InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

TEAM Italy

in saeculis authenticus

Guidelines and Recommendations for E-mail Records Management and Long-term Preservation

Massimiliano Grandi

Archival Association of British Columbia (AABC)
Annual Conference

17 April 2010
Vancouver, BC, Canada

A PERVASIVE PHENOMENON

- In 2009 280.2 billion e-mails were sent each day
- "E-mail" is a term which indicates a system by means of which many different information objects are carried
- The word e-mail, as the more traditional "snail mail", also refers to the same objects which are transported
- Objectives:
    1. to gain intellectual control over this huge mass of material
    2. to appropriately handle e-mail records associated with e-mails

SOME PROBLEMS

1. Complex objects, composed of many parts (e.g., attachments, digital signatures, alternative versions of the same content)

2. The intellectual structure of an e-mail can be completely different from the physical one

3. Great number of inbound and outbound flows

4. The contents of different e-mails can be easily mixed through the available editing functions
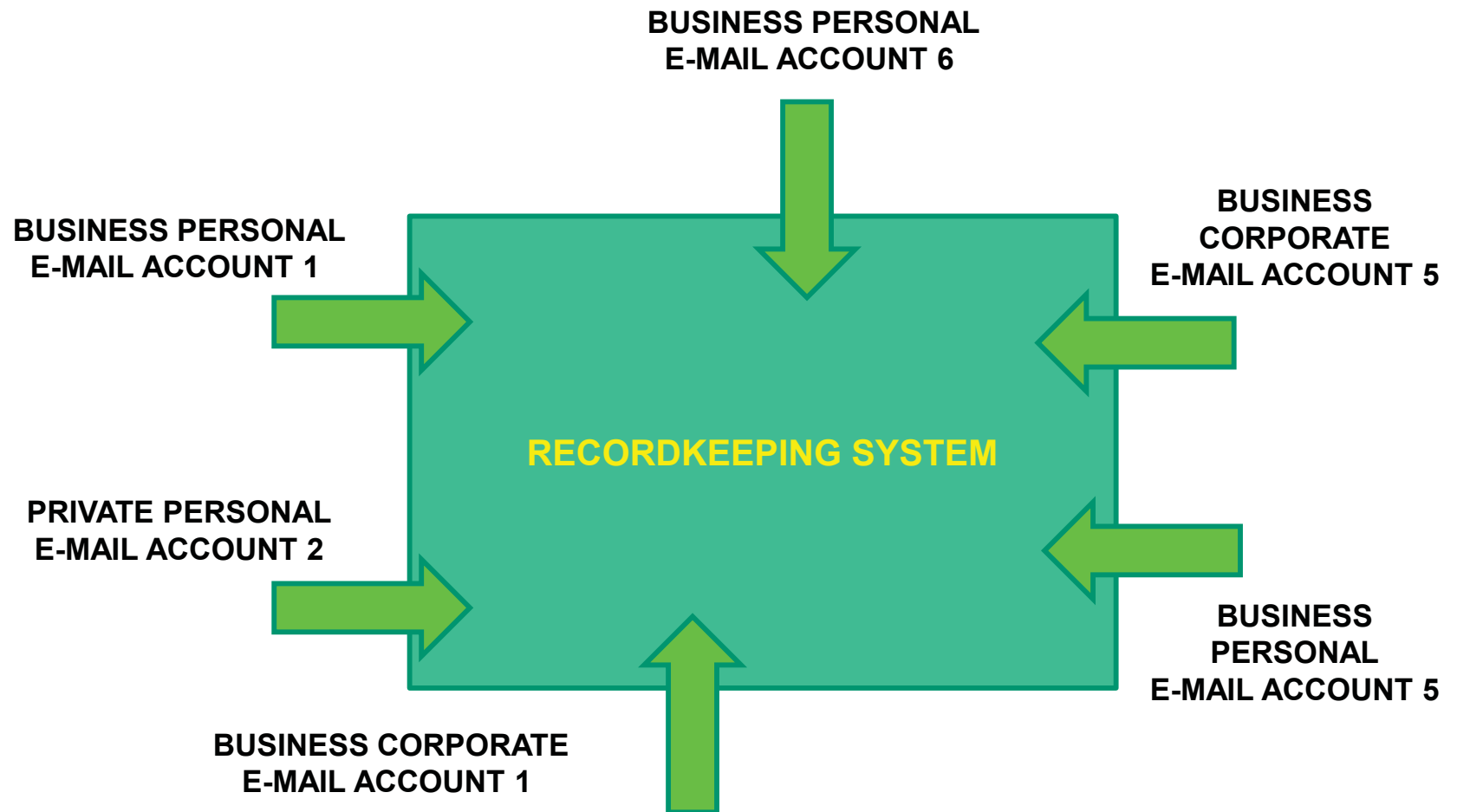
5. An e-mail can include any file format

