Agrisemantics WG

Berlin, 22 March 2018
http://bit.ly/Agrisem_Berlin18

Agenda

11:00 - 11:15 Introduction to the WG group

11:15 - 12:00 Work on draft recommendations

12:00 - 12:20 Invited talk by Luiz Bonino on GO FAIR & discussion: Towards an Agrisemantics GO FAIR Implementation Network?

12:20 - 12:30 Next steps

Collaborative notepad

http://bit.ly/2pvUsQx

Introduction to Agrisemantics

Output of the group (June 2018)

Requirements and recommendations for the semantic component of an e-infrastructure for data in agriculture

Deliverables
1. A report on **Semantics Landscape for Agricultural Data**
2. A set of **use cases** and requirements
3. A document on **Recommendations** for the future of semantics for agricultural data

Where are we?

Done: Landscaping report, collection of use cases (20)
Ongoing: analysis of use cases

To do: Recommendations for the future of semantics for agricultural data - software, functionalities, semantic assets to enhance data interoperability in agriculture.

Group closes on 9 Dec 2018 (deadline extended)

From use cases & discussions

Two recurring points:

1. Make semantic technologies/methodologies more accessible to non-ontologists
2. Provide suite(s) of tools that integrate tasks, e.g. metadata production, format conversion, access to existing resources… all along workflow

Towards recommendations: http://bit.ly/2pvUsQx

A GO FAIR Implementation Network for Agrisemantics?

Presentation of GO FAIR & IN by Luiz Bonino

Discussions

https://www.go-fair.org

What's next?

- Finalize a V1 of the recommendations (by June)
- Repurpose recommendations to other targets (e.g. tech stakeholders)
- Disseminate recommendations
- Derive an article from the landscape (deliv. 1)
- Think of Agrisemantics beyond RDA WG (end in Dec)

Use case analysis - pointers

Use cases map - http://bit.ly/UC_mindmap

Need MindMup2.0 for Google Drive

Use Case folder: http://bit.ly/UC_folder

With original use cases

Deliverable 2 Top folder: http://bit.ly/Del2_folder

From use cases & discussions

Two recurring points:

1. Make semantic technologies/methodologies more accessible to non-ontologists
2. Provide suite(s) of tools that integrate tasks, e.g. metadata production, format conversion, access to existing resources… all along workflow

1) make semantic tech accessible

1. When Using SR to model data (consumption)
2. Work on GUI!
3. Make semantic technologies/methodologies more accessible to all roles in the data management workflow
4. No pipeline of tools but suits of tools

Efforts on user interfaces

From knowledge elicitation to consumption

- Knowledge capture: domain expert, ontologist
- Formalization: ontologist
- Re-engineering / alignment: ontologist
- Validation: domain expert
- Consumption: data manager, domain expert

Tools should include data types relevant for the domain integrate methodology

Group 1 - editing

- Adapt editing tools to users (domain experts vs. ontologists vs. data managers)
- Integrate tools used at different phases of the editing process: eliciting knowledge, formalization, reuse and alignment
- Allow to navigate from one "phase" to another and back
- Editing tools should integrate methodology, quality checking & validation

Group 2 - alignment

1) think of reusing others' 2) if none exists then align!

- 1-to-1 alignment is unrealistic, use a hub to interconnect ontologies
- Validation is needed: develop GUI to visualize and manipulate mappings

Group 3: Consumption

- Develop GUIs & training for non ontologists consumers
- More modularity: editing tools should encourage/warn to consume others' terminologies to populate ontologies
- Improve and sustain low-level resources and create new ones to fill gaps because they are widely consumed
- Consumers subscribe to SR registries that provide notifications (push) of updates
- Develop metrics to assess resources usage
- Promote a standard to represent mapping between pre-semantic resources and SR (promote tools using it)

Group 4: Discoverability

- A service providing identifiers to semantic structures (Map of Standards, AgroPortal?) using URIs when relevant and/or DOI (allows usage tracking, citation…)
- Approach the infrastructure providers (DataCite, crossRef…) and tool providers to align and enforce common metadata models

Agrisemantics WG

IGAD meetings Berlin, 20 March 2018

This presentation:
http://bit.ly/Agrisem_Berlin18

Agrisemantics WG

Berlin, 21 March 2018

IGAD Session

Ongoing: towards recommendations (June)

Two recurring points from use cases and discussions:

1. Make semantic technologies/methodologies more accessible to non-ontologists
2. Provide suite(s) of tools that integrate tasks, e.g. metadata production, format conversion, access to existing resources… all along workflow

A sample from the discussions

- User-friendly GUIs & training for using SRs with data (in DBs, spreadsheets…)
- Promote identifiers for SRs: a URI when relevant and/or a DOI
- Common metadata models enforced by tools
- Tools that support visualization & human validation of mappings
- Improve and sustain low-level resources and create new ones to fill gaps (GACS)

Where are we?

