Library and Information Services in Astronomy IV
July 2-5, 2002, Prague, Czech Republic
B. Corbin, E. Bryson, and M. Wolf (eds)

Digital History: Problem of Creation of Resources

Zdeněk Uhlíř
National Library of the Czech Republic, Klementinum 190, 11001 Prague 1, Czech Republic
Zdenek.Uhlir@nkp.cz

Abstract. This paper is based on the reference background of historical librarianship, so that it does not address astronomical problems per se. In this respect we can see the historical astronomical material at three different levels: first, as factual data preserved from the past or as that we can take today as factual data; second, as evidence of the acquisition and explication of this factual material, i.e. history of astronomy as science; and third, as the inclusion of astronomical progress into historical material in general, and, in another sense, into the development of intellectual thought. At present we are witnessing a transition from an environment of predominantly printed material into an electronic-digital environment. In connection with this important change in communication there is also a new conceptualization of information and knowledge. This must be responded to by a conversion of existing sources and the creation of new ones. This wider horizon will require fundamental changes in the domain of the historical librarian. Creation of historical resources in the electronic-digital environment has three aspects: first, it is a permanent creation of digital image copies; second, building a catalogue as an access point common for both original and subsidiary documents; third, production of full-text databases of bibliographic and factual databases as well as comments-monographs. Owing to the fact that there does not exist in the electronic-digital environment any document in the strict sense of the word, it increases the importance of the context of data and the information itself becomes an interpretation. Also, due to the fact that the subject itself appears as a theme at more than one level, the knowledge becomes an interpretation at each next level. Thus the resources in the electronic-digital environment have the potential for an indirect, indefinite utilization. So it is not enough to just represent data in an objective way, but it is also necessary to prepare them such that they allow interpretation in various contexts (the sphere of information) and also for various disciplines (the sphere of knowledge). It demands both inter-disciplinary and trans-disciplinary approaches not just a concentration on a certain discipline and specialization. From all of this follows a need of changing both the library and the information preparation for the information network of the electronic-digital environment.

Every scientific and special discipline is based on information resources; the organization of every such discipline is conditioned by the possibility of working with resources. Information resources in this way become relatively

independent problems which must appear in two aspects or contexts: firstly, a connection with the respective discipline itself, secondly, the relation among information resources in general regardless to their subject divisions. Besides the scope of individual scientific and special professional disciplines there appears a special area of librarianship, information science, keeping archives etc., which is, in comparison with the preceding consideration, partly an auxiliary one, partly interdisciplinary and transdisciplinary. The approach to these interdisciplinary and transdisciplinary concerns consequently appears in a twofold light: establishing a service for special scientific and professional disciplines and partly defining its own field, which transcends these disciplines. In this sense is the work with information resources a relatively complex discipline in which the qualified, sophisticated, and advanced information service is connected with basic and applied research. This is the first thing which makes the work with information resources so important.

But information resources cannot be determined by themselves, defined in an independent, objective manner. Information resources are cultural artifacts in a wide sense. So here is another problem associated with them: their duration in time, their preservation. We speak about their permanent, documentary value and, by implication, permanent use. But it is quite obvious that we can talk with confidence only about a long-term use: information resources as cultural artifacts are exposed not only to physical or material deterioration but also to obsolescence and loss of value. So there are the questions of both continuity and discontinuity. From the physical or material point of view we could think about their restoration, while from the conceptual point of view, i.e. adequacy of paradigm and discourse, theory and methodology etc., about their transformation. It means that we must both keep the data themselves and adjust their information value within a changing intellectual context. Hence it follows that the work with information resources not only involves facts but also the possibility of change over time. Thus librarianship or information science must take on another dimension, the aspect of, historical librarianship. The fact is that a subject defined by information resources is affected by its cultural and historical dimensions. This is another thing which makes the work with information resources so important.

