

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Work Session on Statistical Metadata (METIS)
(Geneva, Switzerland, 10-12 March 2010)

METIS METADATA CASE STUDIES: WHAT ARE THEY GOOD FOR?

Paper prepared by Alistair Hamilton, Australian Bureau of Statistics,

I. INTRODUCTION

1. This paper arises from an observation that, at least in the experience of the ABS, Metadata Case Studies (from the ABS and elsewhere) have proven useful for a variety of purposes, many of which may be not be “expected” primary usages.
2. This prompted questions such as
 - what additional secondary uses have other agencies encountered from their case studies?
 - would it be worth deliberately supporting some of these secondary purposes in the approach taken to structuring and maintaining case studies?
3. Recognising and promoting appropriate secondary uses may have the potential to increase the “value proposition” an agency sees as arising from a decision to commit resources to preparing and maintaining a Case Study.
4. Section II explores the stated and apparent “expected” uses of the case studies. It concludes that readers who are neither particularly interested in a common (statistical) metadata framework, nor self identified members of the METIS community, appear not to be targeted currently.
5. Section III describes four categories of secondary use experienced within the ABS. Within each of these categories two or more specific examples of use have arisen (ie none have been “one off” use cases).
6. The final section explores options for moving forward.
7. One option would be to continue to focus only on the primary uses, while perhaps giving those responsible for preparing and maintaining case studies a little more information on these uses and how, broadly, the content to be provided might support them.

8. A second possibility would be to briefly review actual and potential secondary uses from a broader perspective and then more actively promote such uses and/or provide broad guidance on how best to manage and support them.

II. WHAT ARE THE CURRENTLY “EXPECTED” USES OF CASE STUDIES?

9. Preparing and maintaining a relatively up to date, informative Case Study requires a significant (although not overwhelming) investment of time. A search failed to discover readily accessible documentation that focuses directly on the uses and benefits expected to flow from that investment.

10. Nevertheless, some information is documented explicitly. Other information can be deduced from the way case study material has been utilised within METIS.

11. The evolution of case studies, in a structured form, can be traced back to the initiation of the METIS Common Metadata Framework (CMF) in 2004. From inception [1], the design of the CMF included an element (Part D) related to implementation of statistical metadata systems.

12. The initial design of this element did not specifically limit it to case studies. To the extent more general “practitioner oriented” implementation guides, checklists etc might be developed and made available they could also be accommodated in Part D.

13. Nevertheless, case studies were settled upon as a good means of starting to populate Part D.

14. This context led to a number of design decisions in regard to the case study template. For example

- Many elements in the template had deliberate connections with other Parts within the CMF (eg the corporate context, relationships with the statistical business process, application of metadata standards). This allowed those whose interest was primarily attracted by other parts of the CMF to find which elements of case studies might contain information and examples of interest to them.
- More generally, there was an emphasis on consistency of structure across case studies, to the extent practicable, so they could be considered as a suite, and compared in various regards.
- Content from the “lessons learned” element of the case studies was subsequently copied for incorporation within Part A.

15. The intention that the case studies form part of the CMF, and support the other parts, is clear. This implies one expected audience is those with an interest in the CMF.

16. This does not in itself lead to a detailed sense of the expected usage of the case studies, as the audience for the CMF, and their expected usage of it, only expressed in general terms. Part A cites the original intent from 2004 being that the CMF would bring together information “to help statistical organizations to develop their metadata systems”[2]. Current wording in the wiki (which hosts the electronic, and primary, expression of the CMF) is more general in stating that the CMF “provides vital knowledge for anyone working with statistical metadata”[3]. This does not refer to an expectation readers will be working within statistical organizations, nor that they will be developing metadata systems.

17. Nevertheless, it seems reasonable to expect anyone who encounters the case studies in the context of the CMF is approaching them from a “framework” perspective. For example, a “Common Metadata Framework” is not a logical place to start looking for information for someone whose interest relates to practice in a specific agency. (If, however, a user who had such a purpose in mind

was already familiar with the CMF then they might use it as a navigation aide). The connection with the CMF suggests users who are likely interested in structured, broadly comparable information across multiple agencies for purposes such as

- discerning trends, leading practice, common practice
- gauging uptake of various standards and technologies and practical experience with them

18. This usage has been mirrored in recent METIS Work Sessions and Workshops where session organizers undertake a review across case studies and draw out key themes.