Done: Landscaping report, collection of use cases (20)
Ongoing: analysis of use cases
To do: recommendations for the future of semantics for agricultural data - software, functionalities, semantic assets to enhance data interoperability in agriculture.

Group closes on 9 Dec 2018 (deadline extended)

Goal for this session

1. **Validate our summary of requirements**
2. **Initial organization of recommendations**
3. **Plan next steps**

Timeframe (tentative)

Apr: First draft recommendations

May: feedback and revisions

Jun: recommendations published

Jul: dissemination starts

Aug:

Sept:

Oct:

Nov:

Dec 9: group closes

Agenda

9:30 - 10:00 Intro - presentation of use case and group org
10:00 - 10:30 Group work

10:30 - 11:00 Coffee

11:00 - 11:45 Group work
11:45 - 12:30 Presentation of group work (up to 2 min per group)
12:30 - 13:00 Discussion, next steps

Recommendations

What should be **improved**, **created**, or **stopped** (and how) to leverage the uptake of semantic approaches

How could they be more efficient, adapted, and FAIR?

Think of:

- Technical
- Content
- Practices & skills
- Organisational

Roles:

- Funders
- Application developers
- Resource maintainers

Group breakouts

1. Editing tools (onto/thes creation & merging)
2. Alignments (creation, publishing, reuse…)
3. Consumption (rdfizing data, handling updates...)
4. Discoverability & Availability (metadata, repositories...)
5. Reference resources

Cross-cutting topics:

Training & skills -- Documentation -- Provenance

Use case analysis - pointers

Use cases map - http://bit.ly/UC_mindmap

Need MindMup2.0 for Google Drive

Deliverable 2 Top folder: http://bit.ly/Del2_folder

Use Case folder: http://bit.ly/UC_folder

With original use cases

Group 1: editing tools

- Variety of concept organizations: vocabularies vs thesauri vs ontologies (hierarchical, categories,…)
- User-friendly
  - manipulating structure (bulk edits, drag & drop…)
  - cloud-based (no install)
  - format transformation (e.g. owl ←→ obo)
  - Easy interaction with non semantic experts (e.g. biologist)
- Natively support best practices (e.g. FAIR)
- Quality checks (structure and content)
- Easy term/concept reuse (selection, provenance)
- Tool support (continuity, availability, cost, training, etc.)

Group 2: Tools and services for alignment

- Included in editing environments and repositories
- Support collaborative eval (also, crowd sourced?)
- Create documentation for process and evaluation
- Include SoA alg., exploit metadata
- Extract mappings in large scale (for mining)
- Provide metadata and provenance info
- Discovery services that deal with SPARQL e.p.
- What is the governance behind the provenance?

Group 3: consuming ontos and vocs

- Support their evolution and inform users about that
- Keep ontologies and their underlying terminologies "aligned"
- Good practices: Prevent ontology overload
- Support (funding) useful, low-level resources
- Support conversion of data from csv into RDF
  - in a reproducible manner
  - According appropriate, normative, semantic models (eg INSPIRE)

Group 4: discovery & availability

- Discoverable semantic resources, concepts, mappings, SPARQL endpoints, data
- Semantic resources should have
    - Rich(er) metadata
    - Metadata automatically generated (when possible)
    - Dereferenceable URIs
- Data should be discoverable by the SR used
- (More) provenance info

Group 5: reference resources

- Cover: socio-economic objects, pests, plant diseases, livestock, grazing activities + data/resource types
- Accessible online (eg, ICASA)
- Usable, ie. with URIs, metadata, structured and formal. Eg, Foodies, AgroRDF
- Reference for application ontologies

