collections to avoid synonyms being created;- Division of roles and responsibilities: a central decision making body coordinates responsibilities in cases of positive competence conflicts; if a collection is proposed for central URI assignment, the central decision body would investigate if there is already a similar collection and, if so, would require the owners of the two collections to either merge or cross-reference the collections;- Central assignment removes any risk of creating homonyms. The URI includes no semantics, it provides no indication of type of resource or ownership.-- URI management is reduced to assigning opaque

	<p>identifiers; this process can be automated.</p> <ul style="list-style-type: none">- The URIs are short and opaque identifiers. <p>Cons:</p> <ul style="list-style-type: none">- Owners of collections need to invoke the central application for each and every resource that is added to a collection;- The lack of human-readable semantics might lead to the creation of ambiguous identifiers as the same opaque identifier could be assigned to two different resources without this being visible to the human eye;- Redirection tables are large, with a redirection rule for each individual URI (#rules=#URIs) creating substantial maintenance overhead;- URI-based redirection requires updating many redirection rules if a collection is moved.
<p>3. Central register of short URIs split by EU institution</p>	<p>The central register assigns prefixes to EU institutions. Institutions then generate a locally unique identifier to follow the mnemonic.</p> <p>Example:</p> <p>http://data.europa.eu/maabcd1234</p> <p>Pros:</p> <ul style="list-style-type: none">- URI management is reduced to assigning opaque identifiers; this process can be automated;- The URIs are short and (almost) opaque identifiers. <p>Cons:</p> <ul style="list-style-type: none">- Owners of collections need to invoke the central application for each and every resource that is added to a collection;- Changing namespaces and URIs: organizational changes and changes of ownership would lead to a change of URIs, or loss of mnemonic meaning if the URIs are maintained by a different organisation;- Requires all institutions to have an approach and tools to generate unique identifiers;- Provides no indication of type of resource while ownership is only visible for someone who knows the mnemonic system;- Synonyms might be created as resources are grouped by institution;- The indicator of ownership is a risk for persistence (i.e. what happens if ownership changes from one institution to another);



Based on this analysis, the following proposal was accepted after discussion in the Persistent URI Task Force:

Build an infrastructure with a central register of URI namespaces and local registers of resources.

III.4. Relationship namespace – register

In section 3.4, it is stated that a given URI namespace is associated with a *single* local register. If two institutions were to claim the ownership of a namespace, the relationship between the namespace and the (potentially multiple) registers would need to be clarified.

Table 17 – Analysis of alternatives: Relationship namespace – register

<p>1. Handle co-ownership of a URI namespace through “fusion” of local registers.</p>	<p>Example (fictitious): DG HR of the European Commission shares the ownership of the URI namespace ‘http://data.europa.eu/id/staff/*’ with DG Personnel of the European Parliament. Each institution uses its own local register for managing staff data. They apply several strategies (e.g. using the same identifying mechanism, federating registries, redirection rules) to guarantee persistence of the URI namespaces.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Persistence: guarantees that changes to local registers do not affect already existing URI sets. <p>Cons:</p> <ul style="list-style-type: none"> - Creation of <i>technological</i> dependencies between EU institutions. In the worst case, the persistent URI service would need to keep track of the location of each individual resource.
<p>2. Ownership of a URI namespace should be singular and correspond to a single register.</p>	<p>Ownership should be singular. If more than one institution have a stake in a namespace, they should delegate the responsibility to one of them. The principle being: one namespace, one register.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Persistence: a single register can implement a strong persistence policy that is managed by a single organisation; - Positive competence conflicts are handled via organisational agreements: all institutions involved should discuss and agree to assign the operational responsibility to one organisation. <p>Cons:</p> <ul style="list-style-type: none"> - Creation of <i>organisational</i> dependencies between EU institutions.

Based on this analysis, the following proposal was accepted after discussion in the Persistent URI Task Force:

Strictly apply the principle one namespace, one register with co-ownership solved through organisational agreements.

