Conversational CoreTrustSeal[1]

[1] https://doi.org/10.5281/zenodo.4033966

Introduction

To apply for the CoreTrustSeal[2] you need to read and respond to the CoreTrustSeal Requirements[3], Extended Guidance[4] and Glossary[5].

This working paper is in no way endorsed by the CoreTrustSeal Board, or indeed anyone else. It's been prepared to put together some thoughts about the CoreTrustSeal Requirements in a way that doesn't quite fit into formal documents, project deliverables, blogs or slides. This text is not intended to disagree with any aspect of the official CoreTrustSeal texts. It presents some key concepts and provides some broad talking points around the CoreTrustSeal and Trustworthy Digital Repositories' data services in fewer than 3000 words. It's being shared because transparency and feedback are always worthwhile.

Trusted/Trustworthy Digital/Data Repository (TDR)

Being trusted is not quite the same thing as demonstrating that you are trustworthy, but the terms are often used interchangeably. The use of digital versus data is a matter of context, or sometimes simple preference. Digital objects can contain structured metadata and other documentation as well as data.

We put things in repositories because we hope they will be managed/cared for/curated in a reliable environment. In a trusted digital repository we expect digital objects to be preserved so that they remain accessible, understandable and usable despite changes to the people, the processes or the technologies (the infrastructures) that consume them.

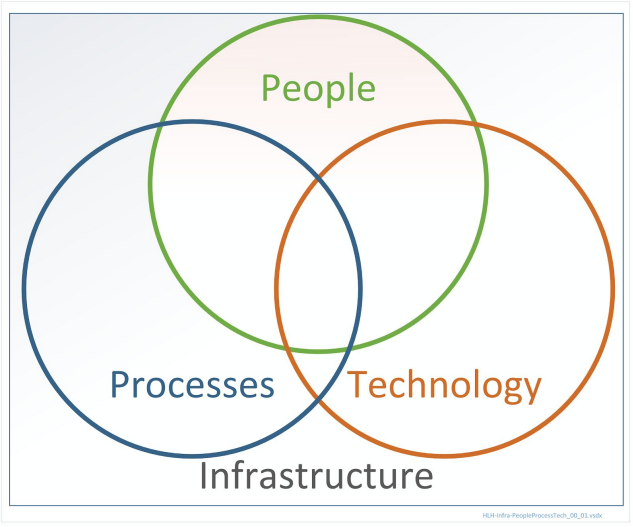

Diagram: Infrastructure. People, Processes, Technology

Applicants for the CoreTrustSeal

If you take responsibility for the digital objects and the actions taken to curate and preserve them, then you can apply for the CoreTrustSeal. Others can help, and their role(s) must be acknowledged, but it is the applicant that takes responsibility and is certified.

[2] https://www.coretrustseal.org/
[3] https://doi.org/10.5281/zenodo.3638211
[4] https://doi.org/10.5281/zenodo.3632533
[5] https://doi.org/10.5281/zenodo.3632563

Audiences

The initial audiences for a self-assessment are the CoreTrustSeal reviewers and Board during a confidential process. Approved CoreTrustSeal assessments are public and will certainly be read by peer repositories and current or future partners in delivering repository data services. Depositors, Users and Funders can read a certified CoreTrustSeal assessment.

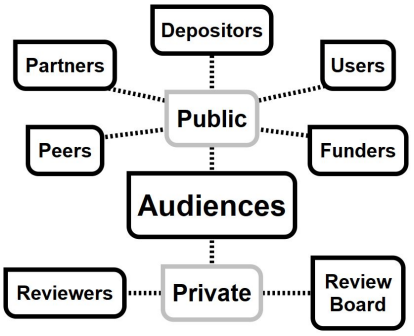

Diagram: Audience

What Do You Do? Who Do You Serve?

The OAIS reference model[6] defines the mandatory repository responsibilities of a repository.

The CoreTrustSeal applies the OAIS definition of the designated community in their Glossary[7]. You may serve multiple, complex communities of users but the CoreTrustSeal asks you to define one or more designated communities whose knowledge base, processes and technologies you understand. You should demonstrate that you identify what they want and deliver what they need, changing as necessary over time.