19. Over time there has emerged recognition of the case studies also serving as a “stand alone” resource, independently of their ongoing role in the context of the CMF.

20. This is reflected in the METIS wiki where the top level entry and orientation document introduces, and provides navigation to, the case studies independently of the CMF [4].

21. This entry and orientation document refers to the wiki as “a place for people working in official statistics to share ideas about statistical metadata”. This could be read as another indication of expected audience and purpose, which would fall somewhere between the 2004 “more restricted” and current “less restricted” audiences discussed above in the context of the CMF.

22. In the “stand alone” context, use cases which target the case study for one particular agency make sense. Some of these use cases might also value a structure that “best tells the story for a particular agency” over a structure that provides commonality and comparability across case studies.

23. The current guidelines for completing case studies balance these interests when they state

The template is flexible. If a particular heading is not relevant, it can be modified to something more appropriate, or deleted. If there appears to be no place for certain information, new headings/sections should be added [5].

24. While case studies standing alone within the METIS wiki suggest that the content of an individual case study may be of primary interest in some use cases, there remains a degree of explicit and implicit expectation in such a scenario that the primary audience will be other people working in official statistics, with a particular interest in statistical metadata, who are familiar with METIS and its wiki.

25. The introduction to the summary of case studies at the 2008 METIS Work Session covered both the CMF related and stand alone aspects when it described the purpose of case studies as

- To get materials for part D: Implementation, but can provide input to all parts of the CMF
- Transfer of knowledge & experience among community; [6]

26. People who are neither particularly interested in a common (statistical) metadata framework, nor self identified members of the METIS community, appear not to be targeted by the current definition of the audience for the case studies.

III. EXAMPLES OF “SECONDARY” USES

27. Some observed uses of case study material beyond the apparent “target” audience and/or primary purposes are described below.

28. Some of these cases involved copying material from case studies into other documents (eg

emails), while in other cases links were provided either to the top level (table of contents) page of a case study or to one or more web pages of relevance within the case study. The extracts or links were often packaged with additional information that

- highlighted the aspects of greatest likely interest and relevance within the extract or link, and/or
- elaborated upon specific aspects the content sourced from the case study.

29. Material sourced from case studies often provided a useful “bigger picture” when discussing a particular aspect in more detail.

30. By listing these uses it is not asserted that these are necessarily appropriate uses which should be supported. Nevertheless, at least within the ABS, there have not been better reference sources which have been readily available and publicly accessible.

31. “Crafting” customised responses to each of these information needs would have been much less efficient and/or would have resulted in much less coherent and detailed information being provided.

32. “Assembling” customised responses by drawing on selected material from case studies, together with additional context specific information, was efficient in those cases where a simply providing an extract from, or links to, the case study documentation was not sufficient on its own.

33. In terms of the ABS website, and explorations of some other NSI websites, while outputs from an agency's statistical metadata system are published there tends to be relatively little consolidated and up to date “publicly facing” information about the statistical metadata system in its own right. (Sometimes there is information about some of the specific processes or applications within the system, but not about the system overall.)

34. Material from the case studies has the advantage of having been designed to provide a coherent perspective to people outside the agency itself.

35. Also, in the ABS context, in accordance with local policies for contributing to wikis, the material has already been reviewed to ensure it is suitable for release outside the agency.

36. Furthermore, in the ABS context, while content contributed to an external wiki must be quality assured in various regards

- this content can be less formal, and
- the process can be more agile

than “officially” publishing equivalent material to the ABS website.

Preparing for international visits

37. This is a use that colleagues from other agencies have mentioned also. The purpose is typically to help to build an initial background understanding of the other agency's environment and approach and to identify topics of interest to follow up further.

38. This use case applies whether someone from another NSI is coming to visit, or whether someone is going to visit another NSI. To avoid complicating expression to cover both possibilities, a scenario of someone coming to visit is described.

39. As the case studies cover broader context such as:

- considerations of the statistical business process;

- IT Architecture;
- software sharing opportunities,

reference to them has proved useful in some cases where the visitor's focus is not specifically related to statistical metadata (e.g. they might not self identify with the “METIS community”). Likewise, once alerted to its existence, the material has proved useful to review “in preparation” by people within the ABS who do not self identify with the “METIS community”.