Because of this the information resource becomes something ambiguous; however, it was originally created with a clear purpose. To categorize the subject and method, in the way in which actual special scientific and professional disciplines do, is one-sided and insufficient. So it is necessary to distinguish here between the "thing" which is the physical content of the information resource and its "meaning" which is determined by an existence of the information resource in a historical context. Retroconversion and mainly transformation of an information resource is not a simple matter at all: the task is to preserve data in an objective physical or material sense, not only to update their information value with regard to their obsolescence, but also to present various meanings in such a way that we may distinguish their continuity and discontinuity relative to prior usage. On this point the role of historical librarianship becomes a dominant one. If nowadays there is concern with the conciliation of the sciences and the humanities, then the natural starting point for it can be found right here. This is the third thing that makes the work with information resources so important.

So, if we come from the reference background of historical librarianship, then we cannot take the definition of the problem of astronomy and of its information resources as given in advance, clearly limited and unambiguously structured, as defining the subject or content. So we can say that in this way astronomy and historical librarianship are defined as peculiar, independent, and separate disciplines. From the standpoint of a subject within the confines of its own limits this would seem to be a conflict, but it is too confining to come from this view. On the other hand further consideration seems to accept this tension because we can view the situation with more flexibility and expanded connectivity. The result is that the astronomical source material and astronomical information resources then can be reconciled in three general areas which will preserve the astronomical reality.

Firstly is the category of a subject. In this area the astronomical information resources are just the resources of data. In other words these are the data or factual data preserved from the past in the sense of what is accepted as data or factual material from the perspective of modern astronomy. Decisive in this case is the current view of scientific correctness. From the view of a permanent, respectively long-term documentary value, which is our first concern, it is necessary to reformat these data for modern usage but not to transform them. Historical librarianship appears here in its auxiliary role as a service, not as an independent discipline formulating its own peculiar problems. For historical librarianship this area is more a technical than a professional matter.

Secondly is the question of method and its transformations. Astronomical information resources are from this perspective involved in acquiring and clarifying data or factual material, i.e. to approach history of astronomy as a science. The past is in this way seen through the prism of the present concept of science, through the prism of present concepts. It is, however, a legitimate one but simplified to a linear, reductionalist approach: the past is judged on the basis of present notions of correctness and everything that is outside that appears either as a quantitatively inadequate, faulty, or a subject no longer of interest and investigation.

Thirdly, is the area of theme. The astronomical information resources are considered here in the wide sense of astronomical themes in historical material in general. Astronomy is taken here to include its pre-scientific and extra-scientific aspects as either an element of an intellectual discourse or a non-intellectual approach. There is a transition from a strict discursive rationality here to a much more unhampered outlook and view of life. But it is a view that in a strict sense is non-scientific, and historical librarianship has a role that is more as an independent discipline formulating its own peculiar problems on interdisciplinary and transdisciplinary levels. Here astronomy as a science is treated not so much in its objective and rigorous aspects but rather in its cultural and historical functions and relations.

Only the coordination of all three areas will constitute a proper approach, one which is not restricted to one dimension and will be able to channel inquiries to take into account considerations of the various intellectual and cultural disciplines in a practical way without imposing too great a limitation of subjects. In addition, we must add to the need of simple retro conversion of information resources also the need of their transformation to the new media. It is obvious

that transformation does mean deviating from the linear notion of the development of science. It means that if we look at the retro conversion maintaining a continuity then the transformation means indicating any discontinuity. The problem of astronomy for historical librarianship is that for categories of historical cultural heritage, thus one must deal with a continuity of ideas resulting from a discontinuity of discourses. It is as if we had a mirror in which is suddenly shown what a methodically educated view could not notice. And so there is a space for posing questions which arise from the tension between a "thing" and a "meaning".

And so historical librarianship must see the problem of astronomical information resources from a perspective other than astronomy itself or than the role of librarianship in its auxiliary and service function. It is obvious that this does not apply as a rule under all circumstances but only under certain defined conditions. These conditions can appear in dealing with several factors at the same time. We can assume that the three most important of these factors are:

First is a pressure on interdisciplinary thought which comes from some state of crisis of the intellectual and scientific discourse. It is obvious that at present it is taking place more or less intensively in all branches of human activity. In this situation we can feel strongly a tension between a "thing" and a "meaning" which should be in some way equalized and balanced.