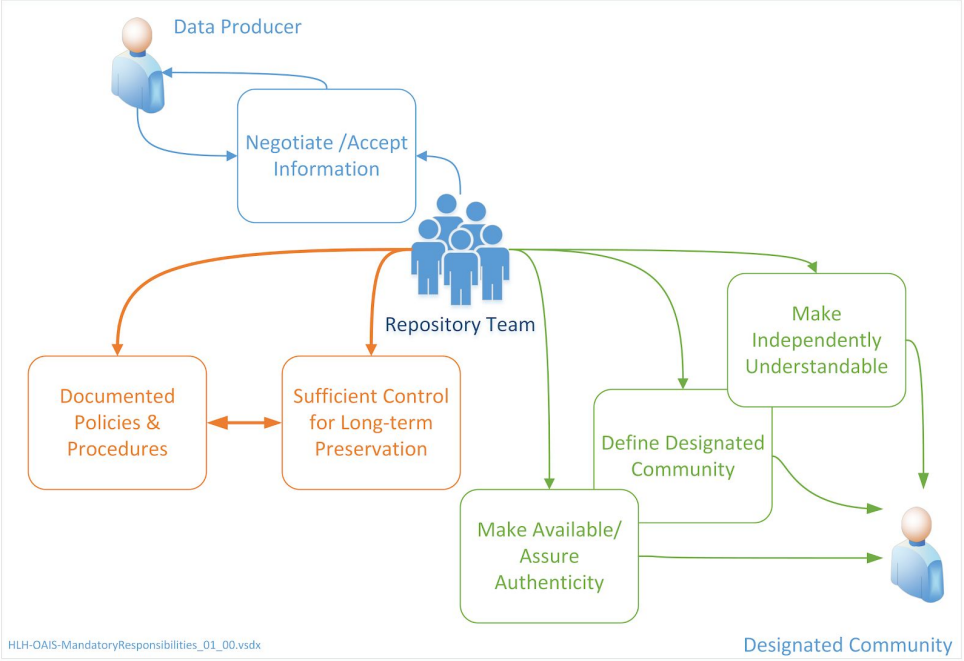

Diagram: a simplified overview of mandatory OAIS responsibilities

[6] https://public.ccsds.org/pubs/650x0m2.pdf
[7] https://doi.org/10.5281/zenodo.3632563

Self-Assessments and Evidence

The self-assessment statements you provide for the CoreTrustSeal are claims which should be supported by evidence. The evidence expected for the CoreTrustSeal is intended to reflect the information that needs to be created and managed when running a quality, efficient, sustainable data service. The CoreTrustSeal can be used as an internal self-assessment and communications tool without applying for certification.

We write business information down. We do this so that people know what we do, so that we can do it consistently, so that we manage change, and so that someone else can do it if we can't do it any more for some reason. We can also use this information as evidence for the CoreTrustSeal.

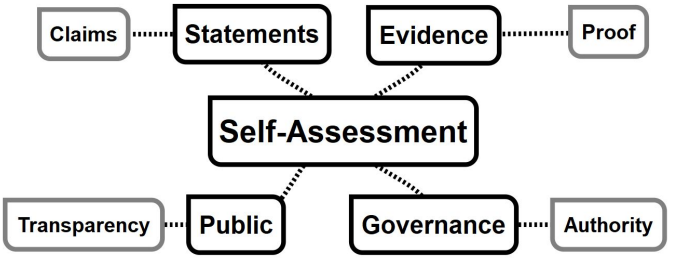

Diagram: Self-Assessment

Clear authority for the creation and management of evidence demonstrates good governance. Public evidence demonstrates transparency.

If you're a small organisation then you might be relying on a limited number of very expert staff. Losing a member of staff is a big risk, so you should write things down. A larger organisation might have multiple staff doing the same job. Inconsistent work from different staff members is a big risk, so you should write things down. If you still think that the information you're being asked to write down is unnecessary or obvious, then consider that you just need to do it once and then review it now and then.

CoreTrustSeal Requirements

R0. CoreTrustSeal Context

A clear context statement supports the interpretation of your self-assessment and evidence for all 16 requirements. You will be asked what type of repository you are, to provide a brief description of yourself and your designated community (see above) and the level of curation you perform. The applicant takes responsibility for meeting all the CoreTrustSeal Requirements, but they may be assisted by others, from host or close partner organisations (insource) to third-party services (outsource). Make it clear what these organisations do for you and how you maintain control of the functions they perform on your behalf.

Organisation

R1. Mission

- Access is explicitly required in your mission
- A TDR goes beyond the bits

Communicate (with the highest level of approval you can get) that an important part of your role is to make sure that data and metadata are assessed, curated and made accessible and usable for the long term. Make it clear that you actively preserve digital objects for your community.