The Task Force recommended that the following remarks should be taken into account:

- As bodies which get a namespace allocated are free to create sub-namespaces, guidelines with “best practices” for the design of local paths should be developed;
- One organisation should not own the exclusive rights for a common namespace, such as “press releases”. This could be avoided by only assigning opaque namespaces (see section III.8 Design rules and URI patterns). Consequently, the request for a mnemonic namespace will require a more thorough analysis than the request for an opaque namespace.

III.5. Governance levels and representation

In section 3.4, a governance structure is proposed with three levels:

- URI Steering Committee with representatives from all EU institutions taking decisions on the policy level
- URI Committee consisting of staff of one or more EU institutions taking decisions on the operational level
- URI Technical Team consisting of staff of one or more EU institutions taking care of the day-to-day operations

Table 18 – Analysis of alternatives: composition of the governance bodies

<p>1. Three level governance structure, with inter-institutional governance at the strategic level only.</p>	<p>The URI policy is set by the steering committee and executed by the URI Committee and URI Technical Team.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Reduction of overhead: only policy decisions (e.g. decision criteria) are taken at the inter-institutional level and less people are involved in the day-to-day management (Committee and Technical Team) <p>Cons:</p> <ul style="list-style-type: none"> - Less control: The institutions have less control over the governance structure, because decisions are taken by the URI Committee on which not all institutions are represented. If an institution does not agree, it needs to appeal to the URI Steering Committee which can take considerable time. - Handling of exceptions requires escalation to the Steering Committee who are only involved infrequently and therefore may need to be brought up to speed, which could slow down the process
---	--

2. Two level governance structure	<p>The URI Committee is expanded to include representatives of all EU institutions. The URI Committee sets policies and executes the policy, supported by the URI Technical Team. Requests for namespaces that clearly fall within the scope and design rules could be handled directly by the Technical Team with reporting after the installation to the Committee.</p> <p>Pros:</p> <ul style="list-style-type: none"> - More control: Direct influence and oversight of all EU institutions on decisions taken on namespaces. <p>Cons:</p> <ul style="list-style-type: none"> - Larger URI Committee is more difficult to manage.
--	---

Based on this analysis, the following proposal was accepted after discussion in the Persistent URI Task Force:

Create a three-level governance structure.

III.6. Responsibility for proposals

In section 3.5.13.4, the process for requesting a namespace starts with an EU institution that manages a local register sending a request to the URI Committee which then instructs the URI Technical Team to elaborate a proposal in consultation with the requesting institution. The Technical Team will take into account the scope criteria and submit a proposal to the URI Committee which then takes a decision on assigning a namespace for the local register.

Table 19 – Analysis of alternatives: responsibility for proposals

1. Scope criteria, URI design rules, and analysis of similar resources and local registers	<p>Each request should indicate how it meets the scope criteria, URI design rules, and whether there are similar local registers maintained by other EU institutions that could claim co-ownership of the URI Namespace.</p> <p>Pros:</p> <ul style="list-style-type: none"> - The URI Committee can take decisions on the information from the requesters with reference to the scope criteria which allows for clear decision in most cases. <p>Cons:</p> <ul style="list-style-type: none"> - Analysis of similar resources may be incomplete, because the requester may not have a complete overview of similar collections; - Requesters may not be completely familiar with the scope criteria and design rules; therefore the request may not be of sufficient quality which would require the URI Committee to ask for additional information.
2. Requester provides just a	<p>The Technical Team analyses whether the resource collection falls within the scope criteria, reviews the local URI policy if it exists or</p>

description of the collection	<p>helps setting up a URI policy if it does not exist, and investigates whether there are similar collections already under the central policy or elsewhere. The Technical Team reports to the URI Committee with a proposal whether and how to implement a new namespace.</p> <p>Pros:</p> <ul style="list-style-type: none"> - The Technical Team has a good overview across the central and local URI policies; - Local collection managers can receive guidance and support from the Technical Team; - Reporting to the URI Committee can be standardised and URI Committee decision taking can be more efficient. <p>Cons:</p> <ul style="list-style-type: none"> - More work for the Technical Team.
--------------------------------------	--

Based on this analysis, the following proposal was accepted after discussion in the Persistent URI Task Force:

Give Technical Team the responsibility to elaborate a proposal to the URI Committee based on the description of the collection and in consultation with the requester.