INTEGRATION WITH THE RECORD-KEEPING SYSTEM

- The only strategy that is succesful and makes sense

- Research projects (e.g., DAVID, DIGITALE DUURZAAMHEID), standards (e.g., MoReq2, DOD), policies and most of the relevant literature recommend it

- Focus on e-mail records

- "Electronic mail, in and of itself, is not a single record series" (D.A. Wallace, 1998)
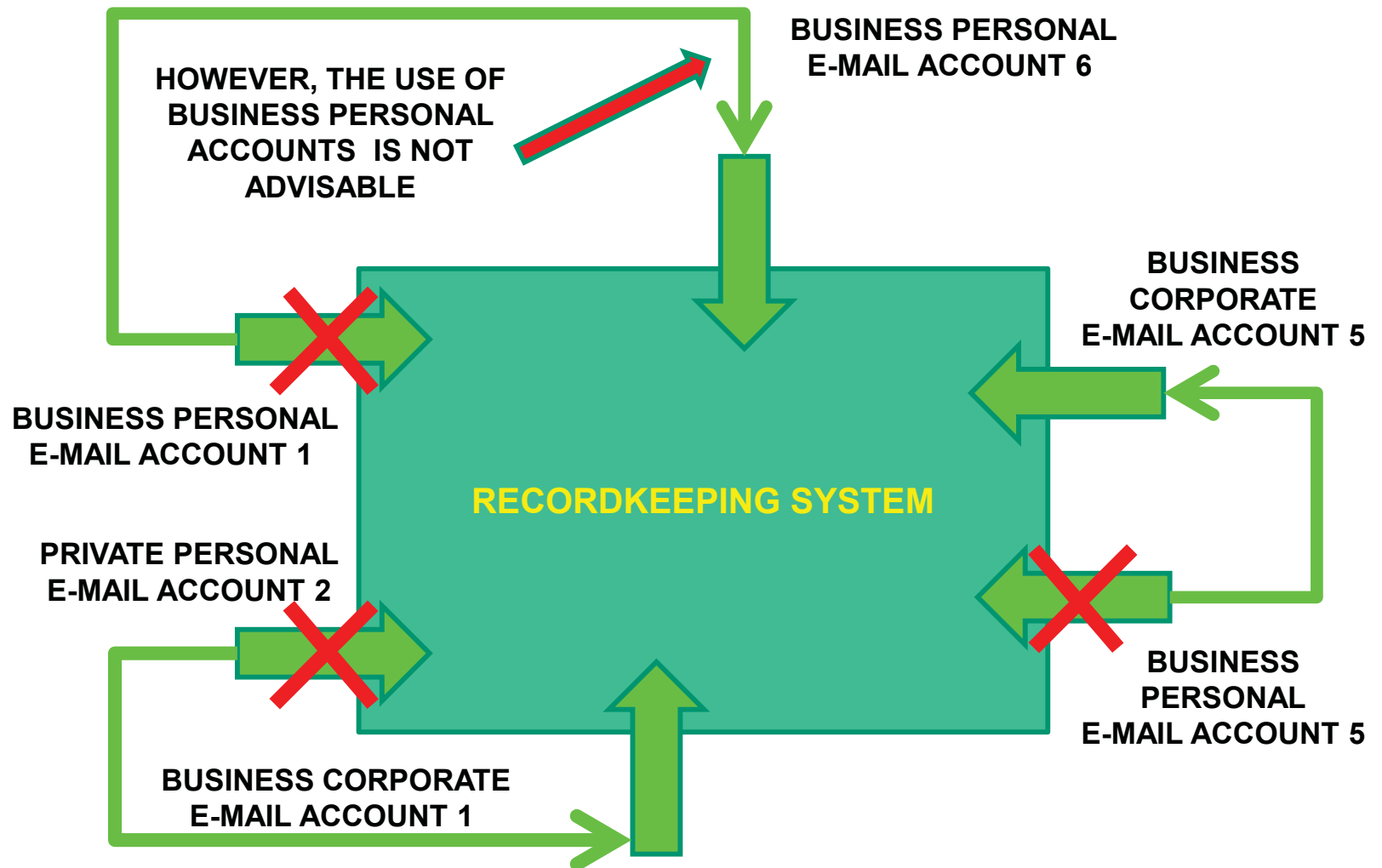
# GOVERNING FLOWS/1