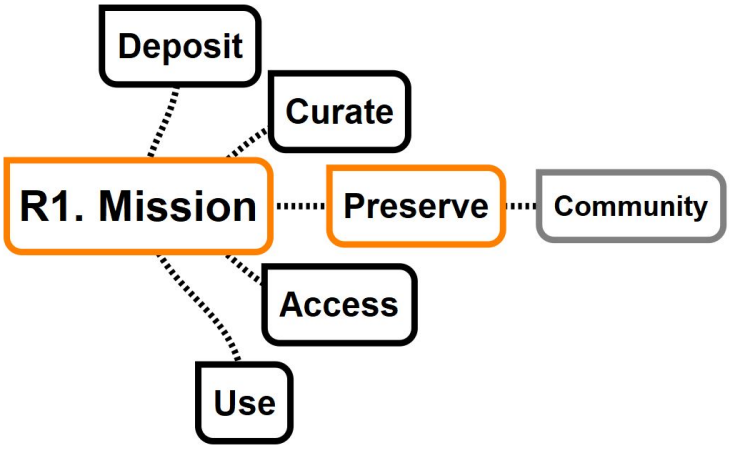

Diagram: R1. Mission

R2. Licences

- Listing licences by type and what they're used for

A repository should be able to list licences for the conditions of deposit that give them the rights to curate and preserve, and list licences which clarify the rules for access and use. We should be able to explain the licences we apply to different users and to different data and metadata, and why. We should be clear on whether these are open and standard licences or if they are locally defined.

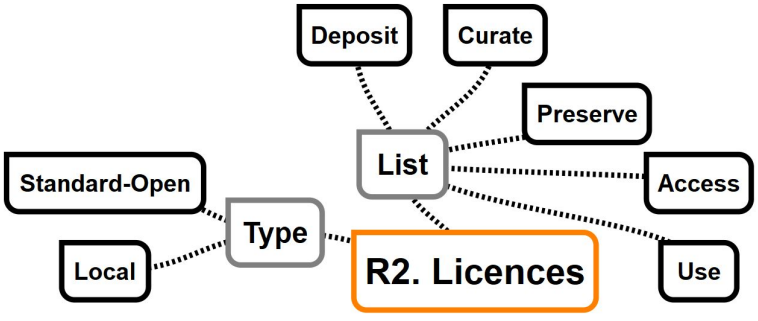

Diagram: R2. Licences. The basics

- Right Management: permissions, prohibitions & duties

The basics of listing licences by their purpose and data type is probably enough to get the CoreTrustSeal. But rights management is a complex issue which goes beyond standard prose licences. If we want to streamline processes or enable machine-actionability then we need to think about rights in clearer terms of permissions, prohibitions & duties and to consider the actors that they apply to, whether depositor, repository or user.

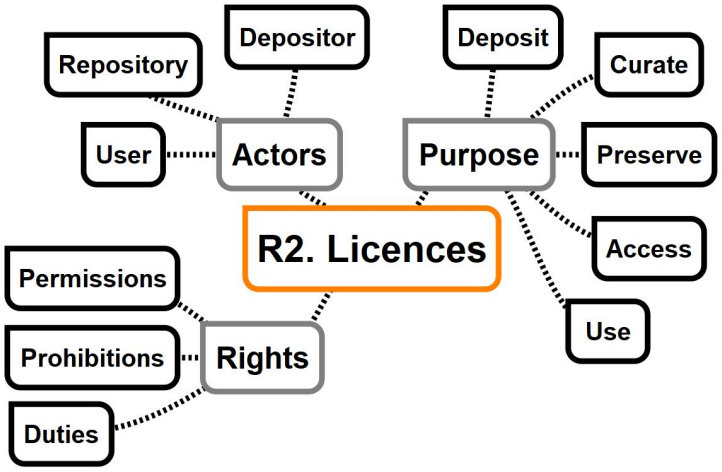

Diagram: R2. Licences. Rights Management

R3. Continuity

- The organisation persists so the data persists
- Business continuity, disaster recovery and succession

We may have the processes and the skills to make sure that objects are accessible and usable for the long term. But what could impact the services we deliver? Do we have enough resources to continue to exist? What do we do if something reduces our services or stops them completely? How do we get back on track? In the worst case scenario could someone else offer the same level of curation and service for the digital objects?