The Task Force recommended that the following remarks should be taken into account:

- The scope criteria can serve as clear, lightweight criteria to indicate whether a request for a persistent URI namespace is in scope for central governance. These scope criteria needs to applied flexibly where possible and strictly where necessary.
- The current scope criteria are accepted. However, there must be additional service level guarantees linked to Persistent URI namespaces. The Task Force agreed to make this a future work item.

III.7. Co-ownership of namespaces

In section 3.5.2, it is noted that co-ownership of a namespace is handled on the organisational level.

Table 20 – Analysis of alternatives: co-ownership of a URI namespace

1. Handle request for co-ownership in the URI Committee	<p>An institution that wants to co-own a namespace addresses its request to the URI Committee which then mediates between the existing owner and the institution that requests co-ownership. The URI Committee decides to accept or reject the request for co-ownership, and if accepted, decides on the division of responsibilities and operational management between the two institutions. If any of the parties disagree, the issue is escalated to the URI Steering Committee who will consider the matter and</p>
--	--

	<p>issue a binding decision.</p> <p>Pros:</p> <ul style="list-style-type: none">- Handling of the issue on the policy level where consequences for interoperability and operational integrity can be taken into account. <p>Cons:</p> <ul style="list-style-type: none">- Only indirect negotiations between the two organisations directly involved.
<p>2. Handle request for co-ownership bilaterally between the existing owner and the requesting co-owner</p>	<p>An institution that wants to co-own a namespace contacts the existing owner and communicates its request to co-own the namespace. The two organisations discuss the way they can share the responsibility and operational management of the namespace. If they come to an agreement, they communicate the agreement to the URI Committee. If they cannot agree, they can request for arbitration by the URI Steering Committee who will consider the matter and issue a binding decision.</p> <p>Pros:</p> <ul style="list-style-type: none">- Discussions take place directly between the affected parties;- No necessity for the URI Steering Committee to become involved in cases where the co-ownership can be agreed between the parties. <p>Cons:</p> <ul style="list-style-type: none">- Discussions not visible for the URI Committee, which means that wider interoperability issues may not be taken into account.

Based on this analysis, the following proposal was accepted after discussion in the Persistent URI Task Force:

Handle co-ownership on the organisational level, allowing the affected organisations to agree between them which one of them is responsible for the operational management of the URIs.

The Task Force recommended that the document should clarify that the “one register – one namespace” principle always applies, i.e. that if two organisations want to share a namespace, they should merge their registers, at either the organisational or the technology level.

III.8. Design rules and URI patterns

In section 4.2, the following general pattern is specified:

Study on a common policy for the governance and management of persistent URIs by EU institutions



During the work of a subgroup of the Persistent URI Task Force, a number of alternatives were discussed as outline in the tables below.