40. In a couple of cases the material from the relevant case study has proven the single most coherent and practical source of background information and discussion starters for visits that are not specifically focused on the topic of statistical metadata systems.

Information for vendors

41. Many vendors have a relatively strong understanding of “metadata” in the context of Business Intelligence, Enterprise Resource Planning, Business Process Management and other “generic” business applications. Based on this experience they can assume that addressing “metadata” in the context of a statistical production process equates to essentially the same thing.

42. If “statistical metadata” is then explained from first principles there can be the opposite problem, with vendors unable to clearly link this explanation back to “metadata” in the context they understand it. This may lead to missing potential opportunities for specifically statistically oriented processes, systems, data and metadata to

- integrate better with more generic business processes which are common to NSIs and other enterprises
- better harness, and integrate with, the capabilities of commercial applications and platforms, and thus improve workflows within the NSI and beyond

43. Reference to case study material is not ideal for the purpose but can provide practically oriented information with

- connections to architectural details that many vendors will recognise, and
- business process descriptions that help vendors understand the context for statistical metadata systems.

44. References to case studies (particularly the one for the NSI concerned) appear more likely to be assimilated by vendors in a reasonable period of time than directing them to the CMF overall.

45. The fact case studies are in the public domain is also very useful in this context. Where vendors are, e.g., seeking more context for metadata related questions in an RFI (request for information) references to case study material reduce questions around whether a query from particular vendor has been responded to with too much or too little additional information.

Metadata practitioners beyond the “METIS community”

46. The ABS, and it appears many other NSIs, are having more regular and purposeful contacts with communities beyond the realm of “official statistics” who share common interests in various regards. This includes engaging with the metadata practitioners from those communities who may have an orientation toward fields such as:

- “traditional” resource discovery and resource management metadata
- structural content rendered via RDF, OWL etc to support semantic web applications
- metadata for data archives, for geospatial data, for various forms of scientific data

47. Similarly to the case for vendors, each practitioner tends to have a “metadata world view” centred on their area of primary interest. Areas of overlapping interest between two practitioners:

- may fail to be identified at all because each describes the area of overlap using different terms and in a different business context;
- may not lead to fruitful engagement because neither party understands the issue in the context of the other
 - for example, someone who is passionately interested in resource discovery may not appreciate someone with an interest in statistical metadata seeing discovery as a means to an end (the end being people making informed use of the statistical outputs they discover)
 - similarly, the person with the passionate interest in statistics may not appreciate their statistical output being viewed as “just another resource” to be discovered along side others

48. Once again, content from metadata cases studies is not necessarily ideal for the purpose but can provide good background material for a discussion on similarities and differences, including identification of areas of common, or otherwise interrelated, interest. This can lead to targeting specific areas for more detailed exchange of information and for “keeping each other in the loop” over time. This, in turn, can lead to active cooperation in regard to specific areas of clearly understood mutual interest.

Information for potential recruits

49. As supporting material, the local case study has proved useful in a couple of cases when potential recruits from outside the ABS were seeking more information about the activities and context of the team within the ABS that has the lead role for promoting the operations and the development of the statistical metadata system within the ABS.

50. Reference to the case study, of course, is not a substitute for a duty statement related to the vacant position or for a work area information sheet related to the particular unit within the organisation. It does, however, summarise broader organisational context. Considered as one in a suite of case studies from around the world, it also gives a sense of connection with an international community. This prospect can be appealing and inspiring to some candidates for recruitment.

51. Once again advantages arise from the fact that the material is written for readers who are not already familiar with the organisation and from the fact that it is already in the public domain.

52. The potential recruits in question, who have now joined the team, had backgrounds in statistical metadata management (but outside the field of official statistics) and business informatics. It is recognised that references to case studies might generate confusion and alarm for true “generalist” readers.

IV. WHERE TO FROM HERE?

53. The existence of secondary uses does not change the fact that the “expected” uses outlined in Section II are the primary purpose for which case studies are written, and are therefore the primary drivers of how case studies are structured and expressed.

54. It is assumed (although the assumption is open to being contested) that considered, selective and contextualised secondary use of case study content is not viewed by METIS as something that should be actively discouraged under all circumstances.

55. This might leave appropriate secondary use being considered as harmless but irrelevant as far as METIS is concerned. If that is the case then no further consideration need be given to secondary uses.