Diagram: R3. Continuity

Organisational resilience is important for data resilience. Business continuity is about keeping things going in the face of adversity. Disaster recovery is about the processes and the time it takes to get services back when they fail. Succession planning is about being able to transition your data and services elsewhere.

Some organisations can tell you where their data will go if they change their mission or cease to exist. A few can tell you what level of service they could hand over. The CoreTrustSeal is looking for evidence that you have considered the worst case scenario. But the CoreTrustSeal does not require that you provide a blueprint for replacing all of your complex, expert-driven data services.

R4. Confidentiality & Ethics

- Ethical research and the protection of sensitive data
- Understanding the legislative and policy context

Any creation or collection of data for use has some ethical implications. Sometimes this involves ethical requirements or standards review boards. Can you list the laws and other regulations about how depositors, repositories and users should handle data? If you're looking after information about identifiable humans or other sensitive data then what additional steps do you take?

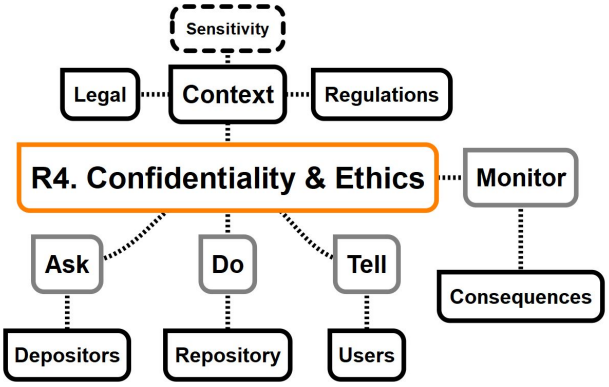

Diagram: R4. Confidentiality & Ethics

What do you ask your depositors about confidentiality and ethics? How do you act as a repository? What do you tell data users about how they should act? Can you tell whether the rules are being followed, and if not then how do you act?

The CoreTrustSeal does not require that we act as the data police. But we need to understand the rules and consider the consequences if they are not met. The CoreTrustSeal is assessing the handling of the deposited digital objects. Despite their importance the data of staff and users is out of scope.

R5. Organisational Infrastructure

- Governance, structure and resources

Organisational infrastructure is complex and varied. The CoreTrustSeal does not set a high bar for detail here. Describe your organisation structure, link to a diagram on the web, note relevant bodies that make decisions and briefly describe your funding model. Ideally this information is public, which demonstrates transparency to all of your stakeholders.

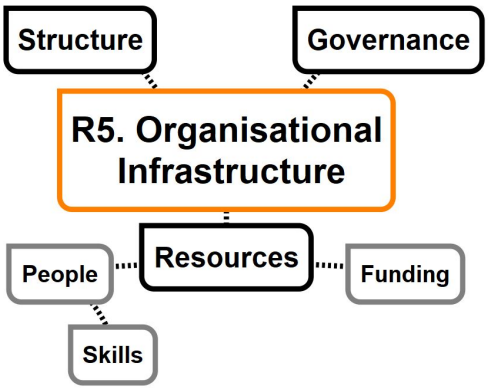

Diagram: R5. Organisational Infrastructure

How is the organisation which looks after this data structured? What sections, departments, roles and hierarchies does it have? How are decisions made? How do you make sure that your teams have the skills and the funding to deliver what you promise?

R6. Expertise

- No organisation is an island.
- Communities of practice

No organisation can have all the skills it needs within their staff, especially if they are small or have a very wide ranging mission. Describe the areas of knowledge you depend on and explain how you engage with and participate in wider groups of experts. It's valuable to be a follower and a leader; often both in different areas.

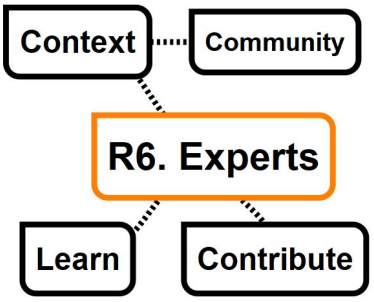

Diagram: R6. Experts

Object Management

R7. Integrity and Authenticity

- Avoiding unintended change
- Managing and communicating intended change

How do you avoid deleting or damaging what you're looking after? If you're changing something, or a copy of something, is it clear why? How do you record and communicate those changes? When and why do you create new versions?