Table 21 – Analysis of alternatives: subdomain

1. {subdomain} as empty string	<p>{subdomain} is an empty string for new URI sets placed under common governance.</p> <p>Examples:</p> <p>http:// europa.eu/data/doc/contract-notice/190403-2009</p> <p>http:// europa.eu/ data/id/vessel/224011550</p> <p>Pros:</p> <ul style="list-style-type: none"> - Stability, independent of organisational changes ;Shorter URI strings; - Central coordination reduces risk of synonym URIs; - Persistent URI service as a single instance for redirection; local registers still necessary to provide a resolution mechanism. <p>Cons:</p> <ul style="list-style-type: none"> - Increases possibility of co-ownership of URI namespaces. This may be irrelevant, depending on whether co-ownership of a URI namespace is allowed.
2. 'data' as {subdomain}	<p>{subdomain} is the string 'data' for persistent identifiers.</p> <p>Examples:</p> <p>http://data.europa.eu/contract-notice/190403-2009</p> <p>http://data.europa.eu/vessel/224011550</p> <p>Pros:</p> <ul style="list-style-type: none"> - Less prone to changes; - Central coordination reduces risk of synonym URIs; - Persistent URI service as a single instance for redirection; local registers still necessary to provide a resolution mechanism. <p>Cons:</p> <ul style="list-style-type: none"> - Increases possibility of co-ownership of URI namespaces. This may be irrelevant, depending on whether co-ownership of a URI namespace is allowed.

3. Subdomain indicating sector or theme	<p>{subdomain} follows a controlled vocabulary indicating the theme or sector.</p> <p>Examples</p> <p>http://publications.europa.eu/id/contract-notice/190403-2009</p> <p>http://markt.europa.eu/id/contract-notice/190403-2009</p> <p>http://transport.europa.eu/id/vessel/224011550</p> <p>http://maritime.europa.eu/id/vessel/224011550</p> <p>Pros:</p> <ul style="list-style-type: none"> - More contextual meaning by indicating theme; - Decreases possibility of co-ownership of URI namespaces. <p>Cons:</p> <ul style="list-style-type: none"> - Any theme-based classification has the risk of being unstable, and will tend towards organisational structures; - Risk of synonyms: URIs for the same resources risk to be created in different collections; - Longer URIs; - Risk of losing central coordination (e.g. Persistent URI service per subdomain).
--	--

Based on this analysis, the following proposal was accepted after discussion in the Persistent URI Task Force:

Place centrally administered URIs under the **data.europa.eu** subdomain.

Table 22 – Analysis of alternatives: URI Pattern

1. Use {/collection*}, {concept} and {type} (in this order)	<p>http://{subdomain}.europa.eu{/collection*}/ {concept}/{type}/{local id}/{version}/ {language}</p> <p>Examples:</p> <p>http://data.europa.eu/contract-notice/doc/190403-2009</p> <p>Pros:</p> <ul style="list-style-type: none"> - Harmonisation and interoperability: a central decision making body coordinates URI namespaces avoiding synonyms and fragmentation of namespaces; - Division of roles and responsibilities: a central decision making body coordinates responsibilities in cases of positive competence conflicts; - The URI provides human-readable semantics while keeping the length of the URI limited; - {concept} is the element that determines to which
--	---

	<p>local register the URI needs to be redirected (via redirection rules);</p> <ul style="list-style-type: none"> - Specifying only <code>http://data.europa.eu/{concept}/</code> could return a landing page for the concept-based register; - Regarding the order of <code>{type}</code> and <code>{concept}</code>: <code>{concept}</code> is the more significant element as this determines to which local register the URI needs to be redirected. The local register needs to take care of the rest (including the <code>{type}</code>); - Specifying <code>http://europa.eu/{concept}/</code> could return a landing page for the concept-based register. <p>Cons:</p> <ul style="list-style-type: none"> - The <code>{type}</code> part is essentially non-functional (see Section Error! Reference source not found.).
<p>2. Only <code>{concept}</code> and <code>{local identifier}</code></p>	<p>This alternative represents a limited version of the URI pattern, i.e. <code>http://data.europa.eu/{concept}/{local id}</code>. The central register assigns a collection identifier to a register for a particular type or resource (e.g. vehicles, vessels, staff members, buildings). Local registers assign unique identifiers to the items in the collection.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Harmonisation and interoperability: a central decision making body coordinates URI namespaces avoiding synonyms and fragmentation of namespaces; - Division of roles and responsibilities: a central decision making body coordinates responsibilities in cases of positive competence conflicts; - Monitor service levels centrally via a <i>central</i> Persistent URI service; - The URI provides human-readable semantics while keeping the length of the URI limited; - <code>{concept}</code> is the element that determines to which local register the URI needs to be redirected (via redirection rules); - Specifying only <code>http://data.europa.eu/{concept}/</code> could return a landing page for the concept-based register. <p>Cons:</p> <ul style="list-style-type: none"> - Loss of local control, creates interdependency: EU institutions need to request ownership over individual URI namespaces for resources that fall within the scope criteria of the inter-institutional URI policy.
<p>3. Zero-semantics shortened</p>	<p>The central URI register assigns random numbers through an algorithm that ensures uniqueness within the context of the local</p>