Group 6: training & skills

- On open tools
- On semantics (e.g., formal logic(s), open vs closed world, reasoning
  - OWL, RDF, RDFS, SPARQL, CL, CLIF, KIF, etc.
- case studies, best practices, lessons learned, shared experiences
- On ontology/schema merging
- On issues related to collaboration, e.g., intellectual properties

Group 7: documentation

- On schema structure (and semantics)
- On best practices for schemas
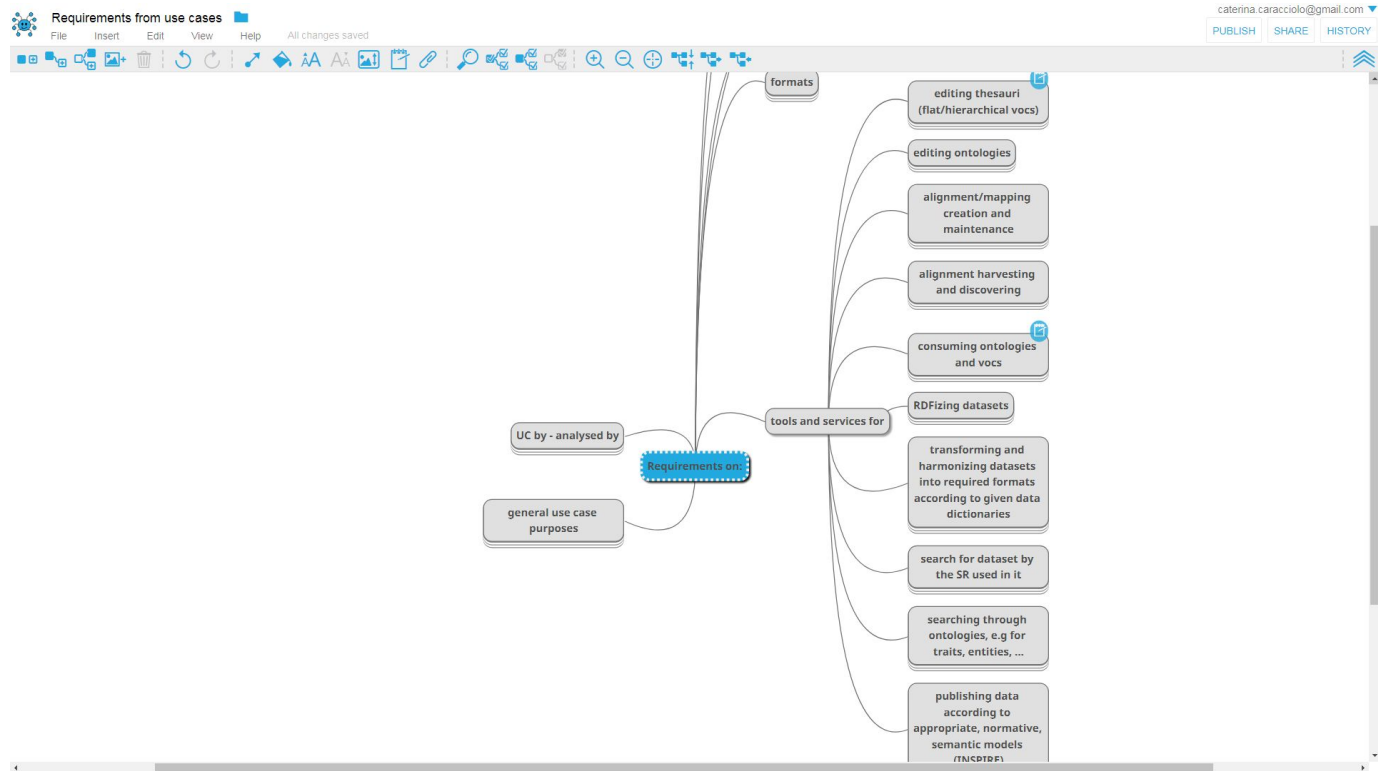- On term/concept status
- Use metadata to build documentation
- On open tools

Group 8: Provenance

- term/concept status
- metadata status
- governance mechanism
- governance status
- how do changes/edits persist?