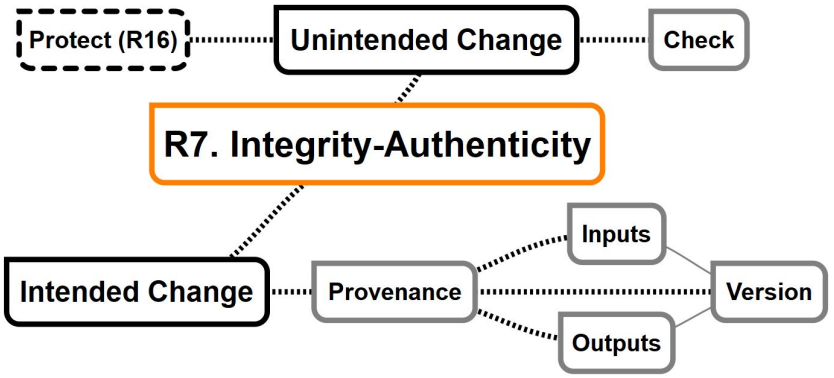

Diagram: R7. Integrity & Authenticity

A clear chain of provenance defines the inputs and outputs for each action on an object and explains why actions were taken. Protecting data from change can be addressed under data security (R16).

R8. Appraisal

- The gatekeeper to the repository
- Objects offered for deposit are Evaluated
- Curation actions are defined

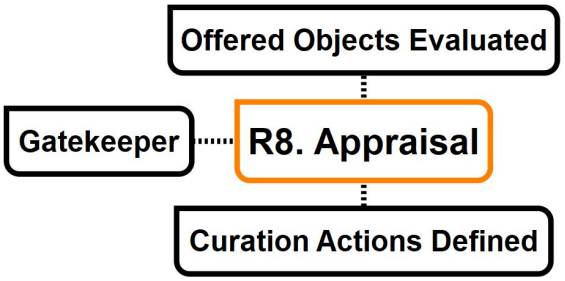

Diagram: R8. Appraisal

What rules do you use to decide what you will and will not accept to look after? Is it clear to the depositor what level of curation and preservation you will offer? How do you decide what steps you will take so that digital objects remain usable?

R9. Documented Storage

- Location, location, location
- Protecting the bits

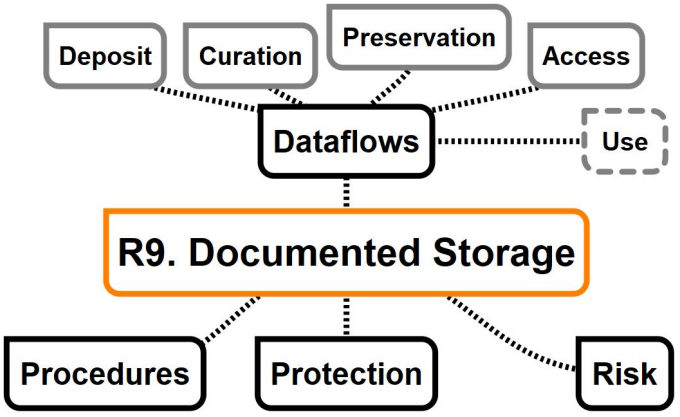

Diagram: R9. Documented Storage

From the moment that you have custody of data you have responsibility for it. Can you describe where the data flows and how is it backed up and copied at each stage of processing and in each storage location? This includes the locations for deposit, curation, long term preservation and access. It also includes any locations where repositories control data use (e.g. secure remote access). How much could be lost if something goes wrong?

R10. Preservation Plan

- Implement and communicate a clear level of preservation.
- If not preserved here, then where?

Preservation goes beyond basic curation. You must understand the needs of your designated community in accessing, understanding and using the data. You also need to demonstrate that you can adjust to changes in that community and those needs over time.

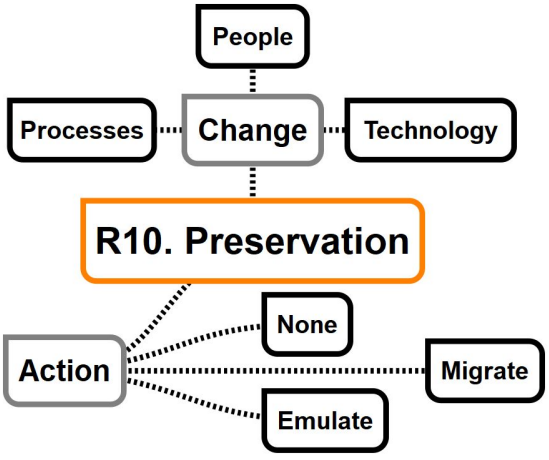

Diagram: R10. Preservation Plan

If the needs of the community of people you serve, or their processes or technologies change then this may provide a reason to take a preservation action on the data, metadata or documentation of a digital object. That action may be to migrate to a new format, or to provide an emulation solution. You may decide that no action is required; the important thing is that your action or inaction is an informed decision.