<p>persistent URI</p> <p>(see also alternatives for governance structure in Section 3.4, Error! eference source not found.)</p>	<p>register. The resulting URI contains no human-readable semantics and only functions as a unique identifier.</p> <p>Example:</p> <p><code>http://data.europa.eu/abcd1234</code></p> <p>Pros:</p> <ul style="list-style-type: none">- Harmonisation and interoperability: a central decision making body coordinates URI namespaces avoiding synonyms and fragmentation of namespaces;- The URIs are short and opaque identifiers;- Under this policy there is no pattern to be followed as all URIs are generated centrally. <p>Cons:</p> <ul style="list-style-type: none">- The URI includes no semantics, it provides no indication of type of resource or ownership. The lack of common design rules would not give confidence to third-parties using the URIs;- Risk of creating synonyms or homonyms due to reduced control over created URIs: The lack of human-readable semantics makes it difficult to prevent the creation of duplicate identifiers (synonyms) or the creation of identifiers that for which the requestor does not guarantee persistence;- There is no indication of what the URI identifies.
<p>4. Mnemonics-based short URI split by EU institution</p> <p>(see also alternatives for governance structure in Section 3.4, Error! eference source not found.)</p>	<p>The central register assigns prefixes to participating institutions. Institutions then generate a locally unique identifier to follow the mnemonic.</p> <p>Example:</p> <p><code>http://data.europa.eu/maabcd1234</code></p> <p>Pros:</p> <ul style="list-style-type: none">- The URIs are short and opaque identifiers. <p>Cons:</p> <ul style="list-style-type: none">- Indication of ownership may lead to changing namespaces and URIs: organizational changes and changes of ownership would lead to a change of URIs;- Requires all institutions to have an approach and tools to generate unique identifiers;- Provides no indication of type of resource or ownership;- Synonyms and homonyms might be created as resources are grouped by institution;- The URI includes limited semantics, it barely provides

	<p>an indication of the type of resource and the ownership. The lack of common design rules would not give confidence to third-parties using the URIs;</p> <ul style="list-style-type: none"> - There is no indication of what the URI identifies.
--	---

Table 23 – Analysis of alternatives for {/collection*}

<p>1. {/Collection*} is an empty string</p>	<p>{/collection*} is an empty string for new URI sets placed under common governance.</p> <p>Examples:</p> <p>http://europa.eu/contract-notice/190403-2009</p> <p>http://europa.eu/vessel/224011550</p> <p>Pros:</p> <ul style="list-style-type: none"> - Less prone to changes; - Shorter URI strings; - Central coordination reduces risk of synonym URIs. <p>Cons:</p> <ul style="list-style-type: none"> - Increases possibility of co-ownership of URI namespaces; - No contextual meaning: no indicated theme, no possibility to reflect a hierarchical navigation pattern in the URI structure.
<p>2. {/Collection*} indicates the local register</p>	<p>{/collection*} is a term indicating a group of concepts that are managed together.</p> <p>Examples</p> <p>http://data.europa.eu/eli/eu/ec/2007/10/23/directive/2007-60-EC</p> <p>http://data.europa.eu/mdr/language/ENG</p> <p>http://data.europa.eu/esco/skills/16747</p> <p>Pros:</p> <ul style="list-style-type: none"> - Contextual meaning: possibility to “navigate” by removing some parts of the URI; - No indication of institutional ownership, allowing responsibilities for the collections to be handed over without change in URIs. <p>Cons:</p> <ul style="list-style-type: none"> - Risk of synonyms: URIs for the same concept may be created in different collections; - Longer URIs; - Risk of losing central coordination.
<p>3. {/Collection*}</p>	<p>{/collection*} follows a controlled vocabulary indicating the</p>