**A FREQUENT SITUATION
CONDUCIVE TO ANARCHY …**

BUSINESS PERSONAL
E-MAIL ACCOUNT 6

BUSINESS
CORPORATE
E-MAIL ACCOUNT 5

BUSINESS PERSONAL
E-MAIL ACCOUNT 1

RECORDKEEPING SYSTEM

PRIVATE PERSONAL
E-MAIL ACCOUNT 2

BUSINESS
PERSONAL
E-MAIL ACCOUNT 5

BUSINESS CORPORATE
E-MAIL ACCOUNT 1

InterPARES Project
Massimiliano Grandi, TEAM Italy
Graduate Research Assistant

GOVERNING FLOWS/2

A SOUND ENVIRONMENT…

HOWEVER, THE USE OF BUSINESS PERSONAL ACCOUNTS IS NOT ADVISABLE

BUSINESS PERSONAL E-MAIL ACCOUNT 6

BUSINESS CORPORATE E-MAIL ACCOUNT 5

RECORDKEEPING SYSTEM

BUSINESS PERSONAL E-MAIL ACCOUNT 1

PRIVATE PERSONAL E-MAIL ACCOUNT 2

BUSINESS PERSONAL E-MAIL ACCOUNT 5

BUSINESS CORPORATE E-MAIL ACCOUNT 1

MASTERING THE KRAKEN/1

- An e-mail must be properly analysed to see which records are associated with each of them

- In theory a single e-mail can
  - Carry two or more records
  - Or a record can be sent by means of more E-mails

- We need to understand the function of each part of an e-mail (e.g., which parts of an e-mail are the attachments record; which part acts as cover letter, if any) and we need to identify and capture metadata