Preservation doesn't have to mean 'forever'. But it should be clear to your stakeholders what circumstances will cause you to change the level of preservation you apply to a digital object.

If you don't have some, or all, the responsibility for preservation it should be clear who does. If no one is taking responsibility for preservation it is vital that this is clear to the stakeholders.

Are you taking the same level of care with every digital object you look after? If not, it should be clear what levels of curation and preservation you apply to which objects.

R11. Quality

- Quality compliance expectations
- Assessing quality
- Curating for quality
- Communicating quality

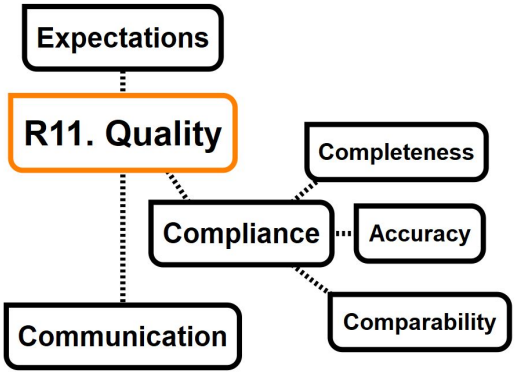

Diagram: R11. Quality

Repositories are important in helping define and apply community quality expectations. But in the end the scientific value aspects of quality are for the data user to evaluate and decide.

The repository may make initial quality assessments at the point of appraisal (R8).

This requirement is about demonstrating that the repository understands community expectations and takes the necessary steps to ensure that the digital objects comply. This involves ensuring that digital objects are comparable to others and are complete and accurate based on some standard(s).

The repository communicates the measures taken, and the level of quality provided at the point of reuse (R14).

R12. Workflows

- Defined curation actions
- Clear responsibilities

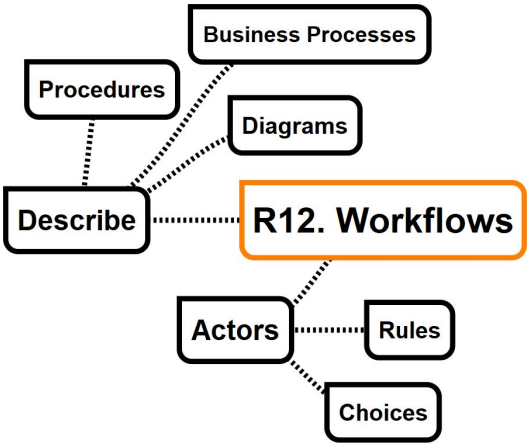

Diagram: R12. Workflows

Taking actions in a clear, consistent way supports both quality and efficiency. How are rules and choices defined? Do you describe the different functions and actions you undertake through business process descriptions, diagrams or standard operating procedures? How do you ensure that the appropriate actors take the right actions on the correct digital objects?

R13. Discovery & Identification

- Providing the context for the discovery of specific objects
- Unique, persistent, resolvable identification

We need enough descriptive information to reach available digital objects and to give credit by acknowledging their use (citation). Each unique object must be associated with its own unique identifier which persists over time and helps us to reach the object (resolution).

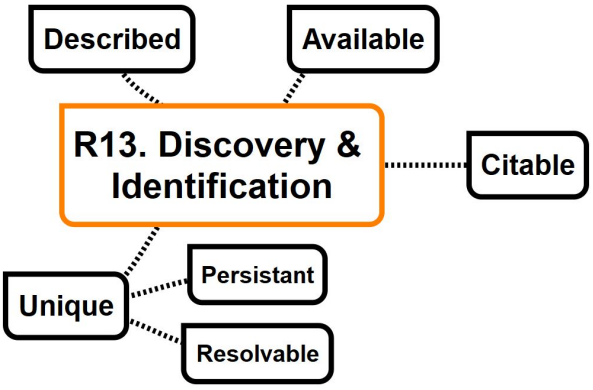

Diagram: R13. Discovery & Identification

How do you assign identifiers to your digital objects and expose them and their metadata to systems which support searching? How do you ensure that users can find the search system itself? How do you support citing of data and metadata to ensure future provenance and the sharing of credit?