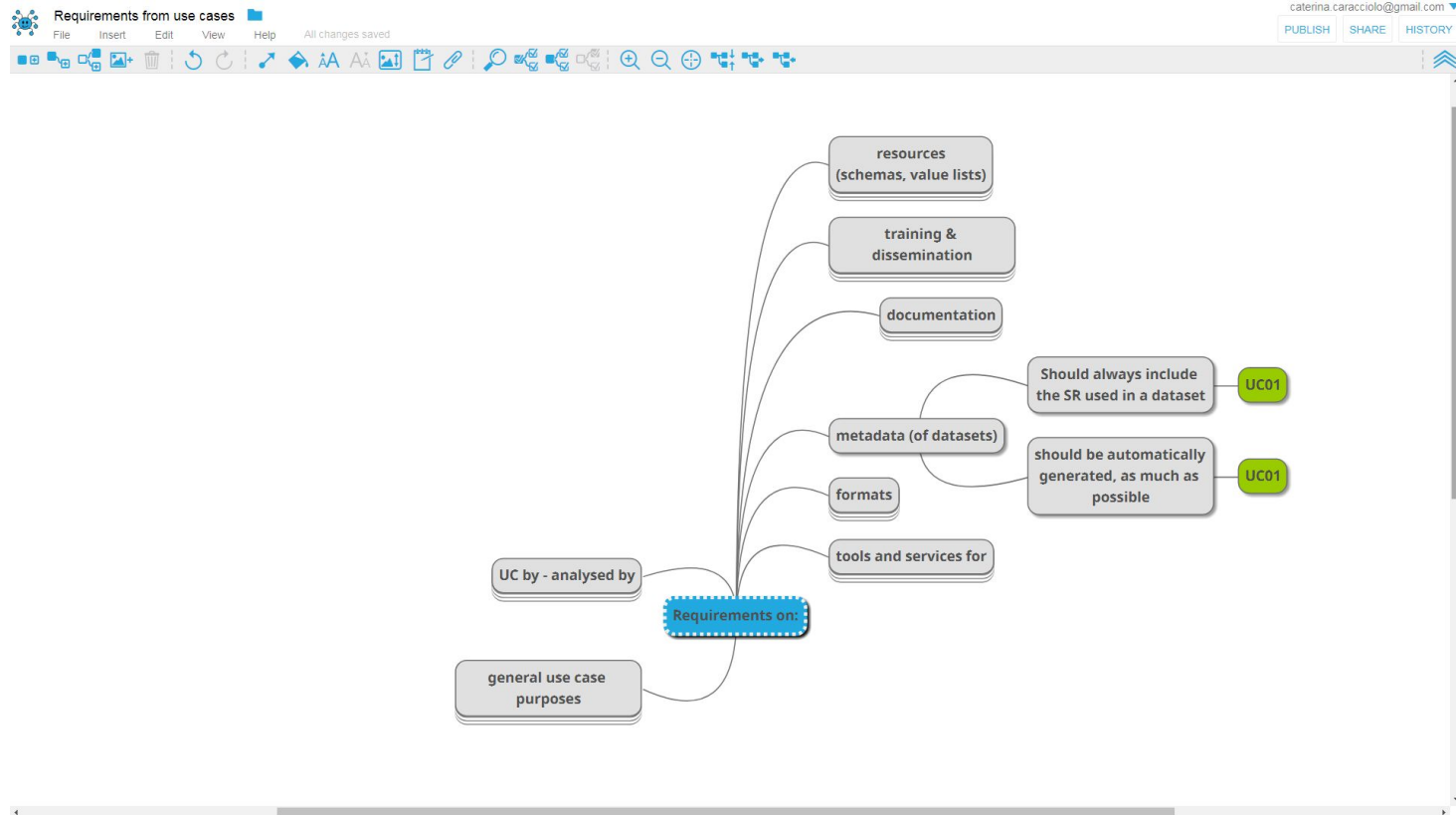# Use case analysis - birds' eye view