- How to manage the resulting immense workload?
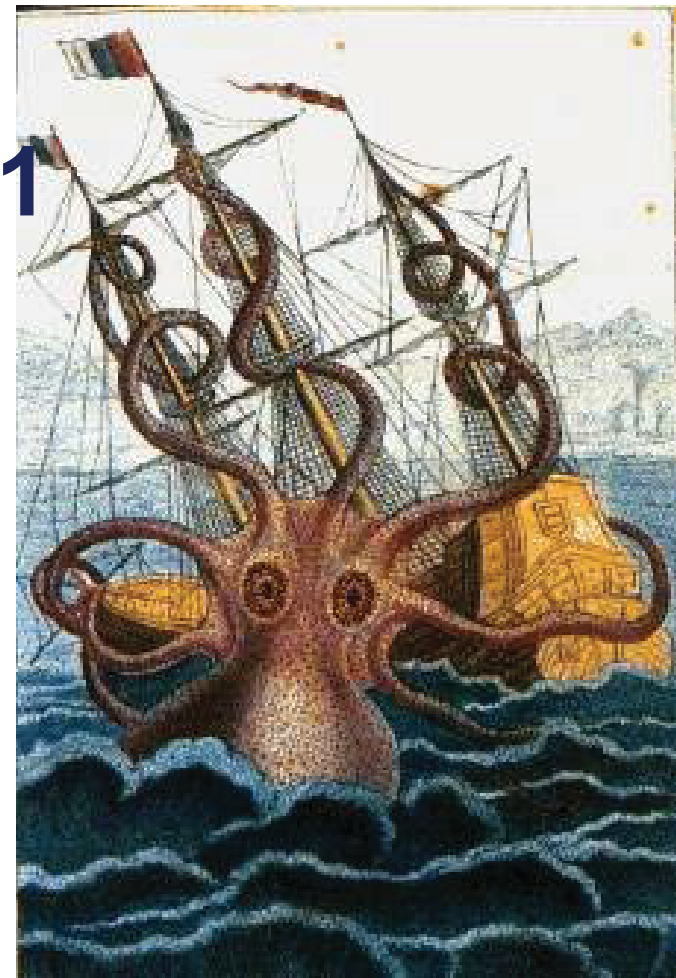
WIKIPEDIA: KRAKEN
http://en.wikipedia.org/wiki/Kraken

MASTERING THE KRAKEN/2

- A strategy based on four pillars
    1. Human supervision
    2. Advanced automation
    3. Management-oriented diplomatics
    4. Business process reengineering

- Humans must have the last word, not only because they are accountable, but also because they can help machines to cope with the ambiguities of interpretation and meaning

MASTERING THE KRAKEN/3

- It is unthinkable to give up automation when you have to handle large numbers of digital records
  - Metadata harvesting
  - Automatic parsing
  - Proposals of classification

- But humans can try to reduce the complexity and irregularity of the objects that must be evaluated

MANAGEMENT-ORIENTED DIPLOMATICS

- Diplomatics is the discipline which deals with the study of the formal aspects of records. Its knowledge can be useful to facilitate human and machine work

- By resorting to diplomatics, it is possible design regular documentary forms intended to characterize e-mail records (above all, but not only, the outgoing ones)

- Reduction of the kaleidoscopic array of configurations by inserting in documents planned to be declared records recurring formal elements that can be easily identified by both machines and human beings

- There are many technological tools (e.g., XML schema): to gain the needed consensus is often the real issue

BUSINESS PROCESS REENGINEERING

- In order to discover which documentary forms are necessary, analyses of business processes, activities and transactions are required
- Through BPR you can, for example:

  1. Discover the most widespread typologies of records

  2. Suggest convenient modifications to documentary forms

  3. Understand which metadata elements and values must be present in a record

  4. Plan workflows, so that – for example – only e-mails containing some definite metadata values or given attachments may be captured in a recordkeeping system

Trying to obtain exhaustiveness at any cost is a mistake

LONG-TERM PRESERVATION

- OAIS Model

- All the research projects have resorted to XML

- The most common strategy is to create XML–based AIP which physically wrap all the components of an e-mail record. However, an AIP need not be a physical object

- Actions for long-term preservation must be based on migration and use of standards formats and software- and hardware independent methods of representation of information.

- Emulation is not a viable option

- Conversion to long-term preservation formats should occur after capture for e-mails selected for permanent retention, but may be delayed until they are moved to the digital repository, of course, provided that all the metadata and the resources required to correctly perform the conversion have been preserved

THANK YOU!

InterPARES Project
Massimiliano Grandi, TEAM Italy
Graduate Research Assistant