R14. ReUse

- Integrating digital objects for re-use
- Understandable, actionable data and metadata
- Delivering impact through data

To support re-use we must design digital objects so that communities can understand them and take action on them by integrating them into future work. This may include re-running, repeating, reproducing or replicating previous work, or doing something entirely new. Through these actions reuse delivers impact.

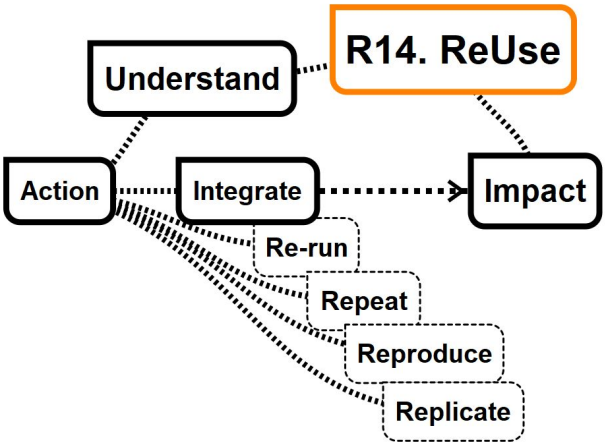

Diagram: R14. ReUse

Providing information relevant to the knowledge and technology base of your intended users is critical if the re-use of data is to be efficient, be reliable and deliver impact. What steps do you take to ensure that users can re-use the digital objects?

Technology & Security

R15. Technical Infrastructure

- The tools of the trade
- Meeting the needs of repositories and users
- Planning for the future
- Recovering from the unexpected

How do you decide and provide the tools needed to meet your users' needs? How do you govern that technical system over time to manage the expected and the unexpected?

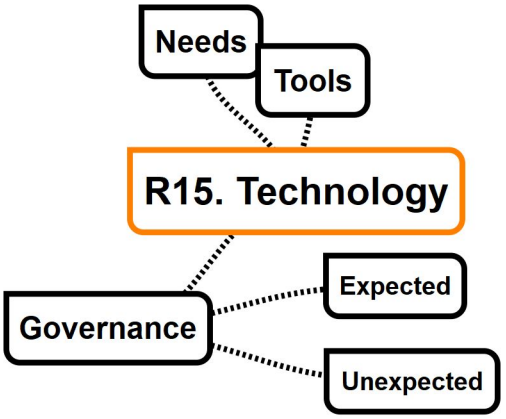

Diagram: R15. Technology

Technology provides the environments and tools for managing, providing access to, and using digital objects. The technologies, standards and processes in place must meet the needs of stakeholders. There should be measures in place to ensure that the technical infrastructure remains fit for purpose over time and for responding to disasters or other business continuity issues.

R16. Security

- Can you protect what you've been trusted with?

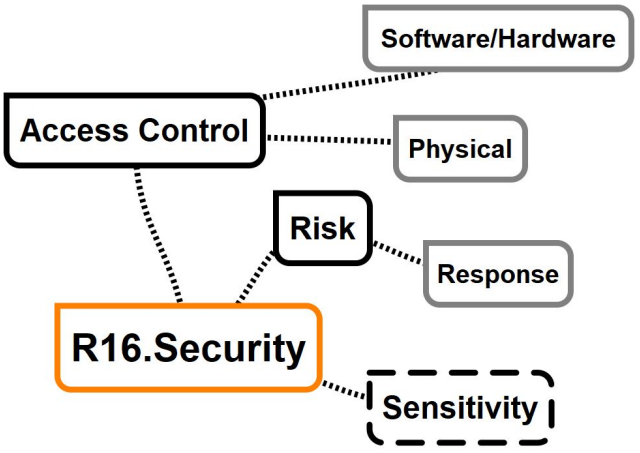

Diagram: R16. Security

All digital objects need some degree of physical and technical (software/hardware) protection. The expectations increase when the data is sensitive for some reason. Access control and rights management measures help to avoid malicious actions and enable permitted actions. Risks should be evaluated based on likelihood and impact, and mitigated where possible. Measures should be in place to respond to any security incidents.