The second of these factors is a non-linear view at human activity, history and culture, in other words absence of an idea of progress. Human knowledge as represented by individual scientific disciplines and their domains appears then out of a model of cumulative complexity. Individual special disciplines then will appear in great quantity as specialization increases. Astronomical information resources appear in this non-linear historical aspect in the same way as astronomy as whole relates to astrology: they differ in domain but coincide in theme. However, the context is changed as it ceases to be important to compare astronomy-science with astrology-quasi-science, or astrology-pseudo-science, and it is necessary to use a broader dimension of examination. Here we can see that astrology is something like an opposite of alchemy. They are not the predecessors of present astronomy and chemistry, respectively, nor are they mere quasi-sciences, or pseudo-sciences. The point is that astrology refers to a macrocosm and alchemy to a microcosm. We do not compare the truth with untruth or correctness with incorrectness but two different conceptualizations of the world which may each be internally consistent and so quite sufficient. However, they cannot be explained simply from themselves at all. So here we have the question of a relation of traditional sciences on one side and discursive or exact sciences on the other. As a consequence the question of information resources appears in different light: it starts to be clear that the value of an information resource does not follow from just providing factual information but also from its pragmatic function. Different paradigms and different discourses appear then from the perspective of their intentions and goals and not from considerations of an inner coherence. This then leads to meta-theoretical questions.

The third factor involves changes in communication and distribution of information. Owing to the fact that the information is a representation of reality, all these changes must lead to a change in the perception and understanding of the world. Today we witness a transfer from written communication and

distribution of information to a form involving communication technologies. The need for conversion and transformation of information resources is quite evident: we can say that while there arises a need for digital history there also a danger of creation of resources in the process.

It is obvious that the condition of mutual fulfillment of all three principal agents is present. The questions which I have indicated until now are not just hypothetical, but they have a fundamental significance for further intellectual work because they point at the paradigm shift which is beginning to emerge, although this is not always obvious. So there is a need to address these questions at a fundamental level. Otherwise it is not possible to deal with them all at the same time. I think that we as librarians would start just from communication and distribution of information as the starting points for digitization.

So from all these questions arises first the problem of the creation of historical resources in an electronic and digital environment. It is important, as we saw before, to develop a new concept of information and knowledge based on interdisciplinary and transdisciplinary considerations. The point is to give a good picture of the change in the nature of a document in the transition from the traditional printed information, communication, and knowledge environment to an electronic and digital one. The creation of historical information resources consequently takes place at three levels which need to be coordinated.

At the first level is making digital images copies of original documents. By this method alone we can transfer historical, original documents from the old, traditional environment to a new one which is electronic and digital. In this way we can get a prerequisite for determining new rational models of information at the level of sensual perception. This aspect is usually neglected, nevertheless it is quite important: sensual perception and rational treatment must be mutually compatible. Digitization in the sense of making images is mainly a technical or technological matter. It needs a standardization for capturing an image as well as loading and depicting formats. An important consideration here, for practical reasons, involves data compression. It is very important to know what purpose the digitized images will serve because this determines how the image should be stored. It is obvious that digital images of original material are made not with the aim of representing quantitative material but rather a visual impression. So we can say that a representation equals a reproduction. The point is that the resulting visual impression is close, or should be close, to the impression created by the original. So in this case there is no difference compared with the traditional environment. The difference becomes in the dynamic situation, i.e. working with a whole set of such images. Information and communication technologies enable an approach other than working in the traditional manner with a book, paper, or similar document. In the transition from an individual precept (one page - one image) to a mass precept (more pages - more images) the representation is still in the reproduction, but it is not a duplication or a transposition. It is a transformation at a sensual level. This must necessarily have serious consequences because even at this level digitization would lose its effect.