56. Nevertheless, for the benefit of those preparing and maintaining case studies, there might be some additional information provided in regard to “expected” primary uses. (Alternatively, if secondary uses are to be supported in addition to primary uses then this additional information for those preparing and maintaining case studies should be prepared after the further evaluation of secondary uses which is proposed below.)

57. For example, while the current guidelines for population of the metadata case study template provide a description of the type of information to be provided in each field, little insight is provided (either in the introduction to the guidelines or in the field by field instructions) on why the information is being collected (eg the purposes to which it is anticipated the information provided might be put).

58. This is not unusual for such guidelines. More information about anticipated uses, however, may provide additional guidance those populating such templates in terms of the level, structure and content of the information they provide.

59. If, however, there is some level of interest in better understanding, from a broader perspective than a single NSI, actual and potential secondary uses of case study material then possibly

- a wiki page could be established for sharing ideas and information, or
- an informal “survey” of case study authors could be undertaken.

(Unlike many other wikis, the METIS wiki doesn't have a tradition of hosting “discussions” in addition to “final” documents. That may, however, be open to change?)

60. Having identified a more complete frame of reference in regard to secondary uses, and possibly an extended categorisation of them, it would then be possible to provide a summary of primary and secondary uses which might

- encourage NSIs to recognise additional value in either submitting a case study (if they have not already done so) or updating and/or extending their existing case study
- encourage NSIs to make additional use of existing case study material to realise additional value from it
- be circulated for information via the Conference of European Statisticians, and possibly via other channels, to raise broader awareness and interest in regard to the case studies within communities that in some sense “border” the METIS community but are not within it

61. It may also be the case that some potential secondary uses of case study material raise concerns for members of METIS. There may be a desire to provide guidelines on recommended practice covering aspects such as

- circumstances where secondary use is either not appropriate or should only occur via direct consultation with the case study author
- citation practices.

62. In any case, secondary use will almost certainly raise a more general question, which has arisen in other contexts already [7], around the formal status of case study content. For example, is a case study's content considered to be

- an official statement on behalf of the relevant NSI, which should therefore undergo the same level of scrutiny and sign off as, for example, an equivalent statement published on the agency's official website, or
- more akin to a “professional paper” in terms of its standing, including being associated primarily with the author (in a professional rather than private capacity) rather than formally with the organisation as a whole
 - this leaves the responsibility with the author to determine the appropriate level of local

consultation and sign off

63. If there is a desire to support secondary uses of case study material similar in nature to those described in Section III then in terms of achieving an appropriate level of

- detail
- currency
- candidness

in regard to case study material it would seem best for it to have the status of a professional entry in a wiki rather than being a formally published statement on behalf of the NSI as a whole

64. Especially if the material contained in case studies is to be more actively promoted for secondary use beyond the boundaries of the METIS community, however, its status must be made clear.

REFERENCES

[1] Juraj Riecan (Chief, Dissemination and Statistical Information Systems, UNECE), “Draft v0.1 of the structure for the framework document”, Unpublished email, 2004.06.22

[2] Foreword to “Part A - Statistical Metadata in a Corporate Context” November 2009 Version accessible via

<http://www1.unece.org/stat/platform/display/metis/Part+A+-+Statistical+Metadata+in+a+Corporate+Context>

[3] <http://www1.unece.org/stat/platform/display/metis/The+Common+Metadata+Framework>

[4] <http://www1.unece.org/stat/platform/display/metis/METIS-wiki>

[5] These guidelines are currently provided within the METIS-wiki

<http://www1.unece.org/stat/platform/download/attachments/3866689/Case+Study+Template+-+October+2009.doc?version=1>

[6] Matjaz Jug, “Summary of case studies”, presentation to METIS 2008

<http://www.unece.org/stats/documents/ece/ces/ge.40/2008/zip.8.e.ppt>

[7] For example, in 20/ADD.3 (“Comments on the development of the Common Metadata Framework”) of the papers discussed at the CES Bureau meeting of 15-16 October 2009 the following country comment is recorded

statistical organizations “should be officially invited at the DG level to provide the case studies”, rather than the current practice of inviting delegates to work sessions on statistical metadata (METIS) to prepare them

<http://www.unece.org/stats/documents/ece/ces/bur/2009/mtg1/20.add.3.e.pdf>

All references and cited URLs were valid as of 2010.02.16. Content of web pages may have been updated subsequent to citation.