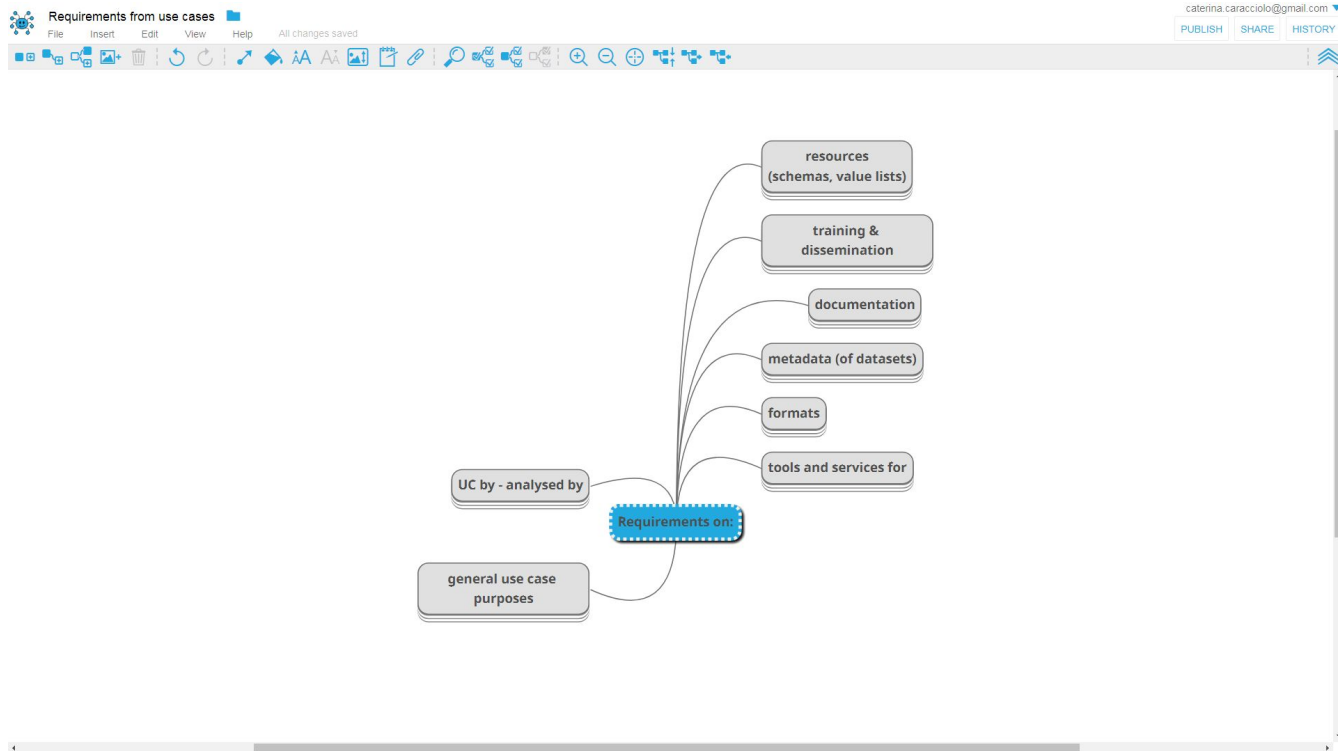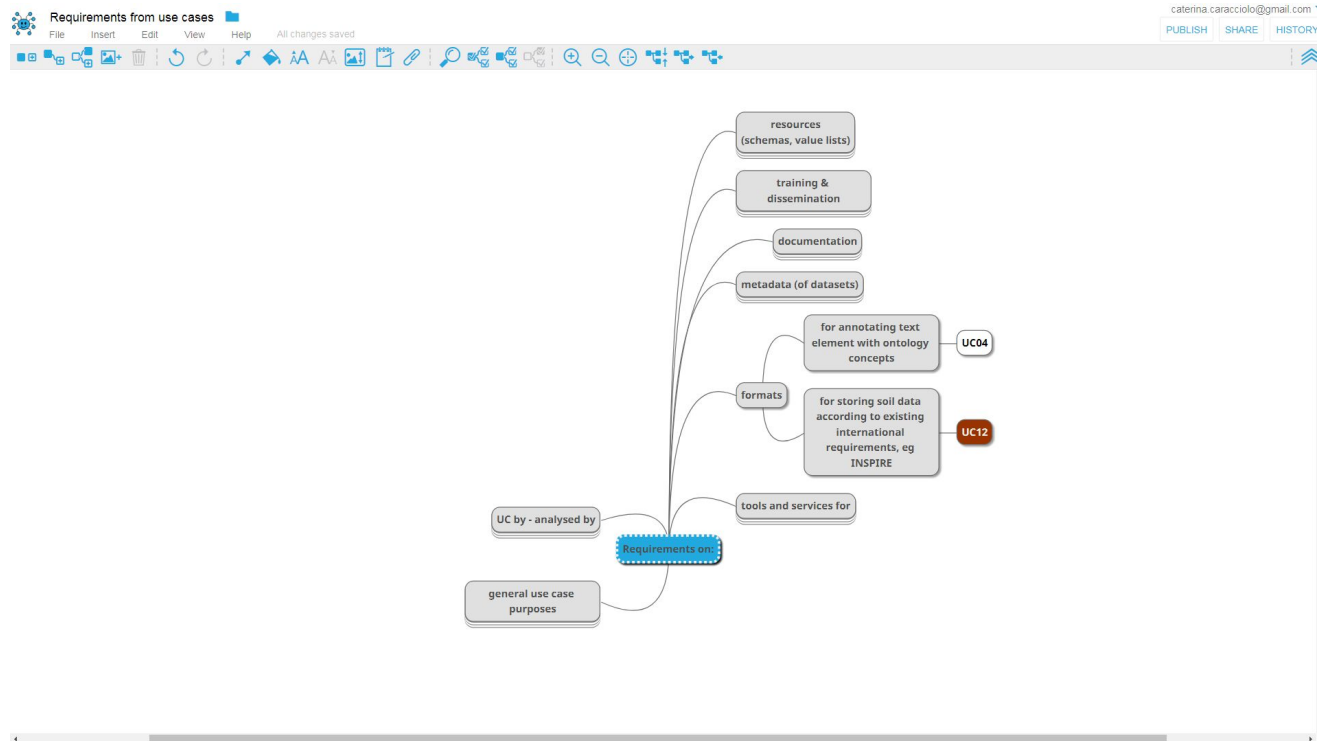
First group - tools and services