To be able to prepare digital copies for wide distribution they must be subjected to information processing. Traditional historical librarianship addresses this through the catalogue by giving a description. It gives a description of

the original historical documents in their physical form because it links a document through a carrier and an entry. In this case it is a catalogue acting as a locator. But now we come to the situation in which the catalogue and the description are related both to the original document and to the digital image, the subsidiary document. So the catalogue is not primarily a description of the documents as physical objects but rather a pathway both to the documents and their digital representations. But the catalogue record is also a representation of the images. The catalogue record merges with the subsidiary document, the digital copy, and in this way it becomes part of the digital document, and on the other hand provides a reference to the original document. So the access catalogue comes into existence. This takes on a very important consequence: in this environment the information is anything but objective because it is really only an interpretation. The library service takes another form; it changes from being a mere service to creating information. This is partly realized through a network of hyperlinks within the framework of a complex digital document and partly within the framework of an information system working with definitions in the actual fields of the catalogue records. The catalogue ceases to function as a mere sum of individual records and becomes a complex of links for searching through the information system based on a variety of request elements and the way in which they are combined. And this increasingly applies the more we use cataloguing instead of descriptive data, authority entries and subject headings, i.e. structured data, attached from the outside. In other words, data that are inherent to the original and subsidiary document, i.e. semi-structured data. In the way in which the information evolves from the explicit to the implicit level, it transforms itself from the objective level to the virtual one. The information resource itself creates another view of the world than the one to which we are accustomed in the traditional printed environment.

If the virtual information environment already exists within the scope of the access catalogue then it is desirable to use its possibilities at different levels as well. It is necessary for the level of information (the original and subsidiary documents) and meta-information (the catalogue record for the catalogue and the information system) to proceed as far as the level of the linked information (the secondary document, i.e. further processing of the primary, original and subsidiary documents). In this way bibliographical, documentary and factual databases are attached to the digital images and to the access catalogue along with full texts regardless of whether these are editions of primary texts or comments to them, related monographs, etc. In this environment we must see the catalogue record as a tertiary document which enables us to orient ourselves in the environment of the primary and secondary documents, and it is a means of sophisticated navigation able to meet a variety of goals. However, at this point we can come to have a different concept of correctness required for the comprehension of science and any professional activity. Correctness ceases to deal with factual contents of the environment. Until now it was in the traditional environment of printed documents, and it now comes to be related to a formal construction of the environment, that is, to the ability to fulfill demands of orientation and navigation. The notion and concept of science is now becoming formulated on another basis than the one that has been traditionally accepted. In the interests of maintaining a continuity of ideas it is necessary to accept discontinuity of discourse and paradigm due to the current changes in

communication and distribution of information. It appears that the role historical librarianship plays in this process is a very important one, and that it not only has an auxiliary, servant role but also a particular knowledge role.

From the description of the digital environment it is obvious that the greatest change is in the concept of what is a document. A document in the traditional sense of the word, that is a permanent, physical item with an unambiguous catalog entry, ceases to exist at all. Due to the fact that a document was normally created to represent, communicate and distribute correctly a particular set of concepts, data, or knowledge, then a document transformed to the digital environment can impart immense consequences. A basic characteristic of the digital environment is that facts are no longer objective reality but rather virtual constructions that only exist electronically. No wonder that, in general, established disciplines accept the digital environment with mixed feelings, even with apprehension, as to its representation of the purely physical document. We can await and welcome with some misgivings such radical changes, nevertheless it is not desirable to avoid changes altogether. It is obvious that we can expect even more change in this field in the future.

Instead of the traditional, printed document which preserves facts for us, there now appears the so-called compound document in the digital environment. This blurs stable documentary boundaries and even overlaps them in that it links permanent and electronic information. At the same time there also exists in information systems and communication technologies the so-called transient document which represents information that is being received in a dynamical manner. The traditional, printed environment can also represent information at different stages of development, but in this case there is a permanent record of what has transpired. However, the transient, compound document exists as only one of many possible stages of the development, and thus it is necessary to interpret the information that it contains because of its unstable and uncertain nature. The digital environment always exists somewhere between certainty and uncertainty. So we again touch meta-theoretical questions.

