

A renewed approach to document, file, and administrative data review: the program vitae

Benoît Gauthier

gauthier@circum.com

Claude-Anne Godbout-Gauthier

cagg@circum.com

Natalie Kishchuk

nkishchuk@kishchuk.ca

Presentation to the 2013 Canadian Evaluation Society Annual Conference, Toronto (June 10, 2013)

The Problem

- Almost all evaluation studies include a review of documents, files, and administrative data.
- Often costly, such reviews sometimes produce limited useful results by lack of organization of the corpus.
- Much more material is available than previously through websites, electronic storage, and information systems: creating easier opportunities to use it as evaluation data

The Literature

- Nick L. Smith And Barbara M. Florini, "The Project Vita as a Documentation and Evaluation Tool for Large-Scale Research and Development Projects", *Evaluation and Program Planning*, Vol. 16, pp. 49-53, 1993
- Lists/database of project activities (similar to a person's curriculum vitae)
- Uses:
 - Program description: flexible and powerful tool
 - Measurement tool: consistency in use is key
 - Other uses

The Tool

- The program vitae is a structured data base of information on program input, activities, outputs, and, possibly, outcomes.
- Maps these chronologically: accretive picture, not only cumulative
- Several units of analysis
 - sub-programs, people, inputs, activities, outputs, possibly outcomes
- Relational nature: possible to link elements together
 - For instance, as part of sub-program X, a person co-wrote article Y with 5 other people.
- Each entry is referenced, for possible future use

Example 1: Context

- Evaluation of the Panel and Secretariat on Research Ethics
 - Entity that stewards the Tri-Council Policy Statement on Ethics of Research with Human Participants
- Summative evaluation conducted close to end of mandate, but prior to release of main result: Version 2 of the TCPS
- Evaluation covered 2002-2008, and activities of many committees, subcommittees and working groups outside of government
 - Project vita useful to document micro-steps towards achievement of main result
- Context of some tension, impatience
 - Project vita also useful to document facts and timelines, neutrally

Example 1: Data Base Structure and Content

- The low-tech version: Excel files and sheets by area of activity
 - E.g, Social science and humanities, clinical trials, Aboriginal research
- Dates as rows, outputs and their characteristics as sets of columns
- Cell entries: document titles, authors, audiences, main content
 - Allowed tracking production of incremental outputs over time, pattern and intensity of activity

Example 1: Supports and Challenges

- Facilitated by:
 - Good centralized record keeping: generally possible to find documents
 - Organization's structured workplan with timelines: could measure production and slippage against expected dates
- Hampered by versionitis

Example 1: CV Usage in the Evaluation

- Synthesis of outputs by expected result area: what was produced, when, by whom, with what contribution to the new TCPS
- Able to document challenges, delays and setbacks and as well as progress.