<p>follows a thematic controlled vocabulary</p>	<p>theme.</p> <p>Examples</p> <p>http://europa.eu/publications/contract-notice/190403-2009</p> <p>http://europa.eu/markt/contract-notice/190403-2009</p> <p>http://europa.eu/transport/vessel/224011550</p> <p>http://europa.eu/maritime/vessel/224011550</p> <p>Pros:</p> <ul style="list-style-type: none"> - Contextual meaning: possibility to “navigate” by removing some parts of the URI; - Decreases possibility of co-ownership of URI namespaces. <p>Cons:</p> <ul style="list-style-type: none"> - Risk of synonyms: URIs for the same resources risk to be created in different collections; - Longer URIs; - Risk of losing central coordination.
--	--

Table 24 – Analysis of alternatives for {type}

<p>1. No {type}</p>	<p>{type} is removed in the URI pattern for new URI sets placed under common governance.</p> <p>Examples:</p> <p>http://data.europa.eu/contract-notice/190403-2009</p> <p>http://data.europa.eu/vessel/224011550</p> <p>Pros:</p> <ul style="list-style-type: none"> - Shorter URI strings - Simplicity; no need to define the type of resource, avoiding discussions such as whether an identifier for a contract notice refers to a ‘real-world’ entity or to a digital object – and resource collections usually contain a single type of resources, e.g. all contract notices are digital objects, all vessels are ‘real-world’ objects. <p>Cons:</p> <ul style="list-style-type: none"> - Not possible to visually distinguish information resources from non-information resources.
<p>2. Only two values for {type}: ‘id’ and ‘doc’</p>	<p>Only two values for {type}: ‘id’ and ‘doc’.</p> <p>Examples:</p> <p>http://data.europa.eu/vessel/id/224011550</p> <p>http://data.europa.eu/vessel/doc/224011550</p> <p>Pros:</p>

	<ul style="list-style-type: none"> - Possible to distinguish information resources from non-information resources; the central URI Application can have a global redirection rule to redirect from */id/* to */doc/* <p>Cons:</p> <ul style="list-style-type: none"> - Longer URI strings - Discussions may arise around the distinction, e.g. whether an identifier for a contract notice refers to a 'real-world' entity or to a digital object
<p>3. Multiple values for {type}: 'id', 'so', 'data', 'def', and 'doc'.</p>	<p>Multiple values for {type}: 'id', 'so', 'data', 'def', and 'doc'.</p> <p>Examples:</p> <p>http://data.europa.eu/id/vessel/224011550 (a vessel)</p> <p>http://data.europa.eu/doc/vessel/224011550 (a description of a vessel)</p> <p><a ;long:6°07'53.09"="" href="http://data.europa.eu/so/coordinates/lat:49°35'55.94">http://data.europa.eu/so/coordinates/lat:49°35'55.94";long:6°07'53.09" (a location)</p> <p>http://data.europa.eu/data/statistics/eu/unemployment/2014 (a statistical dataset)</p> <p>Pros:</p> <ul style="list-style-type: none"> - Possible to distinguish information resources from non-information resources; - Possible to distinguish datasets and spatial objects, although this can be derived from the concept. <p>Cons:</p> <ul style="list-style-type: none"> - Complexity; requires decisions to be taken about the type of resource that could give rise to discussion (e.g. is a building a general real-world object or a specific spatial object); - Longer URI strings; - No functional advantage; only /id/ (identifier for real-world thing) and /doc/ (something that consists of bits and bytes and can be transferred over a network link) are functionally different.