First group - tools and services

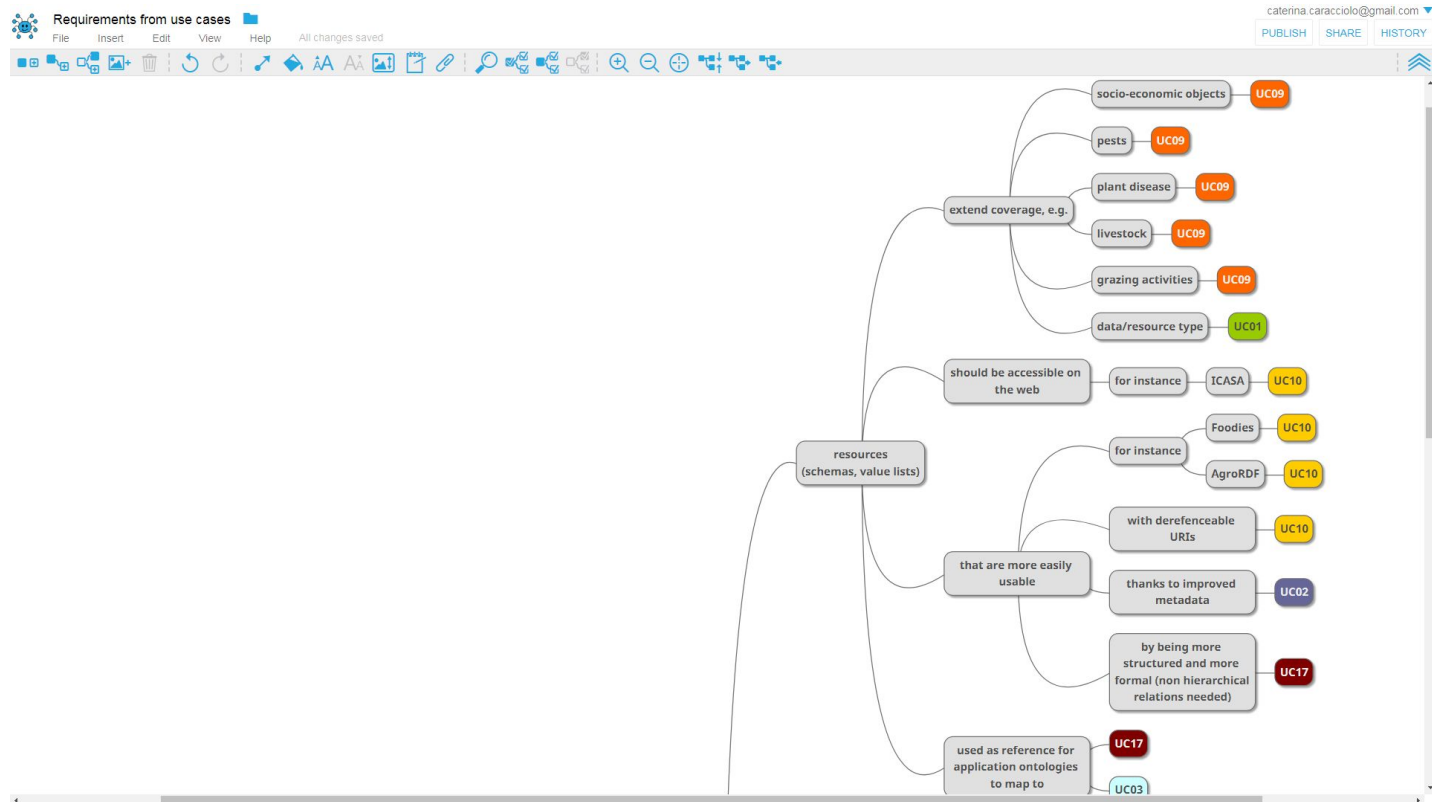# First group - metadata