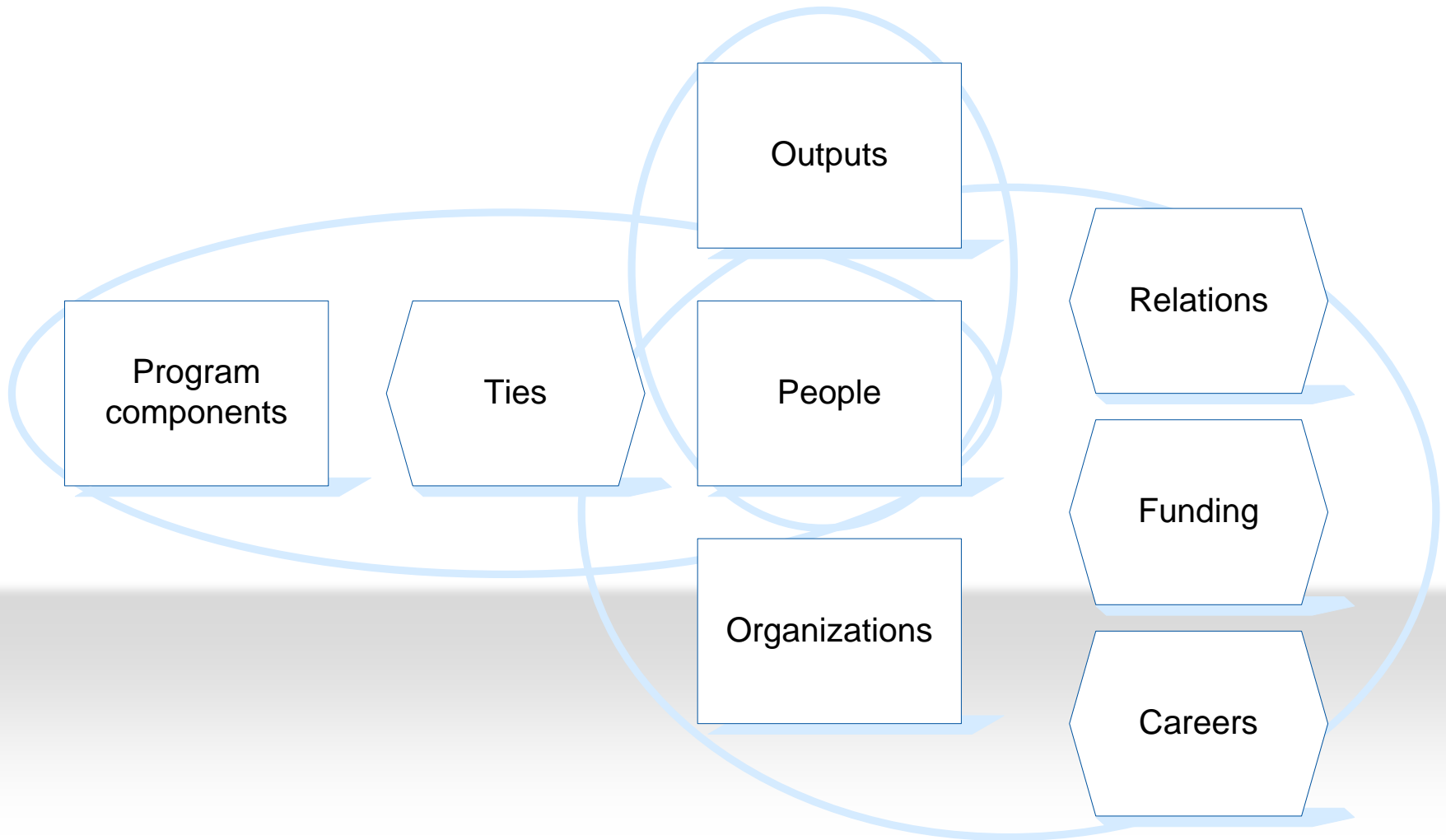
Example 2: Context

- An evaluation of a complex academic program
 - four subprograms
 - several different types of stakeholders
 - 10-year initiative
 - total program budget including partners: about \$80 million
- Corpus
 - 8,850 electronic documents of various natures
 - unknown number of paper documents
- Purpose of the vitae
 - documentation of program activities and outputs
 - some indicators of achievement of objectives
 - development of sampling frames

Example 2: CV Development Process

- **First try**
 - pure relational database
 - setup in about 1 professional day
 - single-user system
- **Second try**
 - relational data base superimposed on a flat db system
 - setup in about 2 professional days
 - single-user system
- **Training of analysts**
 - detailed training manual
 - initial work done jointly with the project manager
- **Corpus management**
 - one data base table to list documents all documents at attribute responsibility
 - electronic documents hyperlinked within the data base
- **Validation mechanisms**
 - some concurrent work
 - mainly ex post facto verifications
 - data cleaning required to standardize input and deal with cut-and-paste issues

Example 2: Data Base Structure



Example 2: CV Content

■ Number and types of documents used

- essentially electronic documents (paper documents provided little additional information)
- of 8,850 initial documents, 4,500 identified as candidates; others of lesser value
- included applications, management notes, data bases supplied by participants, annual reports, resumes, progress reports, etc.
- in the end, information in the data base from 600 documents

■ Data base data elements

- 5,800 people
- 1,500 organizations
- 4 program components
- 3,000 ties between individuals and the program
- 800 relationships between academic and non-academic sectors
- 7,800 career descriptors
- 1,100 funding contributions
- 5,100 tangible and identified program outputs

Example 2: CV Usage in the Evaluation

- Key in the description of the program activities and outputs
- Key in the analysis of trends over the program decade
- Key in support of other evaluation modules
 - valuable information in the preparation of case studies
 - supplied lists of individuals in various stakeholder types for survey purposes (along with e-mail addresses often)
 - permitted an analysis of publications

Example 2: Problems Encountered

■ Data base corruption

- upon reaching about 80% of processed documents
- single-user approach on non-networked computer not conducive to backups
- impossible to recuperate because lost work unidentifiable
- need to start over

■ Data integrity

- conflicting input information
- missing complete objects (e.g. a person not accounted for)
- partial input information (missing information on a given object)
- standardization required extra effort (e.g. institution name)
- program-maintained (aggregate and perceptual) data not always aligned with information collected in the vitae

■ Cost (second try)

- about \$25,000
- about 900 hours of work
- over 4 months

Lessons Learned

■ Strengths

- Flexible tool, various applications
- Thorough (when used consistently)
- Incontrovertible if the data gathering process are of quality

■ Weaknesses

- Can be time consuming
- Mitigating strategy: implement the vita early on in a program and have the program staff take ownership of it
- Requires access to primary data
- Data can be messy, which can lead to confusion, duplicates and overlap

■ Risks

- Possible challenges associated with data entry consistency when multiple people are involved
- May under-represent reality since only documented/sourced facts are assembled
- Losing focus on the end goal (program delivery/results)
- The program vita needs to remain a tool, not an end in itself

■ Opportunities

- Useful for documenting progress toward outcomes when these were not achieved in the evaluation period
- Source of info and lists for other evaluation components
- Best implemented as a management tool, on an on-going basis, rather than as a strict evaluation tool, ex post facto.