Table 25 -- Analysis of alternatives for {concept}

<p>1. Using the nature of the resources in the register as {concept}</p>	<p>The {concept} identifies the nature of the identified resource, e.g. whether it a book, a vessel, a vehicle, a contract notice etc.</p> <p>Examples:</p> <p>http://data.europa.eu/contract-notice/190403-2009</p> <p>http://data.europa.eu/vessel/224011550</p> <p>Pros:</p> <ul style="list-style-type: none"> - Visual indication of the nature of the resources; - Reduces risk of homonyms and synonyms;
---	--

Study on a common policy for the governance and management of persistent URIs by EU institutions

	<ul style="list-style-type: none"> - The nature of the resources is a stable characteristic and will not likely change. <p>Cons:</p> <ul style="list-style-type: none"> - The string is necessarily in a particular language. Non-native speakers of the language chosen may not understand the meaning.
<p>2. Using the name of the register as {concept}</p>	<p>The {concept} aligns with the name of the local register.</p> <p>Note: this is similar to option 2 for {collection}. The difference is that if the register name is taken as the name of the {collection}, the {concept} element is still under central co-ordination, while if it taken as the name of the {concept}, the following sub-concepts are under local control.</p> <p>Examples:</p> <p>http://data.europa.eu/eli/eu/ec/2007/10/23/directive/2007-60-EC http://data.europa.eu/commission-documents/com/2015-xxx http://data.europa.eu/esco/skills/16474</p> <p>Pros:</p> <ul style="list-style-type: none"> - The register contains the local references to the resources and will be stable, especially if based on a wider standard (such as in the case of ELI). <p>Cons:</p> <ul style="list-style-type: none"> - Longer URI strings; - The assignment of sub-concepts takes place outside of central co-ordination, which may lead to synonyms, e.g. when another register creates URIs for the same type of concepts (e.g. a register with educational concepts creating URIs for a set of skills).

Table 26 – Analysis of alternatives: URIs for DOIs

<p>1. Use the DOI handle</p>	<p>Use the namespaces of one existing DOI handle service to mint persistent URIs for resources.</p> <p>Examples:</p> <p>http://doi.org/10.2788/14231; OR http://dx.doi.org/10.2788/14231; OR http://hdl.handle.net/10.2788/14231.</p> <p>Pros:</p> <ul style="list-style-type: none"> - Reuse of existing DOI handle service.
-------------------------------------	---

Study on a common policy for the governance and management of persistent URIs by EU institutions

	<p>Cons:</p> <ul style="list-style-type: none">- The persistent URI is not on the Europa.eu domain, but on the domain of a DOI handle service. This makes the persistent URIs prone to changes because the DOI <i>handle services</i> are not under the control of an EU Institution (even though the Publications Office of the EU is a DOI Registration Agency);- There are several DOI handle services, which would result in duplicate persistent URIs.
<p>2. Use the DOI as a {local id} in a URI namespace on the Europa.eu domain</p>	<p>Mint a persistent URI, using a persistent URI namespace on the Europa.eu domain with the DOI as {local id}. Use redirection rules to redirect HTTP requests to existing DOI handles.</p> <p>Examples:</p> <p>http://data.europa.eu/book/10.2788/14231</p> <p>Pros:</p> <ul style="list-style-type: none">- On the Europea.eu domain, offering additional guarantees of persistence. <p>Cons:</p> <ul style="list-style-type: none">- The resolution mechanism requires a redirection on both the Persistent URI service on Europea.eu and on the DOI handle service.

Based on the analysis in the previous five tables (Table 22 through Table 26), the following proposal was accepted after discussion in the Persistent URI Task Force:

Use single namespace string instead of complex {collection}/{type}/{concept}.