Production, communication, distribution and reception of information in the digital environment are not strictly separate stages, but they penetrate each other and merge. The function of individual special disciplines, mainly as information providers, is a fuzzy one in this environment. Librarianship or historical librarianship transforms from a direct service to an indirect one. In other words the boundary between the production of information and its preservation and distribution begins to disappear. The problem of preserving digital information and creating resources appears just at that boundary where the production of information and preserving and distributing it cease to be distinguishable. This requires a new approach to the whole area of information preservation and dissemination. In fact all scientific and professional disciplines are affected by this.

The creation of historical resources, including those that deal with astronomy, in the digital environment is a relatively clear process with regard to its formal and technical aspects. In this regard the question is mainly how to find the so-called "best practice", i.e. which standards should be used and which organizations ought to be approached for financing . Here we can already see programmes and projects coming into existence in both Europe and America.

There have been, of course, problems of coordination and integration on a global scale, nevertheless these are only a practical problem. We can expect that over the next several years we will see a satisfactory, general solution as efforts to link the world's information centers proceed. Formal and technical questions can solve the basic problem of the creation of resources for digital history, but a greater difficulty can be found in another place.

In our treatment we have already encountered questions which are clearly theoretical and model-dependent. These questions involve the nature of scientific and intellectual discourse. In other words it appears that the main problem is not the formally technical area but rather that of content. This problem can be simply characterized as occuring in the transition from objectivity to virtuality, from fact to construct. This is not only a problem of all scientific and special professional disciplines but of the intellectual endeavors of the mankind in general. It is a problem which can cause discontinuity in such areas as intent and interpretation and thus can create an uncertain or incorrect perspective. So we must be especially sensitive to this when the domain of librarianship or historical librarianship is at this juncture of discourse and thought.

An objective and factual approach comes from the contradiction of phenomenon and substance. Science is currently founded on this contradiction. But we find this approach is not determined by the factual matter itself but by a communication of information and its distribution in the traditional, printed environment. In the digital environment however, information by its very nature exists by implication and that explicit information exists only in a plurality of intentionally created possibilities. There is no firm ground here but rather a lot of various relative positions based on intention and realized by context. The structure of the information field represented by resources for digital history is not linearly cumulative at a given reference level but is broken into an uncertain number of different subject areas. The decisive question for us is to grasp the natures of these areas in a general way and to look into their conceptual structures.

We must understand information resource in the digital environment as a complex cluster of primary, secondary and tertiary documents. We must see, at a fundamental level, a way of exploiting such resources as transient, compound documents. On the basis of the information acquired in this manner there appears the possibility of a twofold approach: either to acquire additional information by a request formulated in a different manner, or to formulate a new question based on the acquired information, which will lead us to the acquisition of a new, primary document, i.e. to revise or amend secondary and tertiary documents, respectively. So in the environment of digital communication and knowledge we constantly move in three different areas - the actual, eventual and potential ones. The actual intent of information is determined by an incoming point of view and is represented in a transient, compound document which was created on the basis of an particular request. Here we can find information support in a way that we can comprehend. The eventual realm of information provided is determined by the incoming intention, or point of view, and is represented in a resulting, transient, compound document which was created on the basis of a multifold request. The result is to widen and enrich the acquired information. Thus traditional information support will result in a more

sophisticated level of support or to an independent source. A potential area of information is determined by the contradiction between the anticipation or expectation or the requestor and a wider or more enriched body of information than requested. It leads back to primary resources and to new processing. So here is where mere knowledge support evolves into providing knowledge because a new question of related information was discovered. It seems to me that the task of the historical librarian as a knowledge professional emerges here, at the level of identifying new thematic circles and forming new themes. This means the detection of related information. The digital environment thus produces a requirement on the librarian to permeate service and scholarly work.