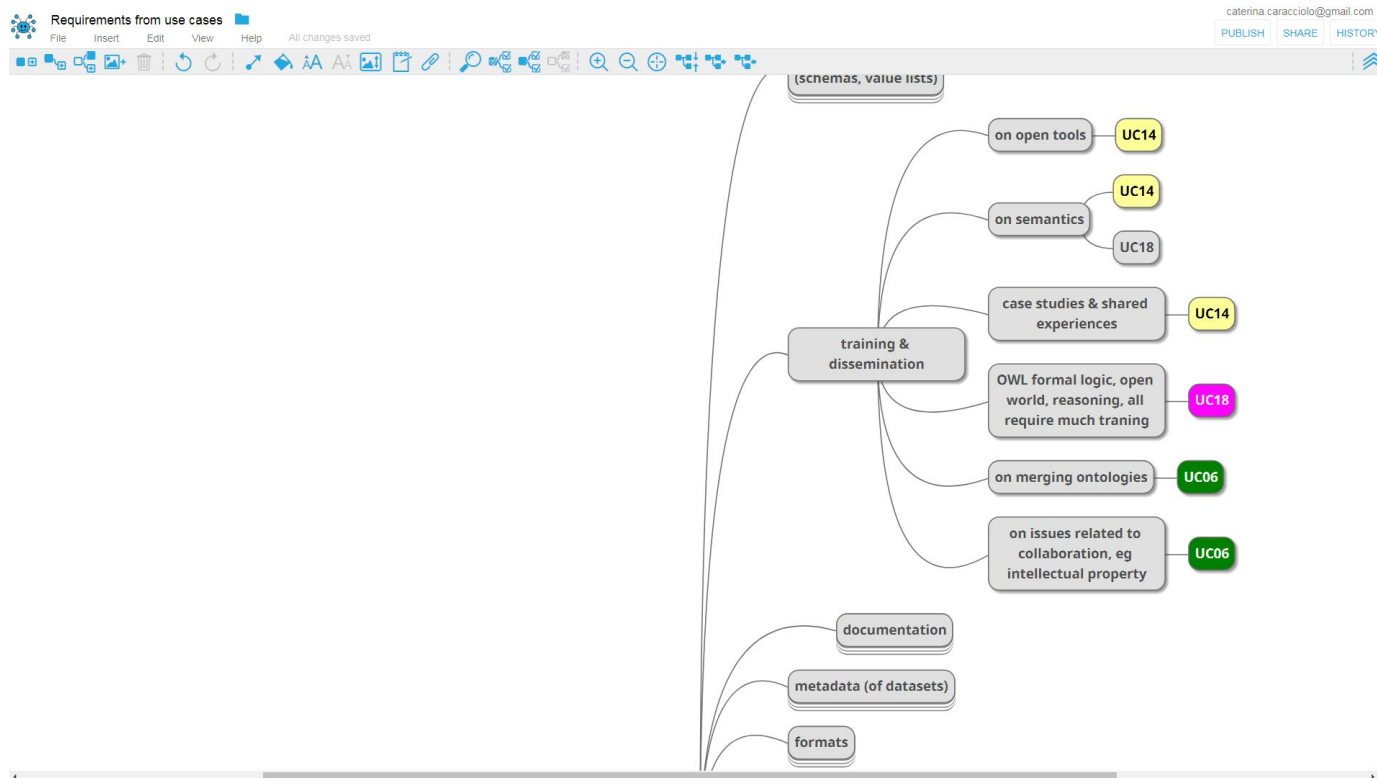www.rd-alliance.org - @resdatall
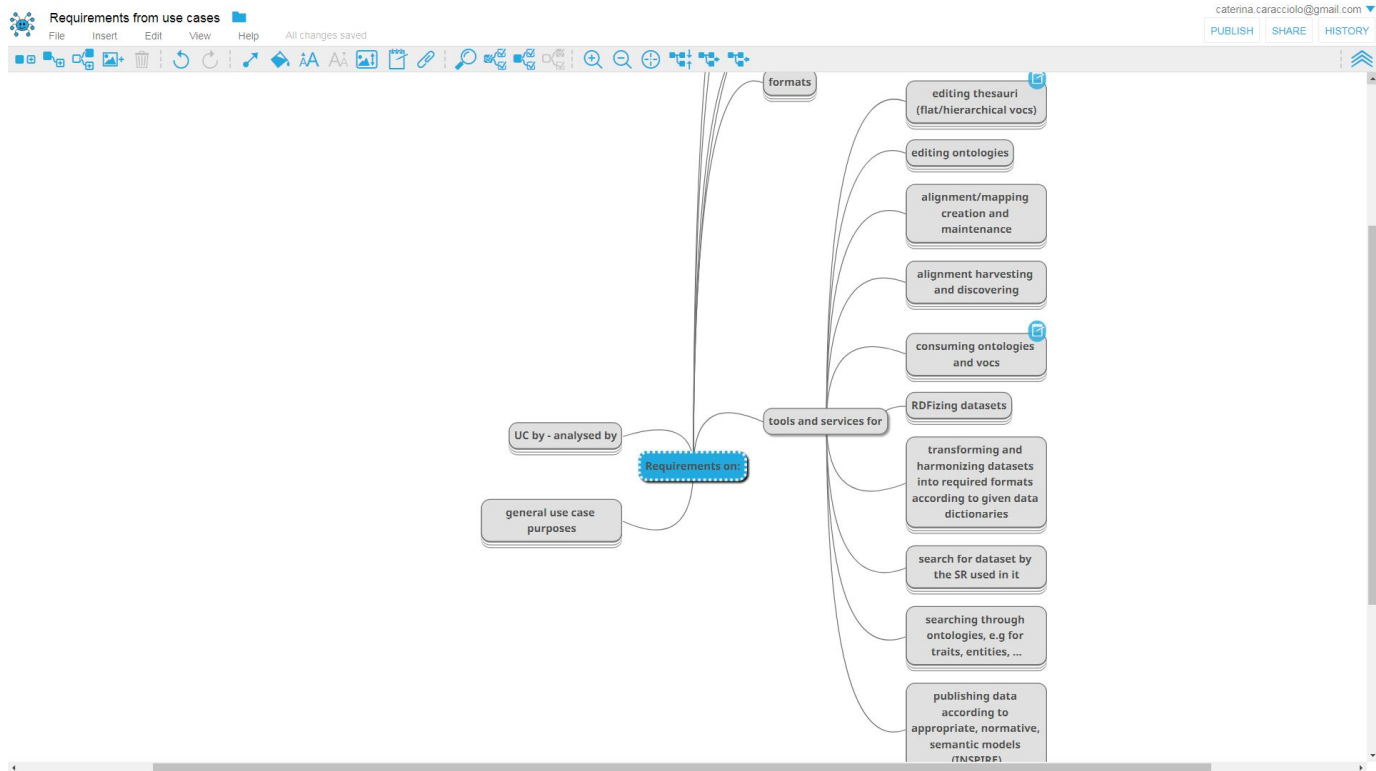
First group - formats

First group - resources (schemas, value lists)

First group - training and dissemination

